Lexikos 22

Lexikos 22

Redakteur

Editor

D.J. Prinsloo

Resensieredakteur

Review Editor

T. Harteveld

African Association for Lexicography

AFRILEX-REEKS 22:2012
AFRILEX SERIES 22:2012

BURO VAN DIE WAT

STELLENBOSCH

Die publikasie van hierdie boek is moontlik gemaak deur 'n ruim skenking van die L.W. Hiemstra-Trust – Opgerig deur Riekie Hiemstra ter herinnering aan Ludwig Wybren (Louis) Hiemstra.

The publication of this book was made possible by a generous donation from the L.W. Hiemstra Trust – Established by Riekie Hiemstra in memory of Ludwig Wybren (Louis) Hiemstra.

Uitgewer Publisher

BURO VAN DIE WAT
Posbus 245
7599 STELLENBOSCH

Eerste uitgawe 2012

Tipografie en uitleg deur Tanja Harteveld en Hermien van der Westhuizen
Proeflees van die Engels deur Lorna Hiles
Bandontwerp deur Piet Grobler
Geset in 10 op 12 pt Palatino

Gedruk en gebind deur Onyx Press
Stewartstraat 61 Goodwood

ISBN 978-0-9814434-4-7
ISSN 1684-4904

Lexikos is elektronies beskikbaar by http://lexikos.journals.ac.za/
Lexikos is available online at http://lexikos.journals.ac.za/

Lexikos is elektronies beskikbaar by Sabinet, AJOL, Ebsco en Proquest
Lexikos is available online from Sabinet, AJOL, Ebsco and Proquest

Indekse Indexes
Arts and Humanities Citation Index®, Current Contents®/Arts & Humanities, Current Contents®/Social and Behavioral Sciences, Journal Citation Reports/Social Sciences Edition, Social Sciences Citation Index®, and Social Scisearch®; Linguistic Bibliography Online; Linguistics Abstracts Online; Linguistics and Language Behavior Abstracts; MLA International Bibliography; R.R.K. Hartmann's Bibliography of Lexicography; Scopus

Inhoud / Contents

Redaksionele doelstellings

Lexikos is 'n tydskrif vir die leksikografiese vakspesialis en word in die AFRILEX-reeks uitgegee. "AFRILEX" is 'n akroniem vir "leksikografie in en vir Afrika". Van die sesde uitgawe af dien *Lexikos* as die amptelike mondstuk van die *African Association for Lexicography* (AFRILEX), onder meer omdat die Buro van die WAT juis die uitgesproke doel met die uitgee van die AFRILEX-reeks gehad het om die stigting van so 'n leksikografiese vereniging vir Afrika te bevorder.

Die strewe van die AFRILEX-reeks is:

(1) om 'n kommunikasiekanaal vir die nasionale en internasionale leksikografiese gesprek te skep, en in die besonder die leksikografie in Afrika met sy ryk taleverskeidenheid te dien;
(2) om die gesprek tussen leksikograwe onderling en tussen leksikograwe en taalkundiges te stimuleer;
(3) om kontak met plaaslike en buitelandse leksikografiese projekte te bewerkstellig en te bevorder;
(4) om die interdissiplinêre aard van die leksikografie, wat ook terreine soos die taalkunde, algemene taalwetenskap, leksikologie, rekenaarwetenskap, bestuurskunde, e.d. betrek, onder die algemene aandag te bring;
(5) om beter samewerking op alle terreine van die leksikografie moontlik te maak en te koördineer, en
(6) om die doelstellings van die *African Association for Lexicography* (AFRILEX) te bevorder.

Hierdie strewe van die AFRILEX-reeks sal deur die volgende gedien word:

(1) Bydraes tot die leksikografiese gesprek word in die vaktydskrif *Lexikos* in die AFRILEX-reeks gepubliseer.
(2) Monografiese en ander studies op hierdie terrein verskyn as afsonderlike publikasies in die AFRILEX-reeks.
(3) Slegs bydraes wat streng vakgerig is en wat oor die suiwer leksikografie of die raakvlak tussen die leksikografie en ander verwante terreine handel, sal vir opname in die AFRILEX-reeks kwalifiseer.
(4) Die wetenskaplike standaard van die bydraes sal gewaarborg word deur hulle aan 'n komitee van vakspesialiste van hoë akademiese aansien voor te lê vir anonieme keuring.

Lexikos sal jaarliks verskyn, terwyl verdienstelike monografiese studies sporadies en onder hulle eie titels in die AFRILEX-reeks uitgegee sal word.

Editorial Objectives

Lexikos is a journal for the lexicographic specialist and is published in the AFRILEX Series. "AFRILEX" is an acronym for "lexicography in and for Africa". From the sixth issue, *Lexikos* serves as the official mouthpiece of the *African Association for Lexicography* (AFRILEX), amongst other reasons because the Bureau of the WAT had the express aim of promoting the establishment of such a lexicographic association for Africa with the publication of the AFRILEX Series.

The objectives of the AFRILEX Series are:

(1) to create a vehicle for national and international discussion of lexicography, and in particular to serve lexicography in Africa with its rich variety of languages;
(2) to stimulate discourse between lexicographers as well as between lexicographers and linguists;
(3) to establish and promote contact with local and foreign lexicographic projects;
(4) to focus general attention on the interdisciplinary nature of lexicography, which also involves fields such as linguistics, general linguistics, lexicology, computer science, management, etc.;
(5) to further and coordinate cooperation in all fields of lexicography; and
(6) to promote the aims of the *African Association for Lexicography* (AFRILEX).

These objectives of the AFRILEX Series will be served by the following:

(1) Contributions to the lexicographic discussion will be published in the specialist journal *Lexikos* in the AFRILEX Series.
(2) Monographic and other studies in this field will appear as separate publications in the AFRILEX Series.
(3) Only subject-related contributions will qualify for publication in the AFRILEX Series. They can deal with pure lexicography or with the intersection between lexicography and other related fields.
(4) Contributions are judged anonymously by a panel of highly-rated experts to guarantee their academic standard.

Lexikos will be published annually, but meritorious monographic studies will appear as separate publications in the AFRILEX Series.

Redaktionelle Ziele

Lexikos is eine Zeitschrift für Fachleute der Lexikographie, die in der AFRILEX-Serie erscheint. "AFRILEX" ist ein Akronym für "Lexikographie in und für Afrika". Von der sechsten Ausgabe an dient *Lexikos* als amtliches Sprachrohr der *African Association for Lexicography* (AFRILEX), u.a. weil das Büro des WAT das gerade angesprochene Ziel mit der Ausgabe der AFRILEX-Serie verfolgt, die Gründungsziele eines solchen lexikographischen Vereins für Afrika zu fördern.

Die folgenden Ziele werden mit den Publikationen der AFRILEX-Serie verfolgt: Man möchte:

(1) ein Medium schaffen für die nationale und internationale Diskussion, besonders aber der Lexikographie in Afrika mit seinen zahlreichen Sprachen dienen;
(2) die Diskussion fördern, unter Lexikographen als auch zwischen Lexikographen und Linguisten;
(3) Kontakt herstellen und fördern zwischen südafrikanischen und ausländischen lexikographischen Projekten;
(4) die Aufmerksamkeit lenken auf die interdisziplinäre wissenschaftliche Praxis der Lexikographie, die Beziehung aufweist zur Linguistik, allgemeinen Sprachwissenschaft, Lexikologie, Computerwissenschaft, zum Management und zu anderen Bereichen;
(5) die Zusammenarbeit auf allen Gebieten der Lexikographie fördern und koordinieren;
(6) die Ziele der *African Association for Lexicography* (AFRILEX) fördern.

Gemäß den Zielsetzungen der AFRILEX-Serie werden:

(1) Beiträge zur lexikographischen Diskussion in der Fachzeitschrift *Lexikos* veröffentlicht;
(2) monographische und andere Studien auf diesem Gebiet als getrennte Publikationen in der AFRILEX-Serie erscheinen;
(3) nur einschlägige Beiträge, die sich ausschließlich mit Lexikographie oder mit fachverwandten Gebieten befassen, für Aufnahme in der AFRILEX-Serie in Betracht gezogen;
(4) Beiträge anonym von einem aus Spezialisten des Faches von hohem akademischen Ansehen bestehenden Ausschuß beurteilt.

Lexikos erscheint jährlich. Ausgewählte monographische Studien dagegen erscheinen gelegentlich als getrennte Publikationen in der AFRILEX-Serie.

Politique éditoriale

La revue ***Lexikos***, destinée aux spécialistes de lexicographie, est publiée dans la collection AFRILEX (acronyme de "lexicographie en Afrique et pour l'Afrique"). Depuis son sixième numéro, *Lexikos* est l'organe officiel de l'*African Association for Lexicography* (AFRILEX), entre autres parce que le Bureau du WAT s'est donné pour objectif de promouvoir le développement d'une telle association lexicographique en Afrique par la publication de la collection AFRILEX.

Les objectifs de la collection AFRILEX sont de :

(1) créer un forum de discussion national et international sur la lexicographie, particulièrement au service de la lexicographie en Afrique, qui représente une grande diversité de langues;
(2) stimuler le débat entre lexicographes, ainsi qu'entre lexicographes et linguistes;
(3) établir et promouvoir le contact avec des projets lexicographiques locaux ou étrangers;
(4) attirer l'attention générale sur la nature interdisciplinaire de la lexicographie, qui touche des domaines comme la linguistique générale, la lexicologie, l'informatique, le management, etc.;
(5) favoriser et coordonner la coopération dans tous les domaines de la lexicographie; et
(6) promouvoir les orientations de l'*African Association for Lexicography* (AFRILEX).

Pour atteindre ces objectifs, la collection AFRILEX

(1) publiera les contributions aux discussions sur la lexicographie dans la revue *Lexikos*, dans la collection AFRILEX;
(2) publiera sous forme de publications séparées dans la collection AFRILEX des monographies et autres travaux dans le domaine de la lexicographie;
(3) ne publiera dans la série AFRILEX que des travaux dans le domaine de la lexicographie, qu'ils traitent de lexicographie pure ou des rapports entre la lexicographie et d'autres disciplines voisines; et
(4) soumettra de manière anonyme toutes les propositions à des experts hautement qualifiés, pour en garantir le niveau académique.

Lexikos est publié annuellement, mais les travaux de qualité exceptionnelle seront publiées sous forme de publications séparées dans la collection AFRILEX.

What Counts as a Proverb? The Case of *NTC's Dictionary of Proverbs and Clichés*

Is'haaq Akbarian, *Department of English Language and Literature, Faculty of Humanities, University of Qom, Iran (i-akbarian@qom.ac.ir)*

Abstract: A dictionary on proverbs should meet its claim in compiling the specific, relevant formulaic entries in order to retrieve the required information as quickly and as successfully as possible. The inconsistency between what such a dictionary claims to include and what it actually includes might stem from not attending to the possible boundaries between the fixed expressions that might as well be considered as guidelines in lexicography. Based on the distinctions between the fixed expressions offered in Gramley and Pätzold (1992) and Simpson (1985), *NTC's Dictionary of Proverbs and Clichés* (Bertram 1996) was investigated as a sample proverb dictionary available, especially useful for non-native speakers. It was found that (a) the above dictionary failed to distinguish between proverbs and other prefabricated expressions and (b) nearly 67 percent out of the total number of the entries in the dictionary were found to be proverbs whereas around 33 percent of them turned out to be non-proverb items. Some lexicographic as well as research implications are also discussed at the end.

Keywords: LEXICOGRAPHY, NTC'S DICTIONARY, PROVERB, PROVERBIAL EXPRESSIONS, CLASSIFICATION SCHEME, FIXED EXPRESSIONS, IDIOM, CLICHÉ, ENTRY

Opsomming: Wat word beskou as 'n spreekwoord? Die geval van *NTC's Dictionary of Proverbs and Clichés*. 'n Spreekwoordeboek behoort sy naam gestand te doen deur die spesifieke, relevante formuleagtige inskrywings op te neem om die gevraagde inligting so vinnig en suksesvol moontlik op te spoor. Die teenstrydigheid tussen wat so 'n woordeboek beweer hy opneem en wat hy inderdaad opneem mag spruit uit die feit dat daar nie aandag gegee word aan die moontlike grense tussen die vaste uitdrukkings, wat as riglyne in die leksikografie beskou kan word, nie. *NTC's Dictionary of Proverbs and Clichés* (Bertram 1996) is ondersoek as 'n voorbeeld van 'n beskikbare spreekwoordeboek wat besonder nuttig vir niemoedertaalsprekers is. Hierdie ondersoek is gebaseer op die onderskeidings tussen vaste uitdrukkings wat aangebied word in Gramley en Pätzold (1992) en Simpson (1985). Daar is bevind dat (a) die bogenoemde woordeboek nagelaat het om te onderskei tussen spreekwoorde en ander vaste uitdrukkings en dat (b) byna 67 persent van die totale aantal inskrywings in die woordeboek spreekwoorde was teenoor ongeveer 33 persent van die items wat geblyk het nie spreekwoorde was nie. Ten slotte word sommige implikasies vir die leksikografie sowel as vir navorsing bespreek.

Sleutelwoorde: LEKSIKOGRAFIE, NTC SE WOORDEBOEK, SPREEKWOORD, SPREEKWOORDELIKE UITDRUKKINGS, KLASSIFIKASIESKEMA, VASTE UITDRUKKINGS, IDIOOM, CLICHÉ, INSKRYWING

Introduction

Thanks to the advances in different fields of study, interdisciplinary approaches to addressing issues of research and practice are highlighted. The resultant contribution of different fields to one another seems to yield better output. From that perspective, lexicography is no exception. As the evidence, technology and computer sciences have contributed much to lexicography in the last decades. The result is that, with the aid of information and communication technology, we have gained within the last quarter of a century what might have been obtained for a century without its help. It has changed using the result of lexicography, i.e. using dictionaries, for native speakers and non-natives learning a language. As Mark Warschauer (2005) observes, it has changed the context and how of language learning and, for that matter, has altered the context and how of using dictionaries as well; we look up the meaning of words in our mobiles nowadays, for instance.

Likewise, lexicography might also benefit from other fields such as linguistics, or more specifically discourse or corpus analysis, and so on. In this paper, we attempt to tackle a fine lexicographic issue from a linguistic point of view. Actually applying a corpus analyst's view to a proverb dictionary might shed some light on the dynamic field of lexicography with regard to proverb dictionaries.

International Journal of Lexicography, as an important outlet for the lexicographic community, has recently welcomed a collection of studies reaffirming the significance of efforts to further develop empirical study into dictionary use (*IJL* 2011). Lew (2011: 3) believes that "As experimental design, methods and techniques get more sophisticated, what we are getting in return is greater, finer, and more useful detail". All this indicates the point that interdisciplinary approaches are emphasized even in lexicography, thus inspiring further, even more innovative and revealing, efforts.

To start with, we must clarify our stance based on which we will consider a proverb dictionary, with probable direct and/or indirect lexicographic implications. Our approach, giving us that stance or foundation to report this study, emerges from the idiom principle of John Sinclair and is put forth much more in details by Gramley and Pätzold (1992). According to the *idiom principle* (Sinclair 1991: 110), "a language user has available to him or her a large number of semi-pre-constructed phrases that constitute single choices, even though they might appear to be analyzable into segments". On that basis, Gramley and Pätzold (1992), discussing words in combination, set out to propose a fixed-expression classification scheme and thus divide or classify such prefabricated phrases, called *multi-word units*, into well-established groups systematically so that any fixed expression might be included under one of the branches as neatly as possible. What makes the groups or units of fixed expressions distinct from one another pertains to a variety of stylistic, situational, formal, semantic, and syntactic aspects.

Classification of fixed expressions

According to the classification scheme of fixed expressions proposed by Gramley and Pätzold (1992: 53), based on one criterion, fixed expressions are divided into two groups, one of which expresses meanings or speech acts, (i.e., acts performed by language) such as promises, warnings, requests, and the like, whereas the other group does not. Another criterion, they put forth, is whether or not 'the expression is equivalent to a whole sentence or free utterance'. Each of the two groups above is then classified into two levels or branches. In this division, the left branch is split into expressions that are utilized in 'set social situations' and those that are not, 'a pragmatic criterion'. The right branch, on the other hand, is subdivided by the semantic criterion of idiomaticity, i.e., a meaning that cannot be deduced from the meaning of the individual words (see Figure 1). The list below can be subdivided further to include other items but for the purposes of the current study we do not go further. The relationship between the items in Figure 1 will be clarified in the following pages. However, it is sufficient here to mention an exemplar expression for proverbs, idioms, collocations, and pragmatic idioms, respectively: (1) *birds of a feather flock together,* (2) *how do you do*, (3) *red herring, and* (4) *meet demand*.

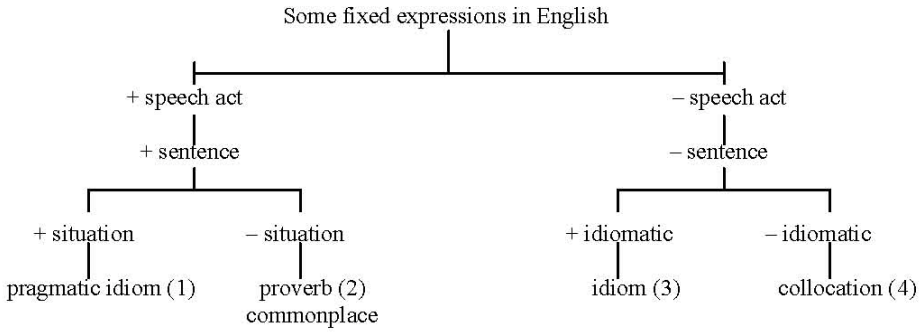


Figure 1: Fixed expressions in English (Gramley and Pätzold 1992: 54)

On the basis of the scheme above, we will only remain more focused on the distinction between a proverb and an idiom since the purpose of the present research is to investigate whether a sample of a compiled proverb dictionary meets these well established criteria or the lexicographers ignore the above-made basics and consequently include items other than proverbs. The classification scheme above is a good and systematic device to distinguish the different kinds of fixed expressions. We also attempt to strengthen the criteria by adding some more points to the scheme above to make it better off in evaluating one specific dictionary as a sample of dictionaries on proverbs, namely, *NTC's Dictionary of Proverbs and Clichés* (Bertram 1996).

Clichés and fixed expressions

What is worth mentioning with regard to the title and therefore the content of the dictionary is the term *cliché*. Howard (1984: 92) states that "*clichés* are routine or stereotypic forms that are found in many areas of life, for example, art, thought, behavior, visual images or urban architecture".

As to the linguistic description of clichés, Luelsdorff (1981) defines them as phrases, clauses and sentences which, owing to very frequent occurrence, have become hackneyed and trite. He divides the native clichés (English items) on the grammatical level into nominal, verbal, and sentence structures. Some language *clichés* are made up of single lexemes, for example, the journalistic items *bombshell*, *brainchild*, *to harmonize*, *to orchestrate* and *scenario*, but most consist of more than two items. Clichés, therefore, are applied to any formulaic item ranging from idioms to proverbs. That is, any fixed expression that is overused. The question that is posed here is whether it is appropriate to juxtapose the terms clichés with proverbs and then include other fixed expressions that are also considered clichés, such as idioms, in a dictionary especially compiled on proverbs.

The charge is that people, in using clichés, do not think when they use expressions such as *acid test*, *psychological moment* or *leave no stone unturned* (Gramley and Pätzold 1992).

Aside from this, there is another aspect to the issue of clichés as well. By using clichés one signals that one has acquired part of the socio-cultural competence of a given speech community (Luelsdorff 1981). Clichés help to create an in-group feeling of sympathy, solidarity and good will. Therefore, they fulfill an important social function (Gramley and Pätzold 1992). Stylistically speaking, published lists of clichés are mostly subjective (Brook 1981).

Proverbs versus other fixed expressions

To justify our position systematically, we have to clarify the issue of fixed expressions. Though in some cases it might be difficult, we try to distinguish the borders between the items in the classification scheme above. To that end, the items are to be contrasted further.

Proverbs are different from pragmatic idioms. The occurrence of pragmatic idioms is determined by a particular social situation. In fact, they need the context of situation to be understood correctly, for example, *single or return* is used at a railway ticket counter. Proverbs, however, do not function as such. They turn up everywhere. Proverbs sum up situations and give advice in short, terse phrases (Bertram 1996). Regarded as "the wit of one and the wisdom of many", the proverb is a terse and witty philosophical saying that conveys a lesson. It "couches conventional wisdom in a poetic capsule, making it esthetically pleasing and memorable" (Yankah 2001: 201).

As for the idioms, meaning is the decisive, if not the only, criterion. The

word forms in an idiom do not constitute lexical units and do not make an isolable contribution to the meaning of the whole, while collocations consist of two word forms which are, at the same time, semantic constituents or lexical units. Idioms therefore show *unitary meaning* and their constituents are termed *formatives* by some linguists, according to Gramley and Pätzold (1992).

There are some means to recognize that an expression is an idiom, setting it off from the other kinds of prefabricated expressions. For one thing, the constituents in an idiom are not semantic constituents or lexical units; they are formatives. Relevant to that is '*recurrent semantic contrast*' which is a test for a semantic constituent (Cruse 1986: 26-9). If *hit* and *pail* are contrasted with *kick* and *bucket* in the expression *to kick the bucket*, it becomes clear that *kick the bucket* is an idiom. The same goes for an adjectival idiom such as *red herring*: in *red book* and *green book*, *red* and *green* are in recurrent semantic contrast whereas in *red herring* and *green herring* they are not. Secondly, many idioms have two meanings, a literal and an idiomatic one. Consider *kick the bucket* or *pull one's leg*. In comprehending these idioms, only the context can give a clue as to which meaning is intended. We are dealing with an idiom if a literal meaning does not make sense in terms of the world as we know it. Thirdly, when an expression is formed in a way that is contrary to the syntactic rules of contemporary English, as in the definite articles in *kick the bucket* and *fly off the handle*, most probably it is an idiom. The definite article normally has the function of indicating that an item has already been mentioned. This condition is not fulfilled in the idioms cited (Dobrovol'skij and Piirainen: 2005). Finally, idioms can be phonologically irregular in that they have an unpredictable stress pattern (Strässler 1982).

Proverbs and *commonplaces* are free utterances or self-contained statements (Norrick 1985). Both can be equivalent to a complete sentence. However in the case of proverbs, shortened versions are quite common. Shortening and other changes — additions, variations, transpositions — do not necessarily affect the intelligibility of proverbs, apparently for they are well known (Akbarian 2010; Norrick 1985).

Proverbs show irregular syntax, for example, *like father, like son* which means 'a son will resemble his father'. Proverbs as well as commonplaces are concerned with general rather than specific meanings, which is why the past tense is not normally found with them. Many proverbs are metaphorical and may pose problems for understanding, while commonplaces are usually literal and easy to process (Norrick 1985: 70). Proverbs are well established, traditional, and recorded in many collections and dictionaries (Smith 1985). They contain 'a good dose of common sense, experience, wisdom and above all truth' (Mieder 1989: 15), 'have no known authors, and cannot be traced to specific sources', that is, they are folklore items (Gramley and Pätzold 1992: 77).

Sometimes, it is difficult to understand what some instances of proverbs really mean. For instance, '*a good husband makes a good wife*' needs to be explained or at least thought about for a while before it makes sense. Still other

proverbs preserve older forms of English words as well as older style and syntax. Some are difficult simply because they are used metaphorically. To exemplify, consider '*a stitch in time saves nine*' that is never used in reference to sewing.

Proverbs as a class cannot be completely frozen (Gläser 1989, as cited in Gramley and Pätzold 1992). Some expressions mark proverbs, such as *(as) they say, it is said, as the proverb goes* that are called proverbial affixes by Norrik (1985). Moreover, variability is a characteristic trait of proverbs; they can be added to, transformed, and abbreviated. Transformations do not change proverbs out of all recognition such as *it is while the iron is hot that it should be struck*. Idioms would become meaningless or allow only a literal reading, if they are treated the same way.

Defining proverb as "a traditional saying which offers advice or presents a moral in a short and pithy manner", Simpson (1985: ix) in the introduction to the dictionary, *The Concise Oxford Dictionary of Proverbs,* states, "Paradoxically, many phrases which are called 'proverbial' are not proverbs as we now understand the term". He observes that this confusion dates from before the eighteenth century, when the term *proverb* also covered metaphorical phrases, similes, and descriptive epithets, and was used far more loosely than it is today. Nowadays, we normally expect a proverb to be cast in the form of a sentence.

However, according to Simpson (1985), some metaphorical phrases such as *to cut off your nose to spite your face* and *to throw the baby out with the bathwater* would be admitted as a proverb since they are rendered in the sentence form and contain some advice, wit, or universal truth, appearing as *Don't cut off your nose to spite your face* and *Don't throw the baby out with the bathwater*.

Significance of the use of proverbs

The proverb is grounded upon years of experience and close observation of life and natural phenomena, and through metaphorical language it may warn, advise, or reprimand by drawing attention to the moral or ethical consequences of human behavior. The proverb may advocate patience, cooperation, and perseverance, and repudiate greed and selfishness. A Turkish proverb says "He who does not listen to proverbs remains screaming for help"; according to Russians, "For the sake of a proverb, a peasant walks to Moscow"; and the Yoruba of Nigeria would say, "The man who knows proverbs reconciles difficulties". All this boils down to the fact that proverbs are prevailing in all the societies. Since they pertain to universal truth, they are welcomed as discourse ornaments. The proverb is a "lamp of the word" among the Arabs and an "ornament of speech" in Iran. The Igbo say that "proverbs are the palm oil with which words are eaten," implying that words are hard to swallow without a proverb lubricant. However, it takes considerable cultural sensitivity to grasp the full semantic nuances of the proverb in social interaction (Yankah 2001: 201-202).

According to Yankah (2001), the lessons often embedded in proverbs

make them tools for moral education. Besides, the element of education in proverb use can be subsumed under the rhetorical function, that is, the proverb as a tool for persuasion in social interaction; the proverb user seeks to alter or reinforce the listener's conviction by referring him to timeless parallels within the proverb universe. By getting the addressee to agree with the moral precept in the proverb, the speaker thereby hopes to win him over. The rhetorical power of the proverb in part derives from its authoritative sources. Yankah (2001) mentions a few instances of this ascription: Among the Punjabi of India, proverbs are "the drum of God", and in Turkey proverbs from the prophets are said to have originated from the Holy Koran and Hadith holy traditions. In several African cultures, proverb authorship is ascribed to elders and ancestors.

Typology of English proverb dictionaries

Prędota (2003: 95-100) suggests that monolingual books of proverbs can be divided into three main categories:

1. Scientific: It has an historic character, including the complete treasure house of English-language proverbs. *The Oxford Dictionary of English Proverbs* (Oxford 1970) is an example.
2. Popular scientific: *The Concise Oxford Dictionary of Proverbs* (Oxford 1998) is an instance of this category, containing the current living core of the English treasury of proverbs.
3. Dictionaries destined for the teaching of foreign languages: This category has two characteristics, namely, (a) the dictionaries offer only a limited number of proverbs and (b) those selected generally belong to the famous and most frequent proverbs. Prędota presents *English Proverbs* (Stuttgart 1988) by Wolfgang Mieder as a good example for its practical lexicographic solutions. It includes 1200 proverbs with clarifying explanations to non-native speakers and dictionary articles kept to a minimum.

The study

There used to be and apparently there still is some confusion as to the distinction between the different types of fixed expressions; for example, some idioms are occasionally taken for proverbs (Simpson 1985). However, the classification scheme proposed above and the criteria established on the basis of the discussion above (Simpson 1985, Gramley and Pätzold 1992) contribute much to the issue at stake in that we have some well-established criteria to distinguish what is what among fixed or prefabricated expressions. This classification scheme can be suggested to lexicographers so that they might filter out what should and/or should not be included in a dictionary of proverbs before they tackle

the issue of arranging the entries. Based on those criteria, this research intends to investigate a sample dictionary, namely, *NTC's Dictionary of Proverbs and Clichés* (Bertram 1996) in which the entries are defined and have realistic examples illustrating meaning, usage, and significance. The motivation for selecting this dictionary is that it is usually the only proverb dictionary available in libraries in Iran. The reason might be that it illustrates the meaning of the proverbs more clearly than any other proverb dictionaries and thus consulted most frequently by Iranian foreign language learners of English. Though the dictionary was published more than a decade ago, it can be regarded as one of the most up-to-date reliable sources since the proverbs contained do not change their nature over time. It is still reprinted as the same, without any revision. As a result, it could be recommended as a good source on proverbial expressions for everyone, especially for foreign language learners, until another revision or a better one comes out. In order to account for a possible lexicographic concern on proverb dictionaries, the following two specific questions guide the present research:

1. *On the basis of the criteria above, does NTC's Dictionary of Proverbs and Clichés include only proverb entries or fail to distinguish between proverbs and other prefabricated expressions and include expressions other than the proverbial ones?*
2. *What percentage of the entries included in the sample dictionary investigated belongs to proverbs?*

Materials

The material for the current study is *NTC's Dictionary of Proverbs and Clichés* (Bertram 1996). This dictionary is a selection of familiar expressions, some of which are also found in other major European languages. Below, the lexicographic microstructure of this dictionary is presented for the current research purposes:

— The entries are alphabetized by the first word in the expression, ignoring punctuation and hyphens.

— The dictionary gives clear definitions of the phrase and good examples of how it is used.

— It provides a literary source if one is known.

— Also, two brief dialogs accompany each expression, showing the typical way that the expression is used.

— Entry heads appear in boldface type, with expressions that are not entries, cited in italic font.

— Only one or two of the possible variants of a proverb are listed in this

dictionary, printed in boldface type.

— Definitions and paraphrases are in roman type, with alternative definitions and paraphrases separated by semicolons.

— A definition or paraphrase may be followed by comments in parentheses that give extra information and list the common clichés derived from the entry heads.

— A Phrase-Finder Index is provided to permit easy locating of an entry head.

The dictionary includes 321 pages. There is an introduction of three pages. The entry heads of the dictionary run from page 1 to page 240 and the Phrase-Finder Index from 241 to 321.

In this research, if a proverb has a variant idiom form originating from the head entry, i.e. proverb entry, the idiom is not regarded as violating the criteria of our investigation. It is excluded from the list of non-proverb items in this sample dictionary. Since that idiom is included and explained within a proverb entry, it is regarded as additional information for the entry, not a separate head entry in itself. Also, a non-proverb entry, such as an idiom, is not regarded as a separate entry if it is already explained as a variant or alternative of another idiom. For example, if the dictionary explains the idiom *a drop in the bucket* and mentions the idiom *a drop in the ocean* as the alternative form for the former and then offers the latter as another entry while cross-referencing to the former for the latter's meaning and exemplification, both are regarded as one entry. This holds true for the proverbs as well; the variant forms of proverbs are regarded as one single entry along with the relevant proverb unless they carry different meanings.

Procedure

To answer the questions posed above, *NTC's Dictionary of Proverbs and Clichés* (Bertram 1996) was considered carefully in accordance with the classification scheme offered above. All the entries were considered by the researcher and the number of proverbs and the non-proverb items were counted. The non-proverb items were separated from the proverbs (See Appendix).

Results and discussion

In accordance with the criteria put forth so far, the present research intends, firstly, to study whether *NTC's Dictionary of Proverbs and Clichés* (Bertram 1996) includes only proverb entries or fail to distinguish between proverbial and other prefabricated expressions and, as a result, includes expressions other than proverb entries, and secondly, what percentage of the entries included in

the sample dictionary belongs to proverbs.

A careful investigation of the above dictionary showed that there were totally 982 entries in it. Out of the total number of entries, 656 entries belong to proverbs and the remaining 326 entries are not proverbs. In other words, 66.80% of the total number of the entries includes proverbs whereas 33.20% of the total number of the entries includes non-proverb items. Approximately, one third of the entries in the dictionary under study does not pertain to proverbs. Yet, we acknowledge the limited sample in our study. Therefore, it is not wise to generalize from our findings to other proverb dictionaries. However, at least this percentage of non-proverb items observed in this dictionary directly supports our position that the compilers of the current dictionary did not distinguish the boundaries between the prevailing fixed expressions such as idioms, proverbs, and the like.

At this stage, we will focus on the non-proverb entries (see Appendix) and try to see, in general, what structural category or pattern they mostly belong to, showing once again that these entries are totally different in kind from proverb entries:

1. Some non-proverb entries are made up of noun phrases expanded by prepositional phrases, e.g. *a chip off the old block*.
2. A few consist of (mixed) prepositional phrase, e.g. *according to someone's lights* or *at sixes and sevens*.
3. A large number of the entries include a structure like '(as) adjective as noun phrase', e.g. *(as) drunk as a lord* or *(as) hard as nails*.
4. Several of them are in the form of a dependent clause that is complete once the independent clause is provided by the language user, e.g. *if (the) worst comes to (the) worst*.
5. A greater number of the non-proverb entries in the Appendix consist of infinitive phrases (expanded by noun phrases or prepositional phrases, etc.), e.g. *to fall between two stools*, *to give credit where credit is due*, or *to let the cat out of the bag*.
6. The structure of a small number of the entries contains adjectives expanded by a prepositional phrase, e.g. *crazy like a fox*.

We have to acknowledge that there are instances of entries including other types of structures not mentioned above. However, it is important to note that some structures cited above are more common than others. For instance, structures 3 and 5 make up the majority of non-proverbs found in this dictionary. Further consideration of the arrangement of the entries in the dictionary reveals that there is a full stop (.) after the entries that are complete sentences. Indeed, the entries, followed by a full stop (.), are free and self-contained utterances. All of these entries are proverbs, not idioms (Akbarian 2010; Norrick

1985); there is no full stop after any entry taken as an idiomatic expression in accordance with the scheme above. We take the entries without a full stop at the end of them to be non-proverbs for the reason that they cannot stand alone even if some of them include verbs. These entries, including verbs but not standing alone, are dependent clauses that need independent clauses. It would have helped the consultants of this dictionary further had the lexicographers marked the entries as such explicitly. The implication for lexicography is that the relevant items or entries comprising a particular type of dictionary should be exclusively included in their respective collections, and not in a collection of a different type of entries. As argued in this research, there are subtle differences between various types of fixed expressions, such as proverbs and idioms. However, if a particular type of expression or entry, e.g. an idiom, is associated with or emerges from proverbs, or for whatever reason, is to be included in a dictionary of proverbs, it should be recognized explicitly as an idiomatic expression. In this manner, users' attention will be explicitly drawn to the entries and the need for a more consistent organization of the entries emerging from the analysis of the dictionaries will be highlighted (Szczepaniak and Lew 2011), as suggested in this paper. It could have double lexicographic and pedagogical effect for the consultants of this type of dictionaries who are mainly language learners; From a lexicographic perspective, the structure of the dictionary would contain clear clues as to the contents and, from a pedagogical perspective, the consultants of such dictionaries will get cognizant of both of these expressions and their distinctive boundaries. The following is suggested as an example of how the items should be arranged and marked:

(As) busy as a cat on a hot tin roof (idiom) Full of lively activity; very busy. □ *I'm afraid I can't go to lunch with you on Saturday; I'll be busy as a cat on a hot tin roof, between working overtime and the two parties I have to go to.*

Practice makes perfect. (proverb) Doing something over and over again is the only way to learn to do it well. □ JILL: *I'm not going to try to play the piano anymore. I always make so many mistakes.* JANE: *Don't give up. Practice makes perfect.*

In order to support the findings of the present study, we further attempted to investigate other dictionaries. Since most of the non-proverb items observed in the dictionary under study were idioms, we investigated *NTC's American Idioms Dictionary*, compiled by Richards A. Spears (2000), to see whether the non-proverb items found in *NTC's Dictionary of Proverbs and Clichés* (Bertram 1996) had been included in *NTC's American Idioms Dictionary* as idiom entries. It is worth emphasizing here that the two dictionaries were published by the same publisher, that is, NTC Publishing Group.

For the purpose of our study, we randomly investigated every tenth entry of the non-proverb entries in the Appendix. We randomly started with number 3. Therefore, entries numbered 3, 13, 23, 33, 43, 53, and so on in our list were

considered (see Appendix). This amounted to one tenth of the total non-proverb items in Appendix. With the non-proverb items numbering 326, this amounted to 33 of the non-proverb items, out of which 19 items were observed in *NTC's American Idioms Dictionary*, compiled by Richards A. Spears (2000) and the remaining 14 were not observed in that dictionary. So the comparison of our results with the dictionary on idioms supported our findings; 57.58% of the non-proverb entries we found in *NTC's Dictionary of Proverbs and Clichés* were definitely idioms whereas the remaining 42.42% of the randomly selected number of non-proverb items were not observed in *NTC's American Idioms Dictionary*. The remaining 14, though unobserved in the above idiom dictionary, are mostly idioms based on the criteria offered in this research. This evidence supported the finding of the current research in that the lexicographers had not heeded the boundaries between the fixed expressions and especially those of idioms and proverbs, thus indiscriminately compiling both proverbs and idioms in a single dictionary of proverbs.

This finding leads one to assume that the other dictionaries compiled on the other kinds of formulaic or prefabricated expressions, such as idioms, clichés, and so on, might be suffering from a similar indiscriminate compilation of various entries of prefabricated expressions or be prone to the confusion as to the distinctions between the fixed expressions as the dictionary investigated in this study. On that basis, despite the limited scope of the study, the researcher assumes that the finding might be of some interest to the lexicographers. However, further research is needed to shed more light on the issue.

What is also worth mentioning here is that clichés are applied to any formulaic item ranging from idioms to proverbs, that is, any fixed expression that is overused. People, in fact, do not think when they use clichés. The question that is posed in this regard is whether it is appropriate to juxtapose the terms clichés with proverbs as the title of a dictionary and then include other fixed expressions that are also considered clichés, such as idioms, in a dictionary especially compiled on proverbs. Clichés and proverbs are different categories in themselves and should be considered or discussed from a different perspective. To put it proverbially, it is highly appropriate to avoid comparing oranges and apples.

Given that the needs of dictionary users are so varied, it is recommended that dictionaries should be compiled with the users' needs foremost in mind, as most experts now agree. However, "very few studies actually address this point directly and in sufficient detail (Lew 2011: 1). To take fuller advantage of the offerings of modern lexicography, awareness-raising should be carried out in learners with regard to what the boundaries among the fixed expressions are. Not mixing the proverbs and idioms in a single dictionary can be recommended as an indirect way to do it.

Implementing the suggestions set forth in this paper by the lexicographers will contribute to the primary function of dictionaries, in our case a *good* dictionary of proverbs; the primary function "is often assumed to be that of pro-

viding immediate assistance in comprehension and production problems (Verlinde, Leroyer, and Binon: 2010). And what makes a dictionary good? In answer to this question, Bergenholtz and Gouws (2010: 119) would say very simply: "It is a dictionary that quickly and securely gives the correct answer to a question posed by the user when consulting a dictionary."

To conclude, our investigation has produced evidence, pointing to the need to reconsider the collection of fixed expressions with regard to the general linguistic boundaries between fixed expressions as illustrated in the scheme put forth by Gramley and Pätzold (1992) and strengthened further in this paper. Our results would lead us to agree with Atkins and Rundell (2008: 4) who observe that "There is an enormous body of linguistic theory which has the potential to help lexicographers to do their jobs more effectively and with greater confidence." Therefore, through interaction with other independent fields of study, e.g. linguistics and technology, lexicography as an independent and dynamic field would ascertain the nature of future dictionaries and improve the access to the data, especially in dictionaries on any type of fixed expressions. The users will securely be guided to the specific place where the relevant data is accommodated so as to retrieve the required information. As Bergenholtz and Gouws (2010: 125) state, "Successful retrieval of information has a well-designed access process as a prerequisite." The scheme above might be of some help.

References

Akbarian, I. 2010. A Spice of Classroom: Incorporating Proverbial Expressions in EFL Classes. *Journal of Asia TEFL 7*(1): 221–238.

Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/ New York: Oxford University Press.

Bergenholtz, H. and R. Gouws. 2010. A New Perspective on the Access Process. *Hermes, Journal of Language and Communication Studies 44*: 103–127.

Bertram, A. 1996. *NTC's Dictionary of Proverbs and Clichés*. Lincolnwood, Illinois: NTC Publishing Group.

Brook, G.L. 1981. *Words in Everyday Life*. London: Macmillan.

Cruse, D. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.

Dobrovol'skij, D. and E. Piirainen. 2005. *Figurative Language: Cross-cultural and Cross-linguistic Perspectives*. Amsterdam: Elsevier.

Gramley, S. and K.M. Pätzold. 1992. *A Survey of Modern English*. London: Routledge.

Howard, P. 1984. Cliché. *The State of the Language*: 87-99. London: Hamish Hamilton.

Lew, R. 2011. Studies in Dictionary Use: Recent Developments. *International Journal of Lexicography 24*(1): 1–4.

Luelsdorff, P. 1981. *What is a Cliché?* Trier: LAUT Papers, Series A, no. 85.

Mieder, W. 1988. *English Proverbs*. Stuttgart: Philipp Reclam.

Mieder, W. 1989. *American Proverbs: A Study of Texts and Contexts*. Bern: Peter Lang.

Norrick, N. 1985. *How Proverbs Mean*. Berlin/New York: Mouton.

Prędota, S. 2003. Dictionaries of Proverbs. Van Sterkenburg, P. (Ed.). 2003. *A Practical Guide to Lexicography:* 94-101. Amsterdam: John Benjamins.

Simpson, J. 1985. *The Concise Oxford Dictionary of Proverbs*. New York: Oxford University Press.

Sinclair, J.M. 1991. *Corpus, Concordance, Collocation.* Oxford: Oxford University Press.

Smith, G. 1985. *The English Companion*. Harmondsworth: Penguin.

Spears, R.A. 2000. *NTC's American Idioms Dictionary*. Lincolnwood, Illinois: NTC Publishing Group.

Strässler, J. 1982. *Idioms in English: A Pragmatic Analysis.* Tübingen: Gunter Narr.

Szczepaniak, R. and Lew, R. 2011. The Role of Imagery in Dictionaries of Idioms. *Applied Linguistics* 32(3): 323–347.

Warschauer, M. 2005. Forword. Arnó Macià, E., A. Soler Cervera and C. Rueda Ramos (Eds.). 2005. *Information Technology in Languages for Specific Purposes: Issues and Prospects*: xiii–xv. New York: Springer.

Verlinde, S., P. Leroyer and J. Binon. 2010. Search and You Will Find. From Stand-alone Lexicographic Tools to User Driven Task and Problem-oriented Multifunctional Leximats. *International Journal of Lexicography 23*(1): 1–17.

Yankah, K. 2001. Proverb. Duranti, A. (Ed.). 2001. *Key Terms in Language and Culture*: 201-203. Malden, MA: Blackwell and American Anthropology Association.

Appendix

1. A blessing in disguise (p. 2)
2. A bolt from the blue (p. 2)
3. A castle in Spain (p. 3)
4. A chip off the old block (p. 4)
5. A drop in the bucket/a drop in the ocean (p. 5)
6. A fine kettle of fish (p. 5)
7. A fish out of water (p. 5)
8. A fool's paradise (p. 6)
9. A penny for your thoughts (p. 11)
10. A square peg in a round hole/ trying to fit a square peg into a round hole (p. 13)
11. A wolf in sheep's clothing (p. 15)
12. According to Hoyle (p. 17)
13. According to someone's lights (p. 18)
14. Add insult to injury (p. 18)
15. (All) other things being equal (p. 19)
16. (As) bald as a coot/(As) bald as a baby's backside (p. 24)
17. (As) black as a sweep (p. 24)
18. (As) black as coal/(As) black as pitch (p. 24)
19. (As) blind as a bat (p. 25)
20. (As) bold as brass (p. 25)
21. (As) bright as a button (p. 25)
22. (As) bright as a new pin (p. 25)
23. (As) busy as a beaver (p. 25)
24. (As) busy as a bee/a busy bee (p. 25)
25. (As) busy as a cat on a hot tin roof (p. 25)
26. (As) clean as a hound's tooth (p. 26)
27. (As) clean as a whistle (p. 26)
28. (As) cold as a witch's tit (p. 26)
29. (As) cold as marble (p. 26)
30. (As) common as dirt (p. 26)
31. (As) cool as a cucumber (p. 26)
32. (As) crazy as a loon (p. 27)
33. (As) dead as a doornail/Deader than a doornail (p. 27)
34. (As) dead as the dodo (p. 27)
35. (As) deaf as a post (p. 27)
36. (As) drunk as a lord (p. 27)
37. (As) drunk as a skunk (p. 27)
38. (As) dry as a bone/Bone-dry (p. 27)
39. (As) dull- as ditchwater/(As) dull as dishwater (p. 28)
40. (As) easy as A, B, C (p. 28)
41. (As) fat as a pig (p. 28)
42. (As) fit as a fiddle (p. 28)
43. (As) flat as a board (p. 28)
44. (As) flat as a pancake (p. 28)
45. (As) free as (the) air (p. 28)
46. (As) gaudy as a butterfly (p. 29)
47. (As) gentle as a lamb (p. 29)
48. (As) good as gold (p. 29)
49. (As) graceful as a swan (p. 29)
50. (As) gruff as a bear (p. 29)
51. (As) happy as a clam (p. 29)
52. (As) happy as a lark (p. 29)
53. (As) hard as nails (p. 29)
54. (As) hoarse as a crow (p. 30)
55. (As) hot as fire (p. 30)
56. (As) hungry as a hunter (p. 30)
57. (As) keen as mustard (p. 30)
58. (As) light as a feather/(As) light as air (p. 30)
59. (As) luck would have it (p. 30)
60. (As) mad as a hatter (p. 30)
61. (As) mad as a March hare (p. 31)
62. (As) meek as a lamb (p. 31)
63. (As) merry as a cricket (p. 31)
64. (As) naked as a jaybird (p. 31)
65. (As) neat as a pin (p. 31)
66. (As) nutty as a fruitcake/Nuttier than a fruitcake (p. 31)
67. (As) old as the hills (p. 31)
68. (As) pale as a ghost (p. 32)
69. (As) pale as death (p. 32)
70. (As) patient as job (p. 32)
71. (As) plain as a pikestaff (p. 32)
72. (As) plain as the nose on one's face (p. 32)
73. (As) pleased as Punch (p. 32)
74. (As) poor as a church-mouse (p. 33)
75. (As) proud as a peacock/(As) vain as a peacock (p. 33)
76. (As) pure as the driven snow (p. 33)
77. (As) queer as a three-dollar bill (p. 33)

78. (As) quick as lightning/ swift as lightning (p. 33)
79. (As) quiet as a mouse (p. 33)
80. (As) red as a cherry (p. 33)
81. (As) red as a poppy (p. 34)
82. (As) red as a rose/Rose-red (p. 34)
83. (As) red as a ruby/Ruby-red (p. 34)
84. (As) red as blood/Blood-red (p. 34)
85. (As) regular as clockwork (p. 34)
86. (As) right as a trivet (p. 34)
87. (As) right as rain (p. 34)
88. (As) scarce as hen's teeth/Scarcer than hen's teeth (p. 35)
89. (As) sharp as a razor/Razor-sharp (p. 35)
90. (As) sharp as a tack (p. 35)
91. (As) silent as the dead/(As) silent as the grave (p. 35)
92. (As) silly as a goose/A goose (p. 35)
93. (As) slippery as an eel (p. 35)
94. (As) slow as molasses in January/ Slower than molasses in January (p. 36)
95. (As) smooth as glass (p. 36)
96. (As) snug as a bug in a rug (p. 36)
97. (As) sober as a judge/(As) grave as a judge (p. 36)
98. (As) soft as down (p. 36)
99. (As) soft as velvet/Velvety-soft (p. 36)
100. (As) solid as a rock/Rock-solid (p. 36)
101. (As) sound as a dollar (p. 37)
102. (As) sour as vinegar (p. 37)
103. (As) steady as a rock/Rock-steady (p. 37)
104. (As) stiff as a poker (p. 37)
105. (As) still as death (p. 37)
106. (As) straight as an arrow (p. 37)
107. (As) strong as a horse/(As) strong as an ox (p. 37)
108. (As) strong as a lion (p. 38)
109. (As) stubborn as a mule/(As) obstinate as a mule (p. 38)
110. (As) sure as death (p. 38)
111. (As) sweet as honey/Sweeter than honey (p. 38)
112. (As) sweet as sugar (p. 38)
113. (As) swift as an arrow (p. 38)
114. (As) swift as the wind/Like the wind (p. 38)
115. (As) swift as thought (p. 39)
116. (As) thick as thieves (p. 39)
117. (As) tight as a drum (p. 39)
118. (As) tight as a tick (p. 39)
119. (As) tough as a (shoe) leather (p. 39)
120. (As) true as steel (p. 39)
121. (As) ugly as a toad (p. 39)
122. (As) ugly as sin (p. 40)
123. (As) warm as toast (p. 40)
124. (As) weak as a baby (p. 40)
125. (As) white as a sheet (p. 40)
126. (As) white as snow/Snow-white (p. 40)
127. (As) wise as Solomon/the wisdom of Solomon (p. 40)
128. At sixes and sevens (p. 41)
129. At the drop of hat (p. 41)
130. At this point in time (p. 42)
131. Babe in the woods (p. 43)
132. Back to the salt mine (p. 43)
133. Bark up the wrong tree (p. 44)
134. Batten down the hatches (p. 44)
135. Be one's own man/Be one's own master (p. 44)
136. Be one's own worst enemy (p. 45)
137. Beard the lion in his den/Beard someone in his den (p. 45)
138. Before you can say Jack Robinson (p. 46)
139. Better left unsaid (p. 47)
140. Between the devil and the deep blue sea/Between a rock and a hard place (p. 48)
141. Between you and me and the bedpost/Between you and me and these four walls (p. 48)
142. Beyond a shadow of a doubt (p. 48)
143. Born on the wrong side of the blanket (p. 50)
144. Born with a silver spoon in one's mouth (p. 51)
145. Bright-eyed and bushy-tailed (p. 51)
146. Bring home the bacon (p. 52)
147. Build castles in the air (p. 52)

148. Burn one's bridges (behind one) (p. 52)
149. Burn the candle at both ends (p. 53)
150. Burn the midnight oil (p. 53)
151. Bury the hatchet (p. 53)
152. Butter wouldn't melt (in someone's mouth) (p. 53)
153. Buy a pig in a poke (p. 54)
154. By the sweat of one's brow (p. 54)
155. By word of mouth (p. 54)
156. Call a spade a spade (p. 55)
157. Cannot call one's soul one's own (p. 56)
158. Cannot hit the broad side of a barn (p. 56)
159. Cannot see the wood for the trees (p. 56)
160. Can't hold a candle to someone (p. 56)
161. Carry coals to Newcastle (p. 56)
162. Cast the first stone (p. 57)
163. Catch-as-catch-can (p. 57)
164. Close enough for government work (p. 59)
165. Come on like gangbusters (p. 60)
166. Come out smelling like a rose (p. 60)
167. Come up roses (p. 60)
168. Come within an inch of something/Come within a hair's breadth of something (p. 60)
169. Conspicuous by one's absence (p. 61)
170. Cool, calm, and collected (p. 62)
171. Cost a pretty penny (p. 62)
172. Crazy like a fox (p. 63)
173. Cry all the way to the bank (p. 64)
174. Damn someone with faint praise (p. 65)
175. Doesn't have the sense God gave geese (p. 69)
176. Drink like a fish (p. 73)
177. Dry as dust (p. 73)
178. Duck soup (p. 73)
179. Eat someone out of house and home (p. 76)
180. Enough to keep body and soul together (p. 77)
181. Escape by the skin of one's teeth/ By the skin of one's teeth (p. 77)
182. Fall between two stools (p. 82)
183. Far from the madding crowd (p. 83)
184. Feel something in one's bones (p. 83)
185. Fiddle while Rome burns (p. 84)
186. Fight fire with fire (p. 84)
187. First see the light of day (p. 86)
188. Get down to brass tracks (p. 91)
189. Get it on the ground floor (p. 91)
190. Get it straight from the horse's mouth (p. 91)
191. Get up on the wrong side of the bed (p. 92)
192. Give credit where credit is due (p. 92)
193. Give someone a dose of his own medicine (p. 92)
194. Give someone the benefit of the doubt (p. 93)
195. Give the shirt off one's back (p. 94)
196. Go from bad to worse (p. 94)
197. Go like the wind (p. 94)
198. Go over/through something with a fine-tooth comb (p. 94)
199. Going to hell in a hand-basket (p. 94)
200. Good riddance to bad rubbish! (p. 96)
201. Grist for the mill/someone's mill (p. 97)
202. (Has the) cat got your tongue? (p. 100)
203. Haul/Rake someone over the coals (p. 100)
204. Have a bee in one's bonnet (p. 100)
205. Have a bone to pick (with someone) (p. 100)
206. Have a chip on one's shoulder (p. 101)
207. Have a finger in every pie (p. 101)
208. Have an ace up one's sleeve (p. 101)
209. Have an axe to grind (p. 101)

210. Have bats in the belfry/Batty crazy/Bats (p. 101)
211. Have better (or other) fish to fry (p. 102)
212. Have one foot in the grave (p. 102)
213. Have second thoughts (p. 102)
214. Have the courage of one's convictions (p. 102)
215. Have too many irons in the fire (p. 102)
216. Have too much of a good thing (p. 103)
217. Here's mud in your eye (p. 109)
218. Hide one's light under a bushel (p. 109)
219. Hit the nail on the head (p. 109)
220. Hoist with one's own petard (p. 110)
221. I would not touch it with a ten-foot pole (p. 113)
222. If the truth were known (p. 116)
223. If (the) worst comes to (the) worst (p. 116)
224. In (at) one ear and out (of) the other (p. 120)
225. In this day and age (p. 121)
226. It is (all) Greek to me (p. 126)
227. It's six of one, half a dozen of another (p. 128)
228. Keep a stiff upper lip (p. 130)
229. Keep one's nose to the grindstone (p. 131)
230. Keep up with the Joneses (p. 131)
231. Kill two birds with one stone (p. 132)
232. Know which side one's bread is buttered on (p. 132)
233. Laugh all the way to the bank (p. 134)
234. Laugh out of the other side of one's mouth (p. 134)
235. Leave no stone upturned (p. 135)
236. Let the cat out of the bag (p. 136)
237. Life is just a bowl of cherries (p. 137)
238. Like a bat out of hell (p. 138)
239. Like a bull in a china shop (p. 138)
240. Like a bump on a log (p. 139)
241. Like death warmed over (p. 139)
242. Like greased lightning (p. 139)
243. Like two peas in a pod (p. 140)
244. Look for a needle in a haystack (p. 142)
245. Make a clean breast of it (p. 144)
246. Make a long story short (p. 144)
247. Make a mountain out of a mole-hill (p. 144)
248. Make a virtue of necessity (p. 145)
249. Make one turn (over) in one's grave (p. 145)
250. Make the best of a bad job (p. 146)
251. Make one's p's and q's (p. 149)
252. Mind your own business (p. 149)
253. Money burns a hole in someone's pocket (p. 150)
254. Much ado about nothing (p. 152)
255. Neither rhyme nor reason (p. 154)
256. Never a dull moment (p. 154)
257. Nineteen to the dozen (p. 157)
258. No sooner said than done (p. 158)
259. Not able to get something for love or money (p. 158)
260. Not able to make head or tail of something (p. 158)
261. Not enough room to swing a cat (p. 158)
262. Not one's cup of tea (p. 159)
263. Not to know someone from Adam (p. 159)
264. Not to let the grass grow under one's feet (p. 159)
265. Not worth a hill of beans (p. 159)
266. Not worth the paper it is written on (p. 160)
267. Nothing to boast about (p. 161)
268. Nothing to write home about (p. 161)
269. Once in a blue moon (p. 163)
270. One's heart is in one's mouth/With one's heart in one's mouth/Have one's heart in one's mouth (p. 166)
271. Paddle one's own canoe (p. 169)
272. Pay the piper (p. 169)
273. Pay through the nose (p. 169)
274. Pour oil on troubled waters (p. 171)
275. Pull oneself up by the bootstraps/Pull oneself up by one's bootstraps (p. 173)

276. Pull the wool over someone's eyes. (p. 173)
277. Put one's best foot forward (p. 173)
278. Put one's foot down (p. 174)
279. Put one's foot in it (p. 174)
280. Put one's shoulder to the wheel (p. 174)
281. Put something on the line (p. 174)
282. Put that in your pipe and smoke it (p. 174)
283. Put your money where your mouth is (p. 175)
284. Rain cats and dogs (p. 176)
285. Read between the lines (p. 176)
286. Read someone like a(n open) book (p. 177)
287. Ring down the curtain (p. 177)
288. Rob Peter to pay Paul (p. 178)
289. Save for a rainy day (p. 179)
290. Seize the bull by the horns/Take the bull by the horns (p. 180)
291. Separate the men from the boys (p. 181)
292. Separate the sheep from the goats (p. 181)
293. Separate the wheat from the chaff (p. 181)
294. Set a thief to catch a thief (p. 182)
295. Share and share alike (p. 182)
296. Ships that pass in the night (p. 182)
297. Sleep like a log (p. 184)
298. Sleep like a top (p. 184)
299. Slow but sure/Slowly but surely (p. 185)
300. Someone cannot see beyond the end of his nose (p. 185)
301. Someone is not out of the woods yet (p. 185)
302. Someone puts his pants on one leg at a time (p. 186)
303. Someone will get his (or hers) (p. 186)
304. Stand the test of time (p. 187)
305. Strain at gnats and swallow camels (p. 188)
306. Take the bit between one's teeth (p. 190)
307. Take the bitter with the sweet (p. 190)
308. Take the rough with the smooth (p. 191)
309. Take the wind out of someone's sails (p. 191)
310. Tar someone with the same brush (p. 191)
311. The biggest frog/toad in the puddle (p. 194)
312. The devil to pay/Someone has the devil to pay/There will be the devil to pay (p. 198)
313. The fat is in the fire (p. 200)
314. The left hand doesn't know what the right hand is doing (p. 203)
315. The less said (about something), the better (p. 203)
316. The lesser of two evils (p. 203)
317. The long and the short of it (p. 203)
318. The last straw/The straw that broke the camel's back (p. 208)
319. The time is ripe (p. 208)
320. The whole ball of wax (p. 210)
321. Tied to one's mother's apron strings/Cut the apron strings (p. 218)
322. Turn back the clock (p. 222)
323. Water over the dam/Water under the bridge (p. 226)
324. What in (the) Sam Hill (p. 227)
325. Wild horses couldn't drag someone away from something (p. 233)
326. Your guess is as good as mine (p. 239)

What is a Dictionary?

Henning Bergenholtz, *Department of Afrikaans and Dutch, University of Stellenbosch, Stellenbosch, Republic of South Africa and Centre for Lexicography, Aarhus University, Aarhus, Denmark (hb@asb.dk)*

Abstract: By comparing different definitions in dictionaries and in theoretical papers we know that lexicographers have very different opinions on the topic: "What is a dictionary?" We see too that lexicographic tools with the same type of content sometimes are called dictionary, sometimes encyclopedia, sometimes lexicon. In reality it is not a large problem if the information in a certain tool is of a high quality as to the intended function. Nevertheless I will give not one but two additional definitions of two different types of information tools. There will also be a discussion on the relevance for the concept of dictionary in connection with the research registration in universities.

Keywords: DICTIONARY, LEXICON, ENCYCLOPAEDIA, LEXICOGRAPHY, INFORMATION TOOL, LEXICOGRAPHIC TOOL, RESEARCH REGISTRATION, LEXICOGRAPHIC DATABASE, MONOFUNTIONAL DICTIONARY, POLYFUNCTIONAL DICTIONARY

Opsomming: Wat is 'n woordeboek? Deur verskillende definisies in woordeboeke en teoretiese artikels te vergelyk, weet ons dat leksikograwe baie verskillende opinies het oor die onderwerp: "Wat is 'n woordeboek?" Ons sien ook dat leksikografiese hulpmiddels met dieselfde soort inhoud soms woordeboeke genoem word, soms ensiklopedieë en soms leksikons. In die praktyk is dit nie 'n groot probleem nie indien die inligting in 'n sekere hulpmiddel van hoë kwaliteit is met betrekking tot die beoogde funksie. Tog sal ek nie een nie maar twee bykomende definisies gee van twee verskillende soorte inligtingshulpmiddels. Daar sal ook 'n bespreking wees oor die relevansie van die konsep woordeboek met betrekking tot die registrasie van navorsing by universiteite.

Sleutelwoorde: WOORDEBOEK, LEKSIKON, ENSIKLOPEDIE, LEKSIKOGRAFIE, INLIGTINGSHULPMIDDEL, LEKSIKOGRAFIEHULPMIDDEL, NAVORSINGSREGISTRASIE, LEKSIKOGRAFIESE DATABASIS, MONOFUNKSIONELE WOORDEBOEK, MEERFUNKSIONELE WOORDEBOEK

1. Do dictionary users know what a dictionary is?

Any native speaker of Danish will normally immediately know what is meant if someone says: *Luk døren!* (close the door!) or *Du har hul i din bluse* (there's a hole in your blouse). The addressee knows what *dør*, *hul* and *bluse* mean and will therefore not consult a dictionary in order to understand these sentences. It's a different story if someone says: *Det vil jeg kalde en eufemistisk omskrivning* (I would call that a euphemistic description). In this case many would ask: "What

does *euphemistic* mean?" Or they would look up *euphemistic* in a dictionary. Such consideration of user requirements was the main reason why, in the first editions, the Nudansk Ordbog (a well-known Danish dictionary) did not include definitions for ordinary words if these could be expected to be immediately understood by all potential users whose mother tongue is Danish. Consequently this dictionary contained no definitions for words such as *ordforråd* (vocabulary), *ordne* (arrange), *ordre* (order) and *ordsprog* (proverb), but the word *ordbog* (dictionary) was not defined either.

However self-evident the argument about the assumed common language knowledge and language ability may be, the question whether a word is known to everybody always remains a problematic issue. For example, *ordbog* (dictionary) was not defined — but *leksikon* (lexicon) was. True, in this case one word is a "Danish" word and the other a foreign word. But is *ordbog* really a word understood by all, and *leksikon* not? The answer can be both yes and no. If we answer "yes", then we are saying that all native speakers will understand if we say: *Please pass me the big blue dictionary over there on the table*. If the answer is "no", we imply that hardly any ordinary language user is able to define the word *ordbog* exactly or accurately. Ordinary language users also do not know exactly whether a lexicon and a word list are the same as a dictionary, or what exactly constitutes the difference (if any) between these three terms. And when such ordinary mother-tongue speakers say they know the answer, it will not be difficult to discover that different people offer quite different definitions. Each one will nevertheless insist that their definition is correct. In this respect they are no different from the experts, nor from the descriptions of the word *ordbog* — which are as numerous as the dictionaries themselves and often problematic or even downright incorrect to boot.

I will return to such problematic definitions later; for the moment, here is the dictionary article from the Nudansk Ordbog from the edition that defines all words:

> **ordbog** en bog med ord der er systematisk ordnet, og med oplysning om fx ordenes stavemåde, ordklasse, bøjning og betydning el. deres oversættelse til et andet sprog (= **dictionary** a book with words in a systematic order and with information about the spelling, word class, inflexion and meaning of the words or their translation to another language) (Nudansk Ordbog 2005).

This definition of *dictionary* is one of the better ones. But it is not optimal. To start with, the phrase "a book with words in a systematic order" is already unclear. Do not all books contain words in a systematic order (so that they correspond to the system that is used to represent the sounds of a language in letters)? It is probably rather a book **about** words or combinations of words. But what is a word? Is a name a word? After all, you can also find names of cities, countries and important people in dictionaries. But if defined in this way, a telephone guide would also be a dictionary, and that would not fit the normal

meaning of *dictionary*. The most serious error in the definition quoted, however, is that it excludes most of the electronic dictionaries. The words and combinations of words described are most certainly arranged systematically in a database, but this system has nothing to do with what the user sees. In many cases, the user sees only a single dictionary article when he has entered a search term and given the "search" command. And if only a single dictionary article is displayed, no system can be discerned.

2. Is it important to know what a dictionary is?

In practice, the title of a lexicographic reference work does not give an unequivocal indication of the type of reference work it is. For example, Politikens Musikordbog (Politiken's Music Dictionary) of 1995 was called a music lexicon by the author who submitted the manuscript for publication. But the publisher changed the title, arguing that the book was being published by the publisher's dictionary department. Had it been published by the textbook department instead, they explained, the book would have been entitled "Lexicon". We cannot say whether the change of title from "lexicon" to "dictionary" had a positive or negative effect on sales. We don't think so. Similarly, there are many reference works with "dictionary" in the title which nobody would call dictionaries, and which do not fit a known definition of "dictionary" either. Moreover, there are many works without the words "dictionary", "lexicon", "word list", "encyclopaedia" or the like in the title, but which must undoubtedly be classified as such. In this regard I would like to refer to an article by Pálfi (2011), which has the meaningful title: "On dictionaries which aren't and non-dictionaries which are". It is a fine article, but although — like many thousands or more on the definition of "dictionary" — it is scientifically interesting, it is neither of immediate practical nor social relevance. In this article I would like to propose a scientifically founded definition and, in addition, discuss practical and social consequences of its terminological application. First of all, it is important to know how a dictionary can be cited. For example, can the common name under which several dictionaries are being sold be used as a title in the references? This applies in the case of the "Music dictionary", Den Danske Musikordbog, which is sold as an electronic dictionary of the Verlag Ordbogen A/S. Actually, this title comprises four dictionaries. They also differ in terms of size; the four hard copies, due to appear in 2012, range from 58, 160, 442 to 444 pages. In the case of the electronic versions, the user receives clear instructions if he clicks on the link "About the dictionary". Here he learns that there are four dictionaries, even though they are not called that in the title, and clear instructions are provided how they can be cited:

> Bergenholtz, Inger i samarbejde Henning Bergenholtz. Database: Richard Almind og Martin Gyde Poulsen: *Betydning af musikudtryk.* Odense: Ordbogen.com 2011. (www.ordbogen.com). (ISBN 978-87-788-2555-1)

Bergenholtz, Inger i samarbejde Henning Bergenholtz. Database: Richard Almind og Martin Gyde Poulsen: *Viden om musikudtryk.* Odense: Ordbogen.com 2011. (www.ordbogen.com). (ISBN 978-87-788-2299-4)

Bergenholtz, Inger i samarbejde Henning Bergenholtz. Database: Richard Almind og Martin Gyde Poulsen: *Find et musikudtryk.* Odense: Ordbogen.com 2011. (www.ordbogen.com). (ISBN 978-87-788-2556-8)

Bergenholtz, Inger i samarbejde Henning Bergenholtz. Database: Richard Almind og Martin Gyde Poulsen: *Fremmedordbog med musikudtryk*. Odense: Ordbogen.com 2011. (www.ordbogen.com). (ISBN 978-87-788-2557-5)

Thus a "dictionary" is not always a "dictionary". Accordingly, we will subsequently propose two different definitions of the term "dictionary":

1. When we refer to the Music Dictionary, "dictionary" means a collection of dictionaries, each of which has completely different dictionary articles.

2. For the individual dictionaries as referred to above, "dictionary" is used in the conventional sense, as shown in the next chapter of this article. That these definitions are often problematic and imprecise is less important in this context; the user senses that this is a dictionary and that it should be cited as shown above.

Far more socially relevant, and certainly relevant to the university, is the use of "dictionary" in the research registration, which all researchers must submit. On the one hand the researchers must document that they are fulfilling their research obligations. On the other, the individual universities have an interest in these research results being submitted because the ministry pays the respective university for each newly published scientific contribution or book, provided that it is a scientific dictionary. Until recently, it was customary to extract only a single dictionary from a database, but in principle 5, 10, 20, 100 or even thousands of dictionaries can be extracted from one and the same database. We have actually done this in several cases at the Centre for Lexicography in Denmark. As mentioned earlier, four dictionaries were extracted from a music database; from an accounting database containing Danish, English and Spanish accounting terms, we have to date published 16 different accounting dictionaries, and this number will grow to 27 in a few years. Whether the publisher sells these as a package or separately is of no importance. But for the registration of the research it does matter. When we register, we have so far submitted 16 different accounting dictionaries which the ministry must recognise. Certainly, the ministry would prefer a definition of "dictionary" which would result in fewer editions.

Something similar would apply if one received a library fee for electronic dictionaries. Such a fee is paid, for example, in Denmark and Germany as a

function of the number of books bought by the respective libraries. Oddly enough it is paid only for printed dictionaries, even though libraries buy more electronic than printed dictionaries. However, if one were to be paid the library fee for electronic dictionaries, the same dilemma would arise as with the allocation of ministerial funds to the universities. I propose emphatically that electronic dictionaries should be dealt with in the same way as printed dictionaries, in other words when the works are completely different, a separate royalty or library fee should be paid.

3. Existing definitions of "dictionary"

If all the scientific contributions related to this problem were gathered together, there would be enough material for several voluminous tomes. I will not quote from the multitude of scientific contributions here, as their definitions do not differ from those I have found on the internet and in some printed dictionaries. Below I will quote a small selection from the many thousands of definitions and mention some points of criticism against each of them, more or less in the same way as those raised against the definition in the Nudansk Dictionary mentioned above. We will see that this definition — despite its problematic imprecision — is better and clearer than most of the other definitions. I did not take particular trouble to look for really good or really poor definitions. I did a Google search with the search string "What is a dictionary?" and then took some of the definitions that appeared on the first three pages. The following discussion therefore does not offer a representative picture of all definitions on the internet. But then that was not the idea either. I want to show what existing definitions reflect as important and in what way they contain problematic or even incorrect statements. Against this background, I want to suggest a definition that is applicable not only to printed or certain types of dictionaries, but takes into account that a dictionary is an information tool and that this tool can be an electronic or printed dictionary. This the first definition does not do; it is brief and easy to understand, but says almost nothing:

(1) Dictionaries are alphabetical lists of words or entries. They differ in the kind and in the volume of information they hold. (http://www.lib.uct.ac.za/infolit/dict1.htm, found November 2011)

The statement that a dictionary is an alphabetical list is, of course, not correct. It is incorrect for at least two reasons: (A) There are dictionaries with systematically arranged dictionary articles; we call this a systematic macrostructure. This means that the dictionary articles are actually not alphabetic; as in a thematic dictionary, they are arranged in such a way that specific topics are presented in a thematic order. Such dictionaries usually have an alphabetical index which enables the user to look up the dictionary articles more easily. (B) The database of an electronic dictionary is specifically not arranged in alphabetic order at all,

but is provided with its own address internally in the database. This may be a number, with each new dictionary article being given a new serial number. But this numbering has nothing to do with the alphabet. It is an internal allocation in a database which has nothing to do with the presentation of the dictionary articles. More important is that the user is usually looking for a particular word and that he receives as the result exactly one and only one dictionary article. If he receives two or more dictionary articles as a result, these dictionary articles may be arranged alphabetically, but they do not have to be. But even if some of the dictionary articles shown are in alphabetical order, the definition is still not correct. After all, the user sees only a small section of the dictionary. The definition is also inadequate because it would turn even a telephone directory into a dictionary. You could call it that, but this would conflict severely with the present terminological language usage and run counter to any existing understanding of "dictionary". Nevertheless, there are many variants of this type of definition:

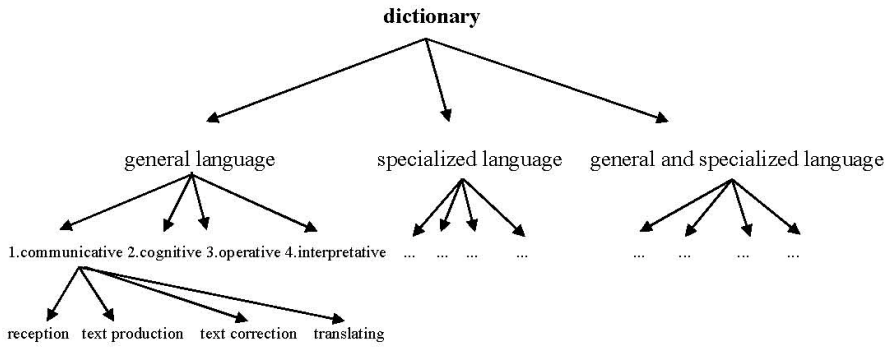
(2) a reference book containing an alphabetical list of words with information about them (http://wiki.answers.com/Q/What_is_the_biggest_dictionary_definition, found November 2011)

All the arguments against definition (1) could be repeated here. It can be described as the shortest and probably also as the most primitive. Nevertheless, definitions such as (1) and (2) and the definition below are those found most commonly. Here an attempt is made to define a dictionary by listing the necessary item types:

(3) A definition for dictionary is a reference that tells you the meanings, parts of speech, sometimes a sentence using the word, and how to pronounce the word. (http://wiki.answers.com/Q/What_is_the_biggest_dictionary_definition, found November 2011)

Such a definition excludes many information tools we normally consider to be dictionaries: a spelling dictionary normally does not contain definitions, no specimen sentences and pronunciation items; according to this definition, this is not a dictionary. A frequency dictionary, which lists only the word form and its frequency, would also not be worthy of the name; neither would a bilingual dictionary, which in most cases has no meaning items. The reason why this widespread and very inaccurate definition has been repeated in so many variants may be that what so many have said and written before is simply repeated blindly. To be a little more positive, one could regard this as the definition of a monolingual polyfunctional dictionary. Or, in other words: It is the type of dictionary that is often regarded as THE dictionary, but it is really a very special dictionary — the kind that linguists and linguists claiming to be lexicographers view as the sole object. It is a very narrow view of a dictionary, which is aimed only at solving communication problems (text production, text reception

and translation), and in this case not even at translation. Such dictionaries are also called "language dictionaries". Many lexicographers regard only dictionaries indicated as (1) for communicative dictionaries in the illustration below as the object, not the other eleven types of dictionaries:



The definition thus aims to fulfil non-cognitive functions in particular, i.e. information tools which aim to transmit knowledge. Dictionaries of this kind are often called lexicons or encyclopaedias. This type accounts for almost two-thirds of all lexicographic works. We will attend to possible differences between the different names of lexicographic information tools later. First we want to quote one more example with a variant of the third definition:

(4) a book consisting of an alphabetical list of words of a language showing their forms and meanings and etymologies (http://wiki.answers.com/Q/What_is_the_definition_of_a_dictionary, found February 2011)

Almost all the critical arguments against definition (3) can be advanced against this definition as well. Yet it is interesting that only the meaning and etymology item types are deemed decisive for the definition of a dictionary here. Besides this, it is also striking that none of the definitions quoted so far even hint that a dictionary is produced as an aid to people having specific information needs. This also applies to the next definition:

(5) A dictionary is a book of words of a particular language and their accepted definitions, origins, parts of speech, pronunciation, spelling and in some cases a sample of their use. Depending on the age and target audience, it may also contain cultural slang and/or other non-traditional words as well. A "language translation dictionary" lists the words of one language and their equivalent words in another language. (http://wiki. answers.com/Q/What_is_a_dictionary, found November 2011).

When it comes to the pinch, one could say that definition (5) takes into account

in one point that a dictionary is an information tool, since "language translation dictionary" is mentioned. But even so, it would in this case be a relatively basic translation dictionary, since only equivalence items are mentioned; no translated collocations, no translated examples and no contrast items. But still. Also excluded in this case is the type that is usually called a biscopal dictionary, a dictionary with words from language A translated into language B and simultaneously from language B into language A. It could also be argued that a biscopal dictionary is not a single dictionary, but two dictionaries. But that is not common practice. It should also be noted that in this definition the expression "a book of words" is used. That's not what it is, of course; it is "a book of dictionary articles". There are lemmas for which different data are provided. The misunderstanding could also be due to the fact that the English term "entry" is used to refer to the lemma as well as to the "articles" in a dictionary. But this does not fully explain the misunderstanding. A dictionary entry ("article") is never called a "word" in English. We will quote one last example, but with the observation that this example, as well as the preceding five examples, is quite typical of all existing definitions:

(6) A dictionary is a reference book that focuses on defining words and phrases, including multiple meanings. The most frequently used dictionary is a language dictionary that includes the majority of frequently used words in a language. (http://www.wisegeek.com/what-is-a-dictionary.htm, found November 2011)

Everything that has been said about (4) and (5) can be repeated with regard to this definition. In addition, there is the statement that a dictionary should always only contain words which are used frequently. This is not correct, of course. There are some dictionaries that do this, e.g. learners' dictionaries, school dictionaries and other general-language dictionaries with fewer than 20,000 lemmas which contain almost only frequently used words. But specialist dictionaries do not do this. For example, a music dictionary does not contain the words most frequently used in music texts; most of the words in texts about music are everyday words. Music dictionaries do not only list the most frequent music terms, but also systematically include those terms which are important in the language of music. Whether these are the most frequent in music texts is not of primary importance. In general-language dictionaries too, the user frequently gets information about less commonly used words, and such words are also found in larger common-language dictionaries. And with good reason, as rarely used words can more often cause problems of reception than frequently used words. Another problem is that this definition focuses exclusively on the idea that a dictionary provides meanings. This excludes bilingual dictionaries, in fact all other dictionaries without items of meaning.

In order to show that somewhat better definitions can also be found on the internet, a final example is quoted. However, such "better" definitions are rare. Although the phrase "a collection of words" is also used in the definition below,

it is also mentioned (besides the usual list of certain item types) that a dictionary is a tool intended to help the users when they have certain information needs:

(7) A **dictionary** (also called a **wordbook**, **lexicon** or **vocabulary**) is a collection of words in one or more specific languages, often listed alphabetically, with usage information, definitions, etymologies, phonetics, pronunciations, and other information; or a book of words in one language with their equivalents in another, also known as a lexicon. According to Nielsen (2008) a dictionary may be regarded as a lexicographical product that is characterised by three significant features: (1) it has been prepared for one or more functions; (2) it contains data that have been selected for the purpose of fulfilling those functions; and (3) its lexicographic structures link and establish relationships between the data so that they can meet the needs of users and fulfil the functions of the dictionary (Wikipedia, found November 2011).

At the end of my article I will contribute my own definition — two, in fact — showing that "dictionary" has two different meanings. First, however, I would like to summarise two points of uncertainty which prompted the criticism of the definitions quoted (and those not quoted) above. Furthermore, there are some points the quoted definitions do not mention, but which do matter. This also applies to the first point, which does not turn up in any of the definitions quoted:

(1) How many dictionary articles must a dictionary have before it can be called a dictionary? There are many printed dictionaries with fewer than 1,000 dictionary articles. On the internet I found a dictionary with only 157 dictionary articles. But how low can the number be? I would say: two. Admittedly, I don't know any dictionaries that small, but in principle this should be the limit if one argues that a dictionary must contain a collection of individual dictionary articles.

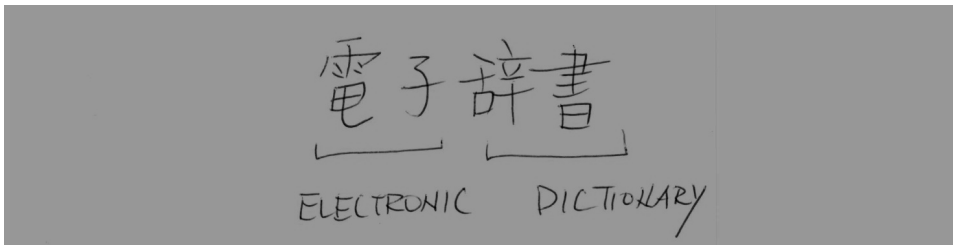
(2) A second issue was referred to indirectly in definition (7), which says: "dictionary, also called a wordbook, lexicon or vocabulary". This is also the solution I would propose. I see "dictionary" as a general term for all kinds of lexicographic reference works — as communication, cognitive as well as information tools which aim to fulfil both main functions. It is right that there is a trend towards terminological use, but not more than a trend. In Den Danske Netordbog (2012) under the dictionary article *ordbog* (dictionary), we describe this trend as follows in a note (translated from Danish):

> A distinction can be drawn between lexicographic reference works which offer assistance in case of problems with a concrete text and other reference works where general or specific knowledge can be

obtained. The first type is often called a dictionary and the second type a lexicon or encyclopaedia, but there is no generally agreed distinction between and naming of these two types.

(3) Are there dictionaries in dictionaries? This question has several aspects. What is interesting is that the question is not asked in any of the definitions of "dictionary" found or otherwise known to me. But dictionaries have many more parts than the lists of lemmas, namely outer texts. This is known and has been described in detail. Some of these outer texts contain brief dictionary articles which correspond directly to the definition of "dictionary", e.g. an alphabetical list of grammatical terms with short explanations of the meanings of these terms. Similarly, lists of currency units or masses and weights also fit any current definition of "dictionary". In other words, you can have several dictionaries within one dictionary. Granted, such outer texts are not usually called "dictionaries". But even biscopal dictionaries are normally referred to as not two dictionaries, but one dictionary if they are bound in the same printed volume, e.g. in a dictionary with two lists of lemmas: Danish–English and English–Danish. Such a dictionary therefore often has a double-barrelled name: Danish–English and English–Danish Dictionary.

That a dictionary can contain a dictionary becomes even clearer when the tradition followed in Japan and other Asian countries is examined. There one can buy a small microcomputer with a number of individual dictionaries installed on it; these are often previously printed dictionaries which have been digitised. Such a computer with up to 400 individual dictionaries is also called a dictionary or, more accurately, an electronic dictionary in Japanese:



4. Proposed definition of "dictionary"

After all the criticisms of existing definitions of "dictionary", it might appear somewhat presumptuous to propose one. I will propose one nevertheless — and not just one, but two, as it has become clear that "dictionary" has not just one, but two distinctly different meanings. It is crucial to explain that a dictionary is an information tool that must satisfy specific requirements:

Dictionary

1. Lexicographic reference work containing dictionary articles related to individual topics or elements of language, and possibly several outer texts as well, which can be consulted if someone needs assistance with text reception, text production or translation or would simply like to know more about a word, part of a word or a combination of words.

2. Lexicographic reference work consisting of several dictionaries, each of which corresponds to the definition of an individual dictionary, i.e. a reference work containing dictionary articles related to individual topics or elements of language, and possibly several outer texts as well, which can be consulted if someone needs assistance with text reception, text production or translation or would simply like to know more about a word, part of a word or a combination of words.

5. References

Bergenholtz, Inger i samarbejde Henning Bergenholtz. Database: Richard Almind og Martin Gyde Poulsen: *Betydning af musikudtryk.* Odense: Ordbogen.com 2011. (www.ordbogen.com).

Bergenholtz, Inger i samarbejde Henning Bergenholtz. Database: Richard Almind og Martin Gyde Poulsen: *Find et musikudtryk.* Odense: Ordbogen.com 2011. (www.ordbogen.com).

Bergenholtz, Inger i samarbejde Henning Bergenholtz. Database: Richard Almind og Martin Gyde Poulsen: *Fremmedordbog med musikudtryk*. Odense: Ordbogen.com 2011. (www.ordbogen.com).

Bergenholtz, Inger i samarbejde Henning Bergenholtz. Database: Richard Almind og Martin Gyde Poulsen: *Viden om musikudtryk.* Odense: Ordbogen.com 2011. (www.ordbogen.com).

Den Danske Netordbog. 2012. = Bergenholtz, Henning under medvirken af Filip Odgaard Bodilsen, Kathrine Brosbøl Eriksen, Helene R. Gudmann, Stine Busk Hedegaard, Helene Halkjær Jensen, Jesper Halkjær Jensen, Christian Kjølhede, Anders Kroman Liin, Jane Nguyen, Heidi Agerbo Pedersen. 2012. Database: Richard Almind, Rasmus Theodor Styrk, Peter Christensen: *Den Danske Netordbog*. Odense: Ordbogen.com. (www.ordbogen.com) (7th edition)

Nudansk Ordbog. 2005. = *Politikens Nudansk Ordbog*. 2005. 19th edition. Copenhagen: Politikens Forlag.

Pálfi, Loránd-Levente. 2011. Om ordbøger, som ikke er ordbøger, og ikke-ordbøger, som er or ordbøger. Arboe, Torben and Inger Schoonderbeek Hansen (Eds.). 2011. *Jysk, ømål, rigsdansk mv. Studier i dansk sprog med sideblik til nordisk og tysk — Festskrift til Viggo Sørensen og Ove Rasmussen udgivet på Viggo Sørensens 70-års dag den 25. februar 2011:* 325-335. Aarhus: Aarhus University.

Politikens musikordbog = Inger Bergenholtz. 1995. *Politikens musikordbog*. Copenhagen: Politikens Forlag.

What is Lexicography?

Henning Bergenholtz, *Department of Afrikaans and Dutch, University of Stellenbosch, Stellenbosch, Republic of South Africa and Centre for Lexicography, Aarhus University, Aarhus, Denmark (hb@asb.dk)*
and
Rufus H. Gouws, *Department of Afrikaans and Dutch, University of Stellenbosch, Stellenbosch, Republic of South Africa (rhg@sun.ac.za)*

Abstract: Within the field of lexicography there are numerous differences when it comes to the interpretation of the term *lexicography* and differences in determining the nature, extent and scope of this term. Although it is widely accepted that lexicography consist of two components, i.e. theoretical lexicography and the lexicographic practice, different definitions of *lexicography* give no unambiguous reflection of this distinction and of the individual components. This paper looks at some prevailing diverse uses and interpretations of the word *lexicography.* This is followed by proposals to ensure a transformative, unified and comprehensive interpretation of this concept.

Keywords: INDEPENDENT DISCIPLINE, INFORMATION TOOL, LEXICOGRAPHER, LEXICOGRAPHIC PRACTICE, LEXICOGRAPHIC THEORY, LEXICOGRAPHIC TOOL, LEXICOGRAPHY, LEXICOLOGY, METALEXICOGRAPHY, SCIENTIFIC LEXICOGRAPHY, TERMINOGRAPHY

Opsomming: Wat is leksikografie? In die leksikografie bestaan daar talle verskille met betrekking tot die interpretasie van die term *leksikografie* asook met betrekking tot 'n vasstelling van die aard, omvang en bestek van hierdie term. Alhoewel dit wyd aanvaar word dat leksikografie uit twee komponente bestaan, naamlik 'n teoretiese en 'n praktiese komponent, bied onderskeie definisies van *leksikografie* geen ondubbelsinnige beskrywing van hierdie verskil en van die onderskeie komponente nie. Hierdie artikel gee aandag aan enkele bestaande maar uiteenlopende gebruike en interpretasies van die woord *leksikografie*. Daarna volg voorstelle ter versekering van 'n transformatiewe verenigde en omvattende interpretasie van hierdie begrip.

Sleutelwoorde: INLIGTINGSWERKTUIG, LEKSIKOGRAAF, LEKSIKOGRAFIE, LEKSIKOGRAFIESE PRAKTYK, LEKSIKOGRAFIESE WERKTUIG, LEKSIKOGRAFIETEORIE, LEKSIKOLOGIE, METALEKSIKOGRAFIE, ONAFHANKLIKE DISSIPLINE, TERMINOGRAFIE, WETENSKAPLIKE LEKSIKOGRAFIE

1. Introduction

The first section of this paper indicates a number of definitions of *lexicography* as found in general sources, specialized dictionaries and scientific publications.

These definitions reflect the prevailing interpretations of this term. It is clear that there are significant differences between some of the interpretations. In order to have a mutual point of departure when talking about lexicography the need is indicated for a unified explanation of *lexicography.* Subsequently this term is discussed in some detail, focusing on different components of lexicography. Taking cognizance of the existing definitions as well as the points raised in the preceding discussion suggestions are made for a new interpretation of the term that encompasses the majority of relevant features.

2. Current views on lexicography

When attempting to define a widely-used concept like *lexicography* it is important to take cognizance of some other existing definitions in order to detect some of the mutual strong and weak points. This section will refer to a few definitions and paraphrases of *lexicography* from three categories of sources, i.e. general sources (including general language printed dictionaries and results from random Google searches), LSP dictionaries dealing with lexicography and, thirdly, scientific discussions in the field of lexicography.

2.1 Definitions in general sources

The definitions or paraphrases given in general sources have to be regarded as important because that is where the non-expert in lexicography finds information regarding the meaning of this word. Someone looking for such a definition, especially those found in random Google searches, has no guarantee as to the authority of the definition or the expertise of the person who formulated it. The quality of these definitions shows significant differences, as can be seen in the following randomly selected examples:

> **Definition no. 1**
>
> the practice of compiling dictionaries (*The New Oxford Dictionary of English*)

This is an extremely unsatisfactory definition that makes no provision for the theoretical component and gives no details regarding the compilation process.

> **Definition no. 2**
>
> lexicography (is) the applied study of the meaning, evolution, and function of the vocabulary units of a language for the purpose of compilation in book form — in short, the process of dictionary making.
> (http://answers.encyclopedia.com/question/lexicography-159511.html)

Although better than the first definition this definition also ignores reference to

a theoretical component. Too strong a focus is placed on meaning — not all dictionaries include items giving the meaning of words.

Definition no. 3

> Perhaps the simplest explanation of lexicography is that it is a scholarly discipline that involves compiling, writing, or editing dictionaries. Lexicography is widely considered an independent scholarly discipline, though it is a subfield within linguistics.
> Many consider lexicography to be divided into two related areas. The act of writing, or editing dictionaries is known as Practical Lexicography. The analysis or description of the vocabulary of a particular language, and the meaning that links certain words to others in a dictionary, is known as Theoretical Lexicography. Theoretical Lexicography is particularly concerned with developing theories regarding the structural and semantic relationships among words in the dictionary. Since it involves theoretical analysis of the lexicon, Theoretical Lexicography is also known as Metalexicography.
> (http://www.wisegeek.com/what-is-lexicography.htm)

A positive aspect of this definition is the distinction it makes between practical and theoretical lexicography. A negative aspect is that it regards lexicography as a subfield within linguistics.

Definition no. 4

> **Lexicography** is divided into two related disciplines:
> - **Practical lexicography** is the art or craft of compiling, writing and editing dictionaries.
> - **Theoretical lexicography** is the scholarly discipline of analyzing and describing the semantic, syntagmatic and paradigmatic relationships within the lexicon (vocabulary) of a language, developing theories of dictionary components and structures linking the data in dictionaries, the needs for information by users in specific types of situation, and how users may best access the data incorporated in printed and electronic dictionaries. This is sometimes referred to as 'metalexicography'.
> (http://en.wikipedia.org/wiki/Lexicography)

This is a fairly well-balanced definition. Reference to dictionaries as practical tools would have enhanced its quality.

Definition no. 5

> Die maak van woordeboeke (The making of dictionaries) (*Woordeboek van die Afrikaanse Taal)*

This article for *leksikografie* in the *Woordeboek van die Afrikaanse Taal* has a cross-reference to *metaleksikografie* where the following definition is found:

> Teoretiese komponent v.d. leksikografie waarin die beginsels en tegnieke v.d.

> leksikografiese praktyk in oënskou geneem en daaroor besin word. (Theoretical component of lexicography which takes stock of the principles and techniques of the lexicographic practice and contemplates them.)

Two synonyms are then given for *metaleksikografie* (=metalexicography), i.e. *leksikografieteorie* (=lexicographic theory) and *leksikologie* (=lexicology).

Like definition 1, definition 5 is poor but the cross-reference to *metalexicography* does help. The definition of *metalexicography* is not bad but the user is ill-advised by the item indicating that it has *lexicology* as synonym.

No definition was given to account for a frequently used reference to lexicography as being an art or a craft — as seen in the title of Landau (2001): *Dictionaries: The Art and Craft of Lexicography* — by explaining the nature of this art or craft, albeit that Definition 4 does refer to *art or craft.* In this paper we do not regard lexicography as an art or a craft albeit that specific skills and talents play an important role in the lexicographic practice. Definitions like 1, 2 and 5 make no distinction between practical and theoretical lexicography and from all these definitions no realistic interpretation of the nature and extent of lexicography is possible. No clarity can be found regarding the relation between lexicography on the one hand and, on the other hand, applied linguistics, lexicology, encyclopedology and terminology. In addition, conflicting comments are found regarding the status of lexicography as an independent discipline. Definition 2 refers to "applied study", definition 5 gives a cross-reference to *metaleksikografie* that has *lexicology* as a synonym. No definition makes a reference to the fact that dictionaries, as products of the lexicographic practice, cover both language for general purposes and language for special purposes. This defies the reality that LSP lexicography, often referred to as *terminography*, also falls within the scope of lexicography. The fact that dictionary typology makes provision for encyclopedic dictionaries, i.e. dictionaries that do not focus on the linguistic but rather the extra-linguistic features, does not come to the fore in any of the definitions. Definitions 3 and 4 make provision for theoretical and practical components of lexicography. The nature of the theoretical component is not clear at all although there is a focus on the analysis of dictionaries and, rightly so, on the development of theories. The independence of lexicography is indicated in definition 3 but the same definition contradicts itself by indicating that lexicography is a subfield within linguistics. The average language user who consults any one of these definitions, with the exception of definition no. 4 which gives a reasonably well-balanced account of *lexicography,* will have an incomplete knowledge of the word *lexicography*, and the user consulting any combination of these definitions will be confused. Clearly the definition of the word *lexicography* in a random selection of sources aimed at the non-expert is totally insufficient.

2.2 Dictionaries of lexicography

Specialized dictionaries of lexicography assist their users in a far better way.

The *Dictionary of Lexicography* (Hartmann and James 1998) defines *lexicography* as:

> The professional activity and academic field concerned with DICTIONARIES and other REFERENCE WORKS. It has two basic divisions: lexicographic practice, or DICTIONARY-MAKING, and lexicographic theory, or DICTIONARY RESEARCH. ...

It continues with a more comprehensive description of both divisions, including a listing of several branches of lexicography, e.g. *author lexicography, bilingual lexicography, encyclopedic lexicography, specialized lexicography, thesaurus lexicography.* By means of a diagram it divides *dictionary research* into four components, i.e. *history, typology, criticism* and *use*, whereas dictionary-making is divided into three components, i.e. *fieldwork, description* and *presentation*.

However, one misses remarks regarding the user, the contents, the structures and the functions of dictionaries and, in a dictionary described on the cover as a "professional handbook", a reference to different theories of lexicography.

In *Lexicography: A dictionary of basic terminology* (Burkhanov 1998) an extensive treatment of *lexicography* is found. The main issues referred to are the "disciplinary status of lexicography, its correlation with other linguistic and non-linguistic disciplines, the scope of *lexicographic description,* methodology of *lexicographic investigation,* typology of *reference works* produced within the framework of lexicography, techniques of *lexicographic presentation* ..." Different perspectives on the position of lexicography are also given (with an indication of the relevant literature; not repeated here): "Lexicography is regarded as a domain of applied linguistics ..., a branch of information science ..., a province of philological and historical study ..., a subject field whose theoretical aspect falls within the realm of theoretical linguistics, whereas its practice pertains to the sphere of applied linguistics ...". Burkhanov also refers to the fact that "lexicography has been successfully developing its own theory." He also argues that "The term 'lexicography' refers to the process, result, and theoretical evaluation, of the making of reference works which represent a wide range of heterogeneous knowledge structures ...".

Burkhanov's description allows the expert user a comprehensive retrieval of information and gives ample guidance in terms of the extent of information transfer in a dictionary. However, the important role of the user and the really important notion of lexicography as an independent discipline does not come to the fore strongly enough.[1]

2.3 Scientific discussions

Whereas the above-mentioned specialized dictionaries of lexicography agree on lexicography having both a practical and a theoretical component, one of the first crucial issues in the discussions of *lexicography* as found in scientific publications is whether the notion of a theory of lexicography is accepted or

not. Different approaches prevail, ranging from Atkins and Rundell (2008: 4) saying, with regard to a theory of lexicography, that they "do not believe that such a thing exists", and Bejoint (2010: 381) saying: "I simply do not believe that there exists a theory of lexicography, and I very much doubt that there can be one", to lexicographers who firmly believe in a lexicographic theory, cf. Wiegand (1989), Bergenholtz and Tarp (2003), Gouws (2011), Tarp (2012). It is important to note that the supporters of a lexicographic theory do not all adhere to the same theory: there are different lexicographic theories but they all acknowledge the fact that the lexicographic practice is complemented by a theoretical component and that lexicography, with dictionaries as its subject matter, should be regarded as an independent discipline.

Wiegand (1984: 13) categorically denies that lexicography is a branch of applied linguistics or of lexicology, but when working outside a specific lexicographic theory, relations between lexicography and other disciplines are often postulated that go against the grain of lexicography as an independent discipline, cf. Kempcke (1982: 44) who says "Eine Wörterbuchtheorie kann nur Teil der Lexikologie sein, …" (A dictionary theory has to be part of lexicology). Lexicographers like Urdang (1963: 594) believe that "Lexicography, in practice is a form of applied linguistics …" and Sinclair (1984: 7) denies the prospect of a theory of lexicography and believes that the relevant theory is to be found in or via the areas of linguistics and information technology. Geeraerts (1987: 1) assumes that lexicography is part of linguistics but can hardly justify it as being a form of applied linguistics:

> As a linguistic discipline, lexicography has rather paradoxical nature. On the one hand, almost everybody will agree to classify lexicography as a form of applied linguistics, but on the other hand, it is virtually impossible to give an adequate reply to the question what linguistic theory lexicography might be an application of.

Some terminologists make a distinction between terminography and lexicography whilst lexicographers adhering to an inclusive lexicographic theory that makes provision for general and specialized lexicography regard *specialized* or *LSP lexicography* and *terminography* as synonyms, cf. Bergenholtz (1995b) and Bergenholtz and Kaufmann (1997).[2]

Tarp (2008: 9-10) distinguishes different types of theory in terms of three sets of distinctions, i.e. *general* and *specific theories, integrated* and *non-integrated theories* and *contemplative* and *transformative theories.* This last distinction is important for the present discussion. A purely contemplative approach only observes existing dictionaries and theoretical models and is rarely if ever put to practice. A transformative approach is innovative and this type of theory does not only interpret and explain lexicographic practice but it transforms it, cf. Tarp (2008: 10).

Within scientific discussions major differences also prevail regarding the scope of lexicography. In the introductory section to the first volume of the

International Journal of Lexicography (IJL) Robert Ilson, the first editor of this journal, gave an indication of the scope the newly established journal would like to cover — a scope that should reflect a specific interpretation of what lexicography is all about. Ilson (1988) says:

> IJL is devoted to examining how people inform one another. In its pages are discussed which items are selected to give information about, what information is given about them, and how that information is used. Our primary concern is with reference works that give lexically relevant information about lexically relevant items. But we realise that the problems facing compilers and users of dictionaries and thesauruses are similar to those facing compilers and users of indexes, encyclopaedias, atlases, and other types of reference work, and our pages are open to the discussion of their problems, too.

This broad scope which Ilson identifies, coincides with current theories pleading for a wider interpretation of the scope of lexicography, cf. Fuertes-Olivera and Bergenholtz (2011).

Engelberg and Lemnitzer (2009: 3) also distinguish between the lexicographic practice, i.e. the cultural practice aimed at the production of dictionaries, and the theoretical domain, directed at lexicography and dictionaries. They clearly state that theoretical lexicography is not part of applied linguistics. According to them lexicography has the following topics as subject matter:

> the structure of dictionaries
>
> the compilation of dictionaries (i.e. lexicography in the actual sense of the word)
>
> the use of dictionaries (including aspects of didactics of dictionary use)
>
> dictionary criticism
>
> the history of lexicography.

Here one misses a reference to lexicographic functions. Looking at these diverse and often conflicting interpretations of *lexicography* this paper works with the assumption that lexicography has a theoretical and a practical component, that different lexicographic theories do exist, and that lexicography is an independent discipline. This point of departure leads to the formulation of a unified and more comprehensive explanation of *lexicography.*

3. A unified and comprehensive approach

The proposals in this paper should be seen in combination with the concurrently written papers of Bergenholtz (2012) and Gouws (2012). Some of the arguments given in these two papers are presupposed in the current paper. The proposals made here are based on and expands the dictionary article of the

lemma **leksikografi** in the Nordic dictionary of lexicography, Bergenholtz et al. (1997).

There are two types of lexicography:

1. The development of theories about and the conceptualization of dictionaries, specifically with regard to the function, the structure and the contents of dictionaries. This part of lexicography is known as metalexicography or theoretical lexicography.
2. The planning and compilation of concrete dictionaries. This part of lexicography is known as practical lexicography or the lexicographic practice.

As seen in the previous sections there is a definite confusion in the metalexicographic discussion regarding the scope of lexicography and the borders with and relations to other disciplines. As previously remarked we regard lexicography as an independent discipline that does show some relation to parts of different other disciplines, e.g. information science and linguistics.

Our approach is not the only one; many people regard lexicography not as an independent discipline but as part of linguistics. Other people see parts of what we regard as lexicography as terminography or encyclopedology. We do not agree with this approach. More detailed motivation can be found in Bergenholtz (1995a) and a brief account thereof in the following paragraph.

In particular we do not see lexicography as part of lexicology — as is the case with some linguists and lexicographers, cf. paragraph 2.3. An approach that sees lexicographic theory as part of lexicology implies that lexicography puts the questions whereas lexicology provides the answers. We do not believe that this is the case in real practical situations. In contrast to their argument it is a fact that many lexicologists exclusively use data from dictionaries in their discussions. In the exact opposite way we regard the relation of terminography to that section of terminology where practical terminology prevails. Contrary to terminologists we regard terminography and subject field lexicography as synonym expressions. They have the same object and aims: to describe specialized fields so that specific information needs of the user can be satisfied, cf. Bergenholtz (1995b).

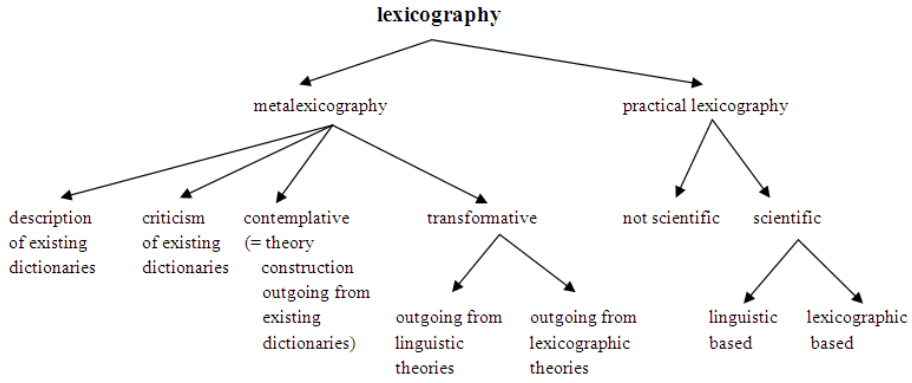
There also is a series of special types of lexicography, e.g. linguistic lexicography, subject field lexicography or corpus lexicography. We don't regard all the prevailing subtypes as necessary or beneficial to lexicography. However, this will not be discussed in detail here.

Linguistic lexicography is usually understood as general language lexicography that needs to achieve communicative functions. Subject field lexicography is typically understood as the monolingual lexicography of different subject fields, where the lexicography needs to achieve a cognitive function. Finally, encyclopedic lexicography is the type of lexicography that includes both linguistic and subject field lexicography.

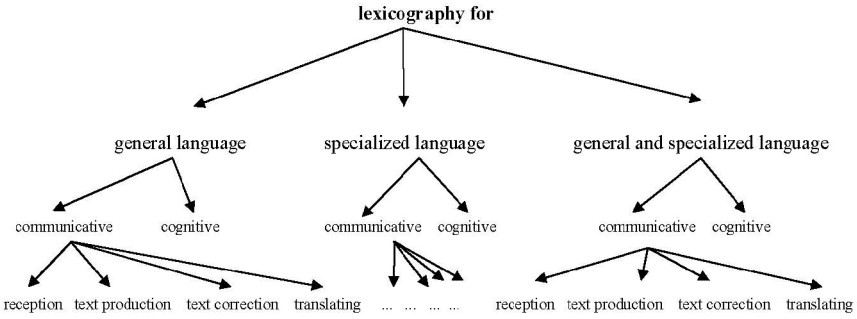
Lexicography is also identified in terms of the number of object languages: monolingual, bilingual or polylingual lexicography. In addition *lexicography* is used as part of a compound term when referring to the source material, e.g. corpus lexicography. But we have never encountered such terms as *informant lexicography* or *citation lexicography* although they could have been constructed accordingly. When the technical aids are put in the centre one refers to e.g. computational lexicography. When focusing on the purpose of the lexicography one talks about e.g. learner lexicography or translation lexicography. Finally the aim of lexicography becomes the documentation of a specific part of language use for future generations by having expressions like *usage lexicography*, i.e. the lexicography that accounts for concrete communicative, cognitive interpretative or operational needs, or *documentation lexicography*, that endeavours to solve a national or a general scientific problem.

There are further distinctions of this type and additional ones can be constructed. Our proposal is not directly related to that. We would rather try to present a general identification of lexicography.

The discussion of some definitions of *lexicography* in the first section of this paper already gives an answer to the question formulated in the title of this paper: Lexicography is the discipline dealing with theories about recently completed and also older existing dictionaries but also about future dictionaries as planned and produced by lexicographers. This simple answer is at the same time right but also too simple. There are different kinds of dictionaries and of lexicographers. This means e.g. that we have a type of lexicography describing, criticizing and making theories outgoing from existing dictionaries, and we have a type of lexicography making theories about how to plan and how to make conceptions for new dictionaries. And we have a branch of lexicography dealing with the concrete conception, planning and editing of a dictionary. Such a conception could be made without any kind of scientific considerations, i.e. by trying to make a new dictionary according to the way of "how it used to be" — the lexicographer makes a dictionary following his/her intuition and by knowing the needs of the intended user. Dictionaries of this type do not necessarily have a low quality, especially if they do not merely copy the "tradition". A splendid example of a dictionary belonging to this type was that of Leth (1800), a priest well familiar with the needs of the young people he was teaching, but not with the then current tradition of making conceptualisations of dictionaries. Another type of lexicography is totally influenced by linguistics and tries to use the best linguistic theories and terms for the planning and compilation of dictionaries. A final type of lexicography argues that lexicography is an independent discipline, perhaps somehow connected to a certain kind of information science or linguistics, but indeed not a subdiscipline of linguistics. Some aspects of these different types of approaches are illustrated in the following figure:



This figure does not show the real dilemma in lexicography. Most lexicographers claim that lexicography is a subdiscipline of linguistics. We do not think so. There are relations to linguistics but also to information science although we realise that, as in the case of lexicography, there are different opinions and definitions of what information science is. For the time being we therefore regard lexicography as an independent discipline, relying on experts coming from different disciplines. In one kind of dictionary the experts are linguists. This is the type of dictionary with the most relevance to linguistics of course. The following figure indicates different approaches regarding the object and functions of lexicography:



Traditionally lexicography had as its main object to deal with communicative information tools for general language dictionaries. We find this a too narrow understanding of lexicography that eschews many very important information tools. We regard lexicographic theory as a discipline not only directed at the production of dictionaries, but in a more general way at the production of information tools. The transformative approach can produce new ideas to ensure theoretically-based products, i.e. better dictionaries and other reference and information tools, and can ensure enhanced information retrieval.

4. Endnotes

1. The *Wörterbuch zur Lexikographie und Wörterbuchforschung/Dictionary of Lexicography and Dictionary Research* will not be discussed here because the published volume has not yet reached the article stretch in which *lexicography* falls. The treatment of *lexicography* in the *Nordisk leksikografisk ordbok* will form a basis of the discussion in paragraph 3.
2. A variety of citations from different authors, reflecting on the diversity in interpretations when it comes to the term *lexicography* can be found in Tarp (2012) and Wiegand (1998: 13-47).

5. References

Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.

Béjoint, H. 2010. *The Lexicography of English.* Oxford: Oxford University Press.

Bergenholtz, H. 1995a. Leksikografi. Hvad er det? Svararsdóttir, Åsta, Guðrun Kvaran and Jón Hilmar Jónsson (Ed.). 1995. *Nordiske studier i leksikografi 3. Rapport fra Konference om leksikografi i Norden. Reykjavik 7–10. juni 1995:* 37-49. Reykjavik: Nordisk forening for leksikografi.

Bergenholtz, H. 1995b. Wodurch unterscheidet sich Fachlexikographie von Terminographie? *Lexicographica* 11: 37-46.

Bergenholtz, H. 2012. What is a Dictionary? *Lexikos* 22: 20-30.

Bergenholtz, Henning, Ilse Cantell, Ruth Vatvedt Fjeld, Dag Gundersen, Jón Hilmar Jónsson, and Bo Svensén. 1997. *NLO. Nordisk leksikografisk ordbok.* Oslo: Universitetsforlaget.

Bergenholtz, Henning and Uwe Kaufmann. 1997. Terminography and Lexicography. A Critical Survey of Dictionaries from a Single Specialised Field. Tarp, S. (Ed.). 1997. Specialised Lexicography and Terminology. *Hermes* 18: 91-125.

Bergenholtz, Henning and Sven Tarp. 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31: 171-196.

Burkhanov, Igor. 1998. *Lexicography: A Dictionary of Basic Terminology* Rzeszów: Wydawnictwo Wyższej Szkoły Pedagogicznej Rzeszów.

Engelberg, Stefan and Lothar Lemnitzer. 2009. *Lexikographie und Wörterbuchbenutzung.* Tübingen: Stauffenburg Verlag.

Fuertes-Olivera, Pedro A. and H. Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. London/New York: Continuum.

Geeraerts, D. 1987. Types of Semantic Information in Dictionaries. Ilson, R. (Ed.). 1987. *A Spectrum of Lexicography: Papers from AILA, Brussels 1984*: 1-10. Amsterdam: John Benjamins.

Gouws, R.H. 2011. Learning, Unlearning and Innovation in the Planning of Electronic Dictionaries. Fuertes-Olivera, Pedro A. and H. Bergenholtz (Eds.). *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 17-29. London/New York: Continuum.

Gouws, R.H. 2012. Who can Really be Called a Lexicographer? *Lexikos* 22: 217-225.

Hartmann, R.R.K. and Gregory James. 1998. *Dictionary of Lexicography.* London/New York: Routledge.

Ilson, R. 1988. Introduction. *International Journal of Lexicography* 1(1): ii.

Kempcke, G. 1982. Lexikologie, lexikographische Theorie und lexikographische Praxis. Agricola, E. et al. (Eds.). 1982. *Wortschatzforschung heute*: 42-61. Leipzig: VEB Verlag.

Landau, S.I. 2001. *Dictionaries: The Art and Craft of Lexicography.* Second Edition. New York/Cambridge: Cambridge University Press.

Leth, J. 1800. Dansk Glossarium. En Ordbog til Forklaring over det danske Sprogs gamle, nye og fremme Ord og Talemaader for unge Mennesker og for Ustuderede. Et Forsøg. Med en Fortale af Professor Rasmus Nyerup. Kiøbenhavn: Trykt paa Hofboghandler Simon Poulsens Forlag hos Bogtrykker Morthorst's Enke & Comp.

Pearsall, J. (Ed.). 1998. *The New Oxford Dictionary of English.* Oxford: Oxford University Press.

Sinclair, J.M. 1984. Lexicography as an Academic Subject. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83 Proceedings*: 3-12. Tübingen: Max Niemeyer.

Tarp, Sven. 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Lexicographica. Series Maior 134. Tübingen: Max Niemeyer.

Tarp, S. 2012. Do We Need a (New) Theory of Lexicography? *Lexikos* 22: 321-332.

Van Schalkwyk, D.J. (Ed.-in-chief). 1994. *Woordeboek van die Afrikaanse Taal. Volume IX.* Stellenbosch: Bureau of the WAT.

Van Schalkwyk, D.J. (Ed.-in-chief). 1996. *Woordeboek van die Afrikaanse Taal*. *Volume X.* Stellenbosch: Bureau of the WAT.

Urdang, L. 1963. Review: Problems in Lexicography: Report of the Conference on Lexicography. *Language* 39(3): 586-594.

Wiegand, H.E. 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983*: 13-30. Lexicographica. Series Maior 1. Tübingen: Max Niemeyer Verlag.

Wiegand, H.E. 1989. Der gegenwärtige Status des Lexikographie und ihr Verhältnis zu anderen Disziplinen. Hausmann, F.J. et al. (Eds.). 1989–1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie*: 246-280. Berlin/New York: Walter de Gruyter.

Wiegand, H.E. 1998. *Wörterbuchforschung*. Berlin: Walter de Gruyter.

E-dictionary Use under the Spotlight: Students' Use of Pocket Electronic Dictionaries for Writing

Atipat Boonmoh, *Department of Language Studies, School of Liberal Arts, King Mongkut's University of Technology Thonburi, Bangmod, Thailand (atipat.boo@kmutt.ac.th/atipat_b@yahoo.com)*

Abstract: This article reports on the utilisation of pocket electronic dictionaries (PEDs) for writing by learners of English at a Thai university. It aims to enrich the study of dictionary use behaviour by investigating, through the use of combined research methods, exactly what happens when students use PEDs for production. The participants in this study included 13 students who were chosen from a group of 1211 students enrolled in a foundation English course at a university in Thailand. Data were collected using a think-aloud protocol, observation, and retrospective interviews as data collection methods. The first eight participants were asked to read a passage in Thai. Using dictionaries in their PEDs, they were asked to write a summary in English (the Water I experiment). The remaining five participants followed the same procedure, but after the summary task was completed, they were asked to review their summaries using the *Oxford Advanced Learner's Dictionary*, and the English–English dictionary in their PEDs (the Water II experiment). The experimental study revealed how the participants tackled the reading passage and wrote summaries, the problems they encountered, and the strategies they used to solve these problems. A graphic representation of the PED consultation process was also proposed. The use of the English–English dictionary in their PEDs helped some participants review their English summaries. It was found that some participants failed to display (extended) knowledge of the PEDs they were using. The investigations revealed several factors that may have hampered dictionary lookup success as well as factors that may have promoted dictionary lookup success.

Keywords: DICTIONARY USE, POCKET ELECTRONIC DICTIONARY, DICTIONARY CONSULTATION, MONOLINGUAL DICTIONARY, BILINGUAL DICTIONARY, WRITING

Opsomming: E-woordeboekgebruik onder die soeklig: Studente se gebruik van elektroniese sakwoordeboeke vir skryfwerk. In hierdie artikel word verslag gedoen van die gebruik van elektroniese sakwoordeboeke (ESW'e) vir skryfwerk deur aanleerders van Engels by 'n Thaise universiteit. Daar word gepoog om die studie van woordeboekgebruiksgedrag te verryk deur ondersoek in te stel na presies wat gebeur wanneer studente ESW'e gebruik vir skryfwerk deur gebruik te maak van gekombineerde navorsingsmetodes. Die deelnemers aan hierdie studie het 13 studente ingesluit wat gekies is uit 'n groep van 1211 studente wat aan 'n universiteit in Thailand ingeskryf was vir 'n kursus in basiese Engels. Data is versamel deur middel van 'n hardopdinkprotokol, waarneming en retrospektiewe onderhoude as dataversamelingsmetodes. Die eerste agt deelnemers is gevra om 'n gedeelte in Thai te lees. Hulle is gevra om 'n opsomming in Engels te skryf (die Water I-eksperiment) terwyl hulle woordeboeke in hulle ESW'e

gebruik. Die oorblywende vyf deelnemers het dieselfde prosedure gevolg, maar nadat die opsommingstaak afgehandel is, is hulle gevra om hulle opsommings te kontroleer deur die *Oxford Advanced Learner's Dictionary* en die Engels–Engels-woordeboek in hulle ESW'e te gebruik (die Water II-eksperiment). Die eksperimentele studie het getoon hoe die deelnemers die leesgedeelte aangepak en opsommings geskryf het, watter probleme hulle teëgekom het en watter strategieë hulle gebruik het om hulle probleme op te los. 'n Grafiese voorstelling van die proses van hoe die ESW'e geraadpleeg is, is ook aangebied. Die gebruik van die Engels–Engels-woordeboek in hulle ESW'e het sommige deelnemers gehelp met die kontrolering van hul Engelse opsommings. Daar is bevind dat sommige deelnemers nie 'n (uitgebreide) kennis kon toon van die ESW'e wat hulle gebruik het nie. Die ondersoeke het verskeie faktore blootgelê wat die sukses van naslaan in die woordeboek sou kon belemmer, sowel as faktore wat die sukses van naslaan sou kon bevorder.

Sleutelwoorde: WOORDEBOEKGEBRUIK, ELEKTRONIESE SAKWOORDEBOEK, WOORDEBOEKRAADPLEGING, EENTALIGE WOORDEBOEK, TWEETALIGE WOORDEBOEK, SKRYFWERK

1. Introduction

Research (e.g. Deng 2005, Midlane 2005, Taylor and Chan 1994) indicates a growing number of pocket electronic dictionary (PED) users in many South and East Asian countries due to the advance of technology, PED ease of use and their portable size. Pocket electronic dictionaries (PEDs) are common in Thailand (Boonmoh and Nesi 2008, Mongphet 2007), and are often advertised in terms of their technological features — what PEDs can do and what hard-copy dictionaries are contained in them — rather than lexicographical features. PEDs are available for sale in major department stores, and PED booths from different manufacturers are normally located next to each other. Their prices are more affordable than in the past. For these reasons, the PED phenomenon can be observed throughout Thailand.

This study has its origin in observations of this phenomenon when the author was employed as a language lecturer at a university in Thailand. It was observed that many students brought PEDs into the classroom and often consulted their PEDs when writing. Although dictionary skills training lessons had been included in foundation English courses at the university and students were encouraged to use any English learners' dictionaries, the students seemed to prefer using PEDs when left to their own devices. Lecturers, however, often complained about the students' language mistakes. Many of them commented that these mistakes could have been made as a result of the PEDs students were using. These teachers seemed to take a negative view when students used PEDs. A discussion with a few colleagues revealed that they did not use PEDs, did not know much about what PEDs can offer, and probably had less knowledge about PEDs than the students.

Information about Thai PEDs is limited and is not often available to Thai lecturers. The PED manufacturers do not promote their products from a lexicographical perspective. Moreover, most of the existing PED studies do not refer to

the Thai context. Therefore, it may not be justified for the teachers to question the quality of PEDs students use without having access to empirical data concerning how students actually use their PEDs. This study aims to find out which processes take place and which procedures are followed when students use PEDs in text production.

2. Literature review

2.1 Research into pocket electronic dictionary use

PEDs have been available for the last thirty years but research into PED use is in its infancy (Jopling 2003, Tono 2001). Most previous studies refer to electronic dictionaries on CD-ROM that were produced by famous publishing houses and can be easily reviewed from a lexicographical perspective. On the other hand, most of the few PED studies (e.g. Taylor and Chan 1994, Deng 2005, Stirling 2005) have been confined to quantitative ownership surveys, and qualitative investigations into teacher and student attitudes and beliefs. A few simple experiments have been conducted, such as a lookup 'race' between PED and print dictionary users (Weschler and Pitts 2000). There have been few studies on how people actually use pocket electronic dictionaries. The few studies that looked closely at what happens when electronic dictionaries are consulted were mainly concerned with the use of learner's dictionaries on CD-ROM (Jopling 2003, Nesi and Haill 2002, Winkler 2001), and an online bilingual dictionary (Liou 2000).

PEDs are popular with students especially in South and East Asian countries. In Taylor and Chan's (1994) survey of 475 Hong Kong students, 18% used PEDs, and 70% of 80 Chinese college students in Deng's (2005) survey were PED users. Most of the 11 EFL students in a UK language school interviewed by Stirling (2005) were in favour of PEDs. Teachers in Midlane's (2005) survey reported students bringing PEDs in the classroom. A recent questionnaire survey conducted by Boonmoh and Nesi (2008) showed that although almost all Thai students (938 out of 1211) reported owning learner's dictionaries in book form, only 102 and 46 respectively stated that they normally used these dictionaries for reading and writing. On the other hand, the number of students who reported owning PEDs (456 students) was found to correspond well with the number of students who reported using them (435 for reading, and 412 for writing). Interestingly, the number of students who reported they wanted to buy PEDs in the future rose to 818 as opposed to 117 students who reported they wanted to buy learner's dictionaries in book form. PEDs are a promising tool for students. Students increasingly prefer PEDs to dictionaries in book form. As Midlane (2005: 125) points out, one aspect of the growth in PED use is because "it had been a bottom-up movement". It is student-led — not led by teachers or lexicographers. Furthermore, the greater use of PEDs may to a certain extent change the nature of classroom learning.

Teachers' views of PEDs, on the other hand, tend to be quite negative. Teachers complain that PED contents are inadequate. Deng (2005) claims that the PEDs his students use in China do not supply English definitions, inflectional forms or examples. Koren (1997) reports that Israeli teachers object to PEDs because they lack "word meanings, word families, parts of speech, tense, usage and idioms, etc." Several of the 11 EFL teachers in the UK interviewed by Stirling (2005) also complained about "inaccurate meanings" and "insufficient examples". Boonmoh and Nesi's (2008) survey of 30 lecturers of English reveals that lecturers are highly critical of the PED as a tool for students, and many refer to the inadequacy of the dictionary information it provides.

It is seen, on the one hand, that the students are overwhelmingly in favour of PEDs. The teachers, on the other hand, have negative attitudes towards PED use as they believe PEDs to be noisy and distracting. They are also more critical of the students' use of PEDs. Knowing only 'which' dictionaries students use, however, may not be sufficient. In order to help teachers to be able to provide authoritative advice on PED purchases, and develop e-dictionary skill training programmes, it is important to find out "exactly what … students are doing with their dictionaries, what they expect from them, and how easily they are satisfied during the process of consultation" (Atkins and Varantola 1998: 115). A number of studies that attempt to uncover how students actually use dictionaries for reading can be found in Liou (2000), Winkler (2001), Wingate (2004) and Nesi and Boonmoh (2009). Liou focuses on online bilingual dictionaries, Winkler on learner's dictionaries on CD-ROM, and Wingate on dictionaries in book form. An attempt to uncover how students use PEDs can be found in Nesi and Boonmoh (2009). In their study, Nesi and Boonmoh investigate how Thai students use their PEDs for reading. The findings suggest that the subjects failed to display dictionary skills and knowledge of the PEDs they were using.

In order to complete the picture of PED use for both receptive and productive purposes, this study aims to report on how PEDs are utilised for writing. The purpose of this study is, therefore, to answer the following two questions:

— How do Thai students use their pocket electronic dictionaries to read a passage in Thai in order to write a summary in English?

— How successful are their PED consultations?

2.2 Methodological options for PED research

Questionnaire research is perhaps the most common method of enquiry into the use of dictionaries. Many studies have been confined to surveys, mainly conducted by means of questionnaires (Deng 2005, Midlane 2005, Sobkowiak 2002, Tang 1997, and Taylor and Chan 1994), since they can be used as a way of obtaining results from a great number of respondents. They can be useful for identifying general trends which might then be examined more closely in smaller, more empirical studies. A questionnaire alone, however, cannot reveal "exactly

what … students are doing with their dictionaries, what they expect from them, and how easily they are satisfied during the process of consultation" (Atkins and Varantola 1998: 115). For this reason, it seems a good idea to triangulate questionnaire data with more qualitative data obtained by other means.

Interviews can be used to elicit opinions, and interactive settings are another step towards gathering more direct evidence of dictionary lookup behaviour. The interview questions can be structured, but can also be flexible in the sense that interviewers may ask further questions related to the interviewees' reply. Although an interview cannot reveal exactly how students actually use PEDs, it may be a useful device to use retrospectively and to obtain data which can be triangulated with those collected in the same study through the use of other techniques. Previous research which employs interviews together with other research instruments includes Diab (1990), Winkler (2001) and Boonmoh (2003).

Observation is an obvious means of collecting data in educational settings, but PED displays are much smaller than computer screens or the printed page, so it is very difficult for teachers or researchers to see what is happening during a student's PED consultation. PED consultation is also a private activity, and one which learners are often inclined to be secretive about (Nesi and Boonmoh 2009, Nesi and Haill 2002). Observing PED use in a natural setting is, therefore, almost impossible. Video recording users, a method employed by Jopling (2003) when investigating the use of CD-ROM dictionaries, is not a practical means of researching PED use. Although 'spy' software — a method used in Liou (2000) — has some potential as a means of observing online dictionary use (through keystroke logging and screenshots), it cannot be loaded into the standard PED.

Another method of investigating dictionary use involves lookup record sheets. Some studies that have employed this method are Atkins and Varantola (1998), Diab and Hamdan (1999), Al-Ajmi (2002), Paisart (2004) and Frankenberg-Garcia (2005). Asking students to record words they look up, however, may not be appropriate for research into how PEDs are really used. This is because the focus of the study would be on the final decision of the students rather than the entire lookup process. Since PED use is much quicker than paper-based dictionary use (Weschler and Pitts 2000) and the speed encourages more lookups, supplying information for the dictionary record sheets would disrupt lookup and reduce the speed of consultation, and as a consequence discourage subjects from looking up words.

Another possible methodology is self-observation, as opposed to observation by the researcher. This can be in the form of retrospection or introspection. Retrospection requires subjects to report their working process after finishing the task; however, the limitations of memory can affect the quality of data reported in this way, especially given the unsatisfactory nature of video recordings of PED use.

Think-aloud data are basically unedited and unanalysed, as subjects are not in any way controlled or directed. There are, however, drawbacks and "this procedure is not a replacement for other research methodologies for investigating mental processes" (Cohen 1998: 39). For example, the process seems to

work better with extrovert subjects (introverts often fail to provide sufficient data), and there is a danger that subjects may modify what they say in order to meet the perceived needs of the researcher. Moreover, if the task is too easy, subjects may not be able to access their own thought processes, because the process of comprehending will be too quick and automatic (Ericsson and Simon 1980: 225, cited in Matsumoto, 1993: 48-49). Researchers must therefore pay particular attention to the selection of texts and tasks, but provided that this is done, think-aloud seems to be one of the most appropriate instruments for PED research.

Dictionary consultation is a private matter and there is no way of discovering what people actually do when they use a dictionary without, to a certain degree, interfering with their natural behaviour. Using a log file can help reveal this, but it is limited to observing dictionary use on CD-ROM or on the Internet. Observation can look at how students use dictionaries in a natural setting but only observable behaviours can be observed. To reveal their mental working processes, the subjects need to verbalise. Asking the subjects to verbalise will inevitably disrupt the subjects' working processes. This study, therefore, will employ a mix of research instruments in order to uncover exactly how participants use their PEDs for writing.

2.3 Pocket Electronic dictionaries in Thailand

Before exploring how PEDs are used, it is necessary to give some idea of the range of features Thai PEDs offer. A Thai PED normally contains at least three paper-based dictionaries: two bilingual English–Thai and Thai–English, and one English monolingual. There are at least four companies that produce PEDs in Thailand, but *TalkingDict* (Group Sense Ltd.) and *CyberDict* (Besta) are the leading brands. Over the past two decades, *TalkingDict* has published more than 20 models and *CyberDict*, which was established a few years later, has published more than 16 models. An investigation of two PED models by Boonmoh (2009), the *Super Smart* by *TalkingDict* and the *CyberDict 3 Advance* by *CyberDict*, found the main difference to be lexicographical features.

It is seen from Table 1 that *CyberDict 3 Advance* contains material from newer and more up-to-date paper-based dictionaries than *Super Smart*. More recent PED models by *TalkingDict* replaced the paper-based English–English *Concise American Heritage Dictionary* with the *Concise Oxford English Dictionary* (11th edition, 2006). It should be noted that the *Concise Oxford English Dictionary* is not intended for learners of English but for native speakers of English. Some newer *TalkingDict* models claim to contain the *Oxford River Books English–Thai Dictionary* and the *Cambridge Advanced Learner's Dictionary* (unstated edition). The contents of the English–Thai and Thai–English dictionaries of these two PED brands, however, remain the same.

Table 1: Comparison of two PED models

Dictionary	*Super Smart*	*CyberDict 3 Advance*
Thai–English	Compiled by lecturers from the Chalermprakiat Center of Translation and Interpretation (undated)	Thiengburanathum, W. 2002. *Thai–English Dictionary*. Library Edition. Bangkok: รวมสาส์น.
English–Thai	Compiled by lecturers from the Chalermprakiat Center of Translation and Interpretation (undated)	Thiengburanathum, W. 1998. *SE-ED's Modern English–Thai Dictionary (Complete and Updated) Desk Reference Edition*, Bangkok: SE-Education.
English–Thai	—	Mallikamas, P., N. Chakrabongse and P. Piammaattawat. 2004. *Oxford River Books English–Thai Dictionary*. Bangkok: River Books.
English–English	*The Concise American Heritage Dictionary*. (1983, Houghton Mifflin)	Hornby, A.S. 2000. *Oxford Advanced Learner's Dictionary*. 6th Edition. Oxford: Oxford University Press.

3. Methodology

3.1 Subjects

The participants in this study included 13 students chosen from a cohort of 1211 students (reported in Boonmoh and Nesi, 2008) enrolled in a foundation English course (Fundamental English II) in the 2007 academic year at a university in Thailand. They were from three faculties: Faculty of Engineering, Faculty of Sciences and Faculty of Industrial Engineering. These 13 subjects were selected because, in an earlier questionnaire survey, they had claimed to possess and use pocket electronic dictionaries, and had indicated their willingness to participate in the experiment. All the participants had passed a foundation English course (Fundamental English I) which included dictionary skills training lessons.

The first eight participants were asked to read a passage in Thai. Using dictionaries in their PEDs, they were asked to write a summary in English (**the Water I experiment**). They used *TalkingDict* or *CyberDict* PEDs with a variable combination of bilingual and monolingual English dictionaries (the *Concise American Heritage Dictionary* for *TalkingDict* PEDs, and the *Oxford Advanced Learner's Dictionary* for *CyberDict* PEDs).

The remaining five participants, who used *CyberDict* models, followed the same procedures, but additionally after the summary task was completed they were asked to review their summaries using the *Oxford Advanced Learner's Dictionary*, the monolingual English dictionary in their PEDs (**the Water II experiment**). The aim of the **Water II experiment** was to see what difference it would make when the participants used the *OALD 6th edition* to write a summary in English.

3.2 Instruments

A think-aloud protocol, observation, record sheet keeping and interviews were employed as data collection instruments in this study. The questionnaires were not only used as part of the surveys (reported in Boonmoh and Nesi 2008) but were also used to select potential participants for the Water experiments. The author's presence during the experiments enabled him to observe the whole 'visible' process of PED consultation. The think-aloud procedure allowed him to explore the participants' mental working processes which are 'invisible' when they are doing the tasks. Finally, the interview enabled the author to ask related questions concerning the participants' previous use of PEDs as well as providing the participants with the opportunity to clarify their working processes and their feelings after completing the tasks. Combining the data from these instruments would make the findings richer and more reliable.

3.3 Procedure

The 13 participants who were initially selected based on the findings of the questionnaire (see the Subject section above), underwent think-aloud training. The training consisted of two periods, i.e. the first was in plenary, and the second was individual (on the day each participant came to do the summary tasks). Participants were asked to read a Thai reading passage "Water" taken from Wikipedia, the free encyclopedia in the Thai version (See Appendix A for the reading passage). Then, they were asked to write a summary in English. This text had been piloted with a comparable group of volunteers, and had proven to be appropriate in terms of topic, difficulty level and length.

The participants were asked to complete the task in individual sessions. The author also observed every word looked up and completed an observation check sheet. The check sheets were used to record the words and meanings the participants looked up and to ask specific questions during the retrospective interviews conducted with each subject at the end of the session. (See Appendix B for interview schedule.)

3.4 Data analysis

Findings were derived from consideration of four data sources: think-aloud protocols, participants' written summaries in English, observation notes and interviews. The data were analysed only in cases where the participants consulted their PEDs to write English summaries. The number of words looked up and the number of successful and unsuccessful lookups were noted. The data from the interviews and the observation check sheets were analysed with reference to the following questions:

1. How did the participants write their summaries?
2. Which words did the participants look up in the PEDs?

3. Did they find the words they looked up? If not, why were they unable to find them?
4. Did they select any words found in the PEDs to use in the summary? If yes, which word did they select and why? If no, why not?
5. What did they do if they did not find the words they looked up or if they did not select any words after the lookup?
6. If they did not choose any word, what did they do?

4. Findings

4.1 The Water I Experiment

Table 2 summarises the participants' approaches to the task of summary writing.

Table 2: Participants' approaches to the summary task

Procedures	Participants							
	A	B	C	D	E	F	G	H
Participants read the passage word by word and looked up equivalents of the L1 words.	✓							
Participants read the whole passage first.		✓	✓	✓	✓	✓	✓	✓
Participants wrote a summary in Thai and then translated it into English.			✓			✓	✓	
Participants searched for English equivalents while writing a summary.				✓	✓			✓
Participants underlined the key words before starting to look them up.		✓						
Participants reviewed the summary before submitting it to the researcher.								✓

Figure 1 is a graphic representation of the different lookup stages during the PED consultation. It represents how the participants started a search, what problem(s) they had, and how they tried to solve the problem(s). The **first step** is concerned with whether or not the participants consulted their PEDs when writing the summary. If the participants did not consult their PEDs, those words or sentences were ignored, regardless of whether the participants used the words correctly or incorrectly. In contrast, if the participants consulted their PEDs, the analysis was continued in order to discover the reason(s) for this.

The **second step** is to ascertain whether the participants found the target word in the dictionary or not. If they found the target word in the dictionary (Yes), a third step was then analysed. The PED consultation was considered successful if the participants found the word they were looking for. If the participants were not able to find the target word through PED consultation, there were three possible reasons for this:

— The participant typed in the incorrect spelling of either the Thai or English word.

— The target word the participant was looking for was not in the PED.

— The search term was a multi-word unit. In some cases such a multi-word unit could have been broken down into separate components (which still conveyed the same sense).

The **third step** concerns the ability to locate an English word or phrase in the entry which is an appropriate equivalent for the Thai search term. Being able to locate appropriate equivalents is considered to be a success at a deeper level. However, deciding whether the participants located appropriate equivalents is not a straightforward business. The participants often did not simply locate the English equivalent and use it in their summary. They usually employed other, more complex strategies. It can be seen that not every lookup can be classified as either appropriate (yes) or inappropriate (no) since for some lookups the participants did not immediately choose which equivalent(s) they would use in the summary. Instead, they chose one of three alternative routes:

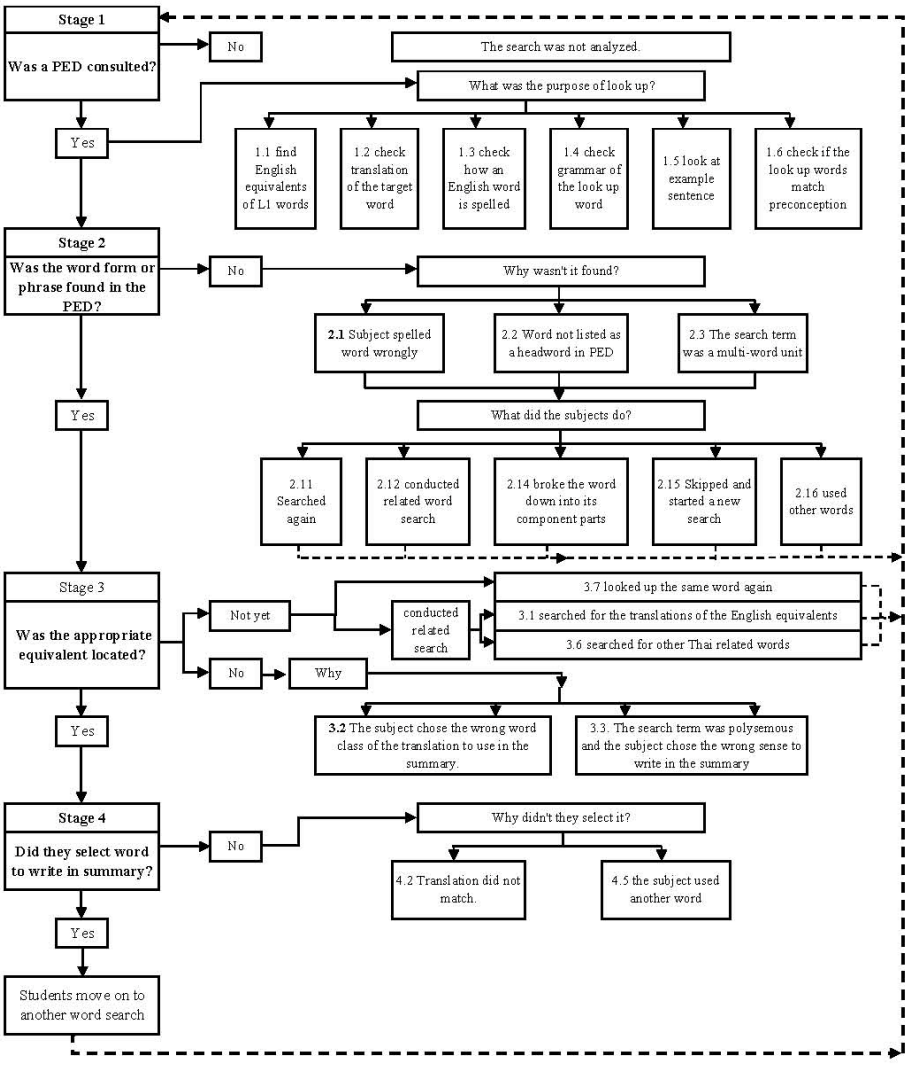
— They might look up the same word again.

— They might look up the translation(s) of the English equivalents (related word type II). For example, a search for the Thai word บึง [bueng] yielded "*n. a bog, a fen, a marsh, a swamp*". B wanted to choose one of the equivalents but was not sure which one was the most appropriate. He therefore searched for the translations of the words BOG, FEN, and MARSH in the default English–Thai dictionary before making a decision.

— They might search for other Thai words that have similar meanings or share the same root (related word type I). For example, B wanted to find the equivalent of the word ปรากฏการณ์ [prakotkan] (*phenomenon*) but made a mistake when typing in this word, so it yielded no results. He then searched for other Thai words ปรากฏ [prakot] (*to appear, to be evident, to be known, to take place*), and used the word wheel facility to scroll down and successfully find the word ปรากฏการณ์ [prakotkan] (*phenomenon*).

The related search type I is different from the related search type II in the sense that the former involves the headword in Thai while the latter involves the headword in English.

Having employed these complex strategies, participants then made a decision regarding which translation equivalent of the search term they should use. If the participant was able to locate an appropriate equivalent, a fourth stage was analysed.



Note The solid line indicates progression throught the stages while the dotted line indicates movement backward to the previous stages

Figure 1: Stages in using PEDs for production

Table 3 summarizes the lookup behaviour of each participant, with a special focus on Stage 1, i.e. overall lookups and their purposes.

Table 3: A summary of lookup behaviour and purposes of dictionary lookups

	TalkingDict Users					CyberDict Users			Total
	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	
Total minutes taken	60	40	43	40	45	36	35	40	
Total number of lookups (*1 + *2 + *3)	85	47	12	18	34	14	12	28	250 (100%)
Purpose of lookup									
To find English equivalents	85	29	10	16	25	14	10	10	199 (79.6%)
To check translations of English equivalents	-	17	2	-	6	-	1	10	36 (14.4%)
To check spelling	-	-	-	2	1	-	1	2	6 (2.4%)
To check if the word matched their preconception	-	1	-	-	2	-	-	3	6 (2.4%)
To check grammar	-	-	-	-	-	-	-	1	1 (0.4%)
To look at example sentences	-	-	-	-	-	-	-	2	2 (0.8%)
Number of words looked up *1	70	24	9	17	22	12	10	16	180
Number of related lookups *2 (out of actual words)	7 (6)	22 (12)	3 (2)	1 (1)	11 (7)	2 (1)	1 (1)	11 (6)	58 (36)
Number of repeated lookups *3 (out of actual words)	8 (6)	1 (1)	0	0	1 (1)	0	1 (1)	1 (1)	12 (10)

The average time spent doing the summary task was 42 minutes. Other participants, however, spent ±7 minutes of the average time in doing the summary task. Considering Table 3 alone, it may be sensible to conclude that participant A made use of the least number of features of his dictionary. He used the PED only to find the English equivalents of the Thai words (using only the default Thai–English dictionary). Participant H made the most use of his dictionary, using both the Thai–English dictionary (for English equivalents) and the English–Thai dictionary (for English translations), exploring the lexicographical features of the PED (i.e. grammar, example sentences), and having various reasons for lookups (e.g. checking spelling).

The total number of lookups ranged from 12 (C and G) to 85 (A). Participant A looked up 70 different words. The remaining participants looked up from as few as nine (participant C) to as many as 24 different words (participant B). It should be noted, however, that although B, E, and H conducted 47,

34, and 28 searches respectively, they only looked up 24, 22, and 16 different words. These three participants conducted many related word searches.

For related lookups, it should be noted that all of A's related lookups were different from those of the three participants (B, E, and H) above. All of A's related lookups were to search for other related Thai words (related search type I). On the other hand, almost all related searches made by B, E, and H were to search for the translations of English equivalents (related search type II). These three participants were more likely than A to be concerned about the appropriate use of the equivalents.

Table 4 summarises the lookup behaviour of each participant, with a special focus on stage 2, i.e. successful and unsuccessful lookups. The overall success rate was 89.6% (224 out of 250 lookups) and the overall failure rate was 10.4% (26 out of 250 lookups).

There were 26 unsuccessful lookups, and the most frequent reason for failure was that the search words were not included in the PEDs (12 lookups), the second most frequent reason was that the search term was a multi-word unit (10 lookups), and the last reason was that search words were incorrectly spelled (4 lookups).

Some of the words and expressions that were not listed in the PEDs were การดำรงชีวิต [kan damrongchiwit] (*the act of maintaining one's life*) and การดำรง [kan damrong] (*the act of maintaining, keeping*). In Thai, derived forms are created by adding a prefix to the stem of a word. Therefore, the search would have been successful if a participant had looked up the root form (the verbal form) ดำรงชีวิต [damrongchiwit] or ดำรง [damrong] in his PED.

A multi-word unit could be broken down into single words. For example, A was unable to find เขตหนาว [khet nao] (*cold area*), which can be separated into เขต [khet] (*area, location*) and หนาว [nao] (*cold*), both of which, of course, are listed.

Table 4: Successful and unsuccessful lookups

STAGE 2	***TalkingDict* Users**					***CyberDict* users**			**Total**
	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	
Did participants find words in their PEDs?									
— Yes	72	45	11	16	31	11	11	27	224
— No	13	2	1	2	3	3	1	1	26
Why didn't they find the words?									
— Participant spelled word incorrectly	1	1	-	-	1	-	1	-	4
— Word not listed as a headword in PED	6	-	1	1	1	3	-	-	12
— Search term was a multi-word unit	6	1	-	1	1	-	-	1	10

What did they do?									
— Searched again	-	-	-	-	1	-	-	-	1
— Conducted related word search	1	-	-	1	1	2	-	-	5
— Broke the word down	5	1	-	-	-	-	-	1	7
— Skipped and started a new search	7	1	-	1	-	-	-	-	9
— Used other words	-	-	1	-	1	1	1	-	4

Table 5 summarises the lookup behaviour of each subject, with a special focus on Stage 3, appropriateness of equivalents. As indicated above, 89.6% of lookups were successful. However, this does not necessarily imply that the participants were able to locate appropriate English equivalents of search words to write in their summaries, nor does it imply that the PEDs contained sufficient headwords, or that the participants possessed good dictionary skills. This author considers locating (and using) appropriate English equivalents to be the most crucial part of the dictionary consultation process, since it directly contributes to the success of the writing of the summary. Although there are several factors contributing to this success, for example, grammatical knowledge and stylistics, these are not relevant to dictionary use. Stage 3 investigates whether or not the participants could locate the most appropriate English equivalents of the Thai words (see Table 5).

As mentioned earlier, not every lookup can be straightforwardly classified as appropriate or inappropriate. Some lookups may take longer to classify than others. It can be seen that out of 224 successful lookups from stage 2, 159 (70.9%) could be classified as appropriate. Forty-three lookups were classified as 'not yet' because the participants did not simply locate the equivalents but conducted further searches before making a decision. Twenty-two lookups (9.8%) resulted in inappropriate equivalents being found.

Table 5: Appropriateness of equivalents

STAGE 3 **Did the participants locate the appropriate equivalents?**	***TalkingDict* users**					***CyberDict* users**			**Total**
	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	
— Yes	64	24	5	14	20	7	7	18	159
— Not yet	-	20	3	-	9	-	2	9	43
— No	8	1	3	2	2	4	2	-	22
If not yet, what did they do?									
— searched for translations of the English equivalents	-	14	2	-	6	-	-	8	30

— searched for other related Thai words	-	5	1	-	1	-	1	-	8
— looked up the same word again	-	-	-	-	1	-	1	1	3
If no, why?									
— wrong part of speech	4	-	-	-	1	-	-	-	5
— word sense	4	1	3	2	1	4	2	-	17

The last two categories ('not yet' and 'no') merit further discussion in this section. Participants A, D, and F did not interrupt their search to consider other alternatives before deciding which equivalents they would use in their summaries. In contrast, B, C, E, and H did not locate the equivalent in the first instance, but used one of three alternative strategies first. Out of 43 lookups, 30 involved searching for translations of the English equivalents, followed by searching for other related Thai words and lastly, looking up the same word in Thai again. This may be the most important of the alternative strategies because it seems to indicate that the participants were concerned about the appropriateness of the English translation. They wanted the equivalent to convey the closest meaning to the Thai headword. An interview with participant B revealed how he used his PED.

Researcher: Can you briefly explain how you used your PED?

B: I used the PED when I didn't know the English equivalent. And when there were many equivalents provided, I would look up all of their meanings in the English–Thai dictionary. I want to check each equivalent because sometimes it could mean something different. For example, the Thai word ซึม [Suem] can be used to mean ซึมเศร้าไปเลย [SuemSaoPaiLoei] (feeling very sad) or it can be used to mean น ำซึมลงดิน [NamsuemLongDin] (water oozes on the ground). They have completely differently meanings so I have to look up all their equivalents in the PED. If I used it incorrectly, the meaning will be different too.

(My translation of B's Interview)

4.2 The Water II Experiment

This section will report on the opinions of the remaining five participants when using an English–English dictionary to review their summaries. The participants went through the same procedures as in the previous task but they were then invited to use the English–English dictionary (the *OALD* 6th) included in their PEDs to revise what they had written. The findings are shown in Table 6.

Table 6 shows that the purpose of lookups for all participants (except A2) was to check if the meanings of English equivalents were the meanings they

intended, and also to check the translations of the English words.

Table 6: Lookup behaviour using an English–English dictionary

Categories	Participant				
	A2	**B2**	**C2**	**D2**	**E2**
1) Purpose of lookups					
Check if the meaning of English equivalent is the one intended		✓	✓	✓	✓
Check translations of the English words		✓	✓	✓	✓
Look for grammatical and usage information	✓				
Compare English words in order to choose the most appropriate one	✓		✓		✓
2) Exploitation of polysemous entries	Fully	Partly	Fully	No	Partly
3) Use of help options provided by PEDs	✓		✓		
4) Changes made to the summary	✓	✓	✓		✓
5) Displayed knowledge of PED (e.g. abbreviations, grammar)	Fully	None	Partly	Partly	Partly
Highlighting a word	✓		✓	✓	✓
Highlighting two words or more	✓		✓		
Using the backspace function	✓		✓		
Comprehension of the abbreviation e.g.	✓			✓	✓
Comprehension of the abbreviation sth	✓				
Distinguishing between countable and uncountable nouns	✓		✓	✓	✓

In terms of polysemous entries, only A2 and C2 exploited all the entry information. It was observed that they always scrolled down to see what other information was available. They also often looked for example sentences. For example, C2 searched for the word *natural*, which yielded eight senses (signposts) i.e. IN NATURE, EXPECTED, BEHAVIOUR, ABILITY, RELAXED, PARENT/CHILDREN, BASED ON HUMAN REASON and IN MUSIC. The think-aloud protocol confirms that C2 explored all these senses before making the decision whether to make changes to the summary.

natural	
adj., noun	
Adj.	
IN NATURE	
1. [only before noun] existing in nature; not	
1 e.g.	
Made or caused by human beings: e.g.	
-- compare SUPERNATURAL	
EXPECTED	
2. normal; as you would expect: e.g.	(line 10)
-- compare UNNATURAL	
BEHAVIOUR	
3. used to describe behaviour that is part of	
the character that a person or an animal	
Was born with: e.g.	(line 15)
ABILITY	
4. [only before noun] having an ability that	
you were born with: e.g.	
RELAXED	
5. relaxed and not pretending to be sb/sth	(line 20)
Different: e.g.	
PARENTS/CHILDREN	
6. [only before noun] (of parents or their	
Children) related by blood: e.g.	
7. [only before noun] (old use or formal)	(line 25)
(of a son or daughter) born to parents who	
are not married	
((SYN)) ILLEGITIMATE e.g.	
BASED ON HUMAN REASON	

PED screen

8. [only before noun] based on human	(line 30)
reason alone: e.g.	
IN MUSIC	
9. used after the name of a note to show	
That the note is neither SHARP nor FLAT.	
The written symbol is : e.g.	(line 35)
Noun	
PERSON	

Figure 2: The main entry for the search for NATURAL

Compare the think-aloud protocol that illustrates C2's exploration of all senses of the word NATURAL:

> คำต่อมา (*the next headword*) **n-a-t-u-r-a-l r-e-c-o**… อ๋ะ! (aha) **natural resource** … ไม่มีจริงๆ ด้วย เอาเป็น (*did not exist in the PED, so*) **natural** พอ. (*should be enough*).. **natural… natural** ที เป็น (*searching for*) **natural** เป็น (*that functioned as a*) **noun**… **compare you would you would expect**… **describe behaviour that is part of the character**… **natural** อะไรเนี ย นิสัยที มีตั งแต่เกิดหรอ(*what is this? Behaviour since you were born?*)... นิสัยที มีตั งแต่เกิด ไม่ใช่(*I don't think it is this sense*).. **human an ability**… **skill** ที มีตั งแต่เกิด(*that you born with*) **ability** ที มีตั งแต่เกิด (*that you were born with*)... นี ไง(*oh..here it is*) **relaxed and not pretending to be… only before noun…**
>
> The participant paused for a while so I interrupted and asked, *what are you thinking?*
>
> (*this sense is used before*) ใช้ได้เฉพาะ **noun** เป็นได้เฉพาะ (*it can be*) noun อื่อ... **of parents or their children relate by blood** อ๋อ! (oh.. I see!) ___ **old use or formal born to parents who are not married… born parents who are not married… only before noun**… ___ …ตัวนี ไม่ใช่ (*not this sense*) **normal music** มาเกี ยวอะไร(*not relevant*) **a person who is** ไม่ใช่... (*no*) เออ (*ahh*) **based on human reason alone** ___ ไม่ใช่... ตัวนี ไม่ใช่(*no.. this not this sense either*)

On the other hand, B2 and E2 sometimes looked at only a few more lines beyond the initial PED screen. D2 was the only participant who only viewed the information that was available on the screen. It was observed that out of five lookups, none of the entries were explored beyond the first few lines.

Only A2 and C2 made use of the help options provided by their PEDs while using the English–English dictionary to review their summaries. All participants except D2 made minor changes to their summaries. Some added a new sentence; some deleted some words; and some used a new word to replace one used previously. They all agreed that the *OALD* was useful, but they still preferred to use the default English–Thai dictionary in the PEDs.

A2 was by far the most able to demonstrate knowledge and familiarity with the *CyberDict* PED form and *OALD* content. She exploited all subentries, used help options provided by the PED, and knew the abbreviations used in the dictionary. She knew how to highlight a word or a group of words in order to conduct cross searches and how to use the backspace button (deleting the preceding character). Inadequate knowledge of the PED features and specific knowledge of the particular conventions of the dictionary (in this case OALD) would lead users to make mistakes. An interview with B2 confirmed that she is the least efficient PED user. During this interview, it was established that she was unable to decode abbreviations, …,

Researcher: *Do you know abbreviations used in this dictionary (OALD)?*

B2: *Can you give me some examples?*

Researcher: *How about this one—sth?* I pointed to the signpost **~ (about/over sth)**

B2: *Oh sth. I don't know.*

Researcher: *You don't know?*

B2: *No.*

Researcher: *That's ok. Now look at the word PARTICULAR in the dictionary. Do you understand what this means in brackets?* **[only before noun]**

B2: *No.*

Researcher: *How about AmE?*

B2: *No, I don't.*

….

Researcher: *Do you know how to highlight a group of words, for example highlighting the word 'look' and 'up' at the same time?*

B2: *I don't know.*

Researcher: *Ok. When you spelt incorrectly, for example the word 'solution', it should be S-O-L-U but you accidentally typed S-O-L-E, do you know how to delete the 'E'?*

B2: *I really don't know. I have been trying to find this button since I first bought it but I can't find it. So I just type the words again.*

Researcher: *Have you read the manual?*

B2: *Yes. But I think I couldn't find it. And the manual was also in English.*

It is clear from the interview that participant B2 lacked 1) skills in using a dictionary in electronic format and 2) knowledge of conventions used in the OALD. She was not aware that a different part of speech would affect the meaning. She did not know abbreviations such as *sth.* And finally, she did not know how to find example sentences.

5. Summary and Conclusions

The findings of the Water experiments revealed that the manner in which the participants tackled the Thai reading passage had an effect on how they used their Thai–English dictionary in their PEDs, as well as on how they composed their summaries. The participants encountered problems at every step of the PED consultations and employed various strategies to tackle these problems. The findings revealed that some participants lacked adequate skills in PED use. It was found that the use of the English–English dictionary (OALD 6) in the PEDs helped some participants in the production task. Some participants lacked knowledge of PED features and also lacked knowledge of PED dictionary conventions.

This study has confirmed previous assumptions in many respects. The methodology utilised in this study enabled the author to discover exactly how many lines the participants looked up in their PEDs, and it was found that they tended to read only the information available on the PED screen. Few participants would scroll down to see more information. The same tendency to consider only the beginning of entries and ignore any other information has been observed by Wingate (2004) and Winkler (2001) with reference to other kinds of dictionaries. Investigating learners' use of print dictionaries, Wingate (2004) found her participants' lookup behaviour to be superficial and partial. Winkler (2001) reported that when using a dictionary on CD-ROM, her participants had difficulty scanning long entries to find particular details. Nevertheless, although the problem of failing to read beyond the first lines of a long entry seems to occur regardless of dictionary type, PED use is particularly problematic because the PED screen is so small that it severely limits the amount of information that is available at a glance.

This study has confirmed Midlane's findings concerning teachers' assumptions about the type of PED dictionary their students used (2005). In the **Water I** experiment, all 13 participants only used bilingual dictionaries to deal with the task, but in the **Water II** experiment the author had to invite five participants to review their summaries using a monolingual learner's dictionary. Although the participants acknowledged that using the English–

English dictionary had helped them write better summaries, they still stated that they would prefer to consult the bilingual dictionaries in their PEDs before turning to the English–English dictionary component. These findings illustrate Laufer and Kimmel's (1997) distinction between "dictionary usefulness" and "dictionary usability". Similar findings that learners considered their monolingual dictionaries to be very useful for language learning, but preferred to use bilingual dictionaries, are also reported by Taylor and Chan (1994) and Nesi (2003).

Although this study did not focus on look-up speed, it might be sensible to say that the speed with which a PED makes a lookup possible encourages participants to look up many words. This confirms the claims made in PED studies (e.g. Koyama and Takeuchi 2003, Weschler and Pitts 2000) that speed encourages more lookups. Stirling (2005) even claims that speed may encourage overuse. This could be the case for some participants in my study, especially Participant A, who conducted 85 lookups for 70 words in the **Water I** experiment (the average number of lookups per participant was 31). Participant A looked up 4 words twice and 2 words three times. This shows that Participant A relied heavily on his PED; in his case, the claim that PED encourages overuse seems to be true.

Some other wider issues related to successful and unsuccessful lookups concern the language proficiency and metacognitive knowledge of the participant. It is clear that the participants had different degrees of language ability although all of them were from the same year of study and the same foundation English course. The criteria in selecting the participants were possession of particular PED models and willingness to participate. The participants' language proficiency and metacognitive knowledge were not tested prior to the experiment, although this would naturally have had some effect on their PED skills and strategies.

Previous research (Oxford 2001, Liou 2000) shows that proficiency correlates with cognitive and metacognitive strategies. Participant A was the least proficient PED user, judging by his incoherent written summary He relied heavily on his PED and the way he approached the two reading tasks did not seem to involve any metacognitive knowledge of strategies. In contrast, B and H seem to have higher language proficiency, judging by their written products. Not only did they make use of their PEDs strategically, but they also involved a wide range of metacognitive strategies in dealing with the summary tasks. These findings correspond to those of previous studies. Liou (2000) found that an advanced student group spent less time, looked up fewer words and better understood the reading task than a lower language ability group. Also, the advanced student group did not only rely on dictionaries but also on other strategies, for example, guessing, making inferences, and using background knowledge. The findings are also in accordance with Fan's findings (2000) that high proficiency learners make fuller use of their dictionaries. Participants in this study who demonstrated greater knowledge

of English also reported using contextual meaning and information about appropriateness more often and regarded them as more useful than lower proficiency learners. Taking these findings into account, it may be appropriate to say that PED skills are to some extent affected by language proficiency and generic language learning abilities.

References

Al-Ajmi, H. 2002. Which Microstructural Features of Bilingual Dictionaries Affect Users' Look-up Performance? *International Journal of Lexicography* 15(2): 119-131.

Atkins, B.T.S. and K. Varantola. 1998. Monitoring Dictionary Use. Atkins, B.T.S. (Ed.). 1998. *Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators*. Lexicographica. Series Maior 88: 83-122. Tübingen: Max Niemeyer Verlag.

Boonmoh, A. 2003. *Problems with Using Electronic Dictionaries to Translate Thai Written Essays into English*. Unpublished M.A. Thesis, King Mongkut's University of Technology Thonburi, Bangkok.

Boonmoh, A. 2009. *The Use of Pocket Electronic Dictionaries by Thai Learners of English*. Unpublished Ph.D. Thesis, University of Warwick, UK.

Boonmoh, A. and H. Nesi. 2008. A Survey of Dictionary Use by Thai University Staff and Students, with Special Reference to Pocket Electronic Dictionaries. *Horizontes de Lingüística Aplicada* 6(2): 79-90.

Cohen, A.D. 1998. *Strategies in Learning and Using a Second Language.* New York: Addison Wesley Longman.

Deng, Y.P. 2005. A Survey of College Students' Skills and Strategies of Dictionary Use in English Learning. *CELEA Journal* 28(4): 73-77.

Diab, T. 1990. *Pedagogical Lexicography: A Case Study of Arab Nurses as Dictionary Users*. Lexicographica. Series Maior 31. Tübingen: Max Niemeyer Verlag.

Diab, T.A. and J. Hamdan. 1999. Interacting with Words and Dictionaries: The Case of Jordanian EFL Learners**.** *International Journal of Lexicography* 12(4): 281-305.

Ericsson, K.A. and H.A. Simon. 1980. Verbal Reports as Data. *Psychological Review* 87(3): 215-251.

Fan, M.Y. 2000. The Dictionary Look-up Behaviour of Hong Kong Students: A Large-scale Survey. *Education Journal* 28(1): 123-138.

Frankenberg-Garcia, A. 2005. A Peek into What Today's Language Learners as Researchers Actually Do. *International Journal of Lexicography* 18(3): 335-355.

Jopling, F. 2003. *Towards a New Methodology for Research on Electronic and Paper-based Dictionary Use: A Pilot Study of Look-up Patterns on Screen and in Print*. Unpublished M.A. Thesis. Coventry: University of Warwick.

Koren, S. 1997. Quality Versus Convenience: Comparison of Modern Dictionaries from the Researcher's, Teacher's and Learner's Points of View [online]. *Teaching English as a Second or Foreign Language* 2(3). Available: http://www-writing.berkeley.edu/TESL-EJ/ej07/a2.html [27 June 2006].

Koyama, T. and O. Takeuchi. 2003. Printed Dictionaries vs. Electronic Dictionaries: A pilot Study on how Japanese EFL Learners differ in Using Dictionaries. *Language Education and Technology* 40: 61-79.

Laufer, B. and M. Kimmel. 1997. Bilingualised Dictionaries: How Learners Really Use Them. *System* 25(3): 361-369.

Liou, Hsien-Chin. 2000. The Electronic Bilingual Dictionary as a Reading Aid to EFL Learners: Research Findings and Implications. *Computer Assisted Language Learning* 13(4-5): 467-476.

Matsumoto, K. 1993. Verbal-report Data and Introspective Methods in Second Language Research: State of the Art. *RELC Journal* 24(1): 32-60.

Midlane, V. 2005. *Students' Use of Portable Electronic Dictionaries in the EFL/ESL Classroom: A Survey of Teacher Attitudes*. M.Ed. Dissertation. Manchester: University of Manchester, Faculty of Education.

Mongphet, S. 2007. Appropriateness of Vocabulary Chosen from Thai–English Electronic Dictionaries in Writing. Unpublished M.A. Thesis, King Mongkut's University of Technology Thonburi, Bangkok.

Nesi, H. 2003. The Virtual Vocabulary Notebook: The Electronic Dictionary as Learning Tool. Paper presented at the *BALEAP Conference, Developing Academic Literacy*, University of Southampton 10–12 April 2003.

Nesi, H. and A. Boonmoh. 2009. A Close Look at the Use of Pocket Electronic Dictionaries for Receptive and Productive Purposes. Fitzpatrick, T. and A. Barfield (Eds.). 2009. *Lexical Processing in Second Language Learners:* 67-81. Clevedon, UK: Multilingual Matters.

Nesi, H. and R. Haill. 2002. A Study of Dictionary Use by International Students at a British University. *International Journal of Lexicography* 15(4): 277-305.

Oxford, R. 2001. Language Learning Styles and Strategies. Celce-Murcia, M. (Ed.). 2001. *Teaching English as a Second or Foreign Language.* Third edition. Boston, MA: Heinle & Heinle.

Paisart, W. 2004. *An Investigation of Dictionary Use to Facilitate Reading Comprehension*. Unpublished M.A. Thesis, KMUTT.

Sobkowiak, W. 2002. The Challenge of Electronic Learner's Dictionaries [online]. *Teaching English with Technology: A Journal for Teachers of English* 2(1). Available: http://www.iatefl.org.pl/call/j_article7.htm#sob [2006, May 27].

Stirling, J. 2005. The Portable Electronic Dictionary — Faithful Friend or Faceless Foe? *Modern English Teacher* 14(3): 64-72.

Tang, G. 1997. Pocket Electronic Dictionaries for Second Language Learning: Help or Hindrance? *TESL Canada Journal* 15(1): 39-57.

Taylor, A. and A. Chan. 1994. Pocket Electronic Dictionaries and Their Use. Martin, W., W. Meijs, M. Moerland, E. ten Pas, P. van Sterkenburg and P. Vossen (Eds.). 1994. *Euralex 1994 Proceedings, Papers Submitted to the 6th EURALEX International Congress on Lexicography in Amsterdam, The Netherlands*: 598-605. Amsterdam: VU University Amsterdam.

Tono, Y. 2001. *Research on Dictionary Use in the Context of Foreign Language Learning: Focus on Reading Comprehension.* Lexicographica. Series Maior 106. Tübingen: Max Niemeyer Verlag.

Weschler, R. and C. Pitts. 2000. An Experiment Using Electronic Dictionaries with EFL Students [online]. *The Internet TESL Journal* 6(8). Available: http://iteslj.org/Articles/Weschler-ElectroDict.html [3 September 2007].

Wingate, U. 2004. Dictionary Use — The Need to Teach Strategies. *Language Learning Journal* 29(1): 5-11.

Winkler, B. 2001. *Students Working with an English Learners' Dictionary on CD-ROM*. Papers from the ITMELT 2001 Conference. Available: http://elc.polyu.edu.hk/conference/papers2001/winkler.htm [6 January 2007].

Appendix A: Thai Reading Passage: Water

นำ ำ(อังกฤษ: water) เป็นของเหลวชนิดหนึ งซึ งถ้าบริสุทธิ จะไม่มีรส ไม่มีกลิ น และไม่มีสี
นำ ำเป็นของเหลวที มีอยู่มากที สุดบนผิวโลก
และเป็นปัจจัยสำคัญต่อการดำรงชีวิตของสิ งมีชีวิตทุกชนิดที มนุษย์รู้จัก เราสามารถพบนำ ำได้ในหลายๆ สถานที
อาทิ ทะเล ทะเลสาบ แม่นำ ำ ห้วย หนอง คลอง บึง และในหลายๆ รูปแบบ เช่น นำ ำแข็ง หิมะฝน ลูกเห็บ เมฆ
และไอนำ ำ

นำ ำมีสมบัติเป็นตัวทำละลายที ดีมาก เราจึงไม่ค่อยพบนำ ำบริสุทธิ ในธรรมชาติ
ดังนั นนำ ำสะอาดที เหมาะสมต่อการบริโภคของมนุษย์จึงเป็นทรัพยากรที มีค่ายิ ง
ในบางประเทศปัญหาการขาดแคลนนำ ำเป็นปัญหาใหญ่ที ส่งผลกระทบต่อสังคม
และเศรษฐกิจของประเทศนั นอย่างกว้างขวาง

นำ ำมีหลายรูปแบบ เช่น ไอนำ ำและเมฆบนท้องฟ้า คลื นและก้อนนำ ำแข็งในทะเล ธารนำ ำแข็งบนภูเขา
นำ ำบาดาลใต้ดินฯลฯ นำ ำเปลี ยนแปลงรูปแบบ สถานะ และสถานที ของมันตลอดเวลา
โดยผ่านกระบวนการกลายเป็นไอ ตกลงสู่พื นดิน ซึม ชะล้างและไหล
ก่อให้เกิดการหมุนเวียนของนำ ำบนผิวโลกเรียกว่าวัฏจักรของนำ ำ

เนื องจากการตกลงมาของนำ ำมีความสำคัญอย่างยิ งต่อการเกษตรและต่อมนุษย์โดยทั วไป
มนุษย์จึงเรียกการตกลงมาของนำ ำแบบต่างๆ ด้วยชื อเฉพาะตัวฝน ลูกเห็บ หมอก
และนำ ำค้างเป็นการตกลงมาของนำ ำที พบได้ทั วโลก แต่หิมะและนำ ำค้างแข็งมีเฉพาะในประเทศเขตหนาว
รุ้งเป็นปรากฏการณ์ที เกิดขึ นเมื อละอองนำ ำในอากาศต้องแสงอาทิตย์ในมุมที เหมาะสม

http://th.wikipedia.org/wiki/%E0%B8%99%E0%B9%89%E0%B8%B3
Accessed on February 5, 2006

My summary

Water is a liquid. Pure water is tasteless, odorless, and has no colour. Water is very important for living things. Water appears mostly in places such as seas, lakes, and ponds. Moreover, it can also appear in the form of snow, rain water or clouds. Water changes its form, state, and place all the time through complex processes. This results in water circulation on and above the surface of the earth. This phenomenon is called the 'Circle of Water'. Although we know that water is very important for living things, we rarely see pure water in nature. As a result, clean water which is suitable for humans is vital. In some countries, there are a lot of serious water problems and they can affect the society and economy of those countries.

Appendix B: Guideline Interview Questions

Reading a text in Thai and writing a summary in English (production)

— Can you explain how you normally use your pocket electronic dictionary for writing?

— For the first task, i.e. writing a summary in English, what did you do before you wrote?

— What were your difficulties in writing this paragraph, sentence, or word?

— I noticed you did this/that. Why?

— Did you encounter problems when you used your own PED dictionary for this task? If so, what were the problem(s)?

— How do you solve the problem(s)?

— Are there any differences when you used your electronic dictionary for this task and in your spare time?

Reading a text in Thai and writing a summary in English (production)

— How did you feel when you used an English–English dictionary to review this task?

— Is there any difference between using a Thai–English dictionary and an English–English dictionary? If so, what are the differences?

— I noticed you did this/that. Why?

— Did you encounter problems when you used your own PED for this task? Is so, what were the problem(s)?

— How did you solve the problem(s)?

— What changes have you made after consulting this dictionary? Why?

Die hantering van neweskikkers en onderskikkers in Afrikaanse woordeboeke

Nerina Bosman, *Departement Afrikaans, Universiteit van Pretoria, Pretoria, Suid-Afrika (nerina.bosman@up.ac.za)*
en
Anna N. Otto, *Departement Taal en Lettere, NMMU, Port Elizabeth, Suid-Afrika (anna.otto@nmmu.ac.za)*

Opsomming: Die diskrepansie tussen die behoefte aan leksikografiese leiding met betrekking tot voegwoorde en die relatiewe onverskilligheid hierteenoor in terme van leksikografiese navorsing en praktyk het tot hierdie artikel aanleiding gegee, waarin die onbevredigende hantering van voegwoorde in Afrikaanse woordeboeke aangedui word en enkele konstruktiewe leksikografiese oplossings vir die hantering van hierdie woordsoortkategorie aan die hand gedoen word.

'n Eerste aanbeveling is dat die lemmata *voegwoord, verbindingswoord, neweskikker*, *onderskikker* en *voegende bywoord* meer diepgaande sintaktiese inligting, met genoeg voorbeelde (ook oor sinsgrense heen) voorsien. Daar behoort kruisverwysings van die spesifieke voegwoordlemmata na hierdie lemmata te wees. Die voorbeelde wat aangebied word, behoort ook tipiese leksikale en grammatiese patrone aan te dui, sowel as of hipotaktiese binding moontlik is of net inlywing. In aanleerderwoordeboeke kan die tipiese leksikale patrone in vet druk verskyn. Sorg moet gedra word in omvattende woordeboeke, soos *WAT*, om vinniger inligtingsherwinning tot gevolg te hê en leksikograwe behoort nie funksies gelyk te stel aan polisemiese betekenisonderskeidings nie. So is daar byvoorbeeld twee lemmas nodig by **of** aangesien dit 'n homoniem is wat duidelik aparte lemmas vereis.

Sleutelwoorde: LEKSIKOGRAFIE, NEWESKIKKER, KORRELATIEWE NEWESKIKKER, ONDERSKIKKER, EENTALIGE WOORDEBOEK, HIPOTAKTIESE BINDING, INLYWING, KOMPLEMENTSINNE, GRAMMATIKALE LEIDING, LINGUISTIESE FUNDERING, WOORDORDE, KLOUSINTEGRASIE, FUNKSIEWOORD

Abstract: The Treatment of Coordinating and Subordinating Conjunctions in Afrikaans Dictionaries. Prompted by the discrepancy between the needs for lexicographic assistance with regard to conjunctions and the relative indifference concerning this in lexicographic research and practice, this study attempts to indicate the unsatisfactory treatment of conjunctions in Afrikaans dictionaries and to offer some constructive lexicographic solutions to the treatment of this part of speech category.

A first recommendation would be that the lemmata *voegwoord* (conjunction), *verbindingswoord* (connective), *neweskikker* (coordinating conjunction), *onderskikker* (subordinating conjunction) and *voegende bywoord* (conjunctional adverb) provide more in-depth syntactic information with enough examples (also across sentence boundaries). There should be cross-references from the specific con-

junction lemmata to these lemmata. The examples provided should indicate typical lexical and grammatical patterns and whether only hypotactic binding is possible or whether incorporation is also possible. In learners' dictionaries the typical lexical patterns can be in bold print. Care should be taken in inclusive dictionaries, like *WAT*, to promote faster information retrieval and lexicographers should not equate functions with polysemous meaning distinctions, e.g. there should be two lemmata for **of** as it is a homonym which clearly requires separate lemmata.

Keywords: LEXICOGRAPHY, COORDINATING CONJUNCTION, CORRELATIVE COORDINATING CONJUNCTION, SUBORDINATING CONJUNCTION, MONOLINGUAL DICTIONARY, HYPOTACTIC BINDING, INCORPORATION, COMPLEMENT SENTENCES, GRAMMATICAL GUIDANCE, LINGUISTIC GROUNDING, WORD ORDER, CLAUSE INTEGRATION, FUNCTION WORD

1. Agtergrond en probleemstelling

Sedert die verskyning van die twee artikels deur Gouws (1992 en 1998) oor neweskikking en die leksikografiese bewerking van neweskikkers, het min nog verander in die hantering van hierdie woorde in Afrikaanse woordeboeke. Indien 'n mens byvoorbeeld kyk na die groot aantal woordordefoute wat sowel moedertaalsprekers as niemoedertaalsprekers maak met betrekking tot hierdie woordsoortkategorie, dwing dit mens om weer indringend te kyk na die rol wat die leksikograaf moontlik kan speel om duideliker leiding te gee in hierdie verband. Soos die navorsers hieronder sal aantoon, is die huidige hantering van bepaalde verbindingswoorde in Afrikaanse woordeboeke tot 'n hoë mate nog ewe onbevredigend as in 1992. Daardeur word hulle taalkundige gehalte in die gedrang gebring (Gouws 1992: 103) en gee hulle nie die regte leiding aan gebruikers nie.

Die feit dat daar nie genoeg aandag geskenk word aan verbindingswoorde nie, kan miskien toegeskryf word aan die tradisionele onderskeid tussen grammatika en leksikon, die woordgerigte benadering en die tipiese mikrostruktuur van leksikale morfeme in Afrikaanse verklarende woordeboeke. Grammatikale aspekte kry relatief min aandag en die hooffokus van die definiensgleuf gaan gepaard met 'n sterk teenwoordigheid van semantiese inligting by veral leksikale morfeme. Alhoewel Béjoint (2000: 6) byvoorbeeld die presisering van betekenis as die belangrikste oogmerk van die leksikograaf beskou en die gemiddelde woordeboekgebruiker veral op soek is na betekenisinligting (Al-Kasimi 1977), is dit natuurlik ook so dat die leksikografiese praktyk met die insluiting van grammatiese inligting wel die noue integrasie van grammatika en betekenis erken. Hartmann (1982: 83) toon in dié verband aan dat woordeboekgebruikers dikwels juis grammatikale inligting by grammatikale morfeme (of funksiewoorde) soos voorsetsels en verbindingswoorde verlang. Sinclair (1984: 4) wys ook daarop dat grammatikale inligting deel van die fundamentele inligting is wat 'n woordeboek aanbied. Daar moet in gedagte gehou word dat verskillende tipes woordeboeke verskil betreffende die hoeveelheid

en diepte van inligting wat deur die teikengebruiker vereis of gevra word — vergelyk Gouws (1989: 209). Gouws merk elders op: "Die aard en omvang van die inligting word bepaal deur die tipe woordeboek maar ook deur die tipe leksikale item wat as lemma optree. (...) die leksikograaf (moet) besef dat verskillende lemmatipes telkens 'n eiesoortige bewerking moet kry" (Gouws 1992: 91).

Soos die navorsers verderaan sal aantoon, verdien verbindingswoorde se grammatikale gedrag meer aandag in Afrikaanse woordeboeke as wat hulle tans ontvang. Aangesien neweskikkers en onderskikkers so verskillend optree, kan hulle volgens Gouws (1992: 93) as aparte kategorieë beskou word en behoort hulle as sodanig in woordeboeke aangedui te word. Hierdie onderskeid word wel (maar nie deurgaans konsekwent nie) getref in die aanleerderwoordeboek *Basiswoordeboek van Afrikaans*, (voortaan *BA*) maar nie in die *HAT, WAT* of *Pharos Verklarende Afrikaanse Woordeboek* (voortaan *PVAW*) nie.

Tipies vir grammatikale morfeme is dit voorts moeilik om 'n betekenisdefinisie en fyner polisemiese onderskeidings vir verbindingswoorde aan te toon. Daar word tereg deur Gouws (1992: 98) gewys op die feit dat verbindingswoorde se belangrikste taalkundige kenmerke en funksie is dat hulle **voeg** (die navorsers se beklemtoning) en dat hoe en wat hulle voeg prominenter aangedui behoort te word in woordeboekartikels as byvoorbeeld betekenisinligting. Hulle **funksie** behoort dus meer prominensie te kry en dit sou selfs dié riglyn vir die strukturering van hulle woordeboekinskrywings kon wees.

Die probleem waarop hierdie artikel fokus, is dat 'n geslote klas soos verbindingswoorde sowel semanties as sintakties 'n bepaalde eiesoortige gedrag vertoon, maar dat hierdie eiesoortigheid nie tans bevredigend in woordeboeke verreken word nie. Ons poog in die artikel om die eiesoortige karakter van newe- en onderskikkers teoreties te belig en die huidige hantering van hierdie woorde in Afrikaanse verklarende woordeboeke te ondersoek. Ten slotte bied ons aanbevelings aan vir verbeterde leksikografiese praktyk.

2. Metodologie

Vir die teoretiese raamwerk wat ons voorstel onderlê, maak ons gebruik van insigte uit meer as een linguistiese benadering, insluitende die beskrywende en strukturele linguistiek, Halliday en Hasan se werk oor kohesie, die funksionele sistemiese grammatika en kognitiewe linguistiek, soos onder andere verteenwoordig deur Halliday en Hasan (1976), Ponelis (1979), Langacker (1987), Matthiessen en Thompson (1988), Gouws (1992), Bosch (1997), Bosch (1998), Verhagen (2001), Taylor (2002) en Halliday en Matthiessen (2004). Hierdie werkwyse is weliswaar taamlik eklekties, maar as sodanig nie ongewoon binne die leksikografiese tradisie nie — veral sover dit die aanvaarding en gebruik van verskillende erkende linguistiese konsepte betref. Op dié manier probeer ons om die linguistiese fundering van ons argument duidelik te maak — iets wat nie noodwendig altyd die geval is binne die leksikografie nie.

Teen hierdie agtergrond sal die hantering van verbindingswoorde in drie verklarende woordeboeke, naamlik *WAT*, *HAT* en *PVAW* en een aanleerderwoordeboek, *BA* krities ondersoek word.

Laastens sal sekere algemene beginsels en riglyne voorgestel word vir 'n alternatiewe en meer optimale leksikografiese hantering van verbindingswoorde as 'n aparte en unieke woordklas. Hierde beginsels sal met behulp van die verbindingswoorde d*at*¸ *of* en *óf … óf* geïllustreer word.

3. Teoretiese kontekstualisering

Tradisioneel word daar in die Afrikaanse taalkunde 'n kategorie *voegwoord* onderskei waaronder neweskikkers, onderskikkers en voegende bywoorde ingedeel word (vergelyk De Villiers 1983: 61). Die tipes "voegwoorde" word in die eerste plek onderskei op grond van die aard van die integrasie tussen die sinne wat verbind word. Daarmee gepaardgaande het die gebruik van 'n spesifieke woord bepaalde implikasies vir die woordvolgorde van die sin wat geïntegreer word.

3.1 Newe- en onderskikkers — terminologiese standpuntinname

In die Afrikaanse linguistiekliteratuur word 'n hele aantal terme gebruik wat essensieel dieselfde **funksie** aandui en dieselfde tipe grammatikale morfeem of funksiewoord beskryf, soos onder andere:[1] bindwoord, verbinder (Ponelis 1979: 313), verbindingspartikel (Bosch 1984: 18), verbindingswoord (Carstens 1997: 263), verbandswoord (Wybenga 1989: 189), en konjunksiemerker (Carstens 1997: 262). Carstens (1997: 263) meld ook terme soos *junctions, junctive expressions, connectives* en *connectors* wat in die Engelstalige literatuur gebruik word. Die bekendste term is sonder twyfel *voegwoord*,[2] alhoewel Ponelis (1979) reeds daarop wys dat *voegwoord* tradisioneel slegs gebruik word waar 'n bysin ingelyf word.

In navolging van Gouws (1998) stel die navorsers voor dat die term *voegwoord* vermy word. Dit word vervang met die meer neutrale term *verbindingswoord*. As gevolg van die fundamentele verskille (sowel semanties as sintakties) tussen die drie soorte verbindingswoorde wat tradisioneel onder dié term ingesluit word, sal ons so ver moontlik die terme *neweskikker* en *onderskikker* gebruik, asook die term *voegende bywoord* waar nodig.

3.2 Kort kenskets

Verbindingswoorde bewerkstellig integrasie tussen konjunkte. Klousintegrasie het enersyds te make met die graad van interafhanklikheid tussen konjunkte en andersyds ook met die logies-semantiese verhouding tussen hulle (Langacker 1987: 373).

Die aspek van verbindingswoorde wat die moeilikste is om te definieer, is hulle semantiese inhoud omdat dit afhanklik is van die hele konstruksie waarvan hulle deel vorm. Feitlik alle verbindingswoorde dui 'n semantiese verhouding tussen tekselemente aan. In die meeste gevalle kan mens poog om hierdie verhouding te definieer — aaneenskakelend, alternerend, teenstellend, kousaal, temporeel, redegewend en so meer. In (1) en (2) word 'n kousale verhouding byvoorbeeld aangedui:

(1) Die krieketwedstryd is gestop want dit het begin reën.

(2) Die krieketwedstryd is gestop omdat dit begin reën het.

Die verhouding tussen die konjunkte kan nie altyd onder woorde gebring word nie — die onderskikker *dat* is hier die prototipiese voorbeeld. Soos later in 'n paragraaf oor leksikografiese bewerking sal blyk, hou die gesprek rondom die semantiese status van verbindingswoorde implikasies in vir die leksikografiese praktyk (Gouws 1992).

Morfologies gesproke vertoon verbindingswoorde geen variasie nie. Samestellings met *dat* soos *voordat, nadat, totdat, omdat,* ensovoorts is hier die uitsondering.

As mens na die groep verbindingswoorde kyk, is twee aspekte meer prominent as ander, naamlik die funksie wat hulle verrig, ('n *voeg*funksie) en die invloed wat hulle op die sintaktiese struktuur van die sin uitoefen — wat op sy beurt die gevolg is van die graad van integrasie wat bewerkstellig word. Hulle hooffunksie is om woorde, frases of klouse te verbind en om kohesie tussen teksdele tot stand te bring.[3]

Daar moet in die leksikografiepraktyk deeglik kennis geneem word van die rol wat verbindingswoorde vervul om koherente stukke teks te produseer — nie alleen binne sinsgrense nie, maar ook in groter teksgehele. (Vir 'n bespreking van die sintaktiese funksies van voegwoorde binne sinsgrense versus diskoersfunksies oor sinsgrense heen, vergelyk Blühdorn (2008).) Sonder hierdie koherensie is 'n teks nie 'n teks nie. In Halliday en Hasan se definisie verteenwoordig verbindingswoorde semantiese skakels tussen elemente wat 'n teks opbou (Halliday en Hasan 1976: 226, 321). Daar is 'n aspek van teksbetekenis — die interpretatiewe verband — wat nie in terme van die betekenis van die dele van die teks beskryf kan word nie. Hierdie verband kan trouens tot stand gebring word sónder die gebruik van spesifieke leksikale items soos verbindingswoorde — koherensie kan bestaan sonder dat dit leksikaal gemarkeer is:

(3) Gee bietjie die tang aan, die spyker sit nogal vas. (voorbeeld van Verhagen 2001: 108).

In (3) moet die luisteraar self die korrekte verband tussen die konjunkte aflei, maar teksproduseerders kan hierdie interpretatiewe rol van die luisteraar verlig deur gebruik te maak van sekere linguistiese middele en verbindings-

woorde is byvoorbeeld so 'n linguistiese middel. Wanneer verbindingswoorde wel gebruik word, noem Verhagen (2001: 110), met verwysing na Matthiessen en Thompson (1988), die koherensierelasie 'n "gegrammatikaliseerde koherensierelatie".

In (3) hierbo kan die kousale verhouding deur *want* geleksikaliseer word:

(4) Gee bietjie die tang aan want die spyker sit nogal vas

maar ook byvoorbeeld deur *naamlik*:

(5) Gee bietjie die tang aan — die spyker sit naamlik nogal vas.

'n Bysin soos

(6) Ek weet dat hy siek is

bevat 'n posisiegleuf wat deur die onderskikker *dat* geleksikaliseer kan word. Afhanklike bysinne soos in (6) vereis 'n onderskikker; skoon bysinne soos

(7) Ek weet hy is siek

nie.

Wanneer verbindings sterk semanties gemarkeer is, deur byvoorbeeld 'n kousale of voorwaardelike verhouding, kan die verbindingswoorde nie weggelaat word nie; hulle moet geleksikaliseer word:

(8) Jy mag nie televisie kyk voordat jy jou huiswerk gedoen het nie.

*(9) Jy mag nie televisie kyk, _____ jy jou huiswerk gedoen het nie.

Sintakties gesproke kan verbindingswoorde gedefinieer word as woorde wat twee of meer elemente of konjunkte (sinne, frases, woordgroepe en woorde) verbind. Die aard van die verhouding tussen die konjunkte bepaal watter verbindingswoord gekies sal word, maar die keuse hou bepaalde sintaktiese gevolge in — meer hieroor in die paragraaf oor onderskikking. Verbindingswoorde kan óf beskou word dat hulle die graad van afhanklikheid tussen sinne **merk**, óf dat hulle optrede die **gevolg** is van die graad van afhanklikheid wat reeds deur die geïntegreerde sin uitgedruk word (Gouws 1998: 94). Op 'n strukturele vlak is dit die afhanklike aard van die bysin, gemerk deur afhanklike volgorde, wat die basis vorm vir die onderskeid tussen neweskikkers en onderskikkers.

Die term **taksis** kan as superordinaat gebruik word om die graad van interafhanklikheid tussen die klouse wat geïntegreer word aan te dui (Halliday en Matthiessen 2004: 374). Die terme **parataksis** en **hipotaksis** word gebruik om twee punte op 'n kontinuum aan te dui. Die graad van klousintegrasie is hier die maatstaf. Aan die een ent van die spektrum is daar neweskikking sonder 'n bindwoord en aan die ander eindpunt inlywing by byvoorbeeld 'n naam-

woordstuk. In die volgende uiteensetting maak ons veral gebruik van Ponelis (1979), Gouws (1998) en Taylor (2002: 430 e.v.) se uiteensettings.

3.2.1 Parataksis

Volgens Ponelis (1979: 603) is parataktiese konstruksies 'n geslote klas wat "bestaan uit twee of meer funksioneel gelyksoortige lede (konjunkte) wat simmetries (omkeerbaar) geskakel word deur 'n verbindingsmiddel (óf jukstaposisie óf 'n neweskikker)". In Afrikaans het alle klouse in 'n parataktiese konstruksie onafhanklike (hoofsin-) volgorde. Die spilwerkwoord neem die tweede posisie in die sin in, langs die onderwerp.

3.2.1.1 **Minimale integrasie** (Neweskikking sonder verbindingswoord)

Twee of meer klouse word gewoon jukstaposisioneel naas mekaar geplaas. Taylor (2002: 430) definieer dit as "a combinational device for lining up linguistic expressions":

(10) Hy het gekom, hy het gesien, hy het oorwin.

Die hoorder of leser lei self af wat die verband tussen die klouse is — kousaal of chronologies, byvoorbeeld. Die feit dat ikonisiteit 'n ontwerpkenmerk van taal is, veroorsaak dat (10) tipies (selfs uitsluitlik) sekwensieel geïnterpreteer sal word.

3.2.1.2 Neweskikking met neweskikker

Neweskikkers verbind sintakties gelyksoortige elemente. Elke klous kan ook onafhanklik van die ander optree:

(11) Hy kan nie die werk doen nie want hy is te oud.

Die klas neweskikkers verskil van die onderskikkers daarin dat daar by neweskikkers 'n verdere subkategorie, die korrelatiewe neweskikkers, onderskei kan word. Die prototipiese neweskikkingskonstruksie is, volgens Gouws (1998), nie-korrelatief en nie-jukstaposisioneel, gevorm deur e*n, maar, want, of* en *dog*. Die tipies Afrikaanse korrelatiewe neweskikkers is: é*n … én, óf … óf*, *nóg … nóg, hetsy … hetsy, sowel … as*, *beide ... en*, *ewemin … as*.[4]

(12) Jan kan óf huis toe gaan óf eers die werk klaarmaak.

Semanties onderskei die korrelatiewe neweskikkers hulle sistematies daardeur dat groter nadruk uitgedruk word deur hulle gebruik (Gouws 1992: 99).

3.2.2 Hipotaksis

3.2.2.1 Onderskikking

Onderskikkers integreer sintakties ongelyksoortige elemente — 'n onafhanklike sin (die *hoofsin*) en 'n afhanklike sin (die *bysin*). Die bysin word nie slegs verstaan in terme van sy verhouding met die hoofsin nie, maar is in der waarheid afhanklik van die hoofsin vir sy interpretasie. In Afrikaans word hierdie afhanklikheid duidelik gemerk — die woordorde word aangepas sodat die spilwerkwoord uitskuif na die sinsgrens:

(13) Hy word nie gestraf nie omdat hy onskuldig **is**.

Voorbeelde van onderskikkers is onder andere: *dat* (en *dat*-samestellings soos *nadat, totdat, voordat)*, *of*, *toe*, *alvorens*, *alhoewel*.

3.2.2.2 Komplementsinne

By komplementsinne word een klous ingebed in 'n ander:

(15) Ek het hulle sien inbreek.

Dit is 'n komplekse kwessie, veral as gevolg van die verskillende soorte sintaktiese konstruksies wat as komplemente kan optree. Infinitiefsinne (met byvoorbeeld *om, deur* en *ten einde* en die modale partikel *te)* tree tipies as sulke komplemente op en soos alle komplemente is hulle noodsaaklike aanvullings by die werkwoord. Vergelyk die volgende voorbeeld:

(16) Ek hoop om jou gou weer te sien.

Die bysin in (16) met sy afhanklike volgorde is 'n komplement by die hoofsin. Mens kan immers nie net sê: **Ek hoop* nie.

Geankerde komplementklouse word deur *dat* ingelei:

(17) Ek hoop dat ons mekaar gou weer sal sien.

Die hoogste graad van integrasie kry ons wanneer twee klouse saamsmelt tot een:

(18) Volvo's is baie duur om te onderhou.

'n Onderskeid word ook getref tussen klousintegrasie en inlywing (Halliday en Matthiessen 2004: 426), soos in

(19) Die gelukkige tye voor die oorlog begin het

waar 'n sin ingebed word by byvoorbeeld 'n naamwoordstuk (en nie by 'n ander klous nie).

Bogenoemde uiteensetting toon aan dat neweskikkers en onderskikkers dieselfde funksie en dikwels selfs ook dieselfde semantiese lading gemeen het, maar dat hulle duidelik nie tot dieselfde woordsoortkategorie behoort nie en dat ook subkategorisering by neweskikkers van belang is.

4. Leksikografiese bewerking

Aangesien onderskikkers sintakties gemerk is (in teenstelling tot neweskikkers) moet woordeboeke van hierdie basiese onderskeid rekenskap gee deur die kategorie waartoe die verbindingswoord behoort, duidelik aan te toon.

In die *WAT, HAT* en *PVAW* word neweskikkers byvoorbeeld nie kategories onderskei van onderskikkers nie — die woordsoort van beide tipes verbindingswoorde word bloot aangedui as *voegwoord*. Maar soos Gouws (1998: 90) met verwysing na Ponelis (1979) opmerk, is die verskille tussen onderskikkers en neweskikkers so ingrypend dat dit onverantwoordelik is om hulle as lede van dieselfde woordsoortelike kategorie te beskou.

Voorts behoort die eiesoortige semantiese aard van hierdie groep woorde ook neerslag te vind in die artikelstruktuur. Dit is te betwyfel of daar op dieselfde manier van polisemiese onderskeidings sprake is as by leksikale morfeme. Semantiese inligting behoort nie as deel van die definiens aangebied te word nie en die kernfokus behoort nie **betekenis**verklaring (ons beklemtoning) te wees nie (Gouws 1992: 99). Dat korrelatiewe neweskikkers byvoorbeeld duidelik nadruk uitdruk, behoort eksplisiet genoem te word. Meer grammatiese inligting as wat tans aangebied word, moet gegee word.

Gouws (1992) het reeds aangetoon dat dat daar nie voorsiening gemaak word vir 'n alternatiewe artikelstruktuur vir verbindingswoorde nie. Funksiewoorde vra om 'n ander artikelinkleding as leksikale morfeme met 'n eiesoortige mikrostrukturele bewerking. In die onderhawige uiteensetting stel ons 'n alternatiewe leksikografiese bewerking van verbindingswoorde in 'n aanleerderwoordeboek voor. Die beginsels wat toegepas word, is volgens ons egter ewe geldig ook vir verklarende woordeboeke. Aangesien die funksie van verbindingswoorde baie prominent aangedui behoort te word, stel ons voor dat die mikrostruktuur van die woordeboekartikel dit reflekteer.

4.1 Makrostrukturele bewerking

Grammatikale inligting kan op verskillende plekke in 'n woordeboek aangebied word. *Macmillan English Dictionary For Advanced Learners* bevat byvoorbeeld heelwat grammatikale inligting op die skutblad en middelteks (tussen *m* en *n*) van die woordeboek en daar is baie hiervoor te sê. In die *WAT* en die verklarende handwoordeboeke word 'n mate van grammatikale leiding gegee in

die gebruikersinligtinggedeelte. Om toegang tot inligting oor veral woordvolgorde-implikasies te vergemaklik, is 'n eerste aanbeveling dat die newe- en onderskikkerlemmas voldoende sintaktiese inligting sal verskaf met voorbeelde wat ook die gebruik oor sinsgrense heen sal illustreer.

Voegwoord as aanduiding van woordkategorie moet vermy word. In die gebruiksinligting kan daar verwys word na die vroeëre gebruik van die term.[5]

Sowel eenwoordneweskikkers as meerwoordige neweskikkers soos *sowel as* moet lemmastatus kry (vergelyk Gouws 1992).

Korrelatiewe neweskikkers moet as lemmas gelys word.

4.2 Mikrostrukturele bewerking

By elke newe- en onderskikker moet daar 'n kruisverwysing wees na die lemmata *neweskikker* en *onderskikker*. By die inskrywings onder hierdie lemmata moet daar genoeg leiding wees rakende die sintaktiese konsekwensies van dié tipe verbindingswoord, soos vervolgens aangedui sal word. (Dit geld ook voegende bywoorde wat nie in hierdie artikel bespreek word nie.) Wat voorbeeldsinne betref, moet daar afgewyk word van die bestaande praktyk. Die sinne moet die sintaksis en die gevolge van die gebruik van 'n newe- alternatiewelik onderskikker baie duidelik illustreer, en nie sinne wees waarin die gebruik van die trefwoord, naamlik *newe-* of *onderskikker*, geïllustreer word nie.

Neweskikker

'n Neweskikker verbind woorde (*penne, potlode en uitveërs*), woordgroepe (*blaffende honde en miaauende katte)* en sinne (*Die hond blaf en die kat miaau*) wat gelyke status het. Neweskikkers affekteer nie die volgorde van die sin wat volg op die neweskikker nie. Die sinne behou hulle onafhanklike woordorde en die eerste werkwoord in die sin verander nie van posisie nie. *Die hond blaf en die kat* ***miaau****.*

Die neweskikkers (enkelwoorde) in Afrikaans is: *en, maar, want, of, dog.*

Kyk ook: **korrelatiewe neweskikker.**

Korrelatiewe neweskikker

Korrelatiewe neweskikkers bestaan almal uit twee lede wat saam as neweskikker optree: *óf ... óf, én ... én, beide ... en, sowel ... as, nóg ... nóg, nie alleen ... nie, maar ook*

Óf Jan óf Piet gaan met Marie trou.

Nóg hy nóg sy suster was by die begrafnis.

Nie alleen het sy die eksamen gedruip nie, maar ook haar kar afgeskryf.

Onderskikker

'n Onderskikker lyf 'n bysin by 'n hoofsin in. Die bysin verkry afhanklike woordorde. Die posisie van die eerste werkwoord in die sin verander — dit skuif na agter:

Ons het geweet dat alles nie pluis ***was*** *nie.*

Aangesien dit vandag bitter koud ***is****, sal daar nie swemles wees nie.*

5. 'n Bespreking van *dat*

5.1 Die hantering van *dat* in *BA, WAT* en *HAT*

5.1.1 *BA*

In *BA* dui die voorbeelde sekere tipiese gebruike van *dat*-klouse en die hoofsinne wat hulle voorafgaan, aan, bv. *Dit is jammer dat die motor langs die pad ingegee het*. Hier word die tipiese patroon van 'n voorlopige onderwerp plus koppelwerkwoord plus adjektief plus res geïllustreer, maar nêrens word aangedui dat die voorlopige *dit* soms opsioneel is nie, bv. *Jammer dat ek laat is*. Twee voorbeeldsinne dui aan dat *dat*-klouse voorafgegaan kan word deur 'n hoofsin wat die spreker se opinie oor iets aandui. Dit sou waarskynlik voordeliger wees om eerder 'n ander tipiese patroon ook te illustreer, waar 'n feit aangedui word, bv. *Dit is 'n feit dat hy diabetes het.* Ander patrone, bv. die negatiewe bevelsvorm word wel geïllustreer, maar die voorbeelde in *BA* het geen vetgedrukte woorde in wat gebruikers kan help om tipiese leksikale patrone te identifiseer nie.

5.1.2 *WAT*

Die hoofprobleem met die hantering van *dat* in die *WAT* is die informasiedigtheid en die probleem om 'n spesifieke tipe gebruik van *dat* op te spoor. Dit sou sin maak om 'n kort uiteensetting van die hoofgebruike te gee, soos wat die *Macmillan Advanced Learner's Dictionary* met lemmas met 'n hoë informasiedigtheid doen. Op hierdie manier kan beide die gebruiker wat slegs 'n vinnige oorsig wil hê en die gebruiker wat meer intensiewe en tydsame studie wil doen, gehelp word. (Terselfdertyd moet verouderde gebruike in 'n nuwe uitgawe verwyder word.) Daar kan byvoorbeeld twee dinge oor *dat* genoem word voor die uitvoerige bespreking, naamlik dat *dat* as 'n onderskikker en aanwysende voornaamwoord gebruik kan word.

5.1.3 *HAT*

Alhoewel *HAT* nie so 'n hoë informasiedigtheid as WAT het nie, sal die gemiddelde gebruiker waarskynlik nie weet wat 'n naamwoordelike bysin en 'n bywoordelike bysin beteken nie en hulle sal waarskynlik slegs baat vind by die voorbeelde wat tipiese gebruik illustreer.

5.2 Voorgestelde bewerking in 'n gevorderde aanleerderwoordeboek

dat — onderskikker [Kyk *onderskikker*]
Dat word as 'n onderskikker gebruik om twee sinne met mekaar te verbind. Die

werkwoord verskuif na agter in die sin, bv. *Ek het nie geweet dat hy in die tronk **was** nie*.

1. Dat word **gebruik om 'n stelling, feit, idee of rede in te lei:** 1.1 gebruik om in 'n stelling te wys wat iemand sê, dink, glo, ens.: *Sy sê dat sy nie lekker voel nie.▪ Ek dink dat sy poging heel goed was.* ▪ *Jan glo dat sy vrou ontvoer is.* ▪ *Daar is gesuggereer dat ek die wasgoed moet was.* 1.2 gebruik om 'n sin te begin wat 'n feit aandui: *Ons kan nie die feit ignoreer dat so baie jongmense werkloos is nie.* ▪*Dat niemand haar in die hospitaal besoek het nie, is 'n skande.* ▪ **Dit is verstommend/verbasend/interessant/belangrik dat:** *Dit is verstommend dat hy elke keer dieselfde fout maak. Dit was belangrik dat hy warm gehou moes word*. 1.3 gebruik om te verduidelik hoekom iemand bly, hartseer, kwaad, ensovoorts voel: *Ek is só bly dat ek jou weer raakgeloop het. Hy was so kwaad dat hy amper die skelm geskiet het. Ek is jammer dat ek laat is.*

   Wanneer **dat** nie 'n sin begin nie, word dit dikwels weggelaat, veral as mens praat: *Ek sê mos ek is nie lus nie.* Let op die verskuiwing van die werkwoord na agter in die sin wanneer **dat** wel gebruik word.

2. gebruik na **so** of **sulke** om die resultaat van iets aan te dui: *Sy woorde was so kwaai dat sy begin huil het.* Kyk **so ... dat** en **sulke ... dat**.

   Die onderskikker **dat** word dikwels weggelaat in uitdrukkings met **so** of **sulke**, veral wanneer mens praat. *Sy was so bang — sy kon nie beweeg nie*.

6. 'n Bespreking van *of* en *óf … óf*

6.1 Die hantering van *of* in *BA, HAT, WAT* en *PVAW*

Om mee te begin, verskil ons van Bosch (1997: 35) wat stel: *"Of* kan beide as neweskikker (Engels *or*) en as onderskikker (Engels *if*, *whether*, *as if*) funksioneer". Selfs meer bedenklik is haar gevolgtrekking nadat sy die neweskikker *of* bespreek het: "Of ... tree op in die tussengebied tussen neweskikking en onderskikking en (is) die enigste voegwoord wat uitgebreid ook suiwer onderskikkende verband kan aandui" (Bosch 1997: 43). Hierdie standpunt, naamlik dat daar net een "voegwoord" *of* is, word weerspieël in die manier waarop *of* in die *WAT*, die *HAT* en *Pharos VAW* hanteer word — 'n werkwyse wat leksikologies, kategoriaal en semanties onaanvaarbaar is (Gouws 1992: 93). Die drie genoemde woordeboeke hanteer die twee *ofs,* wat duidelik lede is van 'n homonimiese paar, as een polisemiese lemma met as kategorie-aanduiding *voegwoord*, of, in die geval van *PVAW*, as 'n *modaliteitswoord* (wat dit ook mag beteken). In die mikrostruktuur van die genoemde woordeboeke word 'n hele aantal polisemiese waardes gelys asof hierdie waardes aan mekaar verwant is. *BA* volg hier baie beter leksikografiese praktyk deur sowel die neweskikker *of* as die onderskikker as aparte lemmas te lys en hulle woordsoortkategorie ook duidelik so aan te dui.

Kruisverwysing in die verklarende woordeboeke onder bespreking is ook problematies. Die lemma **voegwoord** is vanselfsprekend nog nie in die *WAT* opgeneem nie, maar ook in die *HAT* en *PVAW* sal 'n gebruiker min of selfs geen leiding hier kry nie (die lemma *modaliteitswoord* verskyn nie in *PVAW* nie). *BA* gebruik nie die term as aanduider van die woordsoort nie en het daarom ook nie nodig om dit as lemma op te neem nie.

In *WAT, HAT* en *PVAW* word die funksie telkens aangedui as sou dit saamhang met 'n polisemiese onderskeiding. Die *WAT* begin byvoorbeeld elke polisemiese onderskeiding met die frases *ter verbinding van* of *ter inleiding van*. Dit is duidelik dat dit besonder moeilik is om die betekenis van *of* telkens te parafraseer en baie lang en omslagtige verduidelikings volg. Daar word wel gebruik gemaak van die tipiese verbande wat deur verbindingswoorde uitgedruk word. Waar dit moontlik is, word 'n sinoniem of woordgroep gegee — vergelyk byvoorbeeld 1 a, 1 b en 3 a by **of** in *WAT* Deel XI.

Daar moet baie goed besin word oor die uitdrukkings wat in die *WAT* onder die lemma **of** opgeneem is. Die korrelatiewe neweskikker *óf ... óf* behoort aparte lemmastatus te kry, soos alle neweskikkers van hierdie tipe. *Of so* en *of wat* (in die betekenis "by benadering geskat") kan opgeneem word by die neweskikker ***¹of*** (4b) en *of .. of (nie)* by die onderskikker *of (4).*

6.2 Voorgestelde bewerking van *of* in 'n gevorderde aanleerderwoordeboek

¹of neweskikker [Kyk *neweskikker*]

Of word gebruik om woorde, sinsdele of sinne te verbind.

1. 'n Teenstellende of alternatiewe verband word uitgedruk.

a. Daar moet tussen moontlikhede of keuses gekies word:
Drink jy tee of koffie?
Jy kan leer vir die eksamen of televisie kyk.

In 'n lysie word *of* voor die laaste moontlikheid gebruik: *Wil jy tee, koffie of 'n sappie hê?*

Die dele wat verbind word, kan omgeruil word:
Drink jy koffie of tee?
Jy kan televisie kyk of leer vir die eksamen.

Vir groter nadruk kan *of* vervang word met die korrelatiewe neweskikker óf ... óf: *Jy kan óf per bus óf per trein reis.* *Jy kan óf tee óf koffie óf sap drink.*

b. Daar is meer as een moontlikheid, maar die een hoef die ander nie uit te sluit nie.[6]
Elke middag kuier sy by ons of ons by haar.
Kan ek vir jou tee of koffie gee?

Die dele wat verbind word, kan omgeruil word:
Kan ek vir jou koffie of tee gee?

c. Dit dui aan dat die spreker onseker is:
Sy gaan met Jan of Piet trou, ek weet nie mooi met wie nie.

2.a. Dit druk eendersheid van betekenis uit met byvoorbeeld 'n sinoniem of vervangende woorde:
Die vakgebied staan as chemie of skeikunde bekend.

b. Dit korrigeer, verduidelik of verklaar 'n woord of frase, **of liewer**:
Die tekeninge of ruwe sketse is baie haastig gemaak.

3. Dit verbind 'n opdrag of bevel met 'n dreigement:
Staan of ek skiet!

Die twee dele van hierdie soort verbinding kan nie omgeruil word nie: ** Ek skiet of jy staan!* *of* kan hier vervang word met **of anders** *Julle sal nou moet begin, of anders gaan julle nie klaarkry nie.* *Staan, of anders sal ek skiet.* Die deel na *of anders* kan weggelaat word: *Bly stil, of anders …*

4.a. Dit verbind 'n naamwoord en telwoord sodat 'n onbepaalde tydsperiode of hoeveelheid uitgedruk word:
'n dag of twee; 'n stuk of tien; 'n keer of drie

b. Dit verbind met **so** of **wat** sodat 'n onbepaalde tydsperiode of hoeveelheid uitgedruk word:
Gee my net 'n minuut of wat.
Dit het onder die 100 of so papiere op haar lessenaar gelê.

5. Dit word gebruik om iets of iemand in te sluit in 'n negatiewe stelling, **en nie**:
Ek het vandag nog nie nat of droog oor my lippe gehad nie.

IDIOMATIESE UITDRUKKINGS (met hulle verklarings)
Sy het nie kind of kraai nie
Sonder om te blik of te bloos
Ek kan nie kop of stert uitmaak van wat hy sê nie

6. Dit word gebruik om 'n rede te gee vir jou opinie:
Hy het geen idee waar die plek is nie, of hy sou nie so verkeerd gery het nie.

7. Dit dui aan dat iets normaalweg die geval is:
Hy is nooit in 'n wedstryd nie, of hy speel vuil.

²of onderskikker [Kyk *onderskikker*]

Of lei 'n bysin in by 'n hoofsin.

1. Die bysin druk twyfel of onsekerheid uit en beteken "wat is die moontlikheid dat?". Dit volg dikwels na die werkwoord *wonder*:
Hy wonder of hy die eerste rugbyspan sal haal.
Of hy nog president sal word, kan ek nie sê nie.

2. *of* lei 'n vraag in wat bevestigend of ontkennend beantwoord kan word:
Weet jy al of julle die naweek see toe sal gaan?

3. *of* lei bysinne in wat vergelyking uitdruk, **asof**:
   Dit volg tipies na 'n koppelwerkwoord soos *lyk, voel, maak, is, blyk, voorkom, klink, voel, smaak, ruik*.

   Die bysin na die koppelwerkwoord kan soms afhanklike en soms onafhanklike volgorde hê:
   *Hy lyk of hy flou **wil** word.* (afhanklike volgorde)
   *Hy lyk of hy **wil** flou word.* (onafhanklike volgorde)
   maar nie altyd nie.
   Dit voel vir my of die winter vanjaar besonder vroeg gekom het. (net afhanklike volgorde)
   Dit is of ek jou net gister laas gesien het. (net afhanklike volgorde)

4. Saam met *of (nie)* dui dit aan dat iets sal gebeur of so sal wees al gebeur enigeen van die alternatiewe:
   Of julle vandag vertrek of môre, julle sal nog steeds betyds wees.
   Gee julle antwoorde in of julle nou klaar is of nie.
   Die sinne kan verkort word en die *of (nie)* kan weggelaat word:
   Wen of verloor, die span is reeds in die eindstryd. [Vergelyk: *of hulle môre sal wen of verloor ...*]
   Verwaand of nie, ek hou van hom [in plaas van *of hy verwaand is of nie ...*].

[3]**of** neweskikker [Kyk *neweskikker*]

Of verbind twee sinne

1. Die sin met *of* volg na 'n stelling en is 'n vraag:
   *Hy is weg, of **weet** jy dit nie?*

   Let hier op die inversie (omruiling) van die onderwerp en werkwoord na *of*.

2. Die sin met *of* volg na 'n stelling en bevraagteken of korrigeer dié stelling:
   Ons sal ten minste gedeeltelik die pad van die res van Afrika loop — of so meen oorsese sakelui.
   Sy oefen in die geheim. Of so het sy gedink.

6.3 Die korrelatiewe neweskikker *óf ... óf*

Hierdie neweskikker het wel lemmastatus in *HAT* en *BA*, maar nie in *PVAW* of *WAT* Deel XI nie. Soos al die ander korrelatiewe neweskikkers, behoort dit apart hanteer te word. Die belangrikste semantiese inligting wat gegee behoort te word, is dat die keuse beklemtoon word.

7. Gevolgtrekking

Hoewel Gouws (1992) reeds op verskeie leemtes in die hantering van voegwoorde in verskillende eentalige woordeboeke gewys het, het baie min van sy voorstelle tot hul reg gekom in Afrikaanse woordeboeke. In hierdie artikel is van sy voorstelle herbevestig deur 'n teoretiese begronding en daarop voort-

gebou deur onder andere voorstelle vir die hantering in 'n aanleerderwoordeboek van die onderskikker *dat* en die homonimiese verbinder *of*.

Aantekeninge

1. Vergelyk Carstens (1997: 263) vir 'n uitvoerige lys terme.
2. Vergelyk hier byvoorbeeld Carstens en Van de Poel (2010: 338) wat dié term nog in sy tradisionele betekenis en sonder kommentaar gebruik as 'n soort superordinaat vir sowel neweskikkers as onderskikkers en voegende bywoorde, as sou hulle onproblematies tot dieselfde woordsoortelike kategorie behoort.
3. Die gebruik van verbindingswoorde is maar een van die maniere waarop kohesie in 'n teks bewerkstellig kan word. Ander kohesiemerkers is byvoorbeeld verwysing met behulp van anafore, tydadjunkte soos *Iater, toe, eers* en *intussen*, verskillende tipes leksikale kohesie (herhaling, sinonimie, ens.) en diskoersmerkers soos gambiete, byvoorbeeld.
4. Gouws (1992: 95) onderskei in der waarheid vier subtipes hier.
5. Ons is bewus daarvan dat hierdie term in skoolhandboeke en deur die Departement van Basiese Onderwys gebruik word. Dit is onses insiens egter reeds lankal nodig dat daarop gewys moet word dat baie van die terminologie wat deur die Departement gebruik word verouderd is en dikwels ook verkeerd verstaan word deur die outeurs van sommige skoolhandboeke. Die voorbeeld van *BA* in hulle "Wenke vir die gebruiker" kan eerder nagevolg word.
6. In 'n elektroniese woordeboek kan die verskil in intonasie tussen eksklusiewe disjunksie 1(a) en nie-ekslusiewe disjunksie 1(b) met 'n klankgreep geïllustreer word. In 1(b) sal daar neutrale (gelyke) klem op die twee konjunkte wees, in 1(a) sal elk van die moontlikhede beklemtoon word.

Verwysings

Woordeboeke

Gouws, Rufus et al. (Reds.). 1994. *Basiswoordeboek van Afrikaans.* Pretoria: J.L. van Schaik.

Labuschagne, F.J. en L.C. Eksteen. 2010. *Pharos Verklarende Afrikaanse Woordeboek.* Kaapstad: Pharos Woordeboeke.

Odendal, F.F. en R.H. Gouws. 2005. *HAT. Verklarende Handwoordeboek van die Afrikaanse Taal*. Kaapstad: Pearson Education South Africa.

Rundell, Michael en Gwyneth Fox (Reds.). 2002. *Macmillan English Dictionary for Advanced Learners*. Oxford: Macmillan.

Schoonees, P.C. (Hoofred.). 1956. *Woordeboek van die Afrikaanse Taal. Tweede Deel: D–F.* Derde Oplaag. Pretoria: Die Staatsdrukker.

Van Schalkwyk, D.J. (Hoofred.). 2000. *Woordeboek van die Afrikaanse Taal. Deel XI: N–O.* Stellenbosch: Buro van die WAT.

Ander bronne

Al-Kasimi, A.M. 1977. *Linguistics and Bilingual Dictionaries.* Leiden: E.J. Brill.

Béjoint, H. 2000. *Modern Lexicography: An Introduction*. Oxford: Oxford University Press.

Blühdorn, H. 2008. Subordination and Coordination in Syntax, Semantics and Discourse. Fabricius-Hansen, Catherine en Wiebke Ramm (Reds.). 2008. *'Subordination' versus 'Coordination' in Sentence and Text:* 59-85. Amsterdam: Benjamins.

Bosch, Barbara. 1984. 'n *Sintaktiese ondersoek na die gebruik van onderskikkende voegwoorde en sinsverbindende woorde in Afrikaans*. Ongepubliseerde doktorale proefskrif, Rhodes Universiteit, Grahamstad.

Bosch, Barbara. 1997. Die voegwoord *of*. *Suid-Afrikaanse Tydskrif vir Taalkunde* 15(2): 35-44.

Bosch, Barbara. 1998. Die onderskikker *dat*: 'n korpus-gebaseerde bespreking. (Deel 1). *Suid-Afrikaanse Tydskrif vir Taalkunde* 16(4): 120-126.

Carstens, W.A.M. 1997. *Afrikaanse tekslinguistiek: 'n inleiding*. Pretoria: J.L. van Schaik.

Carstens, W.A.M. en Kris van de Poel. 2010. *Teksredaksie*. Stellenbosch: SUN MeDIA.

De Villiers, M. 1983. *Afrikaanse grammatika vir volwassenes*. Goodwood: Nasou.

Gouws, R.H. 1989. *Leksikografie*. Pretoria en Kaapstad: Academica.

Gouws, R.H. 1992. Die leksikografiese bewerking van neweskikkers. *Tydskrif vir Geesteswetenskappe* 32(2): 91-103.

Gouws, R.H. 1998. Neweskikkers in Afrikaans: sintaktiese en semantiese opmerkings. Gouws, Rufus en Ilse Feinauer (Reds.). 1998: 89-100.

Gouws, Rufus en Ilse Feinauer (Reds.). 1998. *Sintaksis op die voorgrond*. Pretoria: J.L. van Schaik.

Halliday, M.A.K. en R. Hasan (Reds.). 1976. *Cohesion in English*. Londen: Longman.

Halliday, M.A.K. en C.M.I.M. Matthiessen. 2004. *An Introduction to Functional Grammar*. Derde Uitgawe. Londen: Arnold.

Hartmann, R.R.K. 1982. Das zweisprachige Wörterbuch im Fremdsprachenerwerb. Wiegand, H.E. (Red.). 1982: 73-86.

Hartmann, R.R.K. (Red.). 1984. *LEXeter '83 Proceedings*. Tübingen: Max Niemeyer.

Langacker, R.W. 1987. *Foundations of Cognitive Grammar* Vol. I. Stanford: Stanford University Press.

Matthiessen, C. en S.A. Thompson. 1988. The Structure of Discourse and 'Subordination'. Haiman, J. en S.A. Thompson. (Reds). 1988. *Clause Combining in Grammar and Discourse:* 275-329. Amsterdam: John Benjamins.

Ponelis, F.A. 1979. *Afrikaanse Sintaksis*. Pretoria: J.L. van Schaik.

Sinclair, J.M. 1984. Lexicography as an Academic Subject. Hartmann, R.R.K. (Red.). 1984: 3-12.

Taylor, J.R. 2002. *Cognitive Grammar*. Oxford: Oxford University Press.

Verhagen, A. 2001. Terug naar *want* en *omdat*. Dongelmans, B.P.M., J.A. Lalleman en O.J. Praamstra (Red.). 2001. *Kerven in een rots. Opstellen over Nederlandse taalkunde, letterkunde en cultuur, aangeboden aan Jan W. de Vries bij zijn afscheid als hoogleraar Dutch Studies aan de Universiteit Leiden:* 107-119. Leiden: Stichting Neerlandistiek Leiden.

Wiegand, H.E. (Red.). 1982. *Germanistische Linguistik* 3-6/80. Hildesheim: Georg Olms Verlag.

Wybenga, D.M. (Red.). 1989. *Taalkompas: elke student se gids*. Pretoria: Serva.

Lexicography and the Relevance Criterion

Theo J.D. Bothma, *Department of Information Science, University of Pretoria, Pretoria, South Africa (tbothma@up.ac.za) and visiting professor, Centre for Lexicography, Aarhus, Denmark*
and
Sven Tarp, *Centre for Lexicography, Business and Social Science, University of Aarhus, Aarhus, Denmark (st@asb.dk) and visiting professor, Department of Information Science, University of Pretoria, Pretoria, South Africa*

Abstract: Until recently, lexicography and information science could rightly be considered two disciplines which had developed along parallel lines but with no or very little formal relation between them. Although the two disciplines developed in almost complete isolation from each other, during the last few years it has nevertheless become increasingly clear that they have a lot in common. This trend began within lexicography which started viewing lexicographical works as a special kind of tool designed to be consulted in order to obtain information. Upon this basis, it has been suggested that lexicography should be considered a part of information science and, hence, integrated into it (cf. e.g. Bergenholtz and Bothma 2012, Tarp 2009). It is evident that this integration of two hitherto independent disciplines with long traditions of their own is not something to be solved overnight and neither can it be a unilateral process.

This article will explore the concept of relevance in both disciplines in more detail and show, at the hand of examples from lexicographical tools, how the theoretical frameworks of both disciplines can complement one another. This will be done within the framework of the function theory of lexicography, as discussed in the many works of Tarp and Bergenholtz (e.g. Bergenholtz and Tarp 2002) and others, and relevance theory in information science as defined by Saracevic (1975, 1996), Cosijn and Ingwersen (2000) and others.

Keywords: LEXICOGRAPHY, FUNCTION THEORY, COGNITIVE SITUATIONS, COMMUNICATIVE SITUATIONS, OPERATIVE SITUATIONS, INTERPRETIVE SITUATIONS, PRE-LEXICOGRAPHICAL PHASE, INTRA-LEXICOGRAPHICAL PHASE, POST-LEXICOGRAPHICAL PHASE, INFORMATION SCIENCE, RELEVANCE THEORY, TOPICAL RELEVANCE, COGNITIVE RELEVANCE, SITUATIONAL RELEVANCE, SOCIO-COGNITIVE RELEVANCE, AFFECTIVE RELEVANCE

Opsomming: Leksikografie en die relevansie-kriterium. Tot onlangs kon leksikografie en inligtingkunde tereg gesien word as twee dissiplines wat langs parallelle lyne ontwikkel het, maar met min of geen formele verhouding tussen hulle nie. Alhoewel die twee dissiplines in bykans volkome isolasie van mekaar ontwikkel het, het dit gedurende die afgelope aantal jare al hoe meer duidelik geword dat hulle baie in gemeen het. Hierdie tendens het begin met leksikografie wat begin het om leksikografiese werke te sien as 'n spesiale tipe hulpmiddel ("tool") wat ontwerp is om geraadpleeg te word met die doel om inligting te bekom. Op grond hiervan is daar

voorgestel dat leksikografie as deel van inligtingkunde gesien behoort te word, en gevolglik daarin geïntegreer behoort te word (sien bv. Bergenholtz en Bothma 2012, Tarp 2009). Dit is duidelik dat die integrasie van die twee tot-dusver onafhanklike dissiplines met lang tradisies van hulle eie nie iets is wat oornag opgelos sal kan word nie en dat dit nie 'n eensydige proses kan wees nie.

In hierdie artikel word die konsep van relevansie in beide dissiplines bespreek en word daar aan die hand van voorbeelde van leksikografiese hulpmiddels aangetoon hoe die teoretiese raamwerke van die twee dissiplines mekaar kan komplementeer. Dit sal gedoen word binne die raamwerk van die funksieteorie in leksikografie, soos bespreek in die talle werke van Tarp en Bergenholtz (bv. Bergenholtz en Tarp 2002) en andere, en relevansie-teorie in inligtingkunde soos gedefinieer deur Saracevic (1975, 1996), Cosijn en Ingwersen (2000) en andere.

Sleutelwoorde: LEKSIKOGRAFIE, FUNKSIETEORIE, KOGNITIEWE SITUASIES, KOMMUNIKATIEWE SITUASIES, OPERATIEWE SITUASIES, INTERPRETATIEWE SITUASIES, PRE-LEKSIKOGRAFIESE FASE, INTRA-LEKSIKOGRAFIESE FASE, POST-LEKSIKOGRAFIESE SITUASIE, INLIGTINGKUNDE, RELEVANSIETEORIE, TOPIKALE RELEVANSIE, KOGNITIEWE RELEVANSIE, SITUASIONELE RELEVANSIE, SOSIO-KOGNITIEWE RELEVANSIE, AFFEKTIEWE RELEVANSIE

1. Introduction

Until recently, lexicography and information science could rightly be considered two disciplines which had developed along parallel lines but with no or very little formal relation between them. Information science came into the world as an independent discipline in the first half of the 20th century and has experienced an enormous upsurge during the last decades due to, among other things, the rapid development of the corresponding technologies. Lexicography, on the other hand, is a thousand year-old cultural practice which has inevitably resulted in a large accompanying literature of academic reflections but with no systematic theory building until the 20th century, i.e. more or less the same period as information science was founded and flourished. Although the two disciplines developed in almost complete isolation from each other, during the last few years it has nevertheless become increasingly clear that they have a lot in common. The process was started by the trend within lexicography which in the final analysis viewed lexicographical works as a special kind of tool designed to be consulted in order to achieve information. Upon this basis, it has been suggested that lexicography should be considered a part of information science and, hence, integrated into it (cf. Tarp 2009, Leroyer 2011). While analyzing the problems related to innovation in e-lexicography, Bergenholtz and Bothma (2011: 74) conclude:

> One of the reasons for the lack of innovation in e-lexicography is that lexicography is usually treated as a part of linguistics and lexicographical tools are primarily compiled by specialists with linguistic background. Our main thesis, namely that lexicography is not a part of linguistics but a part of information science does not support this line of thought.

It is evident that this integration of two hitherto independent disciplines with long traditions of their own is not something to be solved overnight and neither can it be a unilateral process. Tarp (2011: 56) formulates the challenge in the following way:

> In reality, what we are dealing with is one big discipline embracing all types of consultation tools designed to meet punctual information needs, a discipline which may be considered an integrated part of information science (...) In this respect, lexicography, on the one hand, *has a lot to contribute* to other theories dealing with punctual consultation tools and to information science in general, and on the other hand, *has a lot to learn* from these theories and this science.

Gouws (2011) supports this vision and stresses that learning from each other also means "unlearning", i.e. the capacity to recapitulate and view old stuff in a new light. The road ahead is still long and full of obstacles as there are still a lot of questions of common interest that have to be analyzed and solved. In this respect, Bothma (2011) as well as Bergenholtz and Bothma (2011) have dealt with the needs-adapted data presentation in e-information tools and have shown how information science may contribute to solve this complex problem which is shared by both disciplines. Similarly, in the following contribution we will discuss the criterion of relevance in a lexicographical light in order to see how this criterion — mainly but not exclusively developed in the field of information science — may be applied to lexicography and which consequences this application may have for the concept of relevance itself.

As we shall see, within lexicography the relevance criterion is already widely used, most frequently when the lexicographical process is analyzed in its various phases and important lexicographical decisions have to be taken in relation to specific dictionary projects and data categories. However, until now no attempt (that we are aware of) has been made to systematize and classify the different types and dimensions of relevance used. This may be considered a theoretical shortcoming with practical implications, especially in the present moment where lexicographical products are gradually being placed on electronic platforms requiring much more scientific stringency in all aspects in order to be high quality.

An obvious problem in the mutual approximation of two disciplines developed in isolation from each other — although sharing an overlapping subject field — is that they almost inevitably express themselves in different terminologies using different terms to express concepts that are more or less identical. For instance, when some lexicographic schools (Wiegand 2000, 2002, Tarp 2008a, 2009) employ the term *data* to denote what is selected and presented by lexicographers in dictionaries, information scientists would call it *information* although both parts would agree that what is finally retrieved from these data or information by the users of dictionaries is *information*. Such differences should not be an obstacle to a still closer collaboration between scholars from the two fields. Hence, in this contribution we have opted for the lexico-

graphic terminology whereas in the next contribution it may be the other way around.

2. The function theory: basic elements and relevance

According to the latest research, lexicography has been a social and cultural practice for about four thousand years and has resulted in, at least, hundreds of thousands of dictionaries, encyclopedias, thesauruses and other types of lexicographical works covering almost all spheres of human activity and knowledge and using various sorts of media, from clay over papyrus and paper to modern electronic platforms. Within this immense discipline various general and specific theories of different scopes have been elaborated, especially during the last decades (cf. Tarp 2010). One of the very few general theories is the theory of lexicographical functions, henceforth referred to as the function theory. This theory is built upon the presumption that dictionaries and other lexicographical work are above all *utility tools* conceived and produced with the genuine purpose of satisfying specific types of human needs, i.e. *information needs*, existing in one or several individuals in society (cf. Bergenholtz and Tarp 2002, 2003, 2004, 2005, Tarp 2008a and b).

In order to determine the nature of the information needs relevant to lexicography, it is first of all necessary to establish a distinction between *global information needs*, i.e. the needs related to a more profound study of a specific subject field (or part of it), and *punctual information needs*, i.e. restricted and limited needs related to a single or limited topic within a larger subject field, or to the solution of specific tasks or problems. In this respect, dictionaries and other lexicographical works are par excellence *consultation tools*, i.e. artifacts designed to be consulted in order to meet punctual information needs in contrast to global information needs which may be satisfied by other types of artifacts or texts produced with a view to being read and studied from one end to another (cf. Tarp 2011).

However, it is important that the relation between the concepts of global and punctual needs should not be viewed as an absolute opposition but rather in a linear perspective as a *relation between the whole and the part*. In this respect, the lexicographical tradition shows several comprehensive works which treat what is considered global in other lexicographical works, as "punctual" and a part of a bigger whole. The big French *Encyclopédie, ou Dictionnaire raisonné des sciences, des arts et des métiers* is one such example. Here the user may find many relatively long articles, e.g. d'Alembert's (1754) article on dictionaries, which provide a "global" introduction to a specific field of knowledge but still considered within the global world knowledge (encyclopedia) and made accessible through a lexicographical structure. Another example is the Chinese *Yongle Dadian* from 1408 which was produced in order to collect and provide access to all knowledge existing in China at that moment and which among its 11,095 volumes included several already existing books on various topics which were

incorporated, i.e. rewritten, in their totality into this gigantic lexicographical work and made accessible through a rhyming system for the characters as well as a complex system of indexes.

According to the function theory, the types of information needs relevant to lexicography should never be considered abstract needs, but specific and even concrete needs which are determined by the types of *potential users* of lexicography's practical works as well as the types of extra-lexicographical *situations* where lexicographically relevant information needs may occur, i.e. needs that may be satisfied by lexicographical tools. Of these *two determining factors*, the most important is the extra-lexicographical situation which also determines how many of the various lexicographically relevant characteristics of a potential type of user are relevant in each case. This also means that user needs should always be understood in their intimate relation to a specific type of situation and that they can never be defined only based upon the characteristics of a potential user however relevant these characteristics may be. At this stage, the function theory works with *four fundamental types of lexicographically relevant situations* (cf. Tarp 2008a):

1. Communicative situations where a need to solve a communication problem may occur.
2. Cognitive situations where a need for knowledge may occur.
3. Operative situations where a need for instructions on how to perform a physical or mental action may occur.
4. Interpretive situations where a need to interpret and understand a sign, signal, symbol etc. may occur.

The *communicative situations* are those that so far have been most studied, and they may be further subdivided into a number of situations such as text production and text reception in the mother tongue (or first language), text production and text reception in a foreign (or not-first) language, translation from mother tongue into a foreign language and vice versa, text revision etc. The needs that may occur, even for the same type of user, in these situations vary a lot: in text reception it may be the need to understand a word, in text production it may be a need for information about a word's syntactic properties, in translation it may be a need for an equivalent etc.

The *cognitive situation*s may also be divided into various sub-situations, e.g. when someone needs to know something in order to perform a task (a journalist writing about Napoleon needs to know his day of birth and then most probably forgets it), when someone for whatever reason wants to know something specific and add it to his or her general or specialized knowledge, and when someone needs to know something specific related to a global study of a specific area of knowledge. The information needed in the two first cases

may probably be of the same type for the same type of user, although for different purposes, whereas the information needed in the third case may also include references that relate the small topic to the bigger area of knowledge to which it belongs.

The *operative* and *interpretive situations* have so far been scarcely studied by lexicography and there are only relatively few lexicographical works that cater for these situations. They are, however, extremely interesting because they are situations which lexicography has in common with other references works — e.g. user guides, how-to's and manuals — where users seek information in form of instructions in order to perform certain tasks or in form of explications of signs, signals, symbols, sounds, lights etc. in order to interpret them and determine whether something important and relevant is taking place and whether it is necessary to take action upon this basis.

The function theory claims that in order to produce high-quality lexicographical works and tools it is necessary to study the extra-lexicographical situations where information needs occur because these needs may vary considerably from one situation to another, even for the same type of user (cf. Tarp 2008b). This, of course, does not mean that the users and their characteristics are disregarded by the theory. A user with specific characteristics may have quite different information needs than another user with other characteristics, even when the needs are occurring in the same type of situation. A mother-tongue speaker or an advanced learner of English will definitely have other needs than a learner at a beginner's level when writing an English text. An expert of biotechnology will need other types of information than a layman when consulting a specialized dictionary in order to get knowledge about a specific topic related to this discipline etc.

What the function theory claims at this point is that the relevant user characteristics depend on the situations where the relevant needs occur. First of all it should not be forgotten that any person may have an infinite number of characteristics of which most are lexicographically irrelevant, e.g. that a potential user is left-handed. Secondly, even the lexicographically relevant characteristics are not always relevant, i.e. relevant in each and every situation when information needs may occur. For instance, Danish users' different proficiency levels in English or knowledge of biotechnology are not relevant at all when they read a novel of Hans Christian Andersen and may need to have some old-fashioned words explained etc. but these levels will be highly relevant when they either want to be better informed about biotechnology or get assistance to produce an English text. To this end, the function theory has elaborated a *list of lexicographically relevant user characteristics* — an open list to which new characteristics can be added — from which the characteristics that are relevant to each type of extra-lexicographical situation can be selected. In this respect, the function theory works with a *set of variables* that have to be taken into account when determining the specific type of information needed in each case (cf. Tarp 2008b).

According to the function theory, the user's information need is met by a corresponding set of *lexicographical data* which have been selected (manually or automatically from a corpus etc.), elaborated and prepared (by the lexicographer) and made accessible for consultation. The selection, elaboration and preparation of the data may be performed following various techniques and methods but in order to produce high-quality lexicographical works it is necessary that these techniques and methods in the final analysis are built upon the *criteria of relevance* determined by the specific type of information need in question.

When an individual person experiences an information need, he or she may then *access the data* and *retrieve the needed information from these data* through a complex cognitive process (which the theory does not study). Finally, the information retrieved may be used in various ways in order to *satisfy the user's specific type of need*, e.g. to solve a communicative or cognitive problem, to store it as knowledge, to perform a task or to interpret a sign, signal, symbol etc. This is, at least, what has happened in lexicography until a few years ago, where lexicographical works have only provided direct access to lexicographically selected, elaborated and prepared data, and not to data that have been prepared and made accessible elsewhere, e.g. in books and archives. Only recently, a few advanced lexicographical tools have tried to *reuse* already existing data made available through a data base or the Internet, and one of the visions today is not only to reuse these data but also to *repackage* and even *recreate* them *adapting* them to the specific information needs of the users in each situation (cf. Bothma 2011 and Tarp 2011).

As it has been indicated above, the function theory does not only study the processes taking place from the moment the user starts a consultation process to the moment where the needed information is retrieved from the lexicographical data, i.e. the intra-lexicographical consultation processes. It also studies the extra-lexicographical processes taking place immediately before and after the consultation process. The reason for this approach is double: on the one hand it is necessary to know in which situation the lexicographically relevant information needs occur in order to determine the nature of these needs, and on the other hand, it is absolutely necessary to evaluate the post-lexicographical process in order to establish an objective criterion for success or failure instead of the subjective one used by questionnaires and the like. In this respect and according to the function theory, a "normal" lexicographical process is made up by the following phases:

1. a pre-lexicographical phase where a user with specific characteristics finding him or herself in a specific extra-lexicographic context or situation:
    a. experiences an information need,
    b. becomes aware of the information need,
    c. and decides to start a lexicographical consultation;

2. an intra-lexicographical phase where the user:
   a. selects the relevant lexicographical information tool,
   b. accesses the relevant data,
   c. verifies that he or she has found the right data, i.e. relevant to the information need in question,
   d. and retrieves the needed information from the data;
3. a post-lexicographical phase where the user:
   a. makes use of the retrieved information in order to solve a communicative or cognitive problem, to store it as knowledge, to perform a task or to interpret a sign, signal, symbol etc.

Here it is important to emphasize that the above ideal process presupposes that the user is actually aware of the information need and decides to take lexicographical action. In fact, when a lexicographically relevant information need occurs for an individual in any pre-lexicographical situation, this individual is automatically transformed into a potential user which may:

1. not be aware of the information need and therefore not proceed to any lexicographical consultation;
2. be aware of the information need but not proceed to any lexicographical consultation because he or she thinks — maybe based upon previous lexicographical experience — that the need cannot be solved by means of a lexicographical consultation or that this consultation may require too much time and trouble;
3. be aware of the information need and proceed to a lexicographical consultation but based upon a wrong idea and understanding of the real nature of the information need; or
4. be aware of the real nature of the information need and proceed to a lexicographical consultation.

In an ideal world, lexicographical tools should be able to meet the user's information needs in all four cases, not only for the *actual user* in case 4. In case 3, for instance, a lexicographical tool could by means of various advanced interactive techniques and methods guide the user in such a way that he or she finally will get the really needed information and not "the right answer to the wrong question". But even in cases 1 and 2 where the potential users for one reason or another do not by themselves start a consultation, there are already lexicographical tools available with solutions for the information needs occurring in specific types of situation in a digital environment. In this case, the above model will have to be transformed as follows:

1. a pre-lexicographical phase where the information need occurs but where the potential user for one reason or another does not take lexico-

graphical action;

2. an intra-lexicographical phase where the lexicographical tool:
   a. detects the problem,
   b. and suggests a solution,
3. a post-lexicographical process where the user accepts the proposed solution and uses it in order to solve the hitherto unknown problem.

As mentioned, such advanced lexicographical tools already exist, e.g. Word's Spelling and Grammar Program which, when activated, detects problems and suggest solutions in relation to written text production in a digital environment.

In such cases, the intra-lexicographical process only takes a few seconds or even less. This should be compared with the above process where the user himself has to take lexicographical action, something which frequently may take several minutes. If, for instance, a professional translator needs to make 50 lexicographical consultations in a normal workday and these take an average of 5 minutes, this means 250 minutes, or more than four hours, of consultation, which for most people make up more than half a workday. This constitutes some rather expensive production costs (Nielsen 2008 calls them "lexicographical information costs") and, as such, a terrible waste of time. In this respect, the *time factor* — translated into a quick and easy data access and information retrieval — becomes another important criterion of lexicographical quality and relevance.

To sum up, this section has provided a brief presentation of the basic elements of the function theory and an introduction to the most important contexts where this theory refers to a *relevance criterion*. Relevance is first and foremost used to determine the types of *information needs* that are relevant to lexicography, i.e. those which can be satisfied by lexicographical works. It is then used to determine the *extra-lexicographical situations* and *user characteristics* relevant to lexicography in general and to a concrete lexicographical work in particular. It is also used to determine the *data categories* as well as the *specific data* needed to satisfy a specific type of information need. And finally, it is used with reference to the *time factor*, i.e. the duration of the intra-lexicographical process in terms of data access and information retrieval.

In all these cases, relevance is centered on the information needs and their satisfaction and combines extra- and intra-lexicographical elements. However, it must be admitted that nowhere in the lexicographical literature known to the authors of this contribution, is it possible to find a definition or a more extended discussion of the criterion of relevance in terms of lexicography.

3. Information science and the relevance criterion

In information science research, Saracevic has, already in 1996, stated that "[n]obody has to explain to users of IR [Information Retrieval] systems what relevance is, even if they struggle (sometimes in vain) to find relevant stuff. People understand relevance intuitively" (Saracevic 1996; see also Saracevic 1975). However, Saracevic and others (e.g. Borlund 2000, Cosijn and Ingwersen 2000, Borlund 2003, Cosijn 2003, Cosijn and Bothma 2005, Ingwersen and Järvelin 2005) agree that there are many dimensions to relevance. Ingwersen and Järvelin 2005: 389, for example, defines relevance as

> the perceived topicality, pertinence, usefulness or utility, etc., of *information sources*, made by *cognitive actor(s)* or algorithmic devices, with reference to an information situation […] It can change dynamically over time for the same actor. Relevance can be of a low order objective nature or of higher order, i.e., of subjective multidimensional nature. Its measurement can be binary or graded.

Information sources could be information objects or humans as information sources (Ingwersen and Järvelin 2005: 386). An information object could be any physical or digital entity in a variety of media that could provide potential information and could refer to documents, texts, images and any other media (Ingwersen and Järvelin 2005: 385 and 386).

Two basic classes of relevance are distinguished, viz.

— objective relevance which is system-based (algorithmic or system relevance), and

— subjective relevance which is user-based and which can then be further subdivided.

Algorithmic or system relevance describes the relationship between a search query and the information objects. It "may be measured in terms of the comparative effectiveness of logical and/or statistical similarity of features inferring relevance"; it is system-oriented "because the success of the relation is entirely dependent on a given procedure or algorithm, and the intent behind it. Both the query and the objects contain identical/similar features, such as words and other strings of signs, image colour or author name" (Cosijn and Ingwersen 2000: 537, 539). A document is therefore relevant in terms of the search string when it contains identical or similar words as used in carrying out the search. Information retrieval systems may be exact match systems (typically Boolean systems) or best match systems. "Traditional Boolean systems facilitate binary relevance judgements, whereas best match systems, or a combination of best match and Boolean systems, are able to rank retrieved information by relevance" (Cosijn and Bothma 2005: 50). It is important to note that relevance ranking in all such systems is still systems-based, i.e. on the basis of algorithms

and statistical analysis the system decides which documents are more relevant than others and should be ranked higher.

All other cases of relevance judgments are user-based, i.e. a user evaluates whether a document is relevant in terms of the information need, and not the retrieval system.

User-based relevance categories as defined in Borlund 2000, Cosijn and Ingwersen 2000, Borlund 2003, Cosijn 2003, Cosijn and Bothma 2005 and Ingwersen and Järvelin 2005 are:

— topical relevance,

— cognitive relevance or pertinence,

— situational relevance and

— socio-cognitive relevance;

— affective relevance.

Each of these types of relevance will be discussed below.

Topical relevance or topicality is defined as "the relation between the topic of the query and the topic of the assessed information objects" (Cosijn and Bothma 2005: 50). Topicality deals with the "aboutness relationship between document contents retrieved and request, as assessed by a person" (Ingwersen and Järvelin 2005: 237). ("Aboutness" refers to what a text or document, an image or any other information object is about and refers to the topic it discusses; it could be an inherent feature of the object as recorded in the metadata, or it could be dependent on the individual who interprets the information object.) Borlund calls this type of relevance "intellectual topicality", to "distinguish the subjective type of topical oriented relevance from algorithmic relevance" (Borlund 2003: 915). According to Ingwersen and Järvelin topical relevance signifies "the relationship between the aboutness of information objects and the aboutness of requests as perceived by an actor (whether task performer, searcher or judge in IR experiments). Owing to the human assessment (interpretation) this type of relevance is of subjective emotional and intellectual nature" (Ingwersen and Järvelin 2005: 391).

Cognitive relevance or pertinence is "measured in terms of the relation between the state of knowledge, or cognitive information need of the user, and the information objects as interpreted by that user. The criteria by which pertinence is inferred are cognitive correspondence, informativeness, novelty and information preferences" (Cosijn and Bothma 2005: 51). "Pertinence represents the intellectual relation between the intrinsic human information need and the information objects as currently interpreted or perceived by the cognitive state of an assessor or user" (Borlund 2003: 915). It is dependent on the novelty value of the information for the user, i.e., to what extent it adds new information in the specific user situation. (In standard English "pertinence" and "relevance" are

synonyms, but in relevance theory in information science both words are regarded as technical terms with distinct meanings, with "relevance" the broader term and "pertinence" a narrower term.)

Situational relevance "describes the relationship between the perceived situation, work task or problem at hand and the usefulness of the information objects as perceived by the user. The criteria by which situational relevance is inferred are usefulness in decision-making, appropriateness of information in problem solving and the reduction of uncertainty" (Cosijn and Bothma 2005: 53). According to Borlund (2000) "the judgement of situational relevance embraces not only the user's evaluation of whether a given information object is capable of satisfying the information need, it offers also the potential of creating new knowledge which may motivate change in the decision maker's cognitive structures. The change may further lead to a modification of the perception of the situation and the succeeding relevance judgement, and in an update of the information need". Situational relevance is therefore understood "as the utility or usefulness of the viewed and assessed information object(s) by pointing to the relationship between such retrieved object(s) and the work task at hand underlying the information need as perceived by the user. Situational relevance is a highly context dependent as well as a potentially dynamic type of relevance" (Borlund 2003: 915).

Socio-cognitive relevance describes "the relationship between the situation, the work-task or problem at hand in a given socio-cultural context on the one hand, and the information objects on the other, as perceived by one or more cognitive agents. The social or organizational domain, or cultural context in which the individual finds himself is defined by a paradigm, which dictates what problem explanations may be found to be acceptable" (Cosijn and Bothma 2005: 53). Cosijn and Ingwersen (2000: 549) state that socio-cognitive relevance "is highly context dependent and associated with organizational strategies or scientific community interaction within".

A further type of relevance is listed by Saracevic (1996), viz. *motivational or affective relevance.* Many researchers, however, currently regard this as "forming a natural part of all the subjective relevance categories" (Ingwersen and Järvelin 2005: 237), as also explained by Cosijn and Bothma, who state that affective relevance can be described "in terms of the relation between the goals, intents and motivations of the user and the information objects. Affective relevance should not be seen as the ultimate subjective relevance in a scale of relevances, but rather as another dimension of relevance judgments that may be associated with the other subjective types of relevance" (Cosijn and Bothma 2005: 55), as well as Borlund, who agrees that "motivational/affective relevance is a characteristic of all of the subjective types of relevance" (Borlund 2003: 915).

The interrelationship between work task performance, search task performance and relevance is illustrated in Figure 1, taken from Cosijn and Bothma 2005: 48.

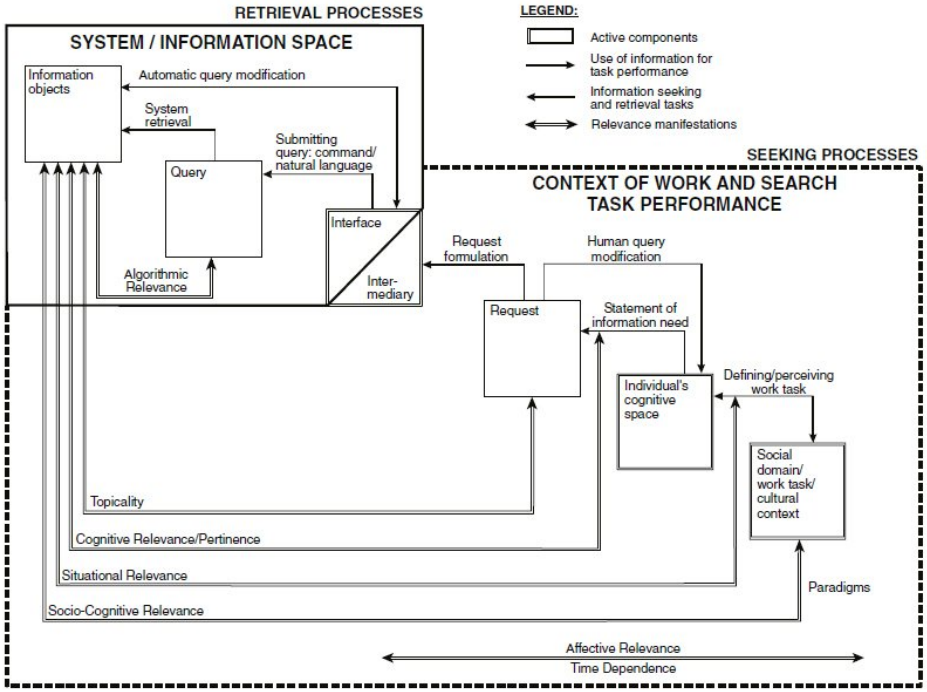


Figure 1: "Interactive Information Retrieval: Work task performance, search task performance and relevance types", as illustrated in Cosijn and Bothma 2005: 48

Relevance theory therefore studies the relationship between a human user (an actor) and an information object (e.g. a text document or other sources, as explained in section 3) in a given situation or context. Specific data may be relevant or not relevant for a specific user in a specific situation based on a combination of objective and subjective criteria. Objective relevance is based on a match between search terms and information objects (in the case of text-based information objects): if the search terms occur in the document, the document is deemed relevant. Whether the document is actually useful in terms of the information need is, however, a different matter. The document may not contain the required amount of detail, or too much detail; the data may be too complex (e.g. aimed at an expert when the user is a lay person) or not complex enough (e.g. a popular discussion of a topic when detailed, technical information is required). The data may be in a language with which the user is not sufficiently familiar. The data may not necessarily be relevant in a given situation or context because the problem that the data addresses or solves are not related to the specific work task the user is involved in at that stage. The data may be biased in terms of a specific theoretical or ideological framework with which

the user may not be sufficiently familiar or with which he or she does not agree. There may therefore be a number of factors that influence a user's perception about how relevant the data are that has been identified by the system as relevant — even though the system may indicate that the data are (objectively) relevant, the user's information need may not necessarily be satisfied.

The question is to what extent these relevance criteria can be mapped to the function theory as discussed in section 2, specifically referring to the three phases of the "normal" lexicographic process. This will be attempted in the following section by discussing a number of examples from various dictionaries.

4. Illustrative example

In this section we will discuss relevance in terms of the information needs which a non-native speaker of English may experience when engaged in written text production and the corresponding lexicographical data required to meet these information needs. However, initially it is necessary to underline that the potential user, i.e. the non-native speaker of English in question, may experience two fundamental types of need when transforming him or herself from a potential into an actual user of a learner's dictionary, i.e. the need directly related to written text production and the need related to the consultation process itself. According to Tarp (2008b: 152-153), a learner of a non-native language (L2) engaged in written text production may have the following information needs directly related to the production process:

- Information about L2 words:
- orthography
  - word class
  - genus (not relevant for English)
  - pragmatic and cultural restrictions
  - inflection
  - word formation
  - syntactic properties
- Information about L2 two collocations
- Information about L2 idioms
- Information about L2 proverbs
- Information about L2 derivates
- Information about L2 synonyms, antonyms, hyponyms etc.

As to the needs exclusively related to the consultation process, the actual user may need access via the native language (especially when he or she is not an advanced learner) as well as information that may verify that he or she has

actually found the right lemma, idiom or proverb (cf. Tarp 2008b: 153):

— Information about L2 words:
— meaning
— orthography (in case of variant)
— word class
— genus (not relevant for English)
— Information about meaning of L2 idioms
— Information about meaning of L2 proverbs

It goes without saying that these two lists may be extended and further specified and that they in any case should be adapted to the language in question (writing in African languages would for instance present other types of information needs), but the above mentioned information needs are without any doubt the most important in terms of problems related to written text production in English as a non-native language. They are attributes of relevance which may vary as a function of the actual user's proficiency level and as such they require specific lexicographical data in order to be satisfied. In this respect, it is important to stress that the above lists contain the hypothetical needs that may occur in a specific *type* of extra-lexicographical situation, i.e. written text production. In a *concrete* situation and subsequent *concrete* consultation, a dictionary user will only very rarely experience all these information needs, but as a rule only one or a few of them in any possible combination. This means, on the one hand, that any printed dictionary with static articles will have to include lexicographical data designed to meet all these needs if it really wants to be an information tool in terms of written text production in a non-native language, and, on the other hand, that electronic dictionaries should consider incorporating an access system that allows only the required data to pop up on the screen in a concrete and specific consultation, i.e. providing dynamic articles adapted to the specific and concrete information needs of the user in any extra-lexicographical situation (cf. Bergenholtz and Bothma 2011, Bothma 2011, Tarp 2011).

In the following we will look at the online *Oxford English Dictionary* as well as one of the "big five" English learners' dictionaries taken at random, the *Cambridge Learner's Dictionary*, and see how it meets the user's information needs in terms of written text production which is one of its declared functions.

On its web site, the first of these two dictionaries describes itself as follows:

> The Oxford English Dictionary (OED) is widely regarded as the accepted authority on the English language. It is an unsurpassed guide to the meaning, history, and pronunciation of 600,000 words — past and present — from across the English-speaking world. As a historical dictionary, the OED is very different

> from those of current English, in which the focus is on present-day meanings. You'll still find these in the OED, but you'll also find the history of individual words, and of the language — traced through 3 million quotations, from classic literature and specialist periodicals to films scripts and cookery books (http://www.oed.com/public/about).

Already from this description it seems clear that the *Oxford English Dictionary* is not the most adequate dictionary to assist a foreign learner of English having text-production problems. If this user, for instance, needs some appropriate collocations with the word *table* and accesses the corresponding article then the dictionary will come up with an article containing almost 35,000 words covering about 90 pages in a MS Word document. Although the needed collocations may be found among all these words, this is a typical example of information overload which may take the user on a long odyssey until the information need is finally satisfied, if ever, unless the user had a cognitive information need and wanted to know as much as possible about the specific word. If one therefore searches for the word "table" in the *OED*, the above-mentioned article will be relevant at the system or algorithmic level, but not at any other level if the user does not have a cognitive information need. However, if the *OED* were to offer the option of searching for and displaying only specific types or subsets of data (e.g. only grammatical features of the word concerned) the retrieved data could be relevant at the topical and situational levels as well. For this to be possible, it would require that the data be structured in such a way that only specified subsets of data be displayed and that the search interface allow the user to specify exactly which subset(s) of data should be displayed (cf. Bothma 2011).

A quite different situation will meet the user if he or she decides to consult the *Cambridge Learner's Dictionary* which, according to its own Introduction, is "the perfect dictionary for the intermediate and upper-intermediate learner of English". Although the concept of intermediate learner is controversial and highly imprecise in terms of lexicography as shown by Tarp (2008b: 138-141), there is little doubt that this concept in one way or another refers to a learner with a limited English vocabulary for which reason access through the native language will frequently be required in relation to text production. In this very important aspect, the *Cambridge Learner's Dictionary* does not provide the necessary tool to assist the user and therefore cannot be considered the "perfect dictionary" for the user group in question. If we abstract from this "little" problem and proceed directly to the dictionary's lemma list, we will among tens of thousands other articles find the following one:

> **anticipate** /æn'tisipeit/ verb [T] to expect something, or to prepare for something before it happens *to anticipate a problem* 0 [+ **that**] *We anticipate that prices will fall next year*.

This article seems to include the necessary data needed by the user to verify

that he or she has arrived at the right article, i.e. meaning and word class of *anticipate*, as there is no orthographic variant and no proverb or idioms of relevance to explain. Apart from that, meaning is provided based upon a controlled and restricted vocabulary of 2,000 common English words making it much easier for the intermediate learner to understand. As to the information needs related directly to text production it provides information about orthography, pronunciation (not relevant for written production), word class and syntactic properties, but not about collocations, derivates (e.g. *anticipation* which can, however, be found in the subsequent article), and synonyms (*expect* is only provided implicitly in the short definition). Neither does it provide information about inflection, but as *anticipate* is a regular verb it may be expected that the envisaged group of intermediate learners will know how to inflect it. Although it exemplifies the explicit syntactic data (T for transitive and that-clause) with small sentences that may help the intended user group to understand these codified data, the main problem is nevertheless the relatively scarce information in this respect compared with some of the other "big five" learners' dictionaries. *Longman Dictionary of Contemporary English*, for example, provides also "*it is anticipated that*" and "*anticipate doing sth*", while the *Macmillan English Dictionary for Advanced Learners* adds "*anticipate (sb/sth) doing sth*", i.e. frequent syntactic constructions that the intermediate learners may need when performing written text production in English.

To sum up, an intermediate learner of English (the foreseen user *type*), when having problems with the specific word *anticipate* in a specific *type* of extra-lexicographical situation, namely written text production, that leads to a subsequent consultation of a printed dictionary, may need information about orthography, word class, syntactic properties, collocations, derivates and synonyms. The corresponding lexicographical data furnished in the dictionary provide solutions for all of these information-need categories, but only orthography, word class and syntactic properties are covered by the provision of explicit data whereas collocations, derivates and synonyms are covered implicitly or in another article (*anticipation*). To this should be added that the important syntactic data are only included partially (*transitive* and normal *that-clause*) whereas highly relevant syntactic data such as "*it is anticipated that*", "*anticipate doing something*" and "*anticipate (somebody or something) doing something*" are completely missing together with data that permit the user to access *anticipate* via his or her native language.

In all dictionary consultations mentioned above, the data would therefore again be relevant at the system or algorithmic level. However, if the user had a broader cognitive information need, his/her information need would in most cases not be satisfied. Since the article in the *Cambridge Learner's Dictionary* is very short, there would be no danger of information overload, but a lack of detail may cause a problem for the user. Therefore, even though the data may be relevant at the topical, cognitive and situational levels, the data offered to the user may be incomplete and may only fulfill the user's information need

partially: the user may not be able to carry on with his/her work task and may have to consult other information sources. This stresses the important role of the compiler of an information tool to ensure that the data taken up in the database is sufficient to solve the specified information needs of the potential user in a specified type of situation: insufficient data are as big a problem as information overload.

5. The lexicographical process and the various dimensions of relevance

In the following examples the above discussion is related to the various phases in the lexicographical process outlined in section 2 (i.e. the process viewed with the eyes of the user) as well as to the dimensions of relevance presented in section 3. We can now say that the broader extra-lexicographical context or situation as well as the specific situation (written text production in a foreign language) in which the user with specific characteristics finds him- or herself are instances of *socio-cognitive relevance* in terms of information science. The situation when an information need occurs due to a communicative problem and the phases where the user becomes aware of this need and decides to take lexicographical action, are all examples of *situational relevance*.

When the user proceeds to the intra-lexicographical phase and, as a first step, selects the appropriate information tool, this selection could be a case of *socio-cognitive relevance* or *situational relevance*. The choice of a specific information tool would have a *socio-cognitive* dimension if the user decides on a specific tool based on his or her knowledge of multiple tools available, is aware of the content, approaches to data selection and presentation, possible biases etc. of each of the tools. However, the choice of tool would usually be based on *situational relevance* as well, if the user simply chooses tool A over tool B without in-depth knowledge of the design and compilation criteria of the creators of the tool, but simply on the basis of the perceived usefulness of the two tools. As to the two information tools chosen and discussed in the previous section, the *Oxford English Dictionary* and the *Cambridge Learner's Dictionary*, both are examples of *topical relevance* because they relate to the topic of the information need, but due to the information overload of the former in terms of text production, only the latter is a case of *situational relevance*.

Whereas all other phases and sub-phases of the overall lexicographical process are instances of s*ubjective*, *user-based relevance*, the sub-phase where the user accesses the data is a clear case of *objective*, *system-based relevance*, either based upon algorithms in an online environment or the user's knowledge of the alphabet used as access route in the printed dictionary.

In the following intra-lexicographical sub-phase the user verifies that the article accessed (*anticipate*) is actually the right one and contains the data required, e.g. data about orthography, word class and syntactic properties. This verification is without any doubt a case of *cognitive relevance* whereas the subsequent retrieval of the needed information provides an example of *situational*

relevance.

Affective relevance could, as indicated in section 3, occur across all dimensions of relevance. This applies to *affective relevance* in the lexicographical process as well. Examples would be where the user decides to ignore an information need based on extraneous factors (e.g. a lack of time or interest in solving the information), a decision not to consult a specific tool based on issues not related to the possibility of the tool providing a solution to the information need, a decision not to use e-information tools because of lack of knowledge about how effectively to access or use the tools, etc.

So far so good! The various pre- and intra-lexicographical phases and subphases described in section 2 seem to be perfectly covered by the various dimensions of relevance treated in information science. However, when it comes to the post-lexicographical phase where the user makes use of the retrieved information in order to solve problems, perform a task, etc., none of the described dimensions of relevance seems to cover the phase, although it must be considered a very import phase because it is here that it is proven in practice if the retrieved information is actually the information needed, cf. "the proof of the pudding is in the eating".

6. Conclusion

In this contribution we have tried to unify criteria from two disciplines which have so far developed in almost complete isolation from each other although they, without any doubt, share a common interest in providing access to data and information. It is not an easy task as both disciplines have their own deep-rooted traditions and different starting points.

Specific sub-disciplines within information science (e.g. information organisation and retrieval) have until now mainly been dedicated to the study and development of systems, technologies and techniques that may provide access to already existing and relatively big documents (books, articles etc.) as well as illustrations and other types of data sources from where information can be retrieved. In this respect, information science works almost exclusively with cognitive needs as they are defined in the function theory.

Lexicography, on the other hand, also studies and develops systems and techniques that allow the users to access relevant data and retrieve the required information. However, when this is said, traditional lexicography differs from information science in at least three ways:

1. the information needs covered by lexicographical works are not only cognitive but also communicative, operative and interpretive;
2. the cognitive needs covered are in most cases (but not always) needs that may be met by relatively small sets of data;
3. the data to which access is provided are, as a rule, not already existing

data but data selected, elaborated and prepared by the lexicographer.

In spite of these obvious differences, history provides various overlapping examples where lexicographical works have been planned and produced according to principles similar to those of information science. The Chinese *Yongle Dadian* (1408) described in section 2 is such an example as it provides access both to completely new data written by the authors and to already existing books. *The New Palgrave Dictionary of Economics Online* (2012) is another example of a lexicographical work which provides access to a huge number of scientific articles written by distinguished economists based upon principles that completely overlap those of information science.

It may therefore not surprise anyone that the discussion in the previous section has shown that the theory of relevance and all its dimensions are perfectly applicable to lexicography, especially when the various phases and subphases of the lexicographical process as it is conceived from the point of view of the user are analyzed in this perspective. The only really big surprise is that the relevance criteria do not seem to apply to the final post-lexicographical phase when the user of the lexicographical tool makes use of the information retrieved in order to solve a communicative or cognitive problem, to store it as knowledge, to perform a task or to interpret a sign, signal, symbol etc. The explanation may be that information organization and retrieval as a sub-discipline of information science, as mentioned, has hitherto mainly worked with large sets of data and information for cognitive purposes and that it is much more difficult to test in practice whether the information eventually stored in the brain is actually the relevant one, whereas this is relatively easy to do when the problem, for instance, is related to written text production. Other sub-disciplines of information science are, however, interested in the use of the data and information, e.g. information (and knowledge) management — the basic tenet is that information (and knowledge) is to be managed so that it can be interpreted for decision making and therefore assist in the work task execution. At this level it again links to lexicography — even though the dimensions of relevance are not used to describe this process.

Why is this comparison important at the theoretical level, for both information science and lexicography?

For information science the comparison is important because it is evident that, in information science, the theory of relevance stops at a crucial stage, viz. before the use of the information. It deals with retrieval issues and judgment of relevance of the retrieved information, but does not deal at all with the actual use of the information to solve problems or help in decision making. In addition, information science deals mainly or solely with cognitive information needs, and the different user situations as defined in the function theory are not distinguished. This research indicates that the theory of relevance in information science should be expanded to include the "post-lexicographical phase", i.e., the actual situation where the information is used, as well as a more finely

grained situation assessment. This is an important theoretical consideration for information science and requires further research in information science.

In lexicography, relevance theory as outlined and illustrated above, provides an elegant theory to understand relevance as a complex phenomenon that may have a profound influence on lexicographers' analysis of users' information needs. Users' information needs are paramount for lexicographers to decide what data are to be shown to the user in any given usage situation, as is evident from function theory (and even general lexicographic practice). Relevance theory implies that lexicographers have to make an even more in-depth study of the users of their products to enable them to understand exactly what may influence a user to use or not use a specific dictionary and to ensure that the lexicographic offering presents to the user only the required data to solve the user's information need in a given situation — no more and no less. If lexicographers understand that there are multiple relevances that may influence a user's judgment they may be able to design the underlying databases and filtering mechanisms in such a way that the end product provides the ideal solution for every user in every usage situation. This requires additional research in both metalexicography and lexicographical practice, especially at the level of database design and the design of filtering mechanisms.

Research in relevance theory therefore enriches the theoretical underpinnings of both information science and lexicography, and has a practical implication for providing better information tools (lexicographic information tools as well as other information tools) to users of such tools.

References

Lexicographical works

Cambridge Learner's Dictionary. Woodford, Kate (Ed.). 2010. *Cambridge Learner's Dictionary*. Cambridge: Cambridge University Press.

Encyclopédie, ou Dictionnaire raisonné des sciences, des arts et des métiers. Diderot, Denis and Jean le Rond d'Alembert (Eds.). 1751-72. *Encyclopédie, ou Dictionnaire raisonné des sciences, des arts et des métiers.* Volume 1-28. Paris: André le Breton, Michel-Antoine David, Laurent Durand and Antoine-Claude Briasson.

Longman Dictionary of Contemporary English. Summers, Della (Ed.). 2003. *Longman Dictionary of Contemporary English*. Harlow: Longman.

Macmillan English Dictionary for Advanced Learners. Rundell, Michael (Ed.). 2002. *Macmillan English Dictionary for Advanced Learners*. London: Macmillan Education.

Oxford English Dictionary. Simpson, John (Ed.). 2012. *Oxford English Dictionary*. Oxford: Oxford University Press. [Online] Available: http://www.oed.com.

The New Palgrave Dictionary of Economics Online. Durlauf, Steven N. and Lawrence E. Blume (Eds.). 2012. *The New Palgrave Dictionary of Economics Online*. Hampshire: Palgrave Macmillan. [Online] Available: http://www.dictionaryofeconomics.com/dictionary.

Yongle Dadian. Xie, Jin (Ed.). 1408. *Yongle Dadian*. Volume 1–11 095. China.

Other literature

d'Alembert, Jean le Rond. 1754. Dictionnaire. Diderot, Denis and Jean le Rond d'Alembert (Eds.). 1754. *Encyclopédie, ou Dictionnaire raisonné des sciences, des arts et des métiers. Tome IV*: 958-969. Paris: André le Breton, Michel-Antoine David, Laurent Durand and Antoine-Claude Briasson.

Bergenholtz, Henning and Theo J.D. Bothma. 2011. Needs-adapted Data Presentation in e-Information Tools. *Lexikos* 21: 53-77.

Bergenholtz, Henning and Sven Tarp. 2002. Die moderne lexikographische Funktionslehre. Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen. *Lexicographica* 18: 253-263.

Bergenholtz, Henning and Sven Tarp. 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31:171-196.

Bergenholtz, Henning and Sven Tarp. 2004. The Concept of "Dictionary Usage". Dollerup, Cai (Ed.). 2004. Worlds of Words. A Tribute to Arne Zettersten. *Nordic Journal of English Studies* 3(1): 23-36.

Bergenholtz, Henning and Sven Tarp. 2005. Wörterbuchfunktionen. Barz, Irmhild, Henning Bergenholtz and Jarmo Korhonen (Eds.). 2005. *Schreiben, Verstehen, Übersetzen, Lernen. Zu ein- und zweisprachigen Wörterbüchern mit Deutsch*: 11-25. Frankfurt am Main/Bern/New York/ Paris: Peter Lang.

Borlund, Pia. 2000. *Evaluation of Interactive Information Retrieval Systems*. Doctoral Thesis. Åbo: Åbo Akademi University Press.

Borlund, Pia. 2003. The Concept of Relevance in IR. *Journal of the American Society for Information Science and Technology* 54(10): 913-925.

Bothma, Theo J.D. 2011. Filtering and Adapting Data and Information in the Online Environment in Response to User Needs. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 71-102. London/New York: Continuum.

Cosijn, Erica. 2003. *Relevance Judgements in Information Retrieval*. Unpublished D.Phil. Thesis. Pretoria: University of Pretoria.

Cosijn, Erica and Theo J.D. Bothma. 2005. Contexts of Relevance for Information Retrieval System Design. Crestani, Fabio and Ian Ruthven (Eds.). 2005. *Context: Nature, Impact, and Role. 5th International Conference on Conceptions of Library and Information Sciences, CoLIS 2005, Glasgow, UK, June 4–8, 2005*: 47-58. Berlin/Heidelberg: Springer.

Cosijn, Erica and Peter Ingwersen. 2000. Dimensions of Relevance. *Information Processing and Management* 36(4): 533-550.

Gouws, Rufus H. 2011. Learning, Unlearning and Innovation in the Planning of Electronic Dictionaries. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 17-29. London/New York: Continuum.

Ingwersen, Peter and Kalervo Järvelin. 2005. *The Turn: Integration of Information Seeking and Retrieval in Context*. Dordrecht: Springer.

Leroyer, Patrick. 2011. Change of Paradigm: From Linguistics to Information Science and from Dictionaries to Lexicographic Information Tools. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 121-140. London/New York: Continuum.

Nielsen, Sandro. 2008. The Effect of Lexicographical Information Costs on Dictionary Making and Use. *Lexikos* 18: 170-189.

Saracevic, Tefko. 1975. Relevance: A Review of and a Framework for the Thinking on the Notion in Information Science**.** *Journal of the American Society for Information Science and Technology* 26(6): 321-343.

Saracevic, Tefko. 1996. Relevance Reconsidered '96. Ingwersen, P and N.O. Pors (Eds.). 1996. *Proceedings of CoLIS 2, the Second International Conference on Conceptions of Library and Information Science: Integration in Perspective, 13–16 October 1996*: 201-218. Copenhagen: The Royal School of Librarianship.

Tarp, Sven. 2008a: Revival of a Dusty Old Profession. *Hermes, Journal of Linguistics* 41: 175-188.

Tarp, Sven. 2008b. *Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography.* Lexicographica. Series Maior 134. Tübingen: Max Niemeyer.

Tarp, Sven. 2009. Beyond Lexicography: New Visions and Challenges in the Information Age. Bergenholtz, Henning, Sandro Nielsen and Sven Tarp (Eds.). 2009. *Lexicography at a Crossroads: Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*. Linguistic Insights. Studies in Language and Communication 90: 17-32. Bern: Peter Lang.

Tarp, Sven. 2011: Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 54-70. London/ New York: Continuum.

Wiegand, Herbert Ernst. 2000. Wissen, Wissenrepräsentationen und Printwörterbücher. Heid, Ulrich, Stefan Evert, Egbert Lehmann and Christian Rohrer (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress. EURALEX 2000, Stuttgart, Germany, August 8–12, 2000*: 15-38. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

Wiegand, Herbert Ernst. 2002. Wissen in der Sprachlexikografie. Ein Plädoyer für einige immer noch notwendige Differenzierungen. Ezawa, Kennosuke, Wilfried Kürschner, Karl H. Rensch and Manfred Ringmacher (Eds.). 2002. *Linguistik jenseits des Strukturalismus. Akten des II. Ost-West-Kolloquiums Berlin 1998*: 265-281. Tübingen: Gunter Narr Verlag.

Advanced Cantonese ESL Learners' Use of a Monolingual Dictionary for Language Production

Alice Y.W. Chan, *Department of English, City University of Hong Kong, Hong Kong (enalice@cityu.edu.hk)*

Abstract: This article reports on the results of a research study which investigated the use of monolingual dictionaries by Hong Kong advanced Cantonese ESL learners in the production of target language sentences. Thirty-one English majors participated in a sentence completion task and a sentence construction task with and without the help of a monolingual dictionary. In the sentence completion task, a full Chinese context and a partial English context were given, whereas in the sentence construction task, only a few English prompts were given. Different self-reporting protocols, including introspective questionnaires, retrospective questionnaires and think-aloud recordings, and a post-task focus-group interview were conducted to tap into the participants' thinking processes during dictionary consultation. The results show that a monolingual dictionary is useful in helping learners produce target language sentences, yet learners encounter different kinds of consultation problems, some of which are related to their general use of dictionaries and others to the language in which their thinking processes are engaged. It is suggested that ESL learners use both monolingual and bilingualized dictionaries in their learning and that ESL teachers design dictionary skills training programs which take into account learners' linguistic competence and actual consultation problems.

Keywords: LEXICOGRAPHY, MONOLINGUAL DICTIONARIES, DICTIONARY USE, DICTIONARY CONSULTATION, USEFULNESS OF DICTIONARIES, LANGUAGE PRODUCTION, SENTENCE COMPLETION, SENTENCE CONSTRUCTION, CANTONESE ESL LEARNERS, PROBLEMS ENCOUNTERED, CONTEXTS GIVEN

Opsomming: Gevorderde Kantonese ESL-aanleerders se gebruik van 'n eentalige woordeboek vir taalproduksie. In hierdie artikel word verslag gedoen van die resultate van 'n navorsingstudie wat die gebruik van eentalige woordeboeke deur gevorderde Kantonese ESL-aanleerders in Hong Kong in die produksie van teikentaalsinne ondersoek het. Een-en-dertig studente met Engels as hoofvak het deelgeneem aan 'n taak waarin hulle sinne moes voltooi en aan 'n taak waarin hulle sinne moes bou met en sonder die hulp van 'n eentalige woordeboek. In die taak waarin hulle sinne moes voltooi, is die volle Chinese konteks en 'n gedeeltelike Engelse konteks verskaf, terwyl daar in die sinsboutaak slegs 'n paar Engelse leidrade gegee is. Verskillende selfrapporteringsprotokolle, insluitende introspektiewe vraelyste, retrospektiewe vraelyste en hardopdinkopnames, asook 'n onderhoud met die fokusgroep nadat die taak afgehandel is, is onderneem om die deelnemers se denkprosesse gedurende die raadpleging van 'n woordeboek te volg. Die resultate toon dat 'n eentalige woordeboek nuttig is om aanleerders te help om teikentaalsinne te produseer, maar aanleerders kom verskillende soorte probleme teë tydens die raadpleging

van 'n woordeboek, waarvan party verwant is aan die algemene gebruik van woordeboeke, en ander aan die taal waarin gedink word. Daar word voorgestel dat ESL-aanleerders sowel eentalige as verklarende woordeboeke met vertalings gebruik wanneer hulle leer en dat ESL-onderwysers programme ontwikkel wat woordeboekvaardighede oordra en wat leerders se taalvaardigheid in ag neem sowel as hulle werklike raadplegingsprobleme.

Sleutelwoorde: LEKSIKOGRAFIE, EENTALIGE WOORDEBOEKE, WOORDEBOEK-GEBRUIK, WOORDEBOEKRAADPLEGING, NUTTIGHEID VAN WOORDEBOEKE, TAALPRODUKSIE, SINSVOLTOOIING, SINSBOU, KANTONESE ESL-AANLEERDERS, PROBLEME TEËGEKOM, KONTEKS VERSKAF

Introduction

The use of a dictionary is regarded as "an indispensable component of home and academic life" (Abecassis 2007: 249). Although dictionaries contain much useful encoding and decoding information, many learners as a foreign language (FL) or second language (SL) cannot make full use of them in their learning and ignore or misread a lot of useful information (Nesi and Meara 1994). They do not possess the dictionary skills needed, and many of them have not received formal dictionary skills training before (Chan 2005). Even language teachers are not necessarily adequately equipped to provide comprehensive dictionary skills training for their students, and they themselves are not fully aware of the potential advantages of a learner's dictionary in the learning of a second or foreign language (Miller 2008). As a result, language learners encounter various difficulties in their use of dictionaries. Among the common ones are their inability to locate the relevant information needed and their difficulties in identifying, for example, the transitivity of a target verb or the countability of a target noun (Chan 2012b).

The most popular kinds of information learners seek from a learner's dictionary are definitions of words or, in the case of bilingual or bilingualized dictionaries,[1] equivalents in the other language for decoding purposes only (Béjoint 1981; Lew 2004). Many dictionary users do not use a dictionary for guiding them on word usage or encoding. If they want to know the syntactic restrictions, register appropriateness or collocations to be able to use the word appropriately, they tend to resort to grammar books instead (Frankenberg-Garcia 2011). Advanced English as a Second Language (ESL) or English as a Foreign Language (EFL) learners use dictionaries also for encoding purposes (Chan 2005), but they still often fail to take advantage of information on grammar and usage, including grammar codes (Chan 2012b; Carduner 2003; Frankenberg-Garcia 2011; Lew and Dziemianko 2006; Summers 1988). Grammatical information in a dictionary is regarded as user-friendly for encoding if it is often consulted and if it results in correct language production (Dziemianko 2006), yet ESL learners most often use examples and definitions in their con-

sultation of dictionaries rather than explicit grammatical information when they want to determine the correct use of a target word (Chan 2012b; Bogaards and Van der Kloot 2002; Dziemianko 2006, in press), leading to some researchers' claim that grammar codes could be eliminated from pedagogical dictionaries (Bogaards and Van der Kloot 2002). While it is true that examples are helpful not only in explaining meaning but also in showing lexico-grammatical patterns (Cowie 1999b), inappropriate generalisations may sometimes be made regarding the use of target words (Chan 2012b), and learners sometimes misidentify examples and indicating useless ones as helpful (Dziemianko in press).

To understand how well dictionaries assist students, it is necessary to conduct experimental tests in user-based lexicographical research (Battenburg 1991) and to examine how successful a dictionary is in the provision of lexical support learners need when they are engaged in various kinds of second language receptive and productive activities (Swanepoel 2000). Despite the fact that interest in empirical user-studies in lexicography is "on the rise" (Lew 2011a: 1), empirical studies into the use of dictionary information for language production are still relatively scarce. The present research aimed at bridging this research gap by examining ESL learners' use of monolingual dictionaries in language production.

The decision to exclusively focus on monolingual dictionaries was not arbitrary. Although many previous research studies found that learners, including advanced ESL or EFL learners, tend to use bilingual or bilingualized dictionaries (Atkins and Varantola 1997, 1998; Baxter 1980; Lew 2004), recent research has found that many advanced Hong Kong Cantonese ESL learners use both bilingualized and monolingual dictionaries in their work or studies, and some use monolingual dictionaries exclusively instead of relying on both (Chan 2011). Monolingual dictionaries are regarded as excellent models for advanced foreign learners (Cowie 1999a). However, just because all dictionary information is presented in the learner's target language, the design and presentation of these dictionaries often necessitate skills that many learners lack (Kernerman 2007). To learners of the language as a second language, monolingual dictionaries may present even more difficulties, as in the production of target language sentences, many ESL learners tend to think in their mother tongue. L1 transfer has been found to be one major source of learner problems not just among lower proficiency learners, but high-proficiency learners also rely on the syntax and vocabulary of their L1 when writing in L2 and/or when encountering difficulties in the production of target language output (Chan 2004b; Bhela 1999; Van Weijen, Van den Bergh, Rijlaarsdam and Sanders 2009). Given that monolingual dictionaries rely on corpora which do not necessarily provide "the English that is really needed or wanted by its users" (Kernerman 2007: 142), it is interesting to investigate the usefulness of these dictionaries to ESL learners when their target language production processes are guided by their native language and when their production processes are not.

Objectives

The objectives of the present study were to (i) explore how advanced ESL learners in Hong Kong used a monolingual dictionary for target language production, (ii) examine the usefulness of a monolingual dictionary for target language production, and (iii) investigate the general problems learners encountered in dictionary consultation as well as those they encountered when different amounts and nature of contexts were given.

Participants

Thirty-one participants, including eight males and twenty-three females, participated in the study. They were all English majors at a local university. Their ages ranged from 20 to 24: One year 1 student, eleven year 2 students, and nineteen year 3 students. Twenty of them had learnt English for 15–19 years and eleven had learnt English for 20 years or more. In view of their English learning backgrounds, they could all be regarded as advanced ESL learners.

Procedures

The participants were asked to do a Sentence Completion Task and a Sentence Construction Task with and without the use of a monolingual dictionary.

Sentence Completion Task

The first part of the Sentence Completion Task was done without the use of a dictionary. It aimed to investigate the participants' prior knowledge of the use of the target words. For each target English word, a sentence context was invented and given in written Chinese. The target words and some parts of the corresponding English sentences were also given (see Appendix A). The participants had to complete the English sentences using the given target words. They were instructed to finish all the questions in this part before they did the second part. Although a Chinese context was given for each sentence, the task was not meant to be a translation one. The provision of the Chinese sentences only gave enough contextual information for the intended use of the target words, and the uncompleted English sentences ensured that the participants would use the desired grammatical patterns in the completion of the sentences.

In the second part of the task, the participants had to complete the sentences again with the help of a monolingual dictionary. They were divided into three groups, with one group using *Cambridge Advanced Learner's Dictionary* 3rd edition (CALD3), one using *Collins COBUILD Advanced Dictionary* 6th edition (COBUILD6), and one using *Longman Dictionary of Contemporary English* 5th edition (LDOCE5). These three dictionaries were chosen because they were, to

the author's knowledge, among the most popular (paper) monolingual dictionaries used in Hong Kong.[2] The participants were instructed to consult the dictionaries for the correct usage of the given words and associated expressions, but they were not allowed to change the answer to any of the questions in the first part after doing the second part. Different self-report protocols, namely Think-aloud Recordings (for the LDOCE5 group), an Introspective Questionnaire (for the COBUILD6 group), and an Instant Retrospective Questionnaire (for the CALD3 group) were used to tap into the participants' thinking processes in the course of dictionary consultation.

Sentence Construction Task

The first part of the Sentence Construction Task was also done without the use of a dictionary to investigate the participants' prior knowledge of the use of the target words. For each sentence, three to four English prompts, one of which being the target word, were given to the participants (see Appendix B). They had to use the given prompts to construct a grammatical and meaningful English sentence, making whatever changes to the prompts deemed necessary but following the order of the prompts. The prompts were given to ensure that the desired grammatical patterns would be followed in the participants' construction of the sentences. They were instructed to finish all the questions in this part before they did the second part.

The second part of the task, which aimed to investigate how the participants used a monolingual dictionary to help them construct English sentences, had the same prompts but the target words were underlined. The participants were required to construct a sentence again by consulting a monolingual dictionary for the correct usage of each of the underlined target words.

The participants were also divided into three groups with different groups using different dictionaries and doing different self-reporting protocols: an introspective questionnaire, a retrospective questionnaire and introspective think-aloud recordings. The LDOCE5 group did the introspective questionnaires, the COBUILD6 group did the retrospective questionnaires, and the CALD3 group did the think-aloud group recordings. The groupings of the participants for this task differed from those for the Sentence Completion Task, so that no participant would use the same dictionary and/or be engaged in the same self-reporting protocol for the two tasks.

Although different dictionaries were used and different protocols were followed by different participants, no attempt was made to compare the effectiveness of the three dictionaries or the performance of the different groups, as the main focus of the study was to uncover learners' problems in extracting dictionary information for language production.

The Sentence Completion Task with the provision of a Chinese context was included because it has often been reported that many Cantonese ESL learners tend to think in Chinese when producing written English output

(Chan 2004b, 2010). They often have a target Chinese word in mind when constructing an English sentence and attempt to convert it in English and fit it into the Chinese context in mind. On the other hand, the Sentence Construction Task without the involvement of Chinese was included because learners sometimes have a target English word as well as its possible collocations or other associated words in mind and try to fit the target word into the context instead of doing mental translations based on a Chinese equivalent.

Target Words

The target words used in both the sentence completion and construction tasks were first piloted with nine non-native English majors whose English proficiency and language backgrounds were comparable to those of the participants. For the Sentence Completion task, a total of thirty-two target words, and thus thirty-two Chinese sentence contexts and corresponding uncompleted English sentences, were piloted. For the Sentence Construction Task, a total of thirty target words and their corresponding prompts were piloted. Only the target words which were found to be difficult for about 70% of the students in the pilot group (i.e. only those which were incorrectly used by about 70% of the pilot group) without consulting a dictionary were selected for the real tasks.

Sentence Completion Task

The target words included in the Sentence Completion Task (a total of 10) were all simple English words familiar to advanced ESL learners rather than new or exotic words, but the uses of the words necessitated by the Chinese contexts were mostly unfamiliar to Hong Kong Cantonese ESL learners.[3] To ensure that the completed English sentences matched the Chinese contexts, the participants had to use the target words correctly with appropriate grammatical associations, such as the correct choice of prepositions, of phrasal verb particles, or verb complementation, etc. The following are some examples of the target words and the associated uncompleted English sentences (see Appendix A for a complete list).

(rush): I don't want to _____ rush into _____ a decision.
(brush): He _____ brushed past _____ me, but he didn't see me.

Sentence Construction Task

The target words included in the Sentence Construction Task (also 10) were also simple English words familiar to advanced ESL learners, but they were all found to be very commonly misused by Hong Kong Cantonese ESL learners (Bunton 1989, 1994; Heaton and Turton 1987; Jenkins 1990). The grammatical

patterns focused on included the choice of prepositions for a target noun, the choice of prepositions after a target verb, and the use of a verb in a correct transitivity pattern, etc. The following are some examples of the target words and the corresponding prompts (see Appendix B for a complete list).

(knowledge): (little) (knowledge) (linguistics)
(assist): (he) (assist) (murder) (her husband)

Self-Reporting Protocols

Both retrospective and introspective self-reporting protocols were used in the dictionary consultation tasks for tapping into the participants' thinking processes during dictionary consultation and for gathering their evaluation of the effectiveness of dictionary information. Since delayed retrospective reports are criticized as being based on participants' memory (Kaivanpanah and Alavi 2008), and narration of what they think they have done after completing a task "may only have a tenuous relationship to the original attended information" (Kasper 2000: 336) and may not relate clearly to any specific observable behavior (Ericsson and Simon 1993), introspective questionnaires were also used in the study to minimize the distortion of information or the effects of participants' having forgotten the information. Introspective think-aloud protocols were used alongside introspective questionnaires to elicit "a real-time process of cognitive activities" (Tono 2001: 68) and to allow learners to process information simultaneously with introspection.

Introspective Questionnaires

The participants in the Introspective Questionnaire groups were required to do a questionnaire immediately after finishing each question in Part II of the respective task. The questionnaires aimed at eliciting their instant and detailed feedback on the way a certain dictionary entry helped them complete or construct a target sentence. They had to report on their feelings when they ended a search (e.g. sure that the decision was correct, not sure whether the decision was correct, sure that they didn't get the right information) and to specify the part(s) of the dictionary entries which they found the information they wanted to look for. These appeared in the questionnaires as forced-choice questions. The participants also had to write out the examples/definitions or other dictionary information they used to make a final decision, to demonstrate how the information showed that their decisions were correct, to account for the reason(s) why they were doubtful about their decisions, to report on the difficulties, if any, they encountered, and so on. These all appeared in the questionnaires as open-ended questions. All the questions were given in written English and all the participants responded in written English.[4]

Introspective Think-aloud Recordings

The participants in the Think-aloud groups recorded the whole of their decision-making processes during dictionary consultation using whatever language they were comfortable with in a sound-proof room. The whole process of searching for the appropriate examples, definitions and/or other dictionary information, determining what dictionary information should be used, and choosing the appropriate dictionary information to guide them to the completion and construction of a target sentence, etc was recorded. Three students doing the Sentence Construction Task chose to speak entirely in English, and the rest (doing either task) chose to speak mainly in a mixed-code of English and Cantonese. The recordings were made using an audio recorder called *Audacity* and converted into wave sound files. A research assistant transcribed all the sound files for analysis and the researcher (i.e. the author) translated the Cantonese utterances into English.

Instant Retrospective Questionnaires

The participants in the instant Retrospective Questionnaire group completed a retrospective questionnaire immediately after finishing the whole task. The questionnaire aimed at investigating the participants' overall assessment of their performance in Part II of the respective task and their general evaluation of the usefulness of the dictionary or the different parts of the entries. They had to identify the information they consulted most, indicate from which part of the dictionary entries they found the answers to most of the questions, comment on the usefulness of the different parts of the dictionary entries in general, evaluate the overall usefulness of the dictionary, specify the frequency with which they encountered difficulties, estimate the extent to which they thought their own uses of the target words were changed after the consultation task, and so on. All the questions were given as forced-choice questions in written English and all the participants responded in written English.

Post-Task Focus Group Interview

A Post-task Focus Group Interview lasting about an hour was conducted within a month of the completion of the two tasks by all the participants. Six students participated in the interview, including one male and five females who used different dictionaries and who were engaged in different self-reporting protocols in the two tasks. Copies of all the dictionaries used and copies of all the task sheets completed by the interviewees were brought by the interviewer to the interviews to refresh the interviewees' memory of their dictionary consultation and decision-making processes and to enable them to quote precise information from the dictionaries. Before the commencement of

the interview, the interviewer gave clear instructions on the expectations of the interview in a mixed code of English and Cantonese. She was then responsible for asking prompting questions centring around the difficulties that the participants encountered, the strategies they used to overcome the difficulties, the part(s) of the dictionary entries they found most useful, the way the information helped them make a decision, and their own use of the target items after dictionary consultation. The reasons underlying their responses were also discussed. All the interviewees spoke in Cantonese except when quoting dictionary examples, dictionary definitions, and/or task-related prompts or contexts. The proceedings of the interviews were recorded using both a video camera and a mini-disk recorder. The interviewer transcribed the whole proceedings of the interview and the researcher translated the Cantonese utterances into English.

Results

Performance of Students and Performance on Target Words

The following section will give a summary of the participants' performance in the sentence completion and construction tasks with and without the use of a dictionary. Their performance will be presented as accuracy rates. A sentence was deemed accurately completed (in the Sentence Completion Task) when the target word was used with the appropriate grammatical associations (e.g. a correct preposition) and the resultant meaning of the sentence was consistent with the Chinese context. A sentence was regarded as accurately constructed (in the Sentence Construction Task) when the target word was used with the appropriate grammatical associations (e.g. correct transitivity pattern) with the corresponding prompts. Irrelevant grammatical mistakes, such as subject-verb agreement, were not taken into account.

Sentence Completion Task

WITHOUT THE USE OF A DICTIONARY

Of the ten words included for investigation, only three words *despair, rush* and *substitute* received an accuracy rate of 30% or above. The rest (70%) received an accuracy rate of 20% or below. Only six students (19.4%) from the three groups could accurately complete four or more sentences (40% or more). Twenty students (64.5%) could accurately complete only one or two sentences or even none (0% – 20%) (See Tables 1 and 2).

WITH THE USE OF A DICTIONARY

Nine sentences (90%) were accurately completed by over half (50%) of the participants. Six sentences (60%) were accurately completed by 80% or more of the

participants. The words which received the highest accuracy rates were *rush* and *brush* (both 93.5%), which showed big leaps from their originals of 48.4% and 16.1%. The word *limit* received the lowest accuracy rate of 48.4%. Only fifteen students (48.4%) could accurately complete the sentence with the use of a dictionary. Twenty students (64.5%) could accurately complete eight or more sentences (80% or more). However, still one student (3.2%) could accurately complete only three sentences or less (30% or less) (See Tables 1 and 2). A paired, two-tailed t-test using Excel 2010 showed that the difference between the participants' overall sentence completion performance without using a dictionary and that using a dictionary was statistically significant at the 0.05 significance level (t= -7.84, p= 0.000259).

Table 1: The accuracy rate of each target word in the Sentence Completion Task with and without the use of a dictionary

Target Word	Percentage of participants who could complete the sentences accurately	
	Without a dictionary (N= 31)	With a dictionary (N= 31)
Rush	48.4%	93.5%
Charge	19.4%	83.9%
Brush	16.1%	93.5%
Frighten	19.4%	80.6%
Deceive	9.7%	58.1%
Limit	19.4%	48.4%
Substitute	45.2%	58.1%
Monument	0%	67.7%
Improve	6.5%	90.3%
Despair	35.5%	83.9%
Total	21.9% (68/310)	75.8% (235/310)

Table 2: Individual participants' performance on the Sentence Completion task with and without the use of a dictionary

Student	Percentage of sentences accurately completed	
	Without a dictionary (N= 10)	With a dictionary (N= 10)
Student 1	10%	70%
Student 2	30%	80%
Student 3	60%	60%
Student 4	10%	80%
Student 5	10%	70%

Student 6	10%	80%
Student 7	0%	80%
Student 8	20%	90%
Student 9	0%	70%
Student 10	10%	50%
Student 11	40%	90%
Student 12	30%	90%
Student 13	50%	90%
Student 14	0%	90%
Student 15	20%	80%
Student 16	10%	30%
Student 17	20%	80%
Student 18	20%	90%
Student 19	20%	80%
Student 20	20%	80%
Student 21	10%	90%
Student 22	0%	40%
Student 23	40%	70%
Student 24	20%	80%
Student 25	60%	100%
Student 26	30%	80%
Student 27	40%	70%
Student 28	10%	80%
Student 29	20%	60%
Student 30	30%	80%
Student 31	30%	70%
Total	21.9% (68/310)	75.8% (235/310)

Sentence Construction Task

WITHOUT THE USE OF A DICTIONARY

Of the ten words included for investigation, only three words (30%) *knowledge, comprise* and *assist* received an accuracy rate of over 50%. Two words (20%), *opposite* and *anticipate* received an accuracy rate of 30% or below. Only nine students (29%) from the three groups could accurately construct five or more sentences (50% or more) correctly. Eleven students (35.5%) could accurately construct only three sentences or even less (30% or less) (See Tables 3 and 4).

WITH THE USE OF A DICTIONARY

Seven sentences (70%) were accurately constructed by 80% or more of the participants. The words which received the highest accuracy rates was *knowledge* (100%), but its original accuracy rate without the use of a dictionary was also quite high (61.3%). The word *guilty* received the lowest accuracy rate of 45.2%. Only fourteen students (45.2%) could accurately construct the sentence even with the use of a dictionary. Twenty students (64.5%) could accurately construct eight or more sentences (80% or more). However, still three students (9.7%) could only accurately construct five sentences or less (50% or less) (see Tables 3 and 4). A paired, two-tailed t-test using Excel 2010 showed that the difference between the participants' overall sentence construction performance without using a dictionary and that using a dictionary was statistically significant at the 0.05 significance level (t= -6.76, p= 0.0000827).

Table 3: The accuracy rate of each target word in the Sentence Construction Task with and without the use of a dictionary

Target Word	Percentage of participants who could construct the sentences accurately	
	Without a dictionary (N= 31)	With a dictionary (N= 31)
Knowledge	61.3%	100%
Reason	32.3%	61.3%
Comprise	61.3%	90.3%
Opposite	3.2%	67.7%
Guilty	41.9%	45.2%
Cure	32.3%	80.6%
Anticipate	25.8%	80.6%
Inform	48.4%	90.3%
Befriend	35.5%	87.1%
Assist	61.3%	83.9%
Total	40.3% (125/310)	78.7% (244/310)

Table 4: Individual participants' performance on the Sentence Construction Task with and without the use of a dictionary

Student	Percentage of sentences accurately constructed	
	Without a dictionary (N= 10)	With a dictionary (N= 10)
Student 1	50%	90%
Student 2	50%	80%
Student 3	40%	80%

Student 4	30%	80%
Student 5	60%	100%
Student 6	70%	80%
Student 7	40%	70%
Student 8	20%	90%
Student 9	30%	100%
Student 10	40%	80%
Student 11	20%	80%
Student 12	50%	70%
Student 13	80%	90%
Student 14	30%	70%
Student 15	30%	100%
Student 16	40%	90%
Student 17	40%	40%
Student 18	10%	90%
Student 19	40%	90%
Student 20	40%	70%
Student 21	30%	50%
Student 22	60%	70%
Student 23	30%	80%
Student 24	60%	60%
Student 25	40%	90%
Student 26	20%	50%
Student 27	40%	100%
Student 28	30%	80%
Student 29	40%	80%
Student 30	40%	70%
Student 31	50%	70%
Total	40.3% (125/310)	78.7% (244/310)

Perception of Usefulness of Dictionary Information

Sentence Completion Task

For the retrospective questionnaire group, 36.4% of the participants (N= 11) thought that the definitions were extremely useful and 90.9% thought that the examples were extremely useful. 45.5% regarded the special features and the explanations as extremely or very useful. The most prevalent difficulty encoun-

tered by the retrospective group was that there were too many definitions and examples (27.3%). Nobody thought that the codes or abbreviations were difficult to understand, that there were no special features to attract their attention, that they could not find the information about word classes, or that the word was missing. On the whole, 81.8% of the participants thought that the dictionaries used were extremely or very useful. 27.3% thought that they had a lot of improvements in the use of the target words after the dictionary consultation whereas the rest thought that they made some improvements. 54.5% of them were confident that 76% – 99% of the target words were constructed correctly after the use of a dictionary. 72.7% thought that it was extremely or very important to have more examples included.

As for the introspective group, 84% (N= 100) (10 participants x 10 consultations) of the decisions were made after the participants had consulted the definitions, and 64% made after the participants had consulted the examples. The word for which most participants were sure that their decisions were correct was *frighten* (90%), and the words which the least number of participants (40%) were confident about were *limit* and *monument*.

Sentence Construction Task

For the retrospective questionnaire group, 50% of the participants (N= 10) thought that the definitions were extremely useful and 90% thought that the examples were extremely useful. 30% regarded the special features as extremely or very useful. The most prevalent difficulty encountered by the retrospective group was that there were too many definitions and examples (50%). Nobody thought that the codes or abbreviations were difficult to understand and only one participant (10%) thought that the usage information was not clear or that the word was missing. On the whole, 80% of the participants thought that the dictionaries were extremely or very useful. 50% thought that they had a lot of improvements in the use of the target words after the dictionary consultation whereas the rest thought that they made some improvements. 90% of them were confident that 76% – 99% of the target words were constructed correctly after the use of a dictionary. 90% thought that it was extremely or very important to have more examples included, and 50% thought that the definitions should be made clear and that more information on the word class of a word should be given.

As for the introspective group, 90% (N= 110) (11 participants x 10 consultations) of the decisions were made after the participants had consulted the examples, 44.5% made after the participants consulted the special features, and 43.6% made after the participants consulted the definitions. The words for which most participants were sure that their decisions were correct were *knowledge* and *befriend* (90.9%), and the word which the least number of participants (72.7%) were confident about was *reason.*[5]

L2 Language Production with the Use of a Monolingual Dictionary

Although the use of dictionaries resulted in the participants' significant improvements in language production, some participants encountered certain difficulties and made incorrect judgment. In this section, the problems that they encountered will be examined by scrutinizing their introspective written reports, think-aloud verbal reports and post-task interview reports. The provision of different amounts of contexts in the two tasks will also be taken into account.[6]

Learner Problems Unique to Language Production with a Full Context in the Native Language

With the provision of a full Chinese context in the Sentence Completion Task, some participants had the tendency to compare the meaning of a target English word with the meaning of the corresponding Chinese expression given in the Chinese context. The amount of correspondence between the two was regarded as very important, and the participants only felt secure when there was an exact match. When no exact match could be found or when the given dictionary meaning had different connotations from the Chinese context, they became daunted. For example, when consulting the entry *frighten* in completing the sentence *The high prices* _____ *many customers* (see Appendix A), Student 4 compared the definition "If you frighten away a person or animal or frighten them off, you make them afraid so that they go away or stay some distance away from you (COBUILD6)" with the Chinese expression *卻步 (frighten off)* literally. He could not decide if the definition could be used as the basis for the completion of the sentence because of his uncertainty about the physical actions involved in the Chinese context.[7]

> I am not sure about *卻步*, whether it means physically away from or just feel afraid, anxious or nervous. (Student 4)

Similarly, the amount of physical contact involved was a major determining factor in the completion of the sentence with the word *brush* (see Appendix A). Some participants were concerned about the amount of physical contact invoked by the given Chinese expression *擦身而過 (brush past)* and based their decisions on their own personal experience and expectations of its use. The definitions of the English word *brush* were, therefore, taken very literally, and comparisons were deliberately made between the amounts of physical contact necessitated by the Chinese expression and the definitions of the English word. The following quotes from the focus-group interview show the participants' deliberate comparisons:

> But it says here *one thing brushes against another, or if you brush one thing against another, the first thing touches the second thing slightly while passing it*, so it matches

> the Chinese expression *擦身而過. …* That is, when they walked past each other they touched each other slightly. (Student 5)
>
> My understanding is that *擦身而過* does not involve any physical contact. (Student 1)
>
> But they touch slightly, so this means that there is really physical contact. (Student 22)

Learners sometimes could not make appropriate deductions from a dictionary definition which had different wording from the given Chinese context and tended to look for unnecessary correspondence. For example, some participants were concerned with the collocation of the word *publicly* with the target word *charge* when seeing the former in one of the definitions of the latter: "When the police charge someone, they formally accuse them of having done something illegal…. If you charge somebody with doing something wrong or unpleasant, you *publicly* say that they have done it." (COBUILD6, emphasis added). Therefore, they were again unsure of the appropriateness of the definition as a model for sentence completion, as the corresponding meaning of *publicly* was not shown in the Chinese context, e.g.:

> I am not sure if the accusing is done 'publicly'. (Student 6)

Words which involved two constituents the order of which was significant to the meaning of the resultant sentences also caused difficulty in making a decision. For example, the dictionary definition and/or examples for the target word *substitute* did not have a clear indication of the order of the constituents. Some participants could not identify from the given information what the replacing element was and what the replaced element was, so they were not sure if the order of constituents given in the dictionary entry was consistent with the order of the corresponding constituents given in the Chinese context, e.g.:

> So what is what, what substitutes for what? … It should be the one in front replacing the one that comes after, … but what substitutes for what? … I am not sure. (Student 12)

Learner Problems Unique to Language Production with a Partial Context in the Target Language

Without a full context in the native language, the cognitive demands required by the Sentence Construction Task with target language prompts resulted in different learning problems.

Because the participants could add any constituent in the construction of a sentence, some of them partially followed the information given in the dictionaries but added intervening constituents which altered the desired sentence

structure. For example, some learners added *you* after the target word *reason* (prompts given: *what – reason – reject – proposal*) and constructed sentences like *What is the reason for you to reject the proposal*, correctly taking the preposition *for* as the complement of *reason* but ignoring the fact that the preposition was supposed to take an *-ing* complement (i.e. *reason for doing something*) rather than a *to*-infinitive complement with an intervening nominal constituent:

> The correct preposition to use after *reason* is *for*, so *what is the reason for you to reject his proposal*. (Student 16)

The ability to distinguish the desired meaning of a target word from its other meanings inconsistent with the given context was also an issue. For example, some participants used *reason to VERB* as a model structure for the target word *reason* and constructed sentences such as *What is the reason to reject the proposal* without noticing that *reason* in *reason to VERB* had a meaning subtly different from *reason* in the given context:

> Let me see…. *The police have reason to believe that he is guilty*, then I can use this one, *reason to*, yes, this *to believe that he is guilty* is an action because of a reason, the reason makes him do that. (Student 14)

> *(Reason) To do sth* in dictionary matches *to reject the proposal in my sentence*. (Student 11)

Some of the given prompts which appeared in a dictionary entry in another form were sometimes followed indiscriminately. In the dictionary definition below, the prompt *illness* was used as part of the post-modifier of *someone* in *someone with an illness*. Some participants followed the pattern literally and constructed "*The doctor cures him with his illness*" without being aware of the incompatibility of using a post-modifier after the pronoun *him*.

> The word *cure* means make well. *To make someone with an illness healthy again*, so my sentence is *The doctor cures him with his illness*, because it means someone with an illness healthy again, and form the sentence it means *the doctor will help the patient with his illness*. (Student 21)

There were cases in which the participants did not know the word class that a certain target word (e.g. *opposite*; prompts given: *go to – restaurant – opposite – cinema*) should belong to in the given partial context. Errors in determining which correct example to rely on occurred when the word class judgment was wrong, e.g.:

> *Opposite… Opposite* can be an adjective… can be a preposition….. If *opposite* is an adjective, then it is possible to have *to* after it, because there is an example *They're completely opposite to each other in every way*. After reading I don't think it is a noun. It shouldn't be an adverb either. Then I need to decide whether it is an adjective or a preposition. Here, *opposite* should be an adjective. It looks

> more like an adjective, so I will use *opposite to*. (Student 20)

General Learner Problems Not Resulting from Amount and Nature of Contexts Given

Learner problems irrespective of the amount and nature of contexts given could also be identified. These could be regarded as problems directly related to the use of a dictionary.

Some participants used the syntactic structure of a synonym as a model for language production despite the presence of conflicting dictionary information. A case in point was the use of the word *comprise* in the Sentence Construction Task (prompts given: *class – comprise – only French students*). The word was regarded as having the same syntactic requirement of its synonym *consist of*. Although there were examples showing the use of *comprise* without the preposition *of* or its coexistence with *of* in a different structure, some participants used the structure of the synonym to confirm their (mis)judgment, e.g.:

> *The course comprises a class book, a practice book and an audio tape*. *This class is comprised mainly of Italian and French students. Italian students comprise 60% of the class*. Then this word should really be of similar meaning to *consist of*. So let me see. If so, then the sentence should be *The class comprises of only French students*, i.e. actually it is quite like *The class consists of only French students* (Student 14).

An expected answer in mind or a previously known usage might override dictionary examples or definitions. Some participants were very adamant about finding information associated with the grammatical patterns which were consistent with their previous knowledge and followed their preconceptions irrespective of what they read from the dictionaries, apparently becoming "blind" to the presence of appropriate examples or definitions, e.g.:

> *A limit of* is the expected answer in my mind. (Student 1)
>
> The example … is *The class is comprised mainly of Italian and French students.* Because there is only this example, I will still use *This class comprises of only French students* (Student 20)
>
> *Comprise* is a verb. … *Comprise a class*. *The course comprises a class*. … Why is *of* not used … *Italian students comprise 60% of the class*. Actually it should be used with *of*. This is a formal use. So it should be *The class comprises of only French students*. (Student 22)

A dictionary headword often contained examples and definitions used in different grammatical patterns with very similar meanings, the subtle differences of which were too small for learners to differentiate. Sometimes a decision was made purely by guessing. Examples included the different complementation patterns of *rush*, such as *"rush into, rush to, rush through"* and of the word *frighten*, such as *"frighten off, frighten away":*

> There are many explanations for *rush*. Let me see…. Oh, they are so similar. Let me see. What's the difference? I think there should be no difference. (Student 19)
>
> Why are they so similar? … Should this be *away* or *off*?... I think it should be *off*. Why do I think so? I just guess. (Student 19)
>
> In the dictionary I can find that the word *frighten* means to make someone feel afraid, scared, and you can also use frighten somebody, something off, that is to make a person or animal so afraid that they go away and do not do something they were going to do. I think the word *off* is optional. I can simply omit it. (Student 20)
>
> I do not know if I should add *into* after the word *deceive*. There is another example in the dictionary with the use of *into* after *deceive*, and I cannot spot the differences between the two. (Student 9)

The grouping of different grammatical patterns under the same definition also created problems. In LDOCE5, *brush + against* and *brush + past* were grouped together under the definition *touch slightly*. An overall definition was given first, followed by an example of the former pattern and an example of the latter. Some participants mistook the definition as applicable only to the first pattern, e.g.:

> *Touch slightly,… to touch someone or something slightly when passing them*…. There is an example here….. But this example *Neil brushed past him in the doorway*. There is no explanation before it. There is suddenly this example…. That's so strange, I am not sure. (Student 13)

Some other problems related to the layout of an entry were also identified. The use of superscripts in LDOCE5 to differentiate words used in different word classes was not salient enough to attract learners' attention. Some learners had wasted much time searching for information of the word used in a different word class, e.g.:

> I saw it. I was stupid. After the word *charge*^{1} there is a very small word *n*. This shows that it is for the use of noun. When I turned over to the next page, I found *charge*^{2}. After *charge*^{2}, there is a very small word showing *v*. That means verb. ... OK, I have wasted a lot of time. (Student 13)

Discussion and Implications

In the foregoing sections, the qualitative and quantitative data obtained from the study have been presented. Some general insights into the usefulness of dictionary information and learners' use of dictionaries will be given in this section, followed by some specific insights into learners' dictionary use in the light of the amount and nature of contexts provided by the two dictionary con-

sultation tasks.

General Usefulness of Dictionary Information in Target Language Production

As can be seen from the improvements shown in the participants' performance in the second parts of the tasks, dictionary information is, to a certain extent, useful in guiding advanced ESL learners to determine the correct usage of a target word in language production. However, even when learners are engaged in language production when usage is the main concern, they do not pay enough attention to explicit grammatical information, such as grammar codes, which summarize the syntactic patterns of a target word. Instead, they rely on examples for the relevant information for language production and deduce the complementation requirements or other grammatical colligation of a target word from the structural patterns shown in the examples.

Learners' use of examples as sample models for language production is, however, often constrained by their dictionary skills and grammatical competence. In line with Dziemianko (in press)'s findings, many learners have difficulties in identifying the correct example from the many given examples and in choosing the corresponding structural pattern. The usage of words which can be used in different grammatical patterns and which invites numerous dictionary definitions and examples is most difficult to identify, as the presentation of the information may be user-unfriendly, and the definitions and/or examples listed may not show the subtle differences between the different usages. Learners' own preconceptions of word usage may also affect their interpretation of examples, and dictionary information seems to be helpful in giving confirmation to learners on their previous knowledge of the usage of a certain word rather than in providing new knowledge. When a certain usage is unfamiliar to learners, or when learners' preconceptions of the uses of a target word deviate from the normative structures, their preconceptions often override the information given in a dictionary entry and result in wrong decisions.

Another common problem is learners' reliance on the grammatical patterns of the synonyms of a target word. As observed in Chan (2012b), ESL learners may fall into the pitfall of incorrectly using the grammatical patterns of a synonym to deduce the usage of a target English word. The results of the present study also suggest that attempts to use a semantically-related word to deduce the syntactic use of a target word often lead to inaccuracies. Many inaccurate target language constructions can actually be seen as the results of influence caused by the syntactic patterns of a word's synonyms, especially when learners take such patterns to reinforce their misconceptions of the usage of the target word. When encountering a seemingly "similar" syntactic pattern, they become blind to the subtle differences between the usage of the target word and its synonyms or to the presence of a dictionary example which suggests an alternative pattern.

Usefulness of Dictionary Information when Learners' Thinking Processes are Led by their Mother Tongue

It is evident from the results of the study that the usefulness of dictionary information is sometimes constrained by the language in which learners' thinking processes are engaged. When learners formulate a thought in a foreign or second language, the word which would most probably come to their minds is the L1 word rather than the target language word that is needed (Laufer and Levitzky-Aviad 2006). An exact match between the meanings of the L1 and L2 words is often sought, so a mismatch in the meanings, albeit slight, may lead to difficulty. Some idiomatic expressions in a learner's native language (e.g. *擦身而過* used in Sentence (3) of the Sentence Completion Task (See Appendix A), which literally means touching the body when passing by), may not have an exact L2 translational equivalent which matches the literal meaning (i.e. touching the body). Monolingual dictionaries, which provide definitions only in one language, will not cater for the possible idiomatic meanings that translations in another language may have, especially when the definitions are meant to give the core meanings of the expressions rather than any associated implied meanings. Even after searching the whole dictionary entry, dictionary users who think in their mother tongue may still be daunted by the "imprecise" information given, thinking that the dictionary entry does not provide the English that they really need or want (Kernerman 2007).

Another problem that often arises from learners' attempts to use an L2 equivalent for an L1 word in mind is their exclusive focus on the meaning of the equivalent and their ignorance of the different structural requirements in the two languages. It is well-known that a pair of L1 and L2 equivalents do not necessary share the same structures, such as having different transitivity patterns for verbs (e.g. The verb *participate* is intransitive in English but its translation in Chinese *參加* is transitive) or different countability for nouns (e.g. English nouns can be countable, uncountable or both, whereas Chinese nouns are not distinguished in number) (Chan: 2004a). The order of constituents associated with a certain word may also differ in different languages. While the word *substitute* (used in Sentence (7) in the Sentence Completion Task (see Appendix A)) requires the same order or constituents (*substitute A for B*) as its Chinese equivalent *代替(A 代替 B)*, its synonym *replace*, which is a more common word familiar to most Cantonese ESL learners, requires a different order (*replace B with/by A*). Such similarities or differences in the syntactic requirements of different target vocabulary items and a certain native vocabulary item will of course not be revealed in a monolingual dictionary entry. Learners' previous knowledge of the target and native languages may intervene with their interpretation of dictionary information or even override the given information, resulting in indetermination or inaccuracy.

Usefulness of Dictionary Information when Learners' Thinking Processes are not Constrained by their Mother Tongue

When learners have in mind a partial context which is entirely in the target language, their familiarity with a word may still be too shallow to enable them to manipulate it appropriately in the context. One obvious problem is their inability to determine the word class of the target word, leading to inappropriate selection of and/or reliance on unacceptable grammatical patterns given in a dictionary entry. Some English words may have comparable contexts when they belong to different word classes. For example, the word *opposite* used in Sentence (4) of the Sentence Construction Task (see Appendix B) can have the pattern *Subject + Linking Verb + opposite (to) + Noun/Noun Phrase* when it is used as an adjective (e.g. *They are* ***opposite to*** *each other in every way*) and as a preposition (e.g. *The bank is* ***opposite*** *the supermarket*). The only difference between the two is the acceptability of a following preposition *to* when the word is used as an adjective. Dictionary definitions, examples or even grammar codes showing such a grammatical pattern could not help learners determine which word class the word should belong to when it is used in the context that the learners have in mind. To a certain extent, the dictionary information which learners base their inappropriate decisions on may even reinforce their common errors rather than help them use a word accurately.

Another problem associated with the use of dictionary information for language production with a partial context in the target language is learners' distortion of the normative sentence structure to accommodate the context. They tend to ignore the syntactic requirements of a target word by extracting or adding a certain constituent or part of a constituent and fit it into an alternative structure. The use of *with an illness* after *cure* (i.e. *cure him* ***with*** *his illness*) upon seeing *someone with an illness*, as well as the adding of the constituent *you* after *reason for* (i.e. *the reason for* ***you to*** *reject the proposal*), is a good illustration. Extracting the correct constituents from a dictionary example or definition and fitting it into a correct context requires a certain level of grammatical competence, without which learners will not be able to process the language successfully. If they cannot identify the subtle differences between the use of a constituent in a dictionary entry and their use of the constituent in their "derived" structure, misapplications of dictionary information will result. Such misapplications are surely not what lexicographers intend to see, but they are exactly the kinds of problems which learners often encounter in their learning of a second or foreign language.

Implications

The findings of the present research enlighten lexicographers about the actual difficulties ESL learners have in their use of monolingual dictionaries for target language production. It can be seen that because of the use of only one lan-

guage (learners' target language) in monolingual dictionaries, learners' language proficiency can become a main obstacle to their extraction and application of relevant information even if their thinking processes are not led by their mother tongue. This seems a paradox, because learners attempt to solve their language problems by consulting a resource which requires certain linguistic competence. In this connection, dictionary skills are of vital importance. "In order to benefit from the achievements of modern lexicography, dictionary users need to be trained how to use the dictionary to solve actual typical problems and questions" (Lew 2011a: 3). Not many students are equipped with the necessary skills which enable them to use a dictionary to the fullest extent, nor are they adequately informed of the bank of information they can get from a dictionary. ESL teachers are advised to design dictionary skills training targeting not just basic skills but also the assimilation of dictionary information and the application of such information to their authentic learning contexts. These training programs should also take into account learners' linguistic competence and their actual consultation problems, such as their misuse of the syntactic requirements of a synonym for the usage of a target word and their ignorance of dictionary information as a result of their preconceptions of target word usage.

ESL learners are advised to use multiple resources in their learning. Using both monolingual dictionaries and bilingualized dictionaries, which systematically take care of the learners' native language (Adamska-Sałaciak 2010), should be a good alternative in learners' paths towards complete mastery of a second language. Learners' mistaken beliefs about the alleged lack of definitions and usage information in bilingualized dictionaries should be dispelled (Chan 2010). They should not just focus on L1 equivalents but should supplement L1 definitions and examples with target language definitions and examples and use either to resolve the ambiguity that might arise from the adoption of usage information provided in the other language.

Dictionary compilers should be aware that many details of a dictionary entry may escape learners' attention, so more highlighting techniques or special features should be used. Extra columns or usage boxes showing subtle differences between certain usages (e.g. *opposite* used as an adjective and as a preposition) may help guide learners' choice of appropriate dictionary information.

Limitations

Notwithstanding the insights discussed above, the nature of the dictionary consultation tasks may have limited the generalizability of the study. The Sentence Completion Task, which attempted to provide a full context for the desired grammatical associations of the target words, may have been treated by some participants as a translation exercise. Their preoccupation with an exact correspondence between the given Chinese contexts and the English expressions in the dictionaries may have been the result of their treatment of the task

as a pure translation exercise.

The use of a few specific paper dictionaries may also have limited the authenticity of the study. In this technology-based new millennium, learners tend to use electronic or online dictionaries for quick and easy reference instead of relying on paper dictionaries. Even those who were used to using paper dictionaries may not have used the assigned dictionaries regularly. As Lew (2011b) points out, users tend to learn the structure and conventions of a dictionary they regularly use, and their reference skills evolve over time. Their unfamiliarity with the dictionaries used in the study may have partially contributed to the problems they encountered.

Conclusion

In this article, I have reported on the results of a study which investigated advanced Cantonese ESL learners' use of a monolingual dictionary in target language production with a full Chinese context and a partial English context. The results suggest that although a monolingual dictionary can help learners identify the correct use of a word, the usefulness of dictionary information is often constrained by the layout of the entries, the presentation of dictionary information, the learners' own preconceptions of word usage, and their deficient dictionary skills. The language in which learners' thinking processes are engaged also affects their use of dictionary information. It is suggested that lexicographers be informed of the needs and problems of dictionary users. With the advent of electronic dictionaries and online dictionaries and the increasing interest in empirical studies of electronic dictionaries (e.g. Chen 2010; Chon 2009; Dziemianko 2010), further studies using electronic or online dictionaries may be illuminating in uncovering other target language production problems facing ESL dictionary users.

Acknowledgements

The work described in this article was fully supported by City University of Hong Kong (CityU Strategic Research Grant Number: 7008005). The support of the university is acknowledged. I would also like to thank all the respondents of the survey for their participation, my research assistant for her administrative help, and the consultants in the City University Statistical Consulting Unit for their expert advice on statistical analyses.

Endnotes

1. A bilingual dictionary gives the target language translation equivalents (e.g. Chinese) of the source language (e.g. English) (Hartmann and James: 1998). A bilingualized dictionary (e.g. OALECD) contains entries which have been translated in full or in part into the target language (e.g. Chinese), but there are also definitions and examples in the source language (e.g. English) (see also Hartmann: 1994; James: 1994).

2. An unpublished result obtained from Chan (2010)'s survey on the use of bilingualized and monolingual dictionaries by 169 Hong Kong Cantonese ESL learners showed that about 38% of learners used Oxford Advanced Learner's Dictionary of Current English (OALDCE), 14%. 20% and 13% of learners respectively used the COBUILD, LDOCE and CALD dictionaries, and the rest used others. OALDCE was not used in the present study for investigation because the definitions and usage information given in that dictionary were used as the basis of the questions in all the dictionary consultation tasks used in the present study and other related studies (e.g. Chan 2012a). Therefore, COBUILD, LDOCE and CALD were chosen.
3. No survey was specifically carried out to investigate or prove that the uses of the words necessitated by the Chinese contexts were familiar or unfamiliar to Cantonese ESL learners in Hong Kong. The claim about the words being unfamiliar was purely made based on the experience of the researcher, who has been teaching English and/or linguistics at different local tertiary institutions for about 20 years.
4. It may be argued that the introspective questionnaires were strictly speaking not introspective, as the participants had to complete one questionnaire *after finishing* each search. The term was used in this article because the questionnaires were done during the implementation of the tasks, so the effects of information distortion or forgetting were minimized. Given that the self-reports were written ones, there was, to the author's knowledge, no better means of obtaining truly introspective feedback other than what was done in the task.
5. Because the results presented in this section were a brief summary of a huge amount of data collected from the study, the actual tables showing the detailed numerical results are not included in this article to save space.
6. In this article, no attempt was made to report all the think-aloud data collected from the study, as a tremendously huge amount of data was generated when the participants recorded the whole of their thinking processes, some of which was not significant enough for reporting. Only the data which generated the insights raised in the article will be included.
7. The introspective quotes included in this article are all reported verbatim from the written reports provided by the participants in the group, as they were required to complete the questionnaires in written English. On the other hand, most of the think-aloud reports and the interview transcriptions are only the author's closest translations of the participants' responses in idiomatic English, as the majority of the participants chose to speak in Cantonese or a mixed-code of Cantonese and English in making the think-aloud recordings and in doing the interview. Where the participants chose to speak in only English, the quotes are also reported verbatim.

References

Dictionaries

Cambridge Advanced Learner's Dictionary. Good, M. (Ed.). 2008. *Cambridge Advanced Learner's Dictionary*. 3rd edition. Cambridge: Cambridge University Press.

Collins COBUILD Advanced Dictionary. Sinclair, J. (Ed.). 2009. *Collins COBUILD Advanced Dictionary.* 6th edition. Boston, MA: Heinle Cengage Learning.

Longman Dictionary of Contemporary English. Fox, C. and R. Combley (Eds.). 2009. *Longman Dictionary of Contemporary English*. 5th edition. Harlow: Pearson Longman.

Other Literature

Abecassis, M. 2007. Is Lexicography Making Progress? On Dictionary Use and Language Learners' Needs. *Lexikos* 17: 247-258.

Adamska-Sałaciak, A. 2010. Why We Need Bilingual Learners' Dictionaries. Kernerman, I. and P. Bogaards (Eds.). 2010. *English Learners' Dictionaries at the DSNA 2009*: 121-137. Tel Aviv: K Dictionaries.

Atkins, B.T.S. and K. Varantola. 1997. Monitoring Dictionary Use. *International Journal of Lexicography* 10(1): 1-45.

Atkins, B.T.S. and K. Varantola. 1998. Language Learners Using Dictionaries: The Final Report on the EURALEX/AILA Research Project on Dictionary Use. Atkins, B.T.S. (Ed.). 1998. *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators.* Lexicographica Series Maior 88: 21-81. Tübingen: Max Niemeyer.

Battenburg, J.D. 1991. *English Monolingual Learners' Dictionaries. A User-oriented Study.* Tübingen: Lexicographica. Series Maior 39. Tübingen: Max Niemeyer.

Baxter, J. 1980. The Dictionary and Vocabulary Behavior: A Single Word or a Handful? *TESOL Quarterly* 14(3): 325-336.

Béjoint, H. 1981. The Foreign Student's Use of Monolingual English Dictionaries: A Study of Language Needs and Reference Skills. *Applied Linguistics* 2(3): 207-222.

Bhela, B. 1999. Native Language Interference in Learning a Second Language: Exploratory Case Studies of Native Language Interference with Target Language Usage. *International Education Journal* 1(1): 22-31.

Bogaards, P. and W.A. van der Kloot. 2002. Verb constructions in learners' dictionaries. Braasch, A. and C. Povslen (Eds.). 2002. *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 13–17, 2002.* Vol. II: 747-757. Euralex 2002, Copenhagen.

Bunton, D. 1989. *Common English Errors in Hong Kong*. Hong Kong: Longman.

Bunton, D. 1994. *Common Social English Errors in Hong Kong: A Guide to Polite English.* Hong Kong: Longman.

Carduner, J. 2003. Productive Dictionary Skills Training: What do Language Learners Find Useful? *Language Learning Journal* 28(1): 70-76.

Chan, A.Y.W. 2004a. Noun Phrases in Chinese and English: A study of English Structural Problems Encountered by Chinese ESL Students in Hong Kong. *Language, Culture and Curriculum* 17(1): 33-47.

Chan, A.Y.W. 2004b. Syntactic Transfer: Evidence from the Interlanguage of Hong Kong Chinese ESL Learners. *The Modern Language Journal* 88(1): 56-74.

Chan, A.Y.W. 2005. Tactics Employed and Problems Encountered by University English majors in Hong Kong in Using a Dictionary. *Applied Language Learning* 15(1&2): 1-28.

Chan, A.Y.W. 2010. Toward a Taxonomy of Written Errors: Investigation into the Written Errors of Hong Kong Cantonese ESL Learners. *TESOL Quarterly* 44(2): 295-319.

Chan, A.Y.W. 2011. Bilingualized or Monolingual Dictionaries? Preferences and Practices of Advanced ESL Learners in Hong Kong. *Language, Culture and Curriculum* 24(1): 1-21.

Chan, A.Y.W. 2012a. The Use of a Monolingual Dictionary for Meaning Determination by Advanced Cantonese ESL Learners in Hong Kong. *Applied Linguistics* 33(2): 115-140.

Chan, A.Y.W. 2012b. Cantonese ESL Learners' Use of Grammatical Information in a Monolingual Dictionary for Determining the Correct Use of a Target Word. *International Journal of Lexicography* 25(1): 68-94.

Chen, Y. 2010. Dictionary Use and EFL Learning. A Contrastive Study of Pocket Electronic Dictionaries and Paper Dictionaries. *International Journal of Lexicography* 23(3): 275-306.

Chon, Y.V. 2009. The Electronic Dictionary for Writing: A Solution or a Problem? *International Journal of Lexicography* 22(1): 23-54.

Cowie, A.P. 1999a. *English Dictionaries for Foreign Learners: A History*. Oxford: Clarendon Press.

Cowie, A.P. 1999b. Learners' dictionaries in a historical and a theoretical perspective. Herbst, T. and K. Popp (Eds.). 1999. *The Perfect Learners' Dictionary (?). Lexicographica Series Maior 95:* 3-13. Tübingen: Max Niemeyer Verlag.

Dziemianko, A. 2006. *User-friendliness of Verb Syntax in Pedagogical Dictionaries of English.* Lexicographica Series Maior 130. Tübingen: Max Niemeyer.

Dziemianko, A. 2010. Paper or Electronic? The Role of Dictionary Form in Language Reception, Production and the Retention of Meaning and Collocations. *International Journal of Lexicography* 23(3): 257-273.

Dziemianko, A. In press. *Noun and Verb Codes in English Monolingual Dictionaries for Foreign Learners: A Study of Usefulness in the Polish Context*.

Ericsson, K.A. and H.A. Simon. 1993. *Protocol Analysis: Verbal Reports as Data.* '93 Edition. Cambridge, MA: Bradford Books/MIT Press.

Frankenberg-Garcia, A. 2011. Beyond L1-L2 Equivalents: Where do Users of Users of English as a Foreign Language Turn for Help? *International Journal of Lexicography* 24(1): 97-123.

Hartmann, R.R.K. 1994. Bilingualised Versions of Learner's Dictionaries. *Fremdsprachen Lehren und Lernen* 23: 206-220.

Hartmann, R.R.K. and G. James. 1998. *Dictionary of Lexicography*. London/New York: Routledge.

Heaton, J.B. and N.D. Turton. 1987. *Longman Dictionary of Common Errors.* Harlow, Essex: Longman.

James, G. 1994. Towards a Typology of Bilingualized Dictionaries. James, G. (Ed.). 1994. *Meeting Points in Language Studies: A Festschrift for Ma Tailai. Working Papers*: 184-196. Hong Kong: Hong Kong University of Science and Technology Language Center.

Jenkins, G. 1990. *English Problem Words.* Hong Kong: The Commercial Press.

Kaivanpanah, S. and S.M. Alavi. 2008. The Role of Linguistic Knowledge in Word-meaning Inferencing. *System* 36(2): 172-195.

Kasper, G. 2000. Data Collection in Pragmatics Research. Spencer-Oatey, H. (Ed.). 2000. *Culturally Speaking: Managing Rapport through Talk across Cultures:* 316-369. London: Continuum.

Kernerman, I. 2007. What's So Good or Bad About Advanced EFL Dictionaries? Gottlieb, H. and J.E. Mogensen (Eds.). 2007. *Dictionary Visions, Research and Practice: Selected Papers From the 12th International Symposium on Lexicography, Copenhagen, 2004*: 139-145. Amsterdam/Philadelphia: John Benjamins.

Laufer, B. and T. Levitzky-Aviad. 2006. Examining the Effectiveness of 'Bilingual Dictionary Plus' — A Dictionary for Production in a Foreign Language. *International Journal of Lexicography* 19(2): 135-155.

Lew, R. 2004. *Which Dictionary for Whom? Receptive Use of Bilingual, Monolingual and Semi-bilingual Dictionaries by Polish Learners of English*. Poznań: Motivex.

Lew, R. 2011a. Studies in Dictionary Use: Recent Developments. *International Journal of Lexicography* 24(1): 1-4.

Lew, R. 2011b. User Studies: Opportunities and Limitations. Akasu, K. and S. Uchida (Eds.). 2011. *ASIALEX 2011 Proceedings. Lexicography: Theoretical and Practical Perspectives:* 7-16. Kyoto: Asian Association for Lexicography.

Lew, R. and A. Dziemianko. 2006. A New Type of Folk-inspired Definition in English Monolingual Learners' Dictionaries and its Usefulness for Conveying Syntactic Information. *International Journal of Lexicography* 19(3): 225-242.

Miller, J. 2008. Teachers and Dictionaries in Australia: Is There a Need to Train the Trainers? *TESOL in Context* 17(2) : 11-19.

Nesi, H. and P. Meara. 1994. Patterns of Misinterpretation in the Productive Use of EFL Dictionary Definitions. *System* 22(1): 1-15.

Summers, D. 1988. The Role of Dictionaries in Language Learning. Carter, R. and M. McCarthy (Eds.). 1988. *Vocabulary and Language Teaching:* 111-125. London: Longman.

Swanepoel, P. 2000. Providing Lexicographic Support for SL Vocabulary Acquisition: What Kind, under What Conditions, for Whom and Why? Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, August 8th–12th, 2000*: 403-418. Stuttgart: Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung.

Tono, Y. 2001. *Research on Dictionary Use in the Context of Foreign Language Learning: Focus on Reading Comprehension*. Lexicographica. Series Maior 106. Tübingen: Max Niemeyer.

Van Weijen, D., H. van den Bergh, G. Rijlaarsdam and T. Sanders. 2009. L1 Use During L2 Writing: An Empirical Study of a Complex Phenomenon. *Journal of Second Language Writing* 18(4) : 235-250.

Appendix A: Sentences used in Sentence Completion Task

1. 我不想草率下決定.
(rush): I don't want to _____ rush into _____ a decision.

2. 警方會控告他謀殺.
(charge): The police will _____ charge him with _____ murder.

3. 我和他擦身而過,但他看不到我.
(brush): He _____ brushed past _____ me, but he didn't see me.

4. 高價使很多顧客卻步.
(frighten): The high prices _____ frightened off _____ many customers.

5. 他騙她簽合約.
(deceive): He _____ deceived him into signing _____ the contract.

6. 我的耐性是有一個限度.
(limit): There is _____ a limit to my patience _____.

7. 這道菜你可以用牛油代替油.
(substitute): You can _____ substitute butter for oil _____ for this dish.

8. 這是一個戰爭死難者的紀念碑.
(monument): This is _____ a monument to _____ the people killed in war.

9. 我們超越了去年的銷售額.
(improve): We _____ improved on _____ last year's sales.

10. 他們對生孩子不抱任何希望.
(despair (verb)): They _____ despaired of having _____ children.

Appendix B: Sentence Construction Task

1. (little) (knowledge) (linguistics)
2. (what) (reason) (reject) (proposal)?
3. (class) (comprise) (only French students)
4. (go to) (restaurant) (opposite) (cinema)
5. (feel) (guilty) (leave the children alone)
6. (the doctor) (cure) (him) (his illness)
7. (anticipate) (see him) (in the library) (tomorrow)
8. (inform) (me) (the latest news)
9. (John) (befriend) (Mary) (when she was lonely)
10. (he) (assist) (murder) (her husband)

Bilingualized Dictionaries with Special Reference to the Chinese EFL Context

Yuzhen Chen, *Department of Foreign Languages, Putian University, Fujian, P.R.C. (spring1918@yahoo.com.cn)*

Abstract: As a type of dictionary with huge popularity among EFL learners in China, the bilingualized dictionary (BLD) deserves more academic and pedagogical attention than it receives nowadays. This article gives an overview of the BLD within the framework of dictionary research, including dictionary history, dictionary typology, dictionary criticism and dictionary use. It first traces, with a special reference to the Chinese EFL context, the origin and historical development of this type of dictionary, and then proposes several approaches to its classification. The strengths and weaknesses of the BLD are evaluated and its role in language pedagogy discussed, followed by an extensive review of the empirical studies of BLD use. Finally, further areas of BLD research are also suggested. It is hoped that such an overview would kindle more research interest in BLDs which is relevant to language pedagogy, dictionary use instruction and lexicographic practices.

Keywords: LEXICOGRAPHY, BILINGUALIZED DICTIONARIES, MONOLINGUAL DICTIONARIES, BILINGUAL DICTIONARIES, ORIGIN, HISTORICAL DEVELOPMENT, DICTIONARY TYPOLOGY, DICTIONARY CRITICISM, DICTIONARY USE, CHINESE EFL CONTEXT

Opsomming: Verklarende woordeboeke met 'n tweetalige dimensie met spesiale verwysing na die Chinese EVT-konteks. As 'n tipe woordeboek wat enorme gewildheid geniet by EVT-aanleerders in China, verdien die verklarende woordeboek met vertalings (*bilingualized dictionary* of BLD) meer akademiese en opvoedkundige aandag as wat dit deesdae ontvang. Hierdie artikel gee 'n oorsig van die BLD binne die raamwerk van woordeboeknavorsing, insluitende die geskiedenis van woordeboeke, woordeboektipologie, woordeboekkritiek en woordeboekgebruik. Dit skets eers, met spesiale verwysing na die Chinese EVT-konteks, die oorsprong en historiese ontwikkeling van hierdie tipe woordeboek, en dan word verskeie benaderings vir die klassifikasie daarvan voorgestel. Die sterk punte en swakhede van die BLD word beoordeel en die rol daarvan in die taalpedagogiek word bespreek, gevolg deur 'n uitvoerige oorsig oor die empiriese studies van die gebruik van die BLD. Ten slotte word verdere gebiede vir BLD-navorsing ook voorgestel. Daar word gehoop dat so 'n oorsig meer belangstelling sal wek in navorsing oor BLD's, wat relevant is vir taalpedagogiek, die aanleer van woordeboekgebruik en leksikografiese praktyk.

Sleutelwoorde: LEKSIKOGRAFIE, VERKLARENDE WOORDEBOEKE MET 'N TWEETALIGE DIMENSIE, EENTALIGE WOORDEBOEKE, TWEETALIGE WOORDEBOEKE, OORSPRONG, HISTORIESE ONTWIKKELING, WOORDEBOEKTIPOLOGIE, WOORDEBOEKKRITIEK, WOORDEBOEKGEBRUIK, CHINESE EVT-KONTEKS

Lexikos 22 (AFRILEX-reeks/series 22: 2012): 139-158

1. Introduction

The most distinctive feature of bilingualized dictionaries (henceforth BLDs) is their entry formula: headword + definition in the same language + gloss in a different language (James 2000), as in English–Chinese BLDs[1]. Such dictionaries are hybrids in nature and seen as "compromise products" between monolingual and bilingual dictionaries, combining the advantages of both (Hartmann 1993). In other words, BLDs are "neither absolutely monolingual nor absolutely bilingual — they are both at once" (Pujol et al. 2006: 198). Many dictionary use surveys in China have shown that BLDs are by far the most popular paper dictionaries with Chinese EFL learners (e.g. Yu 1999, Fan 2000, Lang and Li 2003, Chi 2003, Thumb 2004, Shi and Chen 2007, Li 2009, Chen 2007, 2011a). However, the commercial success of BLDs has not prompted extensive and in-depth research into this type of dictionary. Despite a few studies about the theoretical and practical issues, generally, there remains a lot to be explored in other areas of BLD research, particularly in dictionary use research. Given the prevalence of BLD use in the Chinese EFL context, it is necessary to give an overview of this type of dictionary within the general framework of dictionary research. Such an investigation is also relevant at the world level, as BLDs seem to enjoy only a marginal status in the western lexicographical circle.

2. BLDs: past and present

2.1 Bilingualization as a genre

According to Hartmann and James (2000: 14), the BLD is a type of dictionary based on a monolingual dictionary in which entries have been translated in full or in part into another language. Actually, this definition covers only one category of BLDs, for there are others which are not adapted from a monolingual work but written by the same dictionary writer(s) like the English–Chinese BLDs compiled independently by Chinese scholars. In the broadest sense, BLDs are dictionaries which contain on the right-hand side of an entry the corresponding information in both L1 and L2 to explain the entry headword on the left. This type of dictionary combines features of the monolingual dictionary (e.g. L2 → L2 formulation) with the bilingual dictionary (e.g. L2 → L1 equivalence) in a composite entry formula (e.g. L2 → L2 + L1), and therefore is also called a hybrid.

Compared with monolingual and bilingual dictionaries, BLDs seem to have a relatively short history. Yet long before the emergence of modern BLDs, some features of this genre had already appeared (Chen 2011b). *A Timely Gem Dictionary Tangut–Chinese* (《番汉合时掌中珠》), which was compiled in 1190 by Gulemaocai (骨勒茂才), an ethnic of Dangxiang nationality, was believed to be the world's earliest bilingual glossary with both source and target language explanations (Yong and Peng 2008: 377-378). Archaeological evidence shows

that this dictionary, engraved in wood tablets, contains 414 entries, each of which is arranged in four columns from the right to the left in the order of a) the Chinese phonetic notation of the Tangut word, b) the Tangut word, c) the Chinese equivalent to the Tangut word, and d) the Tangut phonetic notation of the Chinese equivalent. The dictionary can be seen as a very early precursor to the BLD as it includes both L1 information and its counterpart in L2, though only at the phonological level. According to Osselton (1995: 128), the first bilingualized English dictionary intended for language learners is Nathan Bailey's *Orthographical Dictionary* (1727) which provides French and Latin glosses for English headwords.

It should be pointed out that before the advent of modern monolingual learner's dictionaries, BLDs, originally conceived of as antidotes to bilingual dictionaries, had already been very popular in some Asian countries such as Bengal and India. As noted by James (2000: 136), the motivation for the development of a monolingual learner's dictionary was in reaction to the growing popularity of BLDs in India in the early years of the 20th century. However, as it turned out, the monolingual learner's dictionary began to be bilingualized a couple of decades later. The difference is that most modern BLDs provide definitions which have some linguistic basis insofar as the lexical selection is concerned (James 2000: 136). The first bilingualized version of modern monolingual learner's dictionary (English–Bengali) appeared in 1958. In 1970, the first English–Chinese bilingualized product of *Oxford Advanced Learner's Dictionary (OALD)* was published in Taipei, later in Hong Kong, followed by similar adaptations in other languages such as Hindi, Greek, Italian, Spanish, Norwegian, Japanese, and Portuguese (Hartmann 1994, Marello 1998). From the late 1970s onward, the Kernerman Publishers of Tel Aviv, Israel, have produced a special kind of BLDs, known as semi-bilingual dictionaries which provide simple and short glosses for English definitions to speakers of Hebrew, Arabic and some other European, Asian and African languages[2].

2.2 BLDs in China

In China, the early 20th century saw the publication of many English–Chinese BLDs. *The Commercial Press English and Chinese Pronouncing Dictionary* (《华英音韵字典集成》), published in 1902, is believed to be the first of its kind compiled by Chinese scholars (Wang 2010), followed by others such as *An English–Chinese Standard Dictionary* (《英华大辞典》1908), *The English–Chinese Dictionary* (《英汉双解词典》1912), and *The Practical English–Chinese Dictionary* (《实用英汉双解词典》1936). There were also a few early bilingualized versions of American collegiate dictionaries such as *Modern Dictionary of the English Language with Anglo–Chinese Explanation* (《新式英华双解词典》 1919) and *Webster's Collegiate Dictionary with Chinese Translation* (《英汉双解韦氏大学字典》1924). In recent decades, the Chinese dictionary market has been flooded with a wide variety of BLDs, most of which have been adapted from English learner's dictionaries, particularly the "Big

Five". New translated versions would always follow shortly after new editions of these major learner's dictionaries came out. There are also some BLDs based on American collegiate dictionaries such as *Random House Webster's Dictionary of American English*, *Webster's New Collegiate Dictionary* and *The American Heritage Dictionary for Learners of English*. In addition to learner's dictionaries, there are also other types of BLDs such as *DK Oxford Illustrated English–Chinese Dictionary*, *McGraw-Hill Dictionary of Scientific and Technical Terms, Oxford Dictionary of Economics*, and *Longman Dictionary of Language Teaching and Applied Linguistics*, to name but a few.

The category of BLDs that are designed independently by Chinese scholars continue to emerge and serve a wide range of users, such as *A Multi-functional Dictionary for College English Teaching and Learning*, *A New English–Chinese Dictionary with Multiple Usage*, and *Multi-functional English Dictionary with Chinese Translation*. Similarly, there are also other types of BLDs apart from learner's dictionaries, such as *An English–Chinese Dictionary of Business Management*, *A Dictionary of Everyday English Metaphors*, and *A Bilingual Chinese–English Dictionary of English Synonyms*. Compared with those adapted from monolingual learner's dictionaries abroad, the BLDs compiled by Chinese natives are far less known or used.

A random search of any major bookstore in China may give one the impression that BLDs have dominated the dictionary market. Indeed, many surveys on dictionary use by Chinese EFL students have indicated that, among various paper dictionaries, the BLD is by far the most popular and used most frequently. Over a decade ago, Yu (1999) found that both English and non-English majors in his sample of population preferred the use of BLDs, such as *Oxford Advanced Learner's English–Chinese Dictionary* (*OALECD*) and *Longman English–Chinese Dictionary of Contemporary English* (*LECDCE*). The BLD ownership rate by English majors reached 86.0%. Findings by Lang and Li (2003), as well as Kan and Wang (2003), supported that BLDs were most preferred by EFL students, with *OALECD* taking the lead, followed by *LECDCE*. In Shi and Pan (2005), 78.2% of non-English majors owned *OALECD* and in Shi and Chen (2007), the ownership rate of BLDs by English majors reached 87.5%. The popularity of BLDs with EFL learners was further corroborated by Jiang (2007) and Li (2009). In Chen's (2007) large-scale survey, the BLD was found again to be owned by more than half of the respondents and perceived as the most useful type of dictionary for EFL learning. The BLD ownership rate rose to 76.5%, according to Chen's more recent study (2011a). Actually, at the university where the author works as an EFL teacher, each English major owns an *OALECD*.

The BLD also enjoys huge popularity among tertiary-level students in Hong Kong. Fan (2000) reported that a majority of over 1000 respondents often used BLDs and also thought highly of them. Chi (2003) showed that *OALECD* was most popular with her subjects and also used most frequently. As confirmed by Thumb (2004), *OALECD* and *LECDCE* were particularly well received by the Chinese learners of English in Hong Kong. In a more recent study, Chan (2011)

indicated that 79% of the ESL (English as a Second Language) respondents used BLDs either exclusively or together with monolinguals.

Unlike the overwhelming popularity of BLDs over bilinguals and monolinguals in the paper medium, the situation with electronic BLDs is not so clear-cut. Theoretically speaking, all types of paper dictionaries can be converted into electronic forms, be they web-based (e.g. online dictionaries), PC-based (e.g. computer desktop dictionaries), or chip-based (e.g. pocket e-dictionaries). However, owing to the scanty research in the use of electronic dictionaries (except for pocket e-dictionaries) in China, one cannot claim with complete assurance that electronic BLDs are more popular than electronic bilinguals or monolinguals, although dictionary users show a clear preference for electronic dictionaries over paper ones.

As far as pocket e-dictionaries are concerned, despite the differences in the configuration of installed dictionaries, almost all major brands on the market such as CASIO (卡西欧), Wenquxing (文曲星), Nuoyazhou (诺亚舟), Kuaiyitong (快译通), Kuaiyidian（快易典）, Mingren（名人）, and Bubugao (步步高) contain various types of dictionaries, including BLDs, bilinguals and monolinguals. An overwhelming majority of BLDs installed in pocket e-dictionaries are adapted from the "Big Five", and the more expensive the pocket dictionary is, the better version of BLD it includes. Nevertheless, contrary to the case with paper BLDs, there is little evidence to support that students prefer to use electronic BLDs over other types of dictionaries installed in their pocket dictionaries. As for other forms of electronic dictionaries such as online dictionaries, computer desktop dictionaries and cell phone dictionaries, the situation is even more unclear as there are hardly any surveys in this regard.

3. BLDs: dictionary typology

It is notoriously hard to provide an ideal framework for dictionary classification, as it can be based on various criteria. Considering the marginal status of the BLD in the western lexicographical circle, it is not surprising that little effort has ever been made to classify this type of dictionary except for a couple of studies, i.e., James (1994) and Marello (1998).

3.1 James' typology of BLDs

Based on an extensive historical survey of BLDs for a variety of languages, James distinguishes three subtypes of the genre (James 1994, 2000, Hartmann 2001: 77, adapted for the EFL context by the author):

— Learner's dictionary: a monolingual learner's dictionary in the user's target language, with glosses in the user's first language to assist decoding tasks, e.g. an English–English–Chinese dictionary for Chinese learners of English.

— Teaching dictionary: a monolingual dictionary intended for native speakers, adapted for learners by the addition of glosses in their first language (and sometimes the deletion of all or part of the original definitions) to help decoding, e.g. an English (–English) dictionary with Chinese glosses, for Chinese learners of English.

— Learning dictionary: a monolingual dictionary intended for native speakers, with glosses in the user's target language(s) to assist encoding, e.g. a Chinese–Chinese dictionary with English glosses, for Chinese learners of English.

Within the "learner's dictionary", James further identified two subtypes: one is "monolingual adapted", i.e. the bilingualization of an already existing dictionary, the other "originally bilingual", i.e. definitions and equivalents are written by the same writer(s) (which has been the norm in the Indian sub-continent). The "teaching dictionary" may include dictionaries whose original forms are retained but with the addition of L2 glosses, and dictionaries whose original definitions are simplified. With regard to the "learning dictionary", there are also two subtypes: one for native speakers and the other for learners of the second language (James 2000).

Despite his ingenious ideas, James only provides a crude typology for BLDs. The names of the BLD types seem a little confusing and the categorization of BLD subtypes is based on random criteria. As can be seen above, the "learner's dictionary" is further classified according to dictionary compilers; the categorization of the "teaching dictionary" is based on dictionary contents, while the "learning dictionary" is further grouped from the perspective of target dictionary users. It can be argued that these three subtypes can also be categorized according to other criteria. Furthermore, there are other BLDs which may not fit in this framework.

3.2 Marello's categorization of BLDs intended for learners

Marello (1998) explored the bilingualized learner's dictionary with a special reference to A. S. Hornby's works. By examining closely a dozen BLDs adapted from Hornby's dictionary family into various languages, Marello clearly demonstrates that BLDs can differ in their degree of bilingualization as a consequence of using L1 or L2 in a given part of the dictionary article according to the purposes of users and their L2 proficiency levels. Some BLDs were found to have the minimum amount of bilingualization with simple and short glosses added only to entry definitions; some preserve all original information and translate it into L1, thus gaining the maximum amount of bilingualization, and others delete, add or substitute part of entry information and provide translation for headwords and/or examples.

Indeed, as argued by Cowie (1999: 195), "If progress is to be made towards

a precise categorization of bilingualized dictionaries, it is essential to recognize the sources from which they come, as well as the modifications made to the parent works and the purposes these are intended to serve." Marello (1998) presents an analytical framework for the BLDs based on Hornby's works by examining their source and modifications, thus broadening our understanding about BLDs, yet she focuses on only one of the various subtypes of BLDs, i.e. the learner's dictionary. There are also other genres which have been bilingualized, such as pictorial dictionaries and thesauruses. The dictionary examples listed in Section 2.2 also show the wide variety of BLDs available on the Chinese EFL dictionary market. Therefore, the classification of BLDs should be explored from a broader perspective.

3.3 Some tentative approaches to classifying BLDs

It is true that the development of BLDs parallels those in monolingual and bilingual lexicography (Hartmann 1993). There is now a widening range and scope of monolingual and bilingual dictionaries, so is the case with BLDs. Theoretically speaking, a majority of monolinguals and bilinguals can be bilingualized with only a few exceptions such as pronouncing dictionaries, dictionaries of synonyms/antonyms and dictionaries of etymology. Therefore, BLDs can be classified according to the same criteria as those applied to monolinguals or bilinguals. Nevertheless, as a type of dictionary with a hybrid nature, the categorization of BLDs can also be based on its own unique criteria. The author proposes the following tentative approaches to classifying BLDs.

— From the dictionary proper

  BLDs can be distinguished according to the dictionary proper. Like monolinguals and bilinguals, the classification of BLDs can be based on various criteria. For example, we can distinguish BLDs by purpose (descriptive vs. prescriptive, active vs. passive, general-purpose vs. specialized), by function (encoding vs. decoding, academic vs. pedagogical), by predominance of information categories provided (linguistic vs. encyclopedic), by format (alphabetical vs. morphological vs. thematic), by medium (print vs. electronic), by size (from unabridged to gem), and by user type (advanced-level vs. medium-level vs. elementary-level, adult dictionary vs. children's dictionary) (Hartmann and James 2000: 147, Zhang and Yong 2007: 97).

— From the dictionary compiler

  From the perspective of dictionary compilers, there are two categories of BLDs. One is the translated BLD which is based on monolingual dictionaries, like the BLDs adapted from the "Big Five", in which the L1 and L2 information is provided by compilers from different language back-

grounds. The other is the independently compiled BLD in which the L1 and L2 information is given by the same compiler. Within the former category, the BLD can be further divided according to the degree of bilingualization. Some BLDs are the products of full bilingualization in which (almost) all entry information is translated into another language, like *OALECD* and *LECDCE*, while others are semi-bilingual with only part of entry information translated, like those popular in Israel.

— From the style of language presentation

Based on the style of language presentation, there are two kinds of BLDs. One is the traditional BLD characterized by a juxtaposition or simultaneous presentation of L1 and L2 information. For example, the English definition is followed immediately by the Chinese translation, as in most of the English–Chinese BLDs in China. The other type of BLDs, however, separates the L1 and L2 information as an effort to increase the exposure to L2. For example, most BLDs in Israel provide a brief gloss for the L2 definition which is placed after an untranslated dictionary example instead of right after the L2 definition. The print deferred BLD introduced by Pujol et al (2006) is another case in point. Such a BLD is divided into a L2-L2 monolingual part and a L2-L1 bilingual part that are interconnected by means of page numbers. A couple of English–Chinese BLDs are also characterized by a separation of L1 and L2 information, yet in a different way, i.e. to divide each dictionary page into two columns, the left-column being the L2 information and the right-column the corresponding information in the L1.

James (2000: 143-144) observes:

> That, as yet, we have only an imperfect taxonomy of the genre of bilingualized dictionaries is perhaps a result of the stigma still attaching in some quarters to any but monolingual target-language dictionaries for learners, and the notion that bilingualized dictionaries are somewhat of a half-way house, a pandering to learners who have not acquired adequate study skills to master a monolingual dictionary.

Undoubtedly, there is still a long way to go before an ideal taxonomy of BLDs is reached, especially with the rise and popularity of various electronic dictionaries which have made a real impact on the dictionary scene.

4. BLDs: dictionary criticism

4.1 Perceived strengths and weaknesses of BLDs

The trend towards BLDs is in line with the double criticism that, on the one

hand, monolingual learner's dictionaries are too much like native-speaker dictionaries and, on the other hand, straightforward bilingual dictionaries are too far removed from the target language (Hartmann 1991: 79, cited in Fan 2000: 125). By including entry information in both L1 and L2, the BLD is assumed to combine the authenticity and reliability of the monolingual dictionary and the accessibility of native-language equivalents/translations of the traditional bilingual dictionary (Tseng 2005). The unique combination of L2 + L1 information can thus both assist learners with difficulty in understanding the L2 definition and help prevent the undifferentiated equation between L2 and L1 words. In particular, L1 translation in BLD entries can supplement the L2 definition, correct a misconception about it, reinforce user's understanding of the L2 headword, and meet their psychological need to know the equivalent in L1 Furthermore, the BLD entry formula helps prevent the misconception about one-to-one equivalence between L1 and L2 words (Nakamoto 1995, Thumb 2004: 20-21). In brief, the strength of BLDs "derives from their synergy: they bring together the advantages of two types of dictionary that for a long time have been regarded as irreconcilable and, hence, have been used (and sold) separately" (Pujol et al. 2006: 200).

Despite its perceived advantages, the BLD has also attracted some criticism. It is considered as "insufficient with regard to the standards which ought to be set for a true bilingual learner's dictionary" (Zöfgen 1991: 2889), yet this comment is refuted as "off the point" by Nakamoto (1995) as the BLD he discussed (the case of the semi-bilingual) is basically a monolingual work. Due to its unidirectionality, the BLD is of "limited usefulness" except for decoding (Hartmann 1994, Cowie 1999: 195). However, as revealed by a recent large scale survey (Chen in preparation), apart from reading, BLDs are also often referred to for translation, and collocational information in BLD entries is frequently used for encoding purposes. Another perceived shortcoming of the BLD is connected with the neglect or underuse of L2 information on the part of dictionary users (Pujol et al. 2006), as some evidence showed that learners prefer to read L1 rather than both languages (Fan 2000, Laufer and Kimmel 1997, Thumb 2004). Actually, this so-called drawback is more concerned with dictionary users' preference than with the inherent problem of the BLD itself. Besides, some studies (Chen 2011b, in preparation) indicate that most Chinese EFL learners prefer to use both languages in BLD entries.

According to Tseng (2005), the "irreversible inherent shortcoming" of the BLD is a lag behind its monolingual parent work in terms of the currency of the contents. Considering the time needed for a bilingualization project and the active nature of the lexicon, we cannot but agree with Tseng's opinion. Still, this point is irrelevant when it comes to the independently compiled BLDs.

It is argued that during bilingualization intercultural problems are bound to arise, especially when most BLD translators come from different linguistic and cultural backgrounds (Yao, 2004). This pessimism about the role of the translated BLD in intercultural communication was echoed by Zhang (2010)

who held a critical attitude towards such BLDs, asserting that they are based on the cognitive thinking of Western people, split the language system into distinct parts and are organized by a grammar-centered approach. Indeed, the source dictionary of the translated BLD is usually designed without differentiating country-specific learners and thus may fail to meet the specific needs of users from different linguistic and cultural backgrounds.

Apart from the above criticism, another point about the BLD should also be mentioned, i.e. its bulkiness. Since this type of dictionary contains both L2 and L1 information, it is usually big and heavy, which induces inconvenience in use and extra cost. Of course, this is not a problem with electronic BLDs.

It should be noted that, whatever the advantages or disadvantages claimed, they are mostly based on pedagogical intuition or theoretical speculation rather than empirical evidence. More empirical information is needed to argue for or against the use of BLDs. This is the very reason to call for more systematic and in-depth research on the use and usefulness of this type of dictionary.

4.2 BLDs and pedagogical lexicography

Teachers, lexicographers, linguists and users have much to gain from an increased awareness of the role of dictionaries in the language-learning process (Hartmann 1993). The growing market of monolingual learner's dictionaries has attracted extensive interest in pedagogical lexicography. However, within this flourishing field of research, the BLD seems to have drawn much less attention than it deserves. A few remarks about the BLD in language learning have been made, but only in passing (e.g. Atkins 1985, Thompson 1987, Wingate 2002, Lew 2004). Hartmann (1993, 1994) was one of the first to examine the BLD both theoretically and empirically, advocating that one of the priorities of pedagogical lexicography must be to critically evaluate the development of BLDs. In reviewing the types of user-related research, Cowie (1999: 177), based on the four points of focus summarized by Hartmann (1987), adds another two types of enquiry:

— Assessment of the special merits of the so-called bilingualized dictionary as compared with the standard monolingual or bilingual learner's dictionary.

— Consideration of which types of dictionary — monolingual, bilingual or bilingualized — need to be used at various phases of the learning process and for what purposes.

In fact, the evaluation of the BLD did not begin until the end of the last century, Overall, the BLD seems to have been placed in a position between monolinguals and bilinguals, playing a gap-bridging role. Cowie (1999: 195) asserts that the BLD "cannot supplant the monolingual learner's dictionary", but it can

"serve as a bridge between a standard bilingual and a fully-fledged monolingual work". In her investigation into the use of dictionaries for reading comprehension, Wingate (2002: 230) made a passing comment that the BLD might be the ideal solution for the intermediate learners, since they could "facilitate the transition from bilingual to monolingual dictionaries". Lew (2004) found that those entries with two ways of meaning provision had a confusing effect on learners at the lower levels, probably owing to their overcrowded information. However, Laufer and Hadar (1997: 195) argued with empirical evidence that a good BLD is suitable for all types of learners: unskilled dictionary users may rely mostly on the bilingual information. With progress in these skills, the monolingual information will gain relevance and importance.

In China, there is a body of dictionary reviews concerning particular BLD titles, yet they seem to suffer from inconsistent criteria for critical evaluation. As for the role of the BLD in EFL pedagogy, no consensus has ever been reached. Chen (2006) gives a detailed explanation of the rationale behind the BLD and argues that it can cater to the cognition of EFL learners and play a positive role in EFL learning. However, according to Zhang (2010), the BLD cannot meet the needs of Chinese EFL learners, because its source dictionary is designed without differentiating specific needs of learners in different countries. Yao (2004) also advocates a reevaluation of the role that mother tongue and bilingual/bilingualized dictionaries play in EFL learning.

Actually, the BLD assessment would never be complete, valid or convincing without the support of empirical evidence. Theoretical exploration of BLDs should go hand in hand with investigations into the use and usefulness of such dictionaries in language pedagogy.

5. BLDs: dictionary use

On the whole, published research on BLD use is minimal, with only a handful of studies which investigated one or more of the following aspects of dictionary use:

5. 1 Perceptions of BLDs and patterns of use

One of the earliest BLD use studies was reported by Hartmann (1994) who found that users appreciate the juxtaposition of target-language definitions and mother-tongue translations. The BLD can serve a useful function for decoding and provide a bridge between the traditional bilingual dictionaries and monolingual dictionaries aimed at more advanced learners. Though limited in the scope of research and methodology, Hartmann did succeed in sparking more interest in BLD use research.

Some research findings indicate that users tend to read either L1 or L2 rather than both in BLD entries. In her investigation into the use and evaluation

of various BLD information by Hong Kong tertiary-level students, Fan (2000) discovered that L1 equivalents were used significantly more often than L2 definitions though the latter were perceived as more useful than the former. Thumb (2004) confirmed that more students preferred to read L1 rather than L2 or both when using BLD entries. Such a general preference was corroborated by Laufer and Kimmel (1997) who reported that Israeli students tended to use the L1 or the L2 significantly more often than L1 + L2. However, Chen (2011b) revealed a different language preference: most students in her study liked to read both L1 and L2 instead of one of the two languages. Such finding is supported by the author's most recent research (Chen in preparation) which investigated BLD users' lookup behavior in CALL context. The discrepancies between Chen's findings and those of other researchers may be attributed to the subjects involved: Chen targeted at English majors who owned and often used BLDs while students in the other studies were at a lower level of English proficiency and may not be accustomed to BLD use.

Due to its distinctive feature of meaning presentation, the BLD is found to cater to a variety of lookup possibilities, individual preferences and proficiencies. Thumb (2004: 108) manifested that the BLD is "highly usable and useful because of its compatibility with the language needs of learners". Users displayed various individualized lookup patterns: some read only L1 or L2 for all lookups, some alternated between the two languages, some referred persistently to L1 + L2 while a few had a mixed use of L1, L2 and L1 + L2. Laufer and Kimmel (1997) identified five patterns of BLD use while in Chen (2011b), two more patterns turned up. Higher-level students tended to make better use of BLD features than lower-level ones (Fan 2000, Chen 2011b).

It seems that BLD users make only limited use of the dictionary as they tend to ignore the information concerning the habitual and idiomatic use of L2 words (Fan 2000). Among the various entry components, students consulted the context meaning of words most frequently, yet they seldom looked up information related to collocations, pronunciation, frequency, and appropriateness of words, and except for collocation, the above-mentioned information was also considered least useful. Such ignorance of these aspects of word knowledge is a "cause for concern" (Fan 2000: 134).

Efforts have also been made to explore the cognitive strategies during dictionary consultation. Thumb (2004) uncovered seven strategies of BLD use, i.e. ignoring, assuming, minimizing, checking, paraphrasing, stretching, and maximizing. The same strategy used by two different learners could produce different outcomes while the same strategy used to look up different words could produce the same outcome (Thumb 2004: 109). Despite the problems with the methodological issues (Bogaards 2005), Thumb did provide an interesting description of the highly complex and individual lookup strategies that also applied to the use of monolingual and bilingual dictionaries.

A couple of studies provided more detailed information about general BLD use and BLD perceptions. Chen (2011a) is the first one to focus exclusively

on BLD use in the Chinese EFL context, looking into several aspects of dictionary use, such as the degree of popularity of BLDs, the reference needs for BLDs, the reference skills of BLD users, the perceptions and evaluations of BLDs, and the needs of BLD users. Based on Chen (2011a), in Chen (in preparation), the scope of research was widened by covering the differences between paper BLDs and electronic BLDs in patterns of use and dictionary evaluation. Chan (2011) looked into the preferences and practices of Cantonese ESL learners which are mainly associated with BLDs and monolingual dictionaries. The study included investigations into the usual pattern of using a dictionary, the reasons for preferring BLDs over monolingual dictionaries or vice versa, and learners' perception of the disadvantages of both dictionary types.

5. 2 Dictionary effectiveness

A few studies compared the dictionary effectiveness between the BLD and other dictionary types for certain linguistic tasks. Laufer and Melamed (1994), later published as Laufer and Hadar (1997), is the earliest research into BLD use. They found that the BLD was substantially better than the bilingual and monolingual dictionaries for vocabulary comprehension and also significantly better than the monolingual dictionary for vocabulary production. Except for unskilled users, all levels of dictionary users under the BLD condition achieved the best results of task completion. Chen (2011b) yielded similar findings: the BLD was significantly better than the bilingual and the monolingual in terms of its overall effects on the comprehension and production of new words. In addition, BLD use also brought favorable results of vocabulary retention. The overall advantages of the BLD were confirmed again in the author's latest study (Chen in preparation) which involved more vocabulary tasks and retention tests. Furthermore, some problems with BLD use were also identified and discussed.

The merits of the BLD over other dictionary types was also reported by Raudaskoski (2002) who found that those using the BLD made more improvement in their translation performance than those using the bilingual. Unfortunately, the degree of difference in improvement between the two groups is not reported, and no attempt at statistical evaluation is claimed (Lew 2004: 30). The study by Zarei (2010) presented a more complicated picture. He noted that for elementary- and intermediate-level students, the use of the BLD obtained the highest post test scores for both comprehension and production tasks, yet for advanced students, the BLD ranked second in terms of effectiveness for comprehension and third for production.

These studies have cast some light on the effectiveness of different dictionary types in language learning. However, there is a methodological problem undermining the credibility of test results of this kind, i.e. the unbalanced choice of dictionary titles (Marello 1998, Cowie 1999, Tono 2000, Lew 2004). The low degree of comparability between dictionaries involved or the lack of con-

trol of lexicographic presentation that characterizes most of the studies (e.g. Laufer and Hadar 1997, Raudaskoski 2002, Zarei 2010) might make it premature to generalize the results with specific dictionary titles to general dictionary types. Bearing this methodological issue in mind, Lew (2004) designed some balanced entries for different dictionary types and compared their effectiveness for receptive tasks. The findings indicated that BLDs were significantly more effective than the monolinguals, yet, with two ways of meaning provision, such dictionaries may be too crowded and thus confusing to learners at the lower levels. The advantage of the BLD over the monolingual dictionary was supported by Hu (2009) who, in an investigation into the effects of different tasks on incidental vocabulary learning of Chinese EFL learners, reported that the BLD dwarfed the monolingual one on vocabulary retention tests. Both Lew and Hu used minidictionaries specially written for experimental studies. Although the degree of comparability between dictionaries is guaranteed, it can be argued that there may be difference in dictionary use between real-life and artificial conditions.

Laufer and Levitzky-Aviad (2006) compared an experimental L1-L2-L2 dictionary (called bilingual plus), both in paper and computerized forms, with a BLD and a traditional bilingual dictionary for L2 production. The results pointed to a clear advantage of such a dictionary. The authors (Laufer and Levitzky-Aviad 2006: 152) attributed its effectiveness to "the combination of the bilingual and monolingual information which most learners used". It should be mentioned here, though, that the BLD involved in their study was a unidirectional L2-L1 one and was obviously placed at a disadvantage for an L1-L2 translation task.

There are also a couple of studies comparing the effects of BLDs in different media. Chen (2010) revealed that there was no significant difference in dictionary effectiveness between a paper BLD and BLDs stored in pocket e-dictionaries for the comprehension, production and retention of new words. The finding was corroborated by Chen (2012) which involved the use of a computer desktop BLD and its printouts. It seems, as far as BLDs are concerned, outcomes of vocabulary learning are not dependent on the form of dictionary used.

5.3 Language-oriented lookup behavior in CALL context[3]

Several researchers used electronic dictionaries together with built in log files to investigate dictionary users' language-oriented lookup behavior and its effect on vocabulary comprehension, production and retention. With a specially designed CALL dictionary program which incorporated L2 explanation, L1 translation, sound, root and "extra" information, Laufer and Hill (2000) discovered that Hong Kong learners preferred to look up the L2 definition rather than L1 translation while Israeli learners had a reverse preference. Yet, despite such difference between the two learner groups, the use of L1 together with L2 led to good retention. The beneficial effects of L1 + L2 lookup, as noted by

Laufer and Hill (2000: 17), may lie in the richness of semantic encoding, or the prolonged attention that multiple items of information require, or both. In a replication study, Lew and Doroszewska (2009) reported that it was the L1 equivalent, alone or in combination with the L2 definition that was the best predictor of retention as well as being the most popular target of consultation. They (Lew and Doroszewska 2009: 253) also discovered that retention rates were not affected by the sheer amount of dictionary activity, thus leading to the conclusion that it is the quality rather than the quantity of lookups that makes a real difference.

To the author's knowledge, Liu (2007), later published as Chen and Liu (2008), and Chen (in preparation) are the only studies that adopted the CALL methodology to examine dictionary users' lookup behavior in the Chinese EFL context. Liu (2007) incorporated various pieces of dictionary information into a CALL program used for reading comprehension and found that the L1+L2 lookup pattern seemed to be very effective for word retention, a finding similar to Laufer and Hill's (2000). Different from Lew and Doroszewska (2009), Liu identified a significant correlation between the number of word clicks and vocabulary retention scores. Although the study failed to present and discuss the results in a concise and coherent manner, it seems to take the lead in terms of the methodology of dictionary use research in China.

Chen (in preparation) also used log files to record users' lookup behavior and explore its impact on vocabulary learning under CALL context. A general preference for a combined use of L2 definition and L1 translation was observed. Users demonstrated various lookup patterns, among which, the L2 + L1 pattern proved to be most conducive to incidental vocabulary learning. Similar to Lew and Doroszewska's finding (2009), there was no strong correlation between the amount of dictionary activity and the outcomes of vocabulary learning. Furthermore, students at different levels of vocabulary proficiency did not differ significantly in terms of dictionary click behavior.

6. Conclusion

As a type of dictionary with unique features, the BLD is particularly well-received by Chinese EFL learners. Considering its popularity in EFL learning, the research into this type of dictionary is of direct relevance to language pedagogy, dictionary use instruction and lexicographic practices. The investigation into the patterns and strategies of BLD use can provide useful insights and reference for language teaching, particularly for vocabulary pedagogy. By learning about users' needs and reference skills and identifying the problems with BLD use, we can enhance students' awareness of the role of dictionaries in language learning process and improve their dictionary use competence. Furthermore, besides a fundamental knowledge of the history and features of the BLD, the elicited information about users' needs and expectations also has significant bearings on lexicographic practices. This article provides an overview

of the BLD from several branches of dictionary research, including dictionary history, dictionary typology, dictionary criticism and dictionary use. It is hoped that such a review would attract more attention to the BLD from researchers all over the world.

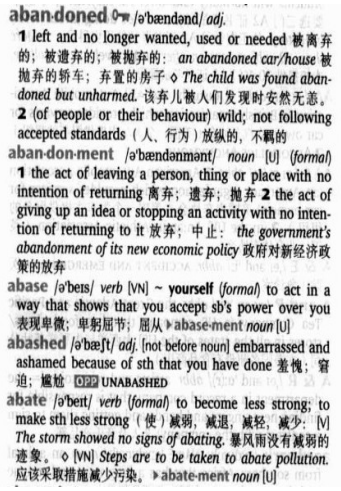
Compared with its large scale of use, the research on the BLD is far from sufficient. There remain many opportunities for further inquiry. As far as dictionary use is concerned, more research should be carried out to investigate the cognitive strategies of BLD use, especially during the lookup process of electronic BLDs. It would be illuminating to examine the effectiveness of BLDs for other language activities besides vocabulary learning, such as text translation and passage writing. It would also be interesting to examine the relation between dictionary use strategy and other vocabulary learning strategies. In addition, more rigorous research design should be adopted to identify specific problems with BLD use. Aside from questionnaires, surveys, interviews, tests and experiments, other methodologies such as observation, self-account, think-aloud protocols, video taping, and server logging etc. could also be usefully exploited.

From a lexicographic perspective, a lot of theoretical and practical issues remain to be solved. Systematic lexicographic principles should be drawn to guide the production or bilingualization of BLDs. How to independently produce high quality country-specific BLDs in response to the language cognition and special needs of BLD users is a huge project that requires effort and collaboration of scholars from different lines. As to the translated BLDs, more research could be done with regard to translation principles and translation skills of headwords and examples, as such issues are the key to BLD bilingualization. It is also worthwhile to explore how to improve the encoding function of BLDs so as to enhance their usefulness for language learning.

As BLDs are used extensively at Chinese colleges and universities, it is also necessary to conduct more studies that are directed at the BLD use instruction. It is of pedagogical significance to find out more about problems and pitfalls of BLD use and provide systematic guidance on the proper use of this type of dictionary. In this regard, more remains to be done.

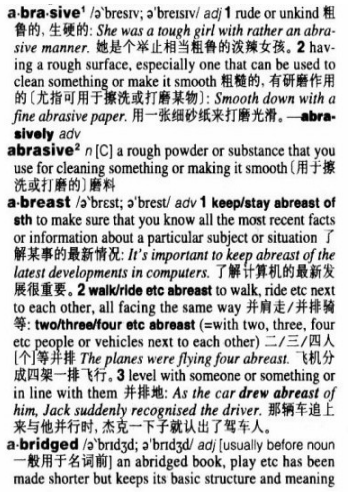
Notes

1. See the following excerpts from the most popular BLDs at the Chinese EFL dictionary market.



aban·doned /əˈbændənd/ *adj.*
1 left and no longer wanted, used or needed 被离弃的；被遗弃的；被抛弃的：*an abandoned car/house* 被抛弃的轿车；弃置的房子 ◇ *The child was found abandoned but unharmed.* 该弃儿被人们发现时安然无恙。**2** (of people or their behaviour) wild; not following accepted standards（人、行为）放纵的，不羁的
aban·don·ment /əˈbændənmənt/ *noun* [U] *(formal)* **1** the act of leaving a person, thing or place with no intention of returning 离弃；遗弃；抛弃 **2** the act of giving up an idea or stopping an activity with no intention of returning to it 放弃；中止：*the government's abandonment of its new economic policy* 政府对新经济政策的放弃
abase /əˈbeɪs/ *verb* [VN] ~ **yourself** *(formal)* to act in a way that shows that you accept sb's power over you 表现卑微；卑躬屈节；屈从 ▸abase·ment *noun* [U]
abashed /əˈbæʃt/ *adj.* [not before noun] embarrassed and ashamed because of sth that you have done 羞愧；窘迫；尴尬 OPP UNABASHED
abate /əˈbeɪt/ *verb* *(formal)* to become less strong; to make sth less strong（使）减弱，减退，减轻，减少 [V] *The storm showed no signs of abating.* 暴风雨没有减弱的迹象。◇ [VN] *Steps are to be taken to abate pollution.* 应该采取措施减少污染。▸abate·ment *noun* [U]

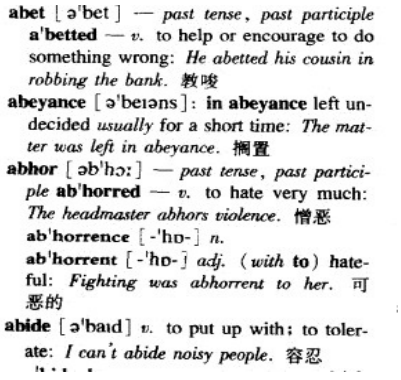
Figure 1: An excerpt from *Oxford Advanced Learner's English–Chinese Dictionary* (7th edition)



a·bra·sive[1] /əˈbreɪsɪv; əˈbreɪsɪv/ *adj* **1** rude or unkind 粗鲁的，生硬的：*She was a tough girl with rather an abrasive manner.* 她是个举止相当粗鲁的泼辣女孩。**2** having a rough surface, especially one that can be used to clean something or make it smooth 粗糙的，有研磨作用的〔尤指可用于擦洗或打磨某物〕：*Smooth down with a fine abrasive paper.* 用一张细砂纸来打磨光滑。—**abrasively** *adv*
abrasive[2] *n* [C] a rough powder or substance that you use for cleaning something or making it smooth〔用于擦洗或打磨的〕磨料
a·breast /əˈbrest; əˈbrest/ *adv* **1 keep/stay abreast of sth** to make sure that you know all the most recent facts or information about a particular subject or situation 了解某事的最新情况：*It's important to keep abreast of the latest developments in computers.* 了解计算机的最新发展很重要。**2 walk/ride etc abreast** to walk, ride etc next to each other, all facing the same way 并肩走／并排骑等：**two/three/four etc abreast** (=with two, three, four etc people or vehicles next to each other) 二／三／四人〔个〕等并排 *The planes were flying four abreast.* 飞机分成四架一排飞行。**3** level with someone or something or in line with them 并排地：*As the car drew abreast of him, Jack suddenly recognised the driver.* 那辆车追上来与他并行时，杰克一下子就认出了驾车人。
a·bridged /əˈbrɪdʒd; əˈbrɪdʒd/ *adj* [usually before noun 一般用于名词前] an abridged book, play etc has been made shorter but keeps its basic structure and meaning

Figure 2: An excerpt from *Longman Dictionary of Contemporary English* (English–Chinese, 4th edition)

2. Semi-bilingual dictionaries are popular in the Middle East and Europe while in China there are only a couple of such dictionaries available, i.e. *Password English–Chinese Semi-bilingual Dictionary* (《半双解英汉词典》) and *Bookman English Dictionary for Speakers of Chinese* (《书林易解英语词典》). The following is an example excerpt.



abet [əˈbet] — *past tense, past participle* **aˈbetted** — *v.* to help or encourage to do something wrong: *He abetted his cousin in robbing the bank.* 教唆
abeyance [əˈbeɪəns]: **in abeyance** left undecided *usually* for a short time: *The matter was left in abeyance.* 搁置
abhor [əbˈhɔː] — *past tense, past participle* **abˈhorred** — *v.* to hate very much: *The headmaster abhors violence.* 憎恶
abˈhorrence [-ˈhɒ-] *n.*
abˈhorrent [-ˈhɒ-] *adj.* (*with* **to**) hateful: *Fighting was abhorrent to her.* 可恶的
abide [əˈbaɪd] *v.* to put up with; to tolerate: *I can't abide noisy people.* 容忍

Figure 3: An excerpt from *Password English–Chinese Semi-bilingual Dictionary*

3. Strictly speaking, most of the studies reviewed in this section did not target at BLD use, yet, since these studies all involved language-oriented lookup preference of dictionary users, they are assumed to be relevant to the review.

References

Atkins, B.T.S. 1985. Monolingual and Bilingual Learners' Dictionaries: A Comparison. Ilson, R. (Ed.). 1985. *Dictionaries, Lexicography, and Language Learning*: 15-24. Oxford: Pergamon Press.

Bogaards, P. 2005. Review of Jenny Thumb. *Dictionary Look-up Strategies and the Bilingualised Learner's Dictionary*. *International Journal of Lexicography* 18(2): 272-274.

Chan, A. 2011. Bilingualised or Monolingual Dictionaries? Preferences and Practices of Advanced ESL Learners in Hong Kong. *Language, Culture and Curriculum* 24(1): 1-21.

Chen, W. 2006. Semi-bilingual Dictionary and Foreign Language Learning — A Case Study of Semi-bilingual Dictionaries. *Foreign Languages in China* 3: 62-66.

Chen, Yuzhen. 2007. A Survey of English Dictionary Use by English Majors in Chinese Universities. *Lexicographical Studies* 2: 120-130.

Chen, Yuzhen. 2010. Dictionary Use and EFL Learning: A Contrastive Study of Pocket Electronic Dictionaries and Paper Dictionaries. *International Journal of Lexicography* 23(3): 275-306.

Chen, Yuzhen. 2011a. A Questionnaire Survey and Experimental Study of English–Chinese Bilingualized Dictionaries. *Lexicographical Studies* 2: 141-158.

Chen, Yuzhen. 2011b. Studies on Bilingualized Dictionaries: The User Perspective. *International Journal of Lexicography* 24(2): 161-197.

Chen, Yuzhen. 2012. Dictionary Use and Vocabulary Learning in the Context of Reading. *International Journal of Lexicography* 25(2): 216-247.

Chen, Yuzhen. In preparation. Dictionary Use and EFL Vocabulary Learning: An Empirical Research of English Majors' Use of Bilingualized Dictionaries in China. Ph.D. Thesis, Xiamen University.

Chen, Y.P. and O. Liu. 2008. A Study of Annotation Information L2 Learners Select and Factors Affecting Word Retention on CALL Model. *Journal of Architectural Education in Institutions of Higher Learning* 17(4): 134-138.

Chi, A. 2003. *An Empirical Study of the Efficacy of Integrating the Teaching of Dictionary Use into a Tertiary English Curriculum in Hong Kong*. Hong Kong: Hong Kong University of Science and Technology.

Cowie, A.P. 1999. *English Dictionaries for Foreign Learners: A History*. Oxford: Clarendon Press.

Fan, M.Y. 2000. The Dictionary Look-up Behavior of Hong Kong Students: A Large-scale Survey. *Educational Journal* 28(1): 123-138.

Hartmann, R.R.K. 1987. Four Perspectives on Dictionary Use: A Critical Review of Research Methods. Cowie, A.P. (Ed.). 1987. *The Dictionary and the Language Learner*: 11-28. Tübingen: Max Niemeyer.

Hartmann, R.R.K. 1991. What's the Use of Learners' Dictionaries? A lecture given at the City Polytechnic of Hong Kong.

Hartmann, R.R.K. 1993. Recent Trends in Pedagogical Lexicography: the Case of the 'Bilingualised' Learner's Dictionary. Crochetiere, A. et al. (Eds.). 1993. *Proceedings of the15th International Congress of Linguistics* Vol.2: 159–162. Quebec: Laval University Press.

Hartmann, R.R.K. 1994. Bilingualised Versions of Learners' Dictionaries. *Fremdsprachen Lehren und Lernen* 23: 206–220.

Hartmann, R.R.K. 2001. *Teaching and Researching Lexicography*. Harlow: Pearson Education.

Hartmann, R.R.K. and Gregory James. 2000. *Dictionary of Lexicography*. Beijing: Foreign Language Teaching and Research Press.

Hu, M.Y. 2009. *L2 Incidental Vocabulary Learning and Retention: Efficacy of Dictionary-assisted Written Exercises*. M.A.. Dissertation, Suzhou University.

James, G. 1994. Towards a Typology of Bilingualized Dictionaries. James, G. (Ed.). 1994. *Meeting Points in Language Studies: A Festschrift for Ma Tailai. Working Papers*: 184-196. Hong Kong: Hong Kong University of Science and Technology Language Center.

James, G. 2000. Bilingualisation as a Genre. James, G. 2000. *Colporul: A History of Tamil Dictionaries:* 450-458. Chennai: Cre-A.. Reprinted in Hartmann, R.R.K. (Ed.). 2003. *Lexicography: Critical Concepts* .Vol. 3: 135–146. London: Routledge.

Jiang, Y.J. 2007. An Inspiration of a Questionnaire on English Majors' Dictionary Use. *Journal of Guangdong Polytechnic Normal University* 4: 107-110.

Kan, H.W. and Y. Wang. 2003. Survey of Dictionary Use by English Majors in Shanghai University. Zeng, D.J. (Ed.). 2003. *Bilingual Dictionary Studies*: 391–399. Shanghai: Shanghai Foreign Language Education Press.

Lang, J. and J. Li. 2003. A Survey of English Dictionary Use. *Journal of Beijing International Studies University* 6: 54–60.

Laufer, B. and L. Hadar. 1997. Assessing the Effectiveness of Monolingual, Bilingual, and "Bilingualised" Dictionaries in the Comprehension and Production of New Words. *The Modern Language Journal* 81(2): 189–196.

Laufer, B. and M. Hill. 2000. What Lexical Information do L2 Learners Select in a CALL Dictionary and How Does it Affect Word Retention? *Language Learning and Technology* 3(2): 58–76.

Laufer, B. and M. Kimmel. 1997. Bilingualized Dictionaries: How Learners Really Use Them. *System* 25(3): 361–369.

Laufer, B. and T. Levitzky-Aviad. 2006. Examining the Effectiveness of "Bilingual Dictionary Plus" — A Dictionary for Production in a Foreign Language. *International Journal of Lexicography* 19(2): 135–155.

Laufer, B. and L. Melamed. 1994. Monolingual, Bilingual and "Bilingualised" Dictionaries: Which are More Effective, for What and for Whom? Martin, W. et al. (Eds.). 1994. *EURALEX 1994 Proceedings, Papers submitted to the 6th EURALEX International Congress on Lexicography in Amsterdam, The Netherlands*: 565-576. Amsterdam: Vrije Universiteit.

Lew, R. 2004. *Which Dictionary for Whom? Receptive Use of Bilingual, Monolingual and Semi-bilingual Dictionaries by Polish Learners of English*. Poznań: Motivex.

Lew, R. and J. Doroszewska. 2009. Electronic Dictionary Entries with Animated Pictures: Lookup Preferences and Word Retention. *International Journal of Lexicography* 22(3): 239–257.

Li, Z. 2009. On the English Majors' Use of Dictionaries. *Journal of Guangdong University of Foreign Studies* 6: 97–100.

Liu, O. 2007. *Research on the Relation between Lexical Information L2 Learners Select in CALL Dictionary and Word Retention*. M.A.. Dissertation, Chongqing University.

Marello, C. 1998. Hornby's Bilingualized Dictionaries. *International Journal of Lexicography* 11(4): 292–314.

Nakamoto, K. 1995. Monolingual or Bilingual, that is not the Question: The Bilingualized Dictionary. Available at http://kdictionaries.com/newsletter/kdn2-2.html4.

Osselton, N.E. 1995. *Chosen Words: Past and Present Problems for Dictionary Makers*. Exeter: University of Exeter Press.

Pujol, D, M., M. Corrius and J. Masnou. 2006. Print Deferred Bilingualized Dictionaries and Their Implications for Effective Language: A New Approach to Pedagogical Lexicography. *International Journal of Lexicography* 19(2): 197–215.

Raudaskoski, S. 2002. Translation, the Key or the Equivalent? *Kernerman Dictionary News* 10: 2–3.

Shi, B. and P. Pan. 2005. Analysis of a Survey on English Learning Dictionary Using.. *Journal of Shanxi Coal-Mining Administrators College* 3: 78-79.

Shi, G.S. and G.H. Chen. 2007. Survey of Dictionary Use by English Majors. *Foreign Language Research* 1: 66–69.

Thompson, G. 1987. Using Bilingual Dictionaries. *ELT Journal* 41(4): 282–286.

Thumb, J. 2004. *Dictionary Look-up Strategies and the Bilingualised Learner's Dictionary: A Think-aloud Study*. Lexicographica. Series Maior 117. Tübingen: Max Niemeyer.

Tono, Y. 2000. On the Effects of Different Types of Electronic Dictionary Interfaces on L2 Learners' Reference Behavior in Productive/Receptive Tasks. Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, August 8th–12th, 2000*: 855-861. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

Tseng, Hugo T.Y. 2005. Reverse Indexing and Customization — Future Trends in Bilingualized Dictionaries. Paper presented at Fourth ASIALEX Biennial Conference, Singapore.

Wang, J.R. 2010. *The Commercial Press English and Chinese Pronouncing Dictionary* — The First large Scale English–Chinese Bilingualized Dictionary. *Publishing Science* 4: 103-106.

Wingate, U. 2002. *The Effectiveness of Different Learner Dictionaries: An Investigation into the Use of Dictionaries for Reading Comprehension by Intermediate Learners of German*. Tübingen: Max Niemeyer.

Yao, N.Q. 2004. Bilingualization: Equivalence and Intercultural Communication. Chan, S.W. (Ed.). 2004. *Lexicography and Translation*: 109-117. Tübingen: Max Niemeyer.

Yong, H. and J. Peng. 2008. *Chinese Lexicography: a History from 1046 BC to AD 1911*. Oxford/New York: Oxford University Press.

Yu, W.C. 1999. Survey of the Use of English–Chinese and Chinese–English Dictionaries. *Journal of Guangzhou Normal University* 12: 88–93.

Zarei, A.A. 2010. The Effect of Monolingual, Bilingual and Bilingualized Dictionaries on Vocabulary Comprehension and Production. *ELT Weekly* 54: 17-40. Available at http://eltweekly.com/ELTWeekly_Issue_54.pdf.

Zhang, Y.H. 2010. Cognitive Semantics and Multidimensional Definition for a New Generation of Bilingual/Bilingualized Learner's Dictionaries. *Journal of Foreign Language Teaching and Research* 42(5): 374-379.

Zhang, Y.H. and H.M. Yong. 2007. *Contemporary Lexicography*. Beijing: Commercial Press.

Zöfgen, E. 1991. Bilingual Learner's Dictionaries. Hausmann, F.J. et al. (Eds.). 1989–1991: *Wörterbücher. Dictionaries. Dictionnaires: An International Encyclopedia of Lexicography*. Vol. 3. 2888-2903. Berlin: De Gruyter.

Pushing Back the Origin of Bantu Lexicography: The *Vocabularium Congense* of 1652, 1928, 2012

Jasper De Kind, *KongoKing Research Group, Department of Languages and Cultures, Ghent University, Ghent, Belgium (jasper.dekind@UGent.be),*

Gilles-Maurice de Schryver, *KongoKing Research Group, Department of Languages and Cultures, Ghent University, Ghent, Belgium; and Xhosa Department, University of the Western Cape, Bellville, South Africa (gillesmaurice.deschryver@UGent.be),*

and

Koen Bostoen, *KongoKing Research Group, Department of Languages and Cultures, Ghent University, Ghent, Belgium; and Université libre de Bruxelles, Brussels, Belgium (koen.bostoen@UGent.be)*

Abstract: In this article, the oldest Bantu dictionary hitherto known is explored, that is the *Vocabularium Latinum, Hispanicum, e Congense*, handed down to us through a manuscript from 1652 by the Flemish Capuchin Joris van Gheel, missionary in the Kongo (present-day north-western Angola and the southern part of the Lower Congo Province of the DRC). The manuscript was heavily reworked by the Belgian Jesuits Joseph van Wing and Constant Penders, and published in 1928. Both works are currently being digitized, linked and added to an interlingual and multimedia database that revolves around Kikongo and the early history of the Kongo kingdom. In Sections 1 and 2 the origins of Bantu lexicography in general and of Kikongo metalexicography in particular are revisited. Sections 3 and 4 are devoted to a study of Van Gheel's manuscript and an analysis of Van Wing and Penders' rework. In Sections 5 and 6 translation equivalence and lexicographical structure in both dictionaries are scrutinized and compared. In Section 7, finally, all the material is brought together.

Keywords: KIKONGO, KONGO KINGDOM, CONGO, ANGOLA, CAPUCHINS, JESUITS, BANTU, LATIN, SPANISH, FRENCH, FLEMISH, AUTHORSHIP, COMPILATION STRATEGY, LANGUAGE, DIALECT, ORTHOGRAPHY, BASE LETTERS, DIACRITICS, PHONETICS, PROTO-BANTU, TRANSLATION EQUIVALENCE, MEANING EXTENSIONS, PARAPHRASES, LOANWORDS, MISNAMINGS, RETRANSLATIONS, LEXICOGRAPHICAL STRUCTURE, MANUSCRIPT, DATABASE

Samenvatting: Het terugschuiven van de oorsprong van de Bantoe lexicografie: het *Vocabularium Congense* uit 1652, 1928, 2012. In dit artikel wordt het oudste gekende Bantoewoordenboek bestudeerd, namelijk het *Vocabularium Latinum, Hispanicum, e Congense*, een manuscript uit 1652 aan ons overgeleverd door de Vlaamse Kapucijn Joris van Gheel, missionaris in Kongo (huidige Noordwest-Angola en het zuidelijk deel van de Neder-Congo provincie van de DRC). Het manuscript werd grondig bewerkt door de Belgische Jezuïeten Joseph van

Wing en Constant Penders, en gepubliceerd in 1928. Beide werken worden op dit ogenblik gedigitaliseerd, aan elkaar gekoppeld en toegevoegd aan een interlinguale en multimediale databasis waarin het Kikongo en de vroege geschiedenis van het Kongo koninkrijk centraal staan. In Delen 1 en 2 worden de oorsprong van de Bantoelexicografie in het algemeen en de Kikongo metalexicography in het bijzonder herbekeken. Delen 3 en 4 zijn gewijd aan een studie van Van Gheels manuscript en een analyse van Van Wing en Penders bewerking. In Delen 5 en 6 worden vertaalequivalentie en de lexicografische structuur in beide woordenboeken bestudeerd en vergeleken. In Deel 7, tenslotte, wordt al het materiaal samengebracht.

Sleutelwoorden: KIKONGO, KONGO KONINKRIJK, CONGO, ANGOLA, KAPUCIJNEN, JEZUÏETEN, BANTOE, LATIJN, SPAANS, FRANS, VLAAMS, AUTEURSCHAP, STRATEGIE VAN DE SAMENSTELLING, TAAL, DIALECT, ORTHOGRAFIE, BASISLETTERS, DIAKRITISCHE TEKENS, FONETIEK, PROTO-BANTOE, VERTAALEQUIVALENTEN, BETEKENISUITBREIDINGEN, PARAFRASES, LEENWOORDEN, VERKEERDE NAAMGEVINGEN, HERVERTALINGEN, LEXICOGRAFISCHE STRUCTUUR, MANUSCRIPT, DATABASIS

1. The origins of Bantu lexicography

In 1964 Benson wrote a remarkable article titled "A Century of Bantu Lexicography". Reading through the recent literature on Bantu lexicography, it seems as if scholars agree that the field, now half a century later, is indeed just 150 years old. In support of his argument Benson starts by retracing the lexicographical efforts of "a pioneer in the field such as Krapf" (p. 65), whose Swahili–English dictionary was published posthumously in 1882, whereas his first manuscript, "a vocabulary which became quite an extensive work" (p. 65), was written in 1844. Also for East Africa, Benson feels that "[a]fter Swahili the major Bantu language meriting consideration is Luganda" (p. 73), for which he starts his account with Le Veux's Luganda–French vocabulary of 1917. For Central Africa, Benson mentions Madan's Lala/Lamba/Wisa–English dictionary of 1913, a Bemba–English dictionary by the White Fathers of 1947, Torrend's English–Bantu-Botatwe dictionary of 1931, Hannan's Shona–English dictionary of 1959, and Scott's encyclopaedic Nyanja–English dictionary which was prepared in about 1870. For Southern Africa, Benson discusses Mabille's Southern Sotho–English dictionary of 1878, Brown's Tswana–English dictionary of the end of the 19th century, Doke and Vilakazi's Zulu–English dictionary of 1948, and McLaren's Xhosa–English dictionary of 1936. For West Central Africa, finally, Benson lists Bentley's Kikongo–English dictionary and grammar of 1887, Van Wing and Penders' Kikongo–French–Flemish dictionary of 1928, and Whitehead's Bobangi–English dictionary and grammar of 1899.

Benson (1964) does not refer to Doke's excellent overview of the "Early Bantu Literature" (1935), published three decades earlier. Doke stresses the invaluable contribution of "[t]he Angola Fathers [who] were the first to give us any monograph in or concerning a Bantu language" (p. 87), singling out Brusciotto as the greatest, being "the discoverer of the Bantu noun class and con-

cord system, and the first recorder of Bantu verbal derivations" (p. 102). Hence the subtitle of Doke's (1935) article: "The Age of Brusciotto". The first four works which Doke discusses all stem from the first half of the 17th century. In 1624 the Portuguese Jesuit Cardoso translates the catechism "Dovtrina Christãa", which is published in Portuguese with interlinear translations into Kikongo, making it the very first text in a Bantu language. Two decades later, in 1643, another catechism, Pacconio and De Couto's "Gentio de Angola" is published, written in Kimbundu with a Portuguese version on the opposite pages. Next comes Brusciotto himself, who is credited with a quadrilingual Kikongo dictionary manuscript as well as a translation of the "Dovtrina Christãa" into Latin and Italian, both in the year 1650. Unfortunately, the quadrilingual Kikongo dictionary is not now extant (Doke 1935: 96), which leads some scholars to doubt whether it was actually compiled (e.g. Van Wing and Penders 1928: xxvii). Conversely, copies of Brusciotto's grammar of Kikongo, published in 1659, are extant and have "earned for him lasting reputation in Bantu language study" (Doke 1935: 97). A manuscript from the same period that has also survived to this date is Van Gheel's (1652) trilingual Latin–Spanish–Kikongo dictionary.

What interests us most here are Brusciotto's 'lost' quadrilingual dictionary manuscript of 1650, and Van Gheel's still-existing trilingual dictionary manuscript of 1652. Given Van Gheel's manuscript survives to this day, it is possible and even necessary to move the origin of the field back to 1652 or, writing in 2012, to state that the field of Bantu lexicography is (at least) 360 years old.

2. Metalexicographical studies on Kikongo

In a way, it is not surprising that the first dictionary of a Bantu language is one for Kikongo (H16), the Kongo kingdom being one of the first Bantu-speaking regions where the Portuguese landed. With a dictionary history of 360 years, one would therefore expect Kikongo lexicography to be a popular and oft-discussed topic in Bantu metalexicographic circles. Yet nothing is further from the truth. In twenty-one years of *Lexikos*, for instance, not a single dictionary aspect of Kikongo lexicography has been discussed. The closest one has come to the Kongo kingdom and its languages and dialects, is via Gabon. Three years ago, Ndinga-Koumba-Binza and Roux (2009) as well as Mavoungou (2009a), each devoted an entire contribution to Civili (H12). Civili, also known as Fiote, belongs to the wider Kongo language cluster — that is, Guthrie's group H10 — and is spoken along the coast in Congo-Brazzaville as well as in adjacent coastal areas in Gabon and Angola's Cabinda, and is associated with the historical Loango kingdom. Moving further afield, "sister languages" of Civili (Mavoungou 2006: 141), namely Yipunu (B43) and Yilumbu (B44), have also been covered to some extent in *Lexikos* (Mavoungou 2002, 2006, 2009). Similarly, in twenty-four years of the *International Journal of Lexicography* (IJL), Kikongo is only mentioned once in passing, in a dictionary review of French in Congo (Rey-Debove 1992: 160), and once in a definition for Kituba (Tsakona

2007: 120). The lingua franca Kituba (H10b) itself, also known as Munukutuba, Monokutuba or Kikongo ya Leta, a pidgin/creole based on Kikongo as lexifier, would be a good candidate to fill the lack of metalexicographical studies on Kikongo, but both *Lexikos* and IJL are silent about this language as well, except for a passing mention in De Schryver (2003: 18).

While lexicographers may not have concerned themselves with metalexicographical studies on Kikongo, dictionary compilers have been quite busy, as the lists of Kikongo reference works in for instance Doke (1945: 17-22) and Hendrix (1982: 45, 96-99, 186-187, 238, 244, 254, 262, 271) attest to.

3. The Vocabularium Latinum, Hispanicum, e Congense (Van Gheel 1652)

3.1 The Capuchin missions in the Kongo and their linguistic works

In the year 1645, the first Capuchins arrived at the port of Mpinda, in Soyo, located in present-day north-western Angola, just south of the Congo River. Their purpose was to spread the Christian faith among the Kongolese population. The missionaries of this first caravan settled in Soyo and Mbanza Kongo (San Salvador), but did not engage in learning the indigenous language, since most of the Africans in these two urban centres already had sufficient knowledge of Portuguese (Hildebrand 1940: 259). Three years later, following the arrival of a second caravan of Capuchin missionaries, they realized the importance of acquiring the native language in order for them to pursue their evangelistic aspirations in the hinterland as well (Hildebrand 1940: 259; Nsondé 1995: 57). This second caravan included such illustrious missionaries as Antonio de Teruel and Girolamo da Montesarchio (Hildebrand 1940: 261), who engaged in the compilation of sermons, vocabulary lists and grammars in Kikongo. Alas, very few of these works have survived.

A later Capuchin caravan to the Kongo included our subject, the Fleming Joris van Gheel. The missionaries had set sail in 1648, but only reached the port of Mpinda in June 1651. After his arrival, Van Gheel was sent into the district of Matari (Van Wing and Penders 1928: xxiii).[1] His stay in Kongo was rather short, since he died on the 17th of December 1652, as a result of having been beaten by villagers for disrupting a ritual and destroying their ritual objects (Nsondé 1995: 127; Thornton 2011). It is during this short period that Van Gheel managed to pen a manuscript which includes, in addition to a number of spiritual and worldly texts appended to the front and back, the trilingual *Vocabularium Latinum, Hispanicum, e Congense*, the oldest surviving source of the Capuchin description of Kikongo.

3.2 The question of authorship

It is generally accepted that Joris van Gheel physically wrote the dictionary,

although the manuscript does not include any sign of authorship. This assumption is based on the fact that the handwriting clearly corresponds to other texts which are known to have been written by Van Gheel (Van Wing and Penders 1928: xxii-xxiii; Thornton 2011). The question of authorship, on the other hand, has been debated ever since the manuscript was discovered. D'Alençon (1914: 42) claims that Van Gheel cannot possibly be the author of the dictionary, considering that his stay was too short to acquire sufficient knowledge of the language. D'Alençon suggests that Van Gheel copied the dictionary merely for his own use. Van Wing and Penders (1928: xxvi-xxvii) refute this argument and point out that no potential original antedating 1652, from which Van Gheel could have copied, has been found. They consider d'Alençon's argument to be a confirmation of Van Gheel's linguistic capacities and of the extreme, though not insuperable, difficulties of the enterprise. Further on, Van Wing and Penders (1928: xxix) seem to nuance their argument, however, and claim that it might also be possible that Van Gheel actually used a vocabulary list of Antonio de Teruel, the Capuchin missionary who was part of the second caravan.

Hildebrand (1940: 263-264), author of a book-length biography of Joris van Gheel, suggests that the Flemish Capuchin copied his dictionary from a vocabulary list previously compiled by the Capuchin prefect Buenaventura d'Alessano, as well as others including Antonio de Teruel and José de Pernambuco.[2] Hildebrand (1940: 259-265) is also the first to mention the considerable influence exerted by Manuel Roboredo on the linguistic enterprises of the Capuchins. Roboredo was a Kongolese priest, child of a Portuguese nobleman and a Kongolese mother who belonged to the royal lineage of King García II of Kongo (Hildebrand 1940: 260). According to Hildebrand (1940: 261-265), it is Roboredo who taught the Capuchins the language, and it is also he who directed most of the compilation of their linguistic works. In fact, Hildebrand is very clear with respect to the authorship of the dictionary in question, as he states:[3]

> Le grand mérite de la rédaction revient à Roboredo, en un certain sens, le dictionnaire est son œuvre. La rédaction a été faite à la demande des Pères; ceux-ci peuvent revendiquer une partie du mérite de la belle entreprise. Le vocabulaire semble le travail collectif des nouveaux missionnaires, surtout d'Antoine de Teruel et de Joseph de Pernambouc, sous la direction de Roboredo … Telle a été la genèse du remarquable vocabulaire latin-espagnol-congolais, que nous connaissons par <u>la copie</u> du P. Georges. (Hildebrand 1940: 264, underlining ours)

Doke (1935), who had had access to an earlier study of Hildebrand (1934), is of the same opinion:

> There can be no doubt, however, that he [Van Gheel] <u>copied</u> a manuscript known to be in existence at the Mission Station of San Salvador before his arrival. Joris was only a beginner, having been under two years in the country at the time of his death. Though the dictionary is probably not the work of a single person, it is practically certain that in the main it is to be ascribed to *Roboredo*, a

> Spaniard whose name is the only one mentioned in the original text. (Doke 1935: 97, underlining ours)

Contemporary scholars support (parts of) this argument, and especially focus on the merits of Manuel Roboredo. Nsondé (1995: 60), for instance, does not neglect the remarkable linguistic capacities of Joris van Gheel — who mastered Latin, Spanish and English before his arrival in the Kongo, in addition to his mother tongue Flemish — but he attributes the majority of the linguistic works of the Capuchins to Roboredo.[4] In this respect, he also mentions the gratitude expressed by Buenaventura d'Alessano, the prefect of the Kongo mission, who openly recognized the merits of Roboredo (Nsondé 1995: 58-59). This view is shared by Thornton (2011), who considers Van Gheel to have copied from a vocabulary list compiled by the Spanish Capuchins José de Pernambuco and Francisco de Veas, with the aid of Roboredo and under the direction of Bonaventura da Sardegna (or da Nuoro). Similar arguments can be found in Bonvini (1996: 140) and Gray (1998), who consider Bonaventura da Sardegna and Manuel Roboredo to be the compilers of the dictionary. Bontinck (1980: 530), on the other hand, singles out José de Pernambuco as the writer of the first vocabulary lists, from which other Capuchins must have copied, such as Antonio de Teruel, Girolamo da Montesarchio and Joris van Gheel. The prefect, Buenaventura d'Alessano, is also often cited in the context of the compilation process, but this may be due to the fact that he reported the event to Rome (Nsondé 1995: 58-59; Thornton 2011).

In Section 3.4, we discuss linguistic evidence indicating that the main dialect represented in the manuscript is the direct ancestor of the Kisikongo variety currently spoken at Mbanza Kongo, the former capital of the Kongo Kingdom, and not the Kisolongo variety spoken along the coast. Given that Roboredo was close to the royal court at Mbanza Kongo, this evidence also supports the hypothesis of his strong contribution to the compilation of the *Vocabularium*.

3.3 The compilation strategy

In Addendum 1, pages 41-42 from the *Vocabularium Latinum, Hispanicum, e Congense* are shown. As may be seen, in this manuscript a lemma sign in Latin is typically followed by, first its translation into Spanish (although at times this slot remains empty), and second one or more translation equivalents in Kikongo. The interspersed metalanguage, which is used to indicate parts of speech and to clarify grammatical points, is presented in (abbreviated) Latin. That missionaries use Latin should not surprise, but the presence of Spanish in Kongo, rather than Portuguese, may surprise. The reason seems to simply boil down to the availability of existing reference works at the Mission Station. Both Hildebrand (1940: 264) and Bontinck (1976: 155-156) suggested that the source text must have been one of the re-editions of De Nebrija's (1492) Latin–Spanish

Dictionarium. In a follow-up study, Bontinck (1980: 531-533) settles for the re-edition of 1581, published in Antequera. On the one hand Bontinck sees some macro- as well as microstructural correlations between De Nebrija's 1581 re-edition and the 1652 manuscript, and on the other he uses the place of publication to go as far as to pinpoint the very missionary — unsurprisingly from Antequera — who must have brought a copy down to the Kongo. That the *Dictionarium* was used as a base sounds rather plausible, but the evidence for a particular edition is less convincing. More or less any of the numerous works of De Nebrija (Wilkinson 2010: 30-38) that had been published by the mid-17th century could have been a candidate, and indeed, Nsondé (1995: 232) refers to the re-edition of 1570. That edition was published in Antwerp, so one could as well argue that it is Joris van Gheel who brought a copy of the *Dictionarium* to the Kongo.[5]

In Addendum 2 the start of the section "C before O" in the 1570 edition of the *Dictionarium* is shown. A comparison with Addendum 1 reveals some similarities, but especially many differences. Pinpointing the exact edition, however, goes beyond the scope of the present article. Yet, what is interesting to note is the strategy itself. Just as the first monograph in a Bantu language was actually a translation (cf. Section 1), so is the first reference work in a Bantu language. The use of an existing dictionary as a kind of template, to be filled in with the local language, seems to have been a common strategy of the time. An example from Mexico is the 16th century *Vocabulario trilingüe*, a trilingual Spanish–Latin–Nahuatl dictionary, incidentally also based on one of De Nebrija's dictionaries, the *Vocabulario de romance en latin* of 1516 (cf. Clayton 2003).

3.4 The language/dialect described

The question of authorship is extremely relevant when it comes to determining the exact variety of Kikongo that is being described in the manuscript, since Kikongo itself does not refer to one single language, but to a large dialect continuum manifesting a family resemblance structure. Neighbouring dialects are mutually intelligible, but dialects at the extreme ends of the chain are not. If Van Gheel copied from another vocabulary list, the variety described in his dictionary does not necessarily represent the varieties of the areas in which he was preaching. Van Wing and Penders, however, make the following, rather contradictory, statement:[6]

> De door hem [Joris van Gheel] opgeteekende taal is die van de streek waar hij werkzaam was; het dialekt van Sogno, wellicht het meest door zijn voorgangers gebruikt, heeft echter de overhand. Deze taal overigens heeft ook P. de Teruel moeten leeren te Mbata, te Nkusu en te Mpemba. (Van Wing and Penders 1928: xxx-xxxi)

While this statement could well be read as an argument favouring the hypothe-

sis that Van Gheel copied from earlier Capuchin work, Van Wing and Penders do not entertain this option and they continue to consider Van Gheel to be the real author of the dictionary.

According to John Thornton (personal communication, January 2012), De Cadornega (1680) mentions that there were three dialects of Kikongo and gives their approximate limits. It is not clear to what extent these dialects correspond to the three major Kikongo varieties spoken in northern Angola today: (i) Kisolongo along the coast; (ii) Kisikongo, also known as Kisansala, spoken in the wide vicinity of Mbanza Kongo; and (iii) Kizombo spoken further east.

Van Wing and Penders are not the only ones who believe that the dialect of Soyo (*Fl. and Fr*. Sogno, *Prt*. Sonho), of which Kisolongo would be the closest descendant, dominates in Van Gheel's manuscript.[7] Bontinck (1976: 156) actually uses the assumed predominance of this dialect as an argument in favour of José de Pernambuco, who stayed in Soyo, to be the compiler of the first vocabulary list. John Thornton (personal communication, January 2012), however, does not believe that it is the coastal dialect of Soyo that is being described, but rather the dialect from Mbanza Kongo (San Salvador), spoken 300 km inland.[8]

In De Kind (2012), a comparative phonological and morphological study between the 17th century Kikongo described in the manuscript and more recent Kisolongo and Kisikongo varieties is carried out. On purely phonological grounds it is not possible to determine which Kikongo variety is described in the manuscript, since only minor differences have been observed in this regard. However, some remarkable differences have been observed regarding the morphology of the Kikongo varieties concerned. The 17th century variety and the Kisikongo variety share innovations regarding prefix loss or reduction which are not shared by the Kisolongo variety. The clearest examples are the prefixes of classes 5 and 10. The former shifted to **e-** both in the 17th century variety and in 19th century Kisikongo, and subsequently disappeared in present-day Kisikongo, but is maintained as **di-** in Kisolongo. The prefix of class 10 is realized as **zi-** in Kisolongo, but is lost in the 17th century variety and in Kisikongo. The sound changes which the augment or pre-prefix underwent also constitute a shared innovation between the 17th century variety and Kisikongo, both having the **e-o-o** type, while Kisolongo exhibits the **e-e-o** type. Both types evolved from the ancestral **e-a-o** type. In sum, based on shared morphological innovations, we can conclude that the variety described in the manuscript is a predecessor of Kisikongo, and not Kisolongo.

3.5 The orthography used

This question of authorship is also relevant to determine on which language the orthography of the manuscript is based. It can, at present, not be answered with complete certainty, but it seems to be both Portuguese and Spanish based. Portuguese was the language spoken by Kongolese priests, such as Roboredo,

who, as we saw, played a pivotal role in the compilation process. At the same time, many of the Capuchin missionaries came from Spain, although several were also Italian. Especially interesting are José de Pernambuco and Francisco de Veas, who participated in the compilation process and who were both Spanish (Thornton 2011). Moreover, the director of the compilation, Bonaventura da Sardegna, was of Italian origin, but studied in Spain (Gray 1998).

4. Le plus ancien dictionnaire bantu/Het oudste Bantu-woordenboek (Van Wing and Penders 1928)

So far, we have neatly kept Van Wing and Penders' Kikongo → French/Flemish dictionary of 1928 (mentioned in Section 1), and Van Gheel's Latin/Spanish → Kikongo manuscript of 1652 apart, even though there is a connection. According to Benson (1964: 77), Van Gheel's (1652) manuscript "was edited and reproduced" by Van Wing and Penders (1928). Merely looking at the direction (into Kikongo in 1652, vs. out of Kikongo in 1928) and languages involved (with Latin and Spanish as source languages in 1652, vs. French and Flemish as target languages in 1928), it should be clear that this cannot be a 'reproduction' by any stretch of the imagination. Compare Addendum 3, which shows a random page taken from Van Wing and Penders' dictionary, with the manuscript pages seen in Addendum 1. In this respect we concur with Doke, who rightly said about Van Wing and Penders' effort:

> Unfortunately the present Editors have not published the manuscript in the form in which it was written, viz. Latin-Spanish-Kongo, but have taken out the 7000 odd Kongo words alphabetically, and then added French and Dutch equivalents. Since the publishing of such a work to-day is not of everyday practical worth, but of great value to students, such a method of handling the manuscript is the opposite of scientific. (Doke 1935: 96)

The *Vocabularium Congense*, in its 1928 incarnation — which Van Wing and Penders titled (in French/Flemish) *Le plus ancien dictionnaire bantu/Het oudste Bantu-woordenboek*, or thus *The Oldest Bantu Dictionary* — remains the more accessible of the two versions, however, so it is important to submit it to an analysis, in order to judge its scientific value.

4.1 The modern Kikongo orthography: base letters

Over and above the changes to the direction and languages involved, an even more obtrusive intervention concerns the adjustment of the Kikongo words to the 'modern' Kikongo orthography. In doing so, several phonemes of the original were obscured and merged in the modern variants. For instance, the grapheme <v> in Van Wing and Penders might refer to <bh> or <u> in the original.

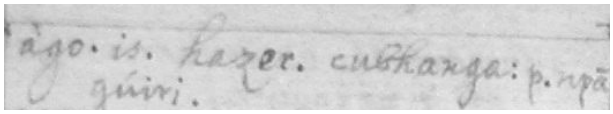
It is extremely doubtful that these two graphemes represented the same phonemes, let alone the same sounds.

In (1) we list the principal changes of Van Wing and Penders (1928: xxxiii-xxxiv) with regard to the orthography, and we discuss some of the problems that result from these changes.

(1)
- a = a
- aa = â, or aCa [with C a consonant]
- b and bh = b or v, according to the modern orthography
- c = k
- cu = kw, in front of a vowel
- ç = s or z, according to the modern orthography
- e = e
- ee = ê, or eCe [with C a consonant]
- gu = g, in front of i or e
- gu = gw, in front of a
- h = i
- i, cf. y
- m' = mu
- mu, followed by a vowel = mw
- nb = mb
- nf = mf
- np = mp
- oe = we
- qâ = kia
- qu = ku
- ss = s
- u, ü = v or w, according to the modern orthography
- y = y or i, according to the modern orthography
- z = z or s, according to the modern orthography

Some of the changes might be considered useful as they clarify the original orthography which was influenced by Portuguese or Spanish and approximate the IPA conventions. The change from <cu> to <kw> in front of vowels should not be considered harmful, nor should the change from <c> to <k>, since <c> always seems to represent the voiceless velar plosive /k/. In modern-day Spanish, the grapheme <c> might refer to the voiceless dental fricative /θ/, when followed by <e> or <i>. The manuscript, however, seems to use the grapheme <z> to represent this voiceless dental fricative, as seen in the Spanish *hazer* 'do, act' in (2).[9]

(2) ago. is. hazer. cubhanga: p. npā gúiri. (*ago* 'to do, to act')

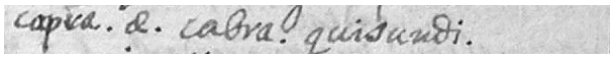


The changes from <gu> to <g> before <i> or <e> and to <gw> before <a> do not imply phonological changes and merely clarify the Portuguese or Spanish orthography. When reading the manuscript, one must thus be conscious of the fact that <gu> before <i> or <e> represents the voiced velar plosive /g/, while <gu> before <a> (or <o>) represents this voiced velar plosive /g/ followed by the voiced labialized velar approximant /w/.[10]

The change from <qu> to <ku> is problematic, since <qu> only represents

/k/ when followed by <i> or <e>. When followed by <a>, <o> or <u>, it represents /kw/, that is the voiceless velar plosive followed by the voiced labialized velar approximant. However, in practice Van Wing and Penders seem to have executed this change correctly regarding the phonetics of <qu>, as shown in (3): <qui> (from the manuscript, 3a) is replaced by <ki> (in Van Wing and Penders, 3b), and not by <kui> or <kwi>.

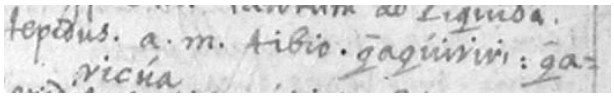
(3a) capra. a. cabra. quisundi. (*capra* 'goat')



(3b) **Ki-sundi, (i),** chèvre, chevreau ; *geit, bokje*.
Ka—, (ia), de chèvre, etc. ; *van een geit, enz*.
— kia mbakala, (i), bouc ; *bok*.

Compare this to the example shown in (4), where <cú> is first replaced with <kú>, and given it is followed by <a>, becomes <kw>. In other words, *qâ ricúa* in the manuscript, becomes *kia arikwa* in Van Wing and Penders.[11]

(4a) tepidus. a. m. tibio. qâqúiriri: qâ= ricúa (*tepidus* 'tepid, lukewarm')



(4b) **Arikwa,** être tiède ; *lauw zijn*.
Kia —, (ia), tiède ; *lauw*.

Other orthographical changes do have an impact on phonetic and/or phonological distinctions. Such is the case with <b> and <bh> becoming <b> or <v>. In most cases <b> remains <b> and <bh> is replaced by <v>, but unfortunately in some cases <bh> is also replaced by <b>. See (5).

(5)

PB reconstruction	Reflex in original	Reflex in VW&P	Translation
*-pátà	ebhata	e-vata	'village'
*-pí	cubhia	-via	'to burn'
*-pʾɪkà	mubhika	mu-bika	'slave, servant'

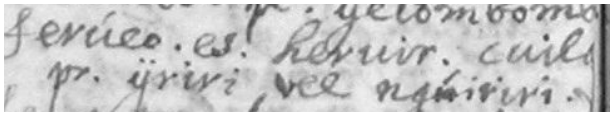
Moreover, the grapheme <v> in Van Wing and Penders may also refer to <u> in the original. This conveys the impression that both graphemes reflect the voiced labiodental fricative /v/. However, when comparing to the Proto-Bantu reconstructions (PB, cf. BLR 3), it becomes clear that <bh> is the unconditioned reflex of *p (everywhere except in front of PB close vowels and behind nasals), while <u> is the conditioned reflex of *b in front of PB close vowels (*i/*u). *b

has both Ø and b as unconditioned reflex (everywhere except in front of PB close vowels, while in postnasal position only b occurs). See the examples in (6).

	PB reconstruction	Reflex in original	Reflex in VW&P	Translation
(6a)	*-páan-	cubhana	-vana	'to give'
	*-pía	ebhia	e-via	'field'
	*-pátà	ebhata	e-vata	'village'
(6b)	*-bìmbà	euimbu	e-vimbu	'corpse'
	*-jíbì	múiúi	mu-ivi	'thief'
	*-bùá	eúúa	e-vwa	'nine'
	*-bú	múuú	mu-vu	'year'
(6c)	*-b`ɪd-	cuila	-ila	'to boil'
	*-bíad-	cúiala	-yala	'to reign'
	*-táb`ɪ	lutai	lu-tai	'branch'

It seems unlikely that both <bh> and <u> in the original represent the voiced labiodental fricative /v/. <bh> never existed as a grapheme in Portuguese or Spanish and its phonetic value cannot be pinpointed with certainty. It is possible that the indication of an aspiration of /b/ was intended, but in the Bantu languages, it is voiceless rather than voiced plosives that are normally aspirated.[12] It is more likely that it represents the voiced bilabial fricative /β/, as is also suggested by Thornton (2011), who mentions the existence of the bilabial fricative in some dialects. It is also attested in Kizombo as a reflex of *p, after a nasal prefix of class 1, for instance in /ŋβaŋgi/ 'creator' (Fernando 2008: 32). However, *p is reflected as /v/ in an intervocalic position, for instance in *-vanga* 'do, make'. Possibly, the dialect in the dictionary did not yet make a distinction between these two sound changes and *p was always reflected as /β/ before a non-close vowel. It seems, nonetheless, problematic to regard <bh> as /β/ with respect to some Spanish words included in the dictionary, in which the <u> grapheme represents the bilabial fricative /β/, as in example (7), *heruir* [erβiɾ].

(7) ferúeo. es. heruir. cuila (*ferúeo* 'to boil')
pr. ijriri uee ngúiriri.

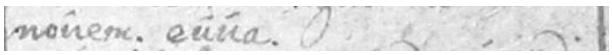


As such, two graphemes (<bh> and <u>) would be used to represent the same sound /β/. This can be explained if we assume that the Spanish words were merely copied from the Latin–Spanish dictionary, and that the Kikongo words were added in with a slightly different orthography, namely the already established Kikongo orthography of the time, which must rather have been based on

Portuguese. Thus, <u> might represent /β/ in Spanish, while <bh> might represent /β/ in Kikongo.

The <u> grapheme, on the other hand, seems to represent several phonetic values. It might represent the voiced labial velar approximant /w/, as it merges with the /w/ sound of several prefixes. It is, thus, used as semivowel. But from a diachronic perspective, the evolution from /w/ in the 17th century to /v/ in the beginning of the 20th century (i.e. the sound reflected in the Kikongo variant to which Van Wing and Penders have adjusted their orthography) seems unlikely. Since /w/ is a 'weaker' sound than /v/, it would be more logical the other way around, a phenomenon called 'lenition' (Crowley and Bowern 2010: 39). It is, therefore, likely that <u> in the manuscript represents both /w/ and /v/. This is corroborated by the fact that no <v> graphemes can be found in the dictionary, which are all included under <u>. Example (8) illustrates different uses of the <u> grapheme in the Kikongo word eúúa, in which the first <ú> might refer to the labiodental fricative /v/ (or perhaps the bilabial fricative /β/, or even something in-between), while the second <ú> probably refers to the semivowel /w/.[13]

(8) noúem. eúúa (*noúem* 'nine')



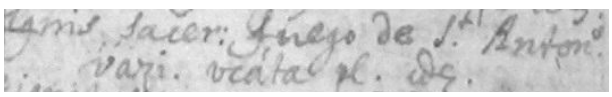
The change of <ç> to <s> or <z>, and of <z> to <z> or <s>, is also likely to cause phonetic changes, but this needs to be studied in further detail.

To summarize this section one can thus say that the orthographic changes executed by Van Wing and Penders, on the level of the basic letters, include changes that clarify, but unfortunately also changes that obscure the phonetic and/or phonological values of the graphemes used.

4.2 The modern Kikongo orthography: diacritic marks

Another remarkable orthography change executed by Van Wing and Penders is their omission of diacritic marks. The precise function of the diacritics in the original is difficult to retrace. One would expect them to represent tone, but this is unlikely for two reasons. First, acute accents, currently associated with high tone in Bantu linguistics, also occur on the Latin and Spanish words, which are definitely not tonal. See (7) and (8) above for Latin examples, and (9) for a Spanish example.

(9) ignis sacer: fúego de S^{t}. Anton°. (*Sp. fúego* 'fire')
uazi . ucáta pl. id[?].

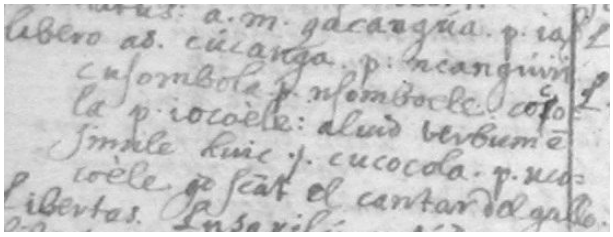


Second, both acute and grave accents occur on the Kikongo data, which would imply a three-tone system, since an unmarked syllable would then be interpreted as mid-tone.[14] This is not found in the contemporary Kikongo varieties or in other Bantu languages (Lumwamu 1973: 25). See (10) for some examples of both acute and grave accents on Kikongo words.[15]

(10) libero as. cúcanga. p. ncangúiri . cusombola. p. nsomboele . cocola p. iocoèle : aluid verbum e simule huic ./. cucocola. p. ncocoèle [?] scāt el cantar del gallo.

(*Kik. ncangúiri* 'I have liberated')

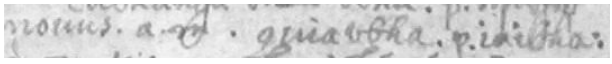
(*Kik. nco-coèle* 'I have crowed')



An analysis of the diacritics on the Latin and Spanish words in the manuscript does not reveal much either. What is significant is that the diacritics in these two European languages only occur on <u>, and exclusively as acute accents, while they occur on more vowels in Kikongo, and also include grave accents and other diacritics. Neither in Latin nor in Spanish do they seem to indicate stress, as they occur on vocalic, consonantal and semi-vocalic uses of <u>. Moreover, this is not consistently done. In (11), for example, *nouus* 'new' is written without any accents, while *noúitas* 'novelty' is written with an acute accent.

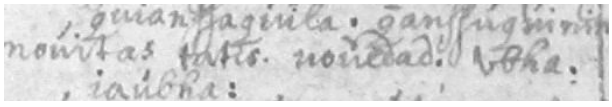
(11a) nouus. a.m. quiaubha. p. iaibha:

(*nouus* 'new')



(11b) noúitas tatis. noùedad. ubha . iaúbha :

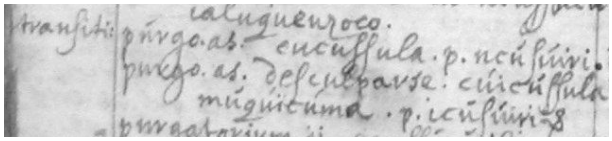
(*noúitas* 'novelty')



Also, both *púrgo* and *purgo* occur, as seen in (12), in which the form with the acute accent represents the transitive form of the verb, 'to purify', as indicated in the margin of the manuscript, while the unmarked form represents the reflexive form, 'to apologize'. Unfortunately, no other instances of such a dif-

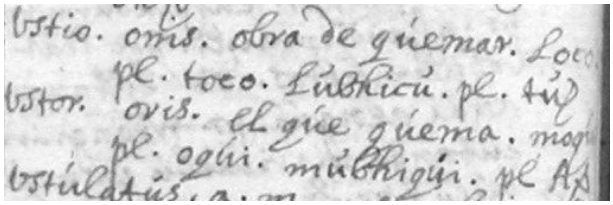
ferentiating function have so far been found.

(12a) púrgo. as. eucussula. p. ncúsúiri. (*púrgo* 'to purify')
(12b) purgo. as. desculparse. cúicússula (*purgo* 'to apologize')
múqúicúma. p. icúsúiri &



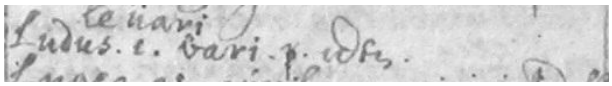
Remarkably, even an unpronounced <u> is occasionally given an acute accent, both in Spanish and in Kikongo, as is illustrated in example (13). The <qú> grapheme in the Spanish *qúemar*, *qúe* and *qúema* is pronounced as the voiceless velar plosive /k/, as it is in the Kikongo *múbhiqúi*. This conveys the impression that the diacritics have not been used in a systematic way.

(13) ustio. onis. obra de qúemar. loco[?] (*ustio* 'act of burning')
pl. toco. lúbhicú. pl. tú&
ustor. oris. el qúe qúema. moqú[?] (*ustor* 'the one that burns')
pl. oqúi. múbhiqúi. pl. a&

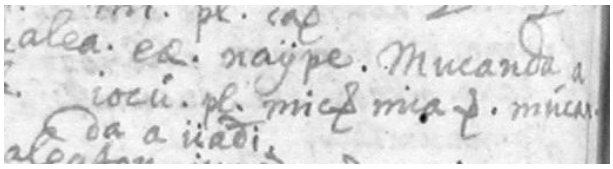


Examples (14) and (15) show that even other diacritics, distinct from the acute and grave accents, are used on the Kikongo words; their meaning is unknown.

(14) ludus. i. ûari. p. id[?] (*Kik. ûari* 'game')



(15) alea. ea. naijpe. Mucanda a (*Kik. üadi* 'game')
iocú. pl. mic& mia& . múcan-
da a üadi.

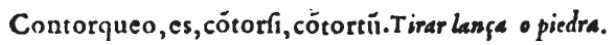


With regard to the diacritic marks on the Latin and Spanish words, one could have hoped that they can be traced back to (one of the editions of) De Nebrija's Latin–Spanish dictionary, but that does not seem to be the case. Example (16), for instance, reproduces the entry for *ferueo* 'to boil' in De Nebrija (1570 [1492]).

(16) Ferueo, es. Feruesco, is, ferbui. *Heruer*. absolutum. (*ferueo* 'to boil')

The diacritic which was seen in Van Gheel's manuscript, see (7) above, is absent from De Nebrija's dictionary. Conversely, De Nebrija's dictionary contains extra diacritic marks not found in the manuscript, such as the macrons on <o> and <u> as seen in (17).[16]

(17) Contorqueo, es, cōtorsi, cōtortū. *Tirar lança o piedra*. (*contorqueo* 'to throw')



To summarize this section one can thus say that the functions of the diacritic marks in Van Gheel's (1652) manuscript, omitted by Van Wing and Penders (1928), are extremely hard to retrace. At this stage we have to conclude that no apparent system was used for the placement of accents and other marks, but further research may, hopefully, invite us to revise this view. The option of vowel length could also be studied further in this regard. It might also be the case that several diacritic systems are intermingled, one belonging to an as-yet undiscovered original, and others belonging to the copies such as the one made by Van Gheel.

5. Translation equivalence in Van Gheel (1652) and Van Wing and Penders (1928)

The difficulties of translating an existing dictionary into another language are well known, especially when having to bridge languages with very different grammatical structures. Several issues are dealt with by Clayton (2003: 101-108), when she discusses the addition of Nahuatl to a 16th century Spanish/Latin template. Earlier, Doke (1935: 87), referring to the Bantu languages in the age of Brusciotto, spoke of "the Latin approach to a treatment of Bantu when grammatical elements are dealt with". As any bilingual (or trilingual, quadrilingual, etc.) lexicographer will be able to confirm wholeheartedly, perfect interlingual correspondence is a chimera. With reference to Zulu, De Schryver and Wilkes (2008) coined the term 'complexicography', and offered some modern (corpus-driven) solutions. Summarising the state of the art, Adamska-Sałaciak recently recognized three potentially interconnected reasons underlying the complexity of interlingual lexicography:

> The lexicons of natural languages are not isomorphic. Reasons for the anisomorphism can be sought on three interrelated planes: language structure, extralinguistic reality, and conceptualisation. Simply put, the relevant differences may

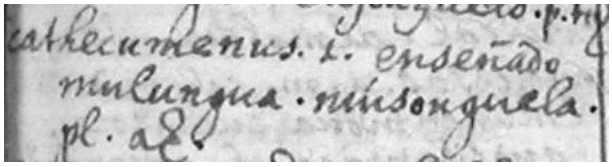
> reside in the language, the world, the mind, or any combination of these. (Adamska-Sałaciak 2011: 1)

No doubt, our Capuchins were faced with exactly these problems when adding in Kikongo to their Latin/Spanish template. It is, therefore, instructive to look at some of the solutions found to combat anisomorphism in the *Vocabularium Latinum, Hispanicum, e Congense* of 1652, and to look at the technique that was used when taking out the Kikongo in compiling *The Oldest Bantu Dictionary* of 1928.

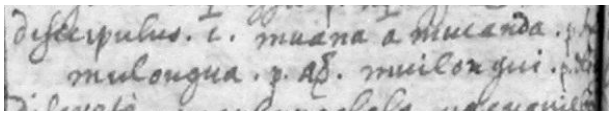
5.1 Meaning extensions

A neat solution for imported (here European) concepts is to resort to extending existing meanings, in combination with the general morphological rules for word formation in a language (here Kikongo), as seen in (18) and (19).

(18) cathecumenus. i. enseñado. (*cathecumenus* 'catechumen')
mulungua . músonguela.
pl. a&.



(19) discipulus. i. muana a mucanda. p[?] (*discipulus* 'disciple')
mulongua. p. a& . muilongui. p[?]



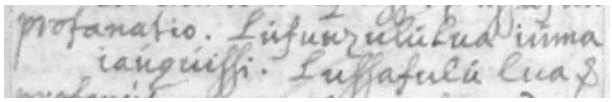
In (18) *mulungua* (sic, rather *mulongua*) and *músonguela* are offered as translation equivalents for *cathecumenus* 'catechumen', both nouns having been put in class 1 (*mu-*), and derived from the verb roots *-longua* 'to learn, to be taught' and *-songuela* 'to advise' respectively. According to the OED a catechumen is "[a] new convert under instruction before baptism", and as in the original Greek (i.e. *κατηχούμενος* 'one being instructed (in the rudiments of religion)'), the Capuchins derived the two Kikongo versions from verb roots equivalent in meaning to the Greek ones. In (19) the second and third translation equivalents for *discipulus* 'disciple' are derived from the same two verb roots as in (18), while the first option *muana a mucanda* literally means 'child of the book', or thus 'student', and by extension 'disciple'. Lexicologically the Capuchins clearly did a rather good job, terminologically they unfortunately introduced an ambiguous term (with *mulongua* being both a 'catechumen' and a 'disciple'),

and lexicographically they have been sloppy: the Spanish equivalent is present in (18) but missing in (19), the plural of the first Kikongo equivalent in (18) is missing but present elsewhere in (18) and (19), and the structural marker preceding plurals is "pl." in (18) but "p." in (19).[17]

5.2 Paraphrases

When the Capuchins did not manage to create a single-word term for a novel concept, they simply combined words paraphrasing the concept, creating a multi-word term, as in (20), where two connectives (*lua* and *ia*) are used.

(20) profanatio. lúfunzulú lua iúma ia úqúissi . lussafulú lua & (*profanatio* 'profanation')

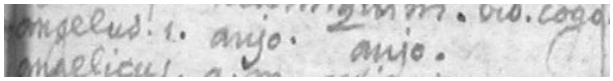


Literally, *lúfunzulú lua iúma ia úqúissi* means 'tarnishing of the thing of sacredness'.

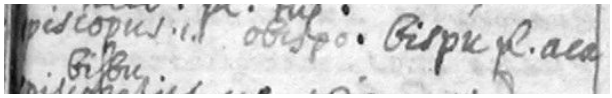
5.3 Loanwords

Unsurprisingly, there are also cases where the Capuchins simply took both the foreign concept and the word itself, with or without phonological adaptation. In (21) and (22) the loanword was taken from Portuguese, while in (23) it was taken from Latin.[18]

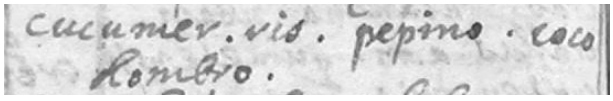
(21) angelus. i. anjo. anjo. (*angelus* 'angel')



(22) episcopus. i. obispo. bispú pl. aca bisbu (*episcopus* 'bishop')



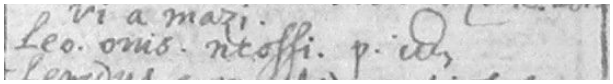
(23) cucumer. ris. pepino. coco hombro. (*cucumer* 'cucumber')



5.4 Misnamings

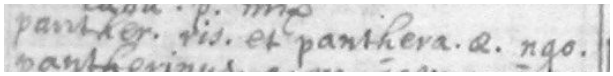
Not only did the Europeans bring elements of their culture to the Kongo area, it is clear that the Kongolese culture consisted of elements unfamiliar to the Europeans as well. This bias might be less visible to the European scholar, as indigenous terms are used to denote foreign concepts. Their meaning is not just extended; their original meaning (at least in Van Gheel's manuscript, as well as in Van Wing and Penders' reversing out) is denied and abandoned for the foreign concept. This becomes especially clear when comparing these terms to other Bantu languages or to the Proto-Bantu reconstructions. For instance, while the Capuchins were familiar with wild animals such as lions, leopards and elephants, they were apparently not familiar with hyenas and jackals. Examples (24) through (26) show that the translations of lion, leopard and elephant correspond to the respective Proto-Bantu reconstructions, while examples (27) and (28) show that there is a mismatch for hyenas and jackals, as these are offered as equivalents for wolves and foxes respectively.

(24) leo. onis. ncossi. p. id. (*leo* 'lion')



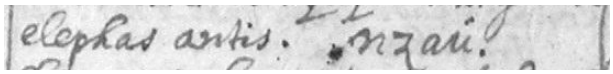
PB reconstruction: *kócì 'lion' or *kòpí 'feline: leopard, lion'

(25) panther. ris. et panthera. a. ngo. (*panther* 'panther, leopard')



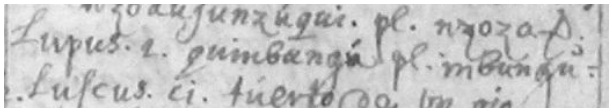
PB reconstruction: *gòì 'leopard'

(26) elephas antis. nzaú (*elephas* 'elephant')



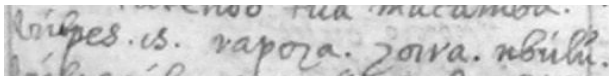
PB reconstruction: *jàjù 'elephant: *Loxodonta africana*'

(27) lupus. i. quimbungú. pl. mbungú. (*lupus* 'wolf')



PB reconstruction: *bùngú 'hyena'

(28) vúlpes. is. rapoza. zorra. nbúlú. (*vúlpes* 'fox')



PB reconstruction: *búdú 'jackal'

Here we have reached a crucial point, and are entering the domain of forensic dictionary analysis (cf. Coleman and Ogilvie 2009). That existing terms may be (re)used to name similar animal species across continents is well known. For instance, the Dutch who settled in the Cape named a certain species of fish they found in the sea *snoek*, drawing an analogy with the fresh water *snoek* they knew from home. The two are however different species, prompting the latest Afrikaans–Dutch dictionary to point out: *"In Afr. verwys 'snoek' na 'n bepaalde soort seevis, nie 'n varswaterroofvis soos in Ned. nie"* (ANNA).[19] In the case of *snoek*, it was one people who used (initially) one language (Old Dutch), to name a new species. Not having a name for the new species, they used a term they already had for a similar fish. This is different from our interlingual Kikongo dictionary. The European-born Capuchins surely had had first-hand experience with wolves and foxes in Europe, and so must have realized that the hyenas and jackals in Africa were different species. Could they then, as suggested at the start of this section, really have taken Kikongo terms in use for other species, to now name animals from Europe? This sounds improbable. More plausible is the situation whereby a native of the Kongo is presented with a description of wolves and foxes, which are unknown to him, to then, based on that evidence, offer terms from his native Kikongo as translation equivalents. If anything, then, the errors noted in (27) and (28) are pointing in the direction of a dictionary compiler whose native language and view of the world are African. In other words, the case in favour of Roboredo as the main compiler of the first Capuchin manuscripts is getting stronger.[20]

A second crucial point concerns the words-and-things method. This method is founded on the basic idea that a community's culture is reflected in its language. It is therefore used to reconstruct the history of a particular region on the basis of vocabulary reconstructed from the languages spoken there (Bostoen 2007: 175). Looking back at examples such as (27) and (28), it should thus be clear that extreme caution must be exhibited in blindly citing 'evidence' from it. Bontinck (1976, 1980), too, pointed this out, and criticized Vansina (1974) for using Van Wing and Penders (1928) very loosely, for instance with respect to his deductions on the presence of certain craftsmen in the Kongolese society, such as "slave traders, wine merchants, butchers, fishmongers, booksellers, shopkeepers, grocers for spices, clothes sellers, perfume dealers, and pharmacists" (Bontinck 1976: 155, Vansina 1974: 149, Van Wing and Penders 1928: 85). Clearly, the same holds for conclusions regarding the Kongolese wildlife, as illustrated above. One cannot conclude that the Kongolese wildlife included wolves and foxes (cf. Kingdon 1997).[21]

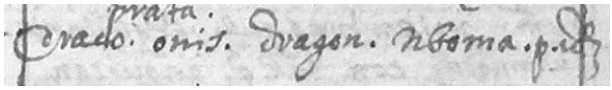
5.5 Retranslations

On top of the anisomorphisms already discussed, Van Wing and Penders added yet another layer of translation inequivalence. In their own words (quoting the French version as it conveys it better than the Flemish):[22]

> En faisant la traduction française et flamande des mots congolais nous avions à tenir compte du sens du mot congolais, tel qu'il nous est connu en congolais moderne et en même temps du sens des mots correspondants en latin et en espagnol donnés par notre auteur. Il arrive parfois que l'auteur rend inexactement en congolais certains mots latins. De la sorte il sera arrivé quelquefois, que nous avons donné une traduction française et flamande qui ne rend pas exactement le sens du mot congolais. (Van Wing and Penders 1928: xvi)

In other words, on top of reversing out the entire dictionary of Van Gheel, Van Wing and Penders also insisted on adding the modern (i.e. end 19th–beginning 20th century) Kikongo meanings, *and* being unhappy with some of the Latin to Kikongo translations, they sometimes additionally translated directly from Latin into French/Flemish, regardless of the Kikongo! At all times, and despite the fact that there is no fixed slot for Latin in their dictionary, one thus actually has to 'imagine' there is an underlying layer of Latin 'driving' the entire enterprise. Van Wing and Penders do not give examples of their claim, but a candidate imbedding several levels is shown in (29).

(29a) draco. onis. dragon. nboma. p. id. (*draco* 'snake; dragon')



PB reconstruction: *bòmà 'snake, python'

(29b) **Mboma,** (id.), espèce de serpent, python, dragon ; *soort slang, reuzenslang, draak.*

With regard to the reversal proper, the Latin lemma sign *draco* and the Kikongo translation equivalent *nboma* traded places, becoming the Kikongo lemma sign *Mboma* and the French/Flemish translation equivalents *dragon/draak*. A better (zoological) knowledge of Kikongo resulted in the fronting of *python/reuzenslang* as translation equivalent; while a retranslation from the Latin (with *dracō* 'snake; dragon') further added *espèce de serpent/soort slang* 'type of snake'. Important here, is that there is no entry for 'python' in the manuscript, nor, of course, for 'type of snake', so Van Wing and Penders' two additional translation equivalents are not the result of reversing out Van Gheel's manuscript.

6. Lexicographical structure in Van Gheel (1652) and Van Wing and Penders (1928)

Van Gheel's dictionary being a manuscript, no typographical variation is pre-

sent. All the information is written in a single file with, for dictionary articles longer than one line, some slight indentation, as seen in the images from the dictionary reproduced above. The only non-typographical structural marker used is the full stop, which delimits both the languages (Latin vs. Spanish vs. Kikongo, whence the full stop is typically attached to the last word of the respective language), and separates synonyms (in Spanish and Kikongo, whence the full stop is typically surrounded by white space). Full stops are also used with abbreviations, and end dictionary articles (though this overlaps with the end of the Kikongo slot). In contrast, and for all its faults, Van Wing and Penders' published dictionary is a rather advanced product for early 20th century Bantu lexicography. Theirs uses typography (bold vs. Roman vs. italics) to separate the three languages, and also uses many more non-typographical structural markers (commas, semi-colons, colons, full stops, long dashes (in lieu of the more usual tildes), ellipsis, the symbol "./.", as well as round and square brackets). Recurrent orthographic markers that structure the text include "N. B.", "*v. g.*", and "dans : *in :*". The latter is especially interesting, as it signals lemma signs which only take on a translatable meaning when combined with other words. Examples are shown in (30) and (31).

(30) **Ifiku,** dans: *in:*
Fwanana mu —, ikala nsinza mu —, être d'égale valeur, être équivalent ; *evenveel waard zijn, evenaren*. ('be worth as much, be equivalent')
Lufwananu lu —, nsinza mu —, équivalence; *gelijkwaardigheid*. ('equivalence')

(31) **Munzonzela,** dans : *in* :
Ka—, (ia), qui coule, coulant ; *vloeiend, loopend*. ('which flows, runny')

This beautifully solves a lexicographic problem in a user-friendly way, by sidestepping the question of lemma-sign status. A variation is shown in (32), where the lemma sign is either *Safiru* 'sapphire' or *etari ria Safiru* 'stone of sapphire'.

(32) **Safiru, etari ria —,** saphir; *saffier*. ('sapphire')

"N. B." stands for *nota bene* 'note well' and "*v. g.*" for *verbi gratia* 'for example', both Latin, and are typically used as shown in (33).

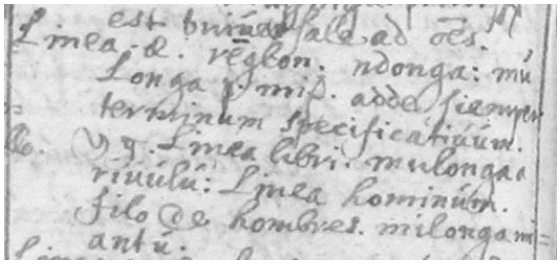
(33) **Andula,** *pr :* **yandwiri**, chauffer, liquéfier, fondre, raréfier; *verwarmen, doen smelten, verdunnen*. ('heat, melt, rarefy')
— riaka, andulula, *pr :* **yandulwiri,** réchauffer ; *opnieuw verwarmen, opwarmen.*
N. B. Pour marquer la répétition on fait suivre le verbe simple du mot : **riaka** ou bien on remplace la lettre **a** finale du radical du verbe par le suffixe **ulula ;** *de herhaling in de werkwoorden wordt aangeduid met* **riaka** *op 't einde van het werkwoord bij te voegen ofwel met de eindletter* **a** *van den* (explaining Kikongo grammar)

stam van het werkwoord te vervangen door het achtervoegsel:
ulula.
v. g. **Vanga,** faire ; *doen*. (exemplification of
Vanga riaka, vangulula, refaire ; *herdoen*. the grammar)

In (33) a general grammatical rule is explained: "To indicate repetition, add the word *riaka*, or change the final vowel *a* to *ulula*".[23] The very same grammatical point and exemplification could of course have been added throughout the dictionary, at many a verb with the potential for repetition (in English *re*-…). It is not clear why Van Wing and Penders decided to include it with this verb only; apart, perhaps, from the fact that it may be the first verb with this feature in the alphabetically-ordered lemma list — but then, no one *reads* a dictionary from A to Z. It also does not seem to be copied from Van Gheel's manuscript, as no such note can be found under *calefacio* 'to heat' or *liquefacio* 'to melt', neither can it be found under *ago* 'to do, to act' (see (2) above) or *facio* 'to do, to make' (the equivalents of Van Wing and Penders' Kikongo *vanga/bhanga*).

Other grammatical points in Van Wing and Penders do find their origin in Van Gheel's manuscript, as may be seen from a comparison of (34a) with (34b) in terms of the clarification "always requires to be specified further".

(34a) linea. a. rēglon. ndonga: mú (*linea* 'line')
longa p. mi&. adde siemper
terminum specificatiúúm.
v. g. linea libri. mulonga a
riúúlú: linea hominúm.
filo de hombres. milonga mi=
antú.



(34b) **Mu-longa, (mi),** (auquel on ajoute toujours le terme spécifique) (line, row, series)
toute espèce de ligne, rangée, rang ; (*vergt een verdere bepaling*),
lijn, rij, reeks.
v. g. **— a rivulu,** ligne d'un livre ; *regel van een boek*.
Milonga mi antu, rangées d'hommes ; *rijen menschen*.

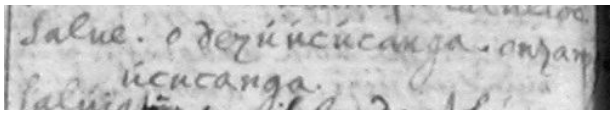
(34c) **Ndonga,** (*id.*), ligne, règle ; *lijn, regel*. (line, ruler)

Example (34) is an excellent illustration of how an entire article from Van Gheel's manuscript was reversed out by Van Wing and Penders. All the information seen in (34b) is derived from (34a), but the reverse is not fully true, as one also needs (34c) in addition to (34b) to complete the information that came from (34a). What is also missing from Van Wing and Penders throughout is the part of speech of each lemma; although one could argue that this is implicit in their treatment (nouns being followed by an indication of how to form their plural, verbs by their first person praeteritum, etc.). Here one dictionary article in the manuscript straightforwardly gave rise to two in the dictionary; one thus deals with divergence.

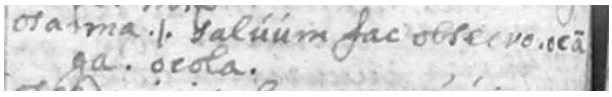
Examples of convergence also abound, whereby different slots from a series of dictionary articles from the manuscript were combined into one by Van Wing and Penders. A clear-cut case is shown in (35a), where the compilers even included a Latin slot, between square brackets, indicating where the information came from in Van Gheel's manuscript, namely *liberare* (cf. (10) above), *salve* (cf. (35b)), and *osanna* (cf. (35c)).

(35a) **Kanga,** *pr :* **nkangiri,** [liberare], délivrer, libérer, sauver ; *bevrijden, verlossen, redden*. ('liberate, save')
O dezu ukukanga, [salve], Jésus te garde ! *Jezus beware u !*
O Nzambi mpungu ukukanga, Dieu tout-puissant te garde ! *God almachtig beware u !*
O kanga, [osanna], sauvez, je vous en prie ; *red, bid ik u*.

(35b) salue. o dezú úcúcanga. onzan[mbi?] úcucanga. (*salue* 'hail')



(35c) osanna ./. salúúm fac observo. ocā ga. ocola. (*osanna* 'hosanna')



The second example in (35a), *O Nzambi mpungu ukukanga*, does not seem to come from Van Gheel and was added by Van Wing and Penders; or else it is an adaptation of the second option in (35b), *onzan[mbi?] úcucanga*.

If one looks at all the other translation equivalents in for instance (10) it should be clear that Van Wing and Penders often had to make use of both divergence and convergence simultaneously, in a first phase taking out each Kikongo word from Van Gheel's manuscript and translating that into French

and Flemish via Latin (divergence), and in a second phase collapsing the material that belongs to single dictionary articles (convergence). Over and above, they added their own material (nonvergence). The result of this approach to compiling a dictionary is that Van Wing and Penders' publication not only looks more dictionary-like but also contains more data. Indeed, what sets Van Wing and Penders most apart visually is their often long lists of combinations; a short version of which is shown in (36).

(36)	**Mu-kanda, (mi),** papier, lettre ; *papier, brief*.	(paper, letter)
	— a antwazi, diplôme ; *diploma, bekwaamheidsbewijs*.	(diploma, degree)
	— a papel, feuille de papier, page ; *blad papier, bladzijde*.	(sheet of paper, page)
	— a Papa, bulle apostolique ; *pauselijke bulle*.	(papal bull)
	— a pergamini, parchemin ; *perkament*.	(parchment)
	— a zioko, — a wadi, cartes à jouer ; *speelkaarten*.	(playing cards)

While the lemma and its translation equivalent, as well as the first and last combination, have been taken from Van Gheel's manuscript, all the combinations in-between have been added by Van Wing and Penders. No wonder Van Gheel's manuscript of 243 pages grew to 361 printed pages in Van Wing and Penders.

7. Bringing everything together: the *KongoKing Database* (2012)

Reading through Van Gheel's (1652) Latin/Spanish → Kikongo manuscript, there can be no doubt about its intended target user: It is an active, encoding dictionary meant to help the missionary produce Kikongo. The main compiler was very likely Roboredo, a Capuchin born in the Kongo. In the front matter to their Kikongo → French/Flemish dictionary, the Belgian Jesuits, Van Wing and Penders (1928: xxxii), are also clear about their goal: It is meant to be a scientific work for both Bantuists and missionaries, hence why they chose Kikongo as a source language, and French and Flemish (the two official languages of Belgium, the colonial power at the time) as target languages. About their effort, the towering Bantuist Malcolm Doke had been scathing, see Section 4 above. In the absence of any other edition of Van Gheel's manuscript, however, it has been the only entry point to it for over 80 years now, and as we saw, it has indeed been (mis)used during that period. With roots in both the 17th century and the turn-of-the-19th-20th century, it is also a valuable dictionary in its own right.

Today, in 2012, there is a renewed interest in getting easy access to this early Kikongo data as a result of the launch of the KongoKing research project. The interdisciplinary KongoKing team wishes to shed new light on the origins, rise and development of the early Kongo kingdom, by combining and coordinating pioneering archaeological fieldwork in Angola and Congo with novel historical linguistics research. To that end, a digital transcription of Van Gheel's

manuscript as well as the digitization of Van Wing and Penders' dictionary has become a necessity. Keeping the need for a long-due critical edition of Van Gheel's manuscript in mind, and the digital reality of the 21st century, we opted for using the dictionary production system TLex (aka *TshwaneLex*, cf. De Schryver 2011). With the aim to allow for cross-searches and with future multimedia extensions in mind, we also opted to work in a single database. TLex has a feature (called linked-view mode) that can automatically connect distinct dictionaries that are stored in a single database, and a common language to enable this is the ideal route. Given both Van Gheel's manuscript and Van Wing and Penders' dictionary have only Kikongo in common, and given Kikongo is the main language of interest to the project, one would be tempted to opt for it as the linking language. However, given the varying Kikongo orthographies of the two reference works, it seems better to abstract to a stable language or formalism. In this respect, we are in luck in that we actually have such a language: it is Latin. Recall that we pointed out in Section 5.5 above that also Van Wing and Penders used Latin as an underlying layer during their compilation. In practical terms, by adding a (hidden) Latin slot to the data of Van Wing and Penders, it is possible to automatically coordinate both dictionaries in an electronic environment, and to visually see the divergences, convergences and nonvergences described in Section 6 above. In metalexicographical terms this amounts to a variation of the hub-and-spoke model (Martin 2004), whereby a hub-language is used to create a series of bilingual dictionaries between it and several spoke-languages, which then allows for a combination of the spokes amongst one another, invisibly through the hub. Latin is our hub-language, but only partly hidden: hidden in Van Wing and Penders, but visible (as the source language) in Van Gheel.

The digitization of Van Wing and Penders has already been completed. Their publication was scanned and OCRed, and then parsed for importation into TLex. In one of the views (TLex allows for any number of dictionary 'views' of the database data) the printed dictionary is mimicked, typography, punctuation and all, though underlying that, extra slots have been provided for Latin (the linking language), as well as for various aspects needed in the KongoKing project such as fields for the addition of the Proto-Bantu forms, various semantic label sets, cross-references to material about corresponding archaeological finds, cross-references to corresponding academic papers, etc.

The digital transcription of Van Gheel's manuscript is ongoing. A major difficulty here is the poor readability of the original, as well as the rather haphazard use of a flat lexicographic structure. This necessitates occasional changes to the DTD (or document type definition, i.e. the dictionary grammar). The positive aspect, though, is that a rigid structure is being imposed onto the manuscript data in the process, with every part of the data ending up in its proper dictionary slot. In addition to the transcribed but now structured material, images of the original entries also accompany each dictionary article. A notes field was also added, used to point out uncertainties, errors, etc. as in a

traditional (i.e. paper) critical edition. A screenshot of the two dictionaries in linked-view mode in TLex may be seen in Addendum 4.

Having first moved the field of Bantu lexicography back with two centuries, it is now exciting to witness the recreation and digitization of the very first extant Bantu dictionary. As a work in progress, it will be made available on the KongoKing website, at which point the oldest Bantu dictionary and its 19th–20th century rework will not only be searchable in five languages, but also searchable using any combination or restriction of lexicographic metalanguage (such as word classes or semantic fields), and it will moreover function as a stepping stone towards new, multimedia data that aims to uncover the Kongo history of what came before the compilation of this first Kikongo dictionary. This fitting digital lexicographic capstone, then, is only the beginning of writing Kongo's early history.

Endnotes

1. Van Wing and Penders (1928: xxiii-xxiv) list a series of villages and rivers visited or encountered by Van Gheel, which are situated in the Matari district. These include Mbata and Ngongo/Ngungu, and also the Inkisi river, regarded as the eastern frontier of the Matari district. These villages and river are all situated in the southern part of the present-day Lower Congo Province of the DRC.
2. Thornton (2011) states the opposite, when he claims that Hildebrand (1940) fully accepts Joris van Gheel to be the real author of the dictionary, following Van Wing and Penders (1928). This does not seem to be justified to us, as Hildebrand (1940) is very clear on this point.
3. Translation: *"Roboredo deserves most credit for the compilation, in a certain way, the dictionary is his work. The Fathers requested the compilation; they can claim part of the credit for the beautiful enterprise. The vocabulary seems the collective work of the new missionaries, especially of Antonio de Teruel and José de Pernambuco, under the direction of Roboredo … This is how the remarkable Latin–Spanish–Kikongo vocabulary came into being, which we know through the copy of F. Joris [van Gheel]."* (Hildebrand 1940: 264, underlining ours)
4. With respect to these 'remarkable' linguistic capacities of Joris van Gheel, Nsondé (1995: 60) cites a letter of Jean François de Rome (Jadin 1975: 1519), praising the thorough knowledge of the language his colleague Van Gheel possesses.
5. In his bibliography, Nsondé (1995: 232) also includes a Catalan–Castillan–Latin adaptation, published in 1587 in Barcelona, of De Nebrija's (1492) *Dictionarium*. It sounds implausible that this work formed the basis of the *Vocabularium Latinum, Hispanicum, e Congense*.
6. Translation: *"The language recorded by him [Joris van Gheel] is the one of the area in which he was active; however, the dialect of Soyo, likely used more often than any other by his predecessors, is dominating. This is the language that also F. de Teruel had to learn in Mbata, Nkusu and Mpemba."* (Van Wing and Penders 1928: xxx-xxxi)
7. But note another contradictory aspect in the material just quoted (and translated in Endnote 6). Both Mbata and Mpemba are situated in the sphere of influence of Mbanza Kongo, rather than Mbanza Soyo, pointing to the Kisikongo/Kisansala variety rather than to Kisolongo.

8. In an e-mail sent to the first author, John Thornton suggested: "*I personally don't think it would be the Soyo dialect, the missionaries moved directly from the coast inland, only a few stayed in Soyo and all the players in the game ended up in the Sansala dialect zone. I have no doubt that Roboredo spoke that dialect also; his cloistername was Francisco de São Salvador, certainly proposing a speaker of the Sansala dialect. Today this dialect is still regarded as the court dialect and probably one that was spoken by the nobility wherever they lived.*"
9. Transcriptions of material from the 1652 manuscript are accompanied by a scanned image of the corresponding section in the manuscript.
10. Unfortunately, the manuscript uses an ambiguous orthography in these cases. The grapheme <gui> is indeed used to represent a /gi/ sound, as noted by Van Wing and Penders, but it is also used to represent /gwi/. Van Wing and Penders also adjusted this, but they did not mention it in their list of implemented orthography changes, reproduced in (1).
11. The macron on the letter <q> when followed by <a> is represented as <qâ> by Van Wing and Penders; see also the change of <qâ> to <kia> in (1).
12. For instance, a common phenomenon is the aspiration of previously nasalized voiceless plosives, such as mp > ph (cf. Kerremans 1980). Due to regressive assimilation of the voiceless plosive, the nasal becomes voiceless, which is then reanalyzed as an aspiration of the voiceless plosive (mp > mph > ph).
13. The phonetic value of <u> in 16th century Latin could also be investigated, as it might shed light on the phonetic value of the same grapheme in the Kikongo data.
14. Another possible interpretation of the combination of both acute and grave accents would be that an accent marks the tone of the accentuated and of all the following unaccentuated syllables. A further accent then reverses the tone for one or more syllables, until the next accent reverses the tone again. However, this convention does not seem to apply for the Kikongo data either, since first syllables are not always marked.
15. The Latin phrase *aluid verbum e simule huic* roughly means: 'another word [formed] by something that resembles it' (with thanks to Lieven Danckaert for the translation). In other words, the two verbs *cocoela* and *cucocola* are not semantically related, but the compiler(s) of the dictionary decided to put them together because of their morphological similarity.
16. For Kikongo, macrons are only found on <q> in the manuscript, cf. Endnote 11, and on <a>, see (2) and further down (35c). In instances such as the latter two, < ā ¶ g> should be read as <ang>.
17. Apart from the inconsistency, there is room for confusion as well given the abbreviation "p." also precedes the first person praeteritum at verbs, as in for instance (2) and (10), which is used interchangeably with "pr.", as in for instance (7).
18. There is a mismatch between the spelling of the singular vs. the spelling of the plural noun in (22).
19. Translation: *"In Afrikaans 'snoek' refers to a certain kind of saltwater fish, not to a predatory freshwater fish as in Dutch"* (ANNA).
20. Interestingly, in Van Wing and Penders' (1928) reversed-out version, the wolves and foxes are still featured, even though the earlier Bentley (1887), to which they had access, got it right talking about hyenas and jackals only.
21. Nor bears and tigers for that matter, given both *ursus* and *tigris* have also both (wrongly!) been given the translation equivalent *ngo*, as in (25). Bears and tigers also feature in Van Wing and Penders, but not in Bentley — cf. Endnote 20.

22. Translation: *"While preparing the French and Flemish translations of the words in Kikongo, we had to take the meaning of each Kikongo word into account, as it is known to us in modern Kikongo, and at the same time consider the meaning of the corresponding words in Latin and Spanish as provided by our author. It happens at times that our author does not exactly render certain Latin words into Kikongo. As a result, there are a few cases where we have provided a French and Flemish translation which does not exactly render the meaning of the Kikongo."* (Van Wing and Penders 1928: xvi)
23. The rule as described is not correct, as it is the repetitive verbal extension *-ulul-* which replaces the reversive-transitive verbal extension *-ul-*.

References

Adamska-Sałaciak, A. 2011. Between *designer drugs* and *afterburners*: A Lexicographic-Semantic Study of Equivalence. *Lexikos* 21: 1-22.

ANNA = Pharos Groot Woordeboek Afrikaans en Nederlands. Martin, W. et al. (Eds.). 2011. *Pharos Groot Woordeboek Afrikaans en Nederlands.* Cape Town: Pharos.

Benson, T.G. 1964. A Century of Bantu Lexicography. *African Language Studies* 5: 64-91.

Bentley, W.H. 1887. *Dictionary and Grammar of the Kongo Language, as spoken at San Salvador, the Ancient Capital of the Old Kongo Empire, West Africa*. London: Baptist Missionary Society and Trübner & Co.

BLR 3 = Bantu Lexical Reconstructions 3. 2005. Available online at: http://www.africamuseum.be/collections/browsecollections/humansciences/blr/. (Last accessed: 5 March 2012)

Bontinck, F. 1976. Remarques marginales à Vansina, 'The Dictionary and the Historian'. *History in Africa* 3: 155-156.

Bontinck, F. 1980. Le *Vocabularium Latinum, Hispanicum et Congense*: nouvelles notes marginales. *Annales Aequatoria* 1: 529-535.

Bonvini, E. 1996. Repères pour une histoire des connaissances linguistiques des langues africaines. *Histoire Épistémologie Langage* 18(2): 127-148.

Bostoen, K. 2007. Pots, words and the Bantu problem: On lexical reconstruction and early African history. *Journal of African History* 48(2): 173-199.

Brown, J. [of Taung]. 1875. *Lokwalo loa mahuku a secwana le seeñeles*. London: London Missionary Society.

Brusciotto, G. [= *It.* Giacinto Brusciotto; *Prt.* Jacinto Brusciato de Vetralha; *Lat.* Hyacinthus Brusciottus a Vetralla]. 1659. *Regulae quaedam pro difficillimi Congensium idiomatis faciliori captu ad grammaticae normam redactae. A F. Hyacintho Brusciotto a Vetralla Concionatore Capuccino Regni Congi Apostolicae Missionis Praefecto.* Rome: Sac. Congreg. de Propaganda Fide.

Cardoso, M. 1624. *Dovtrina Christaã* [sic]. *Composta pelo P. Marcos Iorge da Companhia de IESV Doutor em Theologia. Acrescent ada pelo Padre Ignacio Martinz da mesma Companhia Doutor Theologo. De nouo traduzida na lingoa do Reyno de Congo, por ordem do P. Mattheus Cardoso Theologo, da Campanhia de IESV*. Lisbon: Geraldo da Vinha.

Clayton, M.L. 2003. Evidence for a Native-Speaking Nahuatl Author in the Ayer *Vocabulario Trilingüe*. *International Journal of Lexicography* 16(2): 99-119.

Coleman, J. and S. Ogilvie. 2009. Forensic Dictionary Analysis: Principles and Practice. *International Journal of Lexicography* 22(1): 1-22.

Crowley, T. and C. Bowern. 2010[4] [1987]. *An Introduction to Historical Linguistics*. New York: Oxford University Press.

D'Alençon, É. [*religious name of* François Lecorney; *Fr.* Édouard d'Alençon]. 1914. Essai de bibliographie capucino-congolaise. *Neerlandia Franciscana* 1: 33-42.

De Cadornega, A. de O. 1680. *Hisória geral das guerras angolanas* (anot. e corrigido por José Matias Delgado, 1940, 3 vols.). Lisbon: Agência-Geral do Ultramar.

De Kind, J. 2012. *Taalvariatie in het Kikongo van Angola vanuit historisch perspectief: fonologie en morfologie*. Unpublished M.A. Dissertation. Ghent: Ghent University.

De Nebrija, A. [*Lat.* Aelii Antonii Nebrissensis]. 1492. *Dictionarium latino-hispanicum*. Salamanca.

De Nebrija, A. 1516. *Vocabulario de romance en latin: hecho por el doctissimo maestro Antonio de Nebrissa nueuamente corregido y augmentado: mas de diez mill vocablos de los que antes solia tener*. Sevilla: Joannes Varela.

De Nebrija, A. 1570 [1492]. *Dictionarium Latinohispanicum, et vice versa Hispanicolatinum, Aelio Antonio Ne brissensi interprete, nunc denuo ingenti vocum accessione locupletatum, pristinoque nitori sublata mendarum colluuie restitutum. Ad haec. Dictionarium propriorum nominum, ex probatissimis Graecae et Latinae lingua autoribus, addita ad calcem neoterica locorum appellatione concinnatum.* Antwerp: Ioannis Steelsij.

De Nebrija, A. 1581 [1492]. *Dictionarium, immo quadruplex ejusdem antiqui Dictionarii supplementum, nunc de novo nepotis sui diligentia excussum atque correctum et multipli vocum additione locupletatum, primum quidem dictiones habet Latinas in sermonem Hispanum versas; secundum autem vice versa Hispanas Latinae linguae redditas amplectitur; tertium vero nomina propria indicat regionum, urbium, montiumque et aliarum multarum; quartum et ultimum contraria vice ostendit neotericas ac vulgares regionum ubicumque et aliarum rerum proprias appellationes*. Antequera.

De Nebrija, A. 1587 [1492]. *Lexicon, seu Dictionarium Aelii Antonii Nebrissensis... ex latino sermone in cathalanum et castellanum... conversum; idem etiam ex cathalana et castellana lingua in latinam versum; onomasticon praeterea nominum propriorum... Accessit ejusdem auctoris medicum dictionarium... Adjecimus in fine Valerii Probi... de litteris antiquis opusculum...* Barcelona: Antonio Oliver.

De Schryver, G.-M. 2003. Online Dictionaries on the Internet: An Overview for the African Languages. *Lexikos* 13: 1-20.

De Schryver, G.-M. 2011. Why Opting for a Dedicated, Professional, Off-the-shelf Dictionary Writing System Matters. Akasu, K. and S. Uchida (Eds.). 2011. *ASIALEX 2011 Proceedings. LEXICOGRAPHY: Theoretical and Practical Perspectives. Papers Submitted to the Seventh ASIALEX Biennial International Conference. Kyoto Terrsa, Kyoto, Japan, August 22–24, 2011*: 647-656. Kyoto: The Asian Association for Lexicography.

De Schryver, G.-M. and A. Wilkes. 2008. User-friendly Dictionaries for Zulu: An Exercise in Complexicography. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15–19 July 2008.* Sèrie Activitats 20: 827-836. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

Doke, C.M. 1935. Early Bantu Literature: The Age of Brusciotto. *Bantu Studies, A Journal devoted to the Scientific Study of Bantu, Hottentot and Bushmen* 9(2): 87-114. [A revised version was published in 1959, in *African Studies* 18(2): 49-67.]

Doke, C.M. 1945. *Bantu: Modern Grammatical, Phonetical, and Lexicographical Studies since 1860*. London: (published for the International African Institute by) Percy Lund, Humphries & Co.

Doke, C.M. and B.W. Vilakazi. 1948. *Zulu–English Dictionary*. Johannesburg: Witwatersrand University Press.

Fernando, M. 2008. *An Analysis of Verbal Affixes in Kikongo With Special Reference to Form and Function*. Unpublished M.A. Dissertation. Pretoria: University of South Africa.

Gray, R. 1998. Bonaventura de Sardegna. Anderson, G.H. (Ed.). 1998. *Biographical Dictionary of Christian Missions*. Grand Rapids, MI: W.B. Eerdmans Publishing Company. Available online at: http://www.dacb.org/stories/congo/bonaventura-de-sardegna.html. (Last accessed: 5 March 2012)

Hannan, M. 1959. *Standard Shona Dictionary.* Compiled for the Southern Rhodesia Native Affairs Department. London/New York: Macmillan & St. Martin's Press.

Hendrix, M.K. 1982. *An International Bibliography of African Lexicons*. Metuchen: The Scarecrow Press.

Hildebrand, F. [*religious name of* Jules Raes]. 1934. Nieuwe bijdrage over Joris van Geel. *Kongo-Overzee, tijdschrift voor en over Belgisch Kongo* 1: 103-118.

Hildebrand, F. 1940. *Le martyr Georges de Geel et les débuts de la mission du Congo (1645-1652)*. Antwerp: Capuchin Archives.

Jadin, L. 1975. *L'ancien Congo et l'Angola, 1639-1655, d'après les archives romaines, portugaises, néerlandaises et espagnoles*. Bibliothèque de l'Institut historique belge de Rome.

Kerremans, R. 1980. Nasale suivi de consonne sourde en proto-bantou. *Africana Linguistica* 8: 159-198.

Kingdon, J. 1997. *The Kingdon Field Guide to African Mammals*. Princeton: Princeton University Press.

KongoKing Research Project. 2012. Available online at: http://kongoking.org/. (Last accessed: 5 March 2012)

Krapf, J.L. 1882. *A Dictionary of the Suahili Language* (with an introduction containing an outline of Suahili grammar). London: Trübner & Co.

Le Veux, H. 1917. *Premier essai de vocabulaire Luganda–français d'après l'ordre etymologique*. Maison-Carrée (Algiers): Imprimerie des Missionnaires d'Afrique (White Fathers).

Lumwamu, F. 1973. *Essai de morphosyntaxe systématique des parlers kongo* (Langues et littératures de l'Afrique noire 10). Paris: Klincksieck.

Mabille, A. 1878. *Sesuto–English Vocabulary*. Morija.

Madan, A.C. 1913. *Lala–Lamba–Wisa and English, English and Lala–Lamba–Wisa Dictionary*. Oxford: Clarendon Press.

Martin, W. 2004. SIM*u*LLDA, the Hub-and-Spoke Model and Frames or How to Make the Best of Three Worlds? *International Journal of Lexicography* 17(2): 175-187.

Mavoungou, P.A. 2002. Synopsis Articles in the Planning of a Trilingual Dictionary: Yilumbu–French–English. *Lexikos* 12: 181-200.

Mavoungou, P.A. 2006. A Trilingual Dictionary Yilumbu-French-English: An Ongoing Project. *Lexikos* 16: 121-144.

Mavoungou, P.A. 2009. Place et nécessité des travaux religieux, ethnohistoriques et linguistiques dans le paysage dictionnairique gabonais. *Lexikos* 19: 144-176.

Mavoungou, P.A. 2009a. [Dictionary review of:] ILALOK. *Dictionnaire vili–français/Mpisukulu bi kum' bi tshi vili ku tshi mputu.* 2008. *Lexikos* 19: 524-527.

McLaren, J. 1936[2] [1915]. *A Concise Xhosa–English Dictionary* (2nd edition, revised, edited and set in the new orthography by W.G. Bennie). London/New York: Longmans, Green & Co.

Ndinga-Koumba-Binza, H.S. and J.C. Roux. 2009. The Representation of Vowel Duration in Civili Dictionaries. *Lexikos* 19: 197-206.

Nsondé, J. de D. 1995. *Langues, culture et histoire Koongo aux XVIIe et XVIIIe siècles, à travers les documents linguistiques*. Paris: L'Harmattan.

OED = Oxford English Dictionary. Simpson, J. et al. (Eds.). 2012. *Oxford English Dictionary.* Oxford: Oxford University Press. Available online at: http://www.oed.com/. (Last accessed: 5 March 2012)

Pacconio, F. and A. de Couto. 1643. *Gentio de Angola sufficientemente instruido nos mysterios de nossa Sancta Fé. Obra posthuma, composta pello Padre Francisco Pacconio da Companhia de Iesu. Redusida a methodo mais breve & accomodado á capacidade dos sogeitos, que se instruem pello Padre Antonio do Couto da mesma Companhia*. Lisbon: Domingos Lopes Rosa.

Rey-Debove, J. 1992. [Dictionary review of:] Ambroise Queffélec, avec la collaboration d'Augustin Niangouna: *Le français au Congo* (R.P.C.), Université de Provence, Service des publications, Aix-en-Provence, 1990, 332 p. *International Journal of Lexicography* 5(2): 160-161.

Scott, D.C.R. 1892. *A Cyclopaedic Dictionary of the Mang'anja Language Spoken in British Central Africa*. Edinburg: Foreign Mission Committee of the Church of Scotland.

Thornton, J. 2011. *Roboredo, Kikongo Sermon.* Available online at: http://www.bu.edu/afam/faculty/john-thornton/roboredo-kikongo-sermon/. (Last accessed: 5 March 2012)

Torrend, J. 1931. *An English-Vernacular Dictionary of the Bantu-Botatwe Dialects of Northern Rhodesia*. Mariannhill: Mariannhill Mission Press.

Tsakona, V. 2007. Bilingualisation in Practice: Terminological Issues in Bilingualising a Specialised Glossary. *International Journal of Lexicography* 20(2): 119-145.

Van Gheel, J. [*religious name of* Adriaen Willems; *Fl*. Joris van Gheel; *Fr*. Georges de Gheel; *Lat*. Georgius Gelensis]. 1652. *Vocabularium Latinum, Hispanicum, e Congense. Ad Usum Missionariorû transmittendorû ad Regni Congo Missiones*. Manuscript kept in the National Central Library of Rome: Fundo Minori 1896, MS Varia 274.

Vansina, J. 1974. The Dictionary and the Historian. *History in Africa* 1: 149-152.

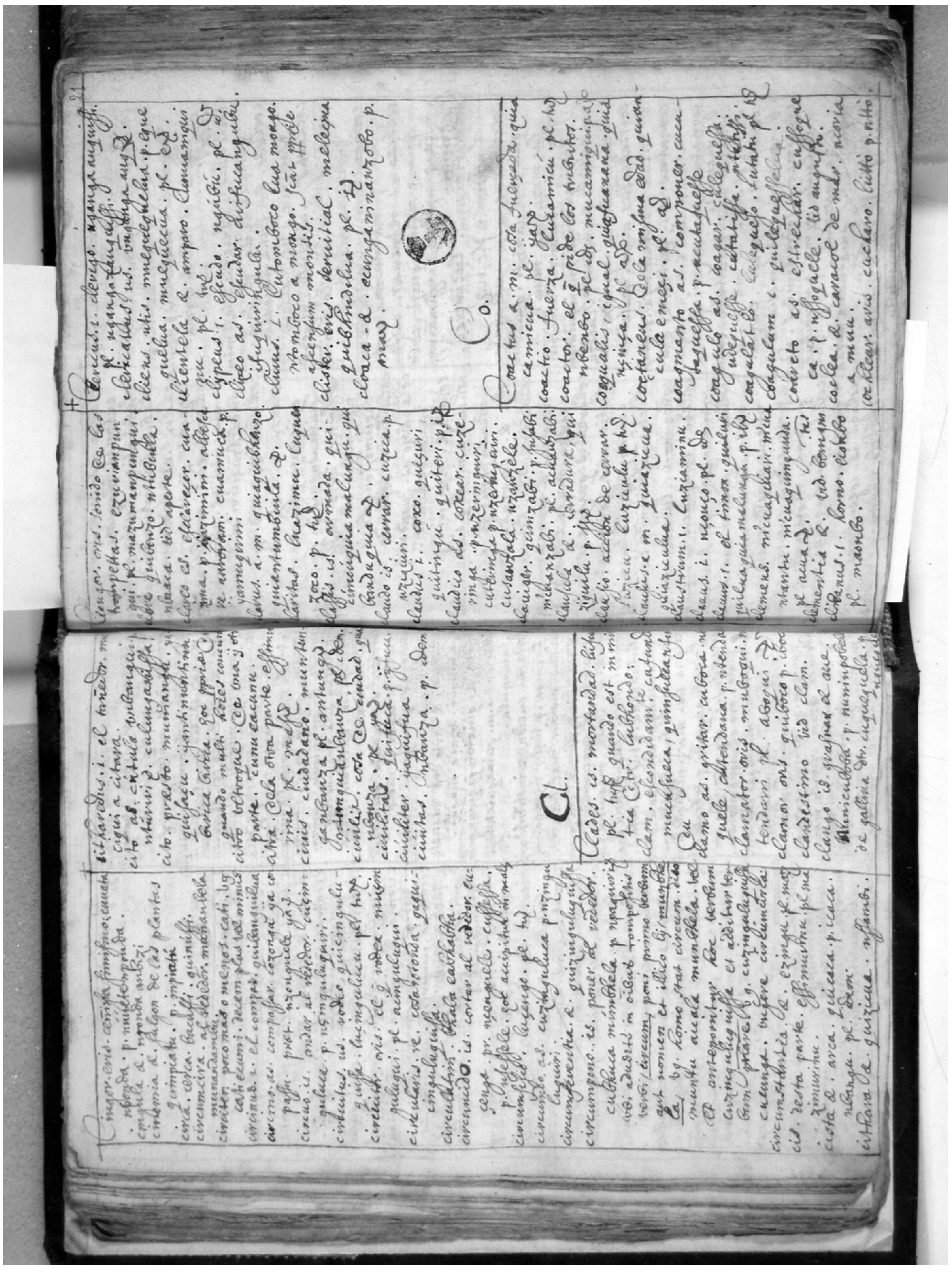
Van Wing, J. and C. Penders. 1928. *Le plus ancien dictionnaire bantu/Het oudste Bantu-woordenboek. Vocabularium P. Georgii Gelensis* (Bibliothèque Congo/Congo Bibliotheek 27). Louvain: J. Kuyl-Otto.

[White Fathers]. 1947. *Bemba–English Dictionary*. Chilubula: White Fathers.

Whitehead, J. 1899. *Grammar and Dictionary of the Bobangi Language, as Spoken over a Part of the Upper Congo, West Central Africa*. London: Baptist Missionary Society & Kegan Paul, Trench, Trübner & Co.

Wilkinson, A.S. (Ed.). 2010. *Iberian Books/Libros ibéricos (IB). Books Published in Spanish or Portuguese or on the Iberian Peninsula before 1601/Libros publicados en español o portugués o en la Península Ibérica antes de 1601*. Leiden: Brill.

Addendum 1: Pages 41-42 from the *Vocabularium Latinum, Hispanicum, e Congense* (Van Gheel 1652)



Addendum 2: Start of the section "C before O" in De Nebrija's *Dictionarium* (edition of 1570)

C ante O

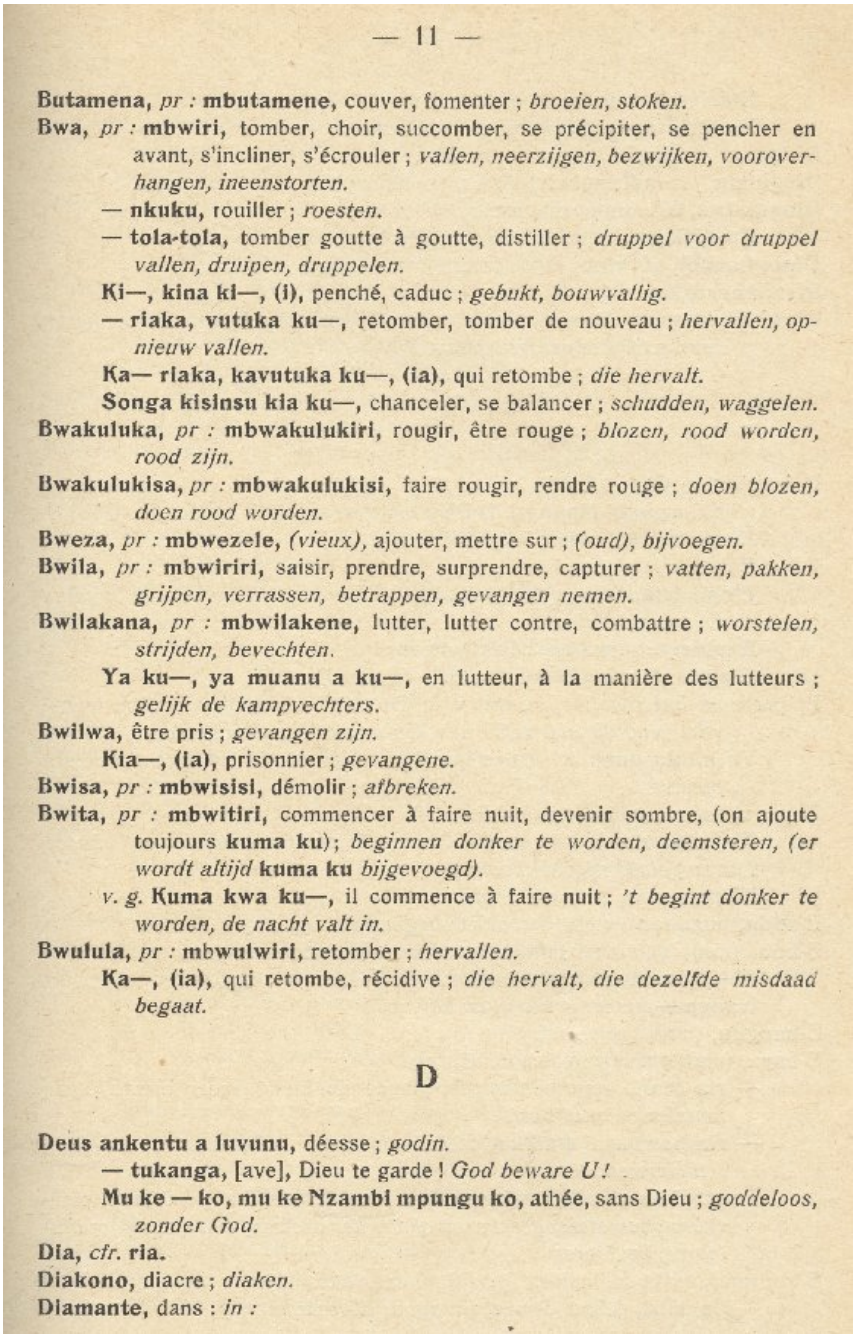
Cnidios coccos, interp. granum gnidium.
Cnipes, cnipos, interp. culex mulio.
Cnipulogos, u. interp. culicilega apis.
Cnoodes, per. ω. in pen. interp. lanuginosus.

C ANTE O.

Coa, siue cola, in talis siue taxillis dicũtur Veneres
Coaceo, cs. vel coacesco, is, coacui. *Por azedarse.*
Coaceruo, as, aui. *Por amontonar o ayuntar.*
Coaceruatio, onis. *Por aquel amontonamiento.*
Coacto, as. frequent. verbum ab eo quod cogo, gis.
Coactor, oris. *Por el cogedor de pechos o monedas.*
Coactus, a, um. participium ab eo quod cogo, gis.
Coactio, nis, vel coactus, us. *Por el constreñimiento.*
‡ Coactilitij calcei. *Alpargatas.*
Coædifico, as, aui. *Por edificar juntamente.*
Coæqualis, æ. *Por cosa ygual con otro cosa.*
Coæqualitas, atis. *Por aquella ygualdad con otra cosa.*
Coæquo, as, aui. *Por ygualar vna cosa con otra.*
Coæquatio, onis. *Por aquella ygualdad.*
Coætaneus, a, um. *Cosa de vna mesma edad y tiempo.* g. Synchronos.
Coagito, as, ex con, & agito, as. *Acoßar, o aguijar.*
Coagitatio, onis. *Aquella obra de acoßar o aguijar.*
Coagmento, as. *Por travar o engrudar vno con otro.*
Coag nentum, i. vel coagmẽtatio, nis. *Aquella travaçõ*
Coagulum, i. *Por el cuajo.* Græce, pyetia, as.
Coagulo, as, aui. *Por cuajar con cuajo.*
Coagulatio, onis. *Aquella obra de cuajar con cuajo.*
Coaleo, es. vel coalesco, is, iui. *Crecer vna cosa con otra*
Coalluo, is, ui, utum. *Por llauar juntamente.*
Coamo, as, aui. *Por amar juntamente con otros.*
Coamator, oris. pro eo qui cum alijs simul amat.
Coangusto, as, aui. *juntamente y con otros estrechar.*
Coangustatio, nis. *Por aquel estrechamiento o angustura.*
Coapto, as, aui. *Por juntamente atauiar o componer.*
Coaptatio, nis. *Aquel ayuntamiento o atauio.*
Coarcto, as, aui. *Por juntamente estrechar o ensangostar.*
Coarctatio, nis. *Por aquel estrecho o ensangostadura.*
Coarguo, is, gui, utum. *Por juntamente redarguir.*
Coasso, as. vel Coaxco, as, aui. *Por entablar con tablas.*
Coassatio, nis, vel coaxatio, nis. *Aquel entablamiento.*
Coaxo, as, aui. verbum factitium. *Por cantar la rana.*
Cobios, u. vel Cobius, ij, per. ω. in prima, lat. dicitur gobius, ij. *Por el gobi.*
Cobios, per. ω. quoq; herba est apud Plinium.
Cocanicus sal. *Por vn genero de sal de Sicilia ista.*
‡ Coccetum, i. edulij genus ex melle & papauere factũ
Coccimela, æ. à Theodoro prunus arbor interp.
Coccus, u. generali significatu interp. granum.
Coccum, i. speciali significatu. *Por la grana para tenir.*
Coccum Cnidium, i. granum ex Cnido peninsula & vrbe.
Coccineus, a, ũ. siue coccinus, a, um. *Por cosa de grana.*
Coccinatus, a, um. *Por cosa vestida de grana.*
Coccyx, ygos. interp. Cuculus auis. *Por el cuclillo.*
Coccyzo, as. *Cantar los cuclillos o los gallos.*
Coccyx, gos. interp. cuculus piscis.
Cochlea, æ. *Por el caracol de comer.* Græ. cochlias, æ.
Cochlearium, ij. locus est vbi cocleæ aluntur.
Cochlea fluuiatilis. *Por el almeja de los rios.*
Cochlea marina. *Por el almeja o caracol dela mar.*
Cochlea, æ. *Por el caracol para subir al altura.*
Cochlear, ris. siue cochlearium, ij. *Por la cuchara.*
Cochlear, per apocopé figura dicitur pro cochleare
Cochlear mensura est quantum cochlear capit.
Cochleæ terrestres dñr limaces testacei. *Los caracoles.*
Cochlea, æ. *Por el engeño para sacar agua de pozo.*
Cochlea, æ. siue Cochleaca. *El rollo piedra lixa del rio.*
Cochleola, æ. dimin. est ab eo quod cochlea, æ.
Cochlias, u. interp. cochlea siue limax.
Cochlis, idos, quoddam genus gemmæ est Plinio.
Cocio, nis. à Gellio appellatur aruspex.
Cocio, per. ω. in prima interp. ploro, lugeo, lamẽtor
Cocitos, interp. luctus, vlulatus, ploratus.
Cocles, itis. i. luscus. *El tuerto menguado de vn ojo.*
Coclos, u. interp. testa, æ. *Por la concha.*
Coctanum, i. *Por el membrillo fruta.*
Coctilis, e. *Por cosa cochia que presto se cuezе.*
Coctiuus, a, um. pro eo quod est coctilis, e.
Coctito, as, aui. frequent. ab eo quod, coquo, is.
Coctus, a, um. participium est à coquo, is, xi.
Coctura, æ. siue coctio, nis. *Por la cozedura.*
Cocton Laconicus, genus quoddam vasis potorij.
Coctonum, i. *Por vn genero de higos pequeños.*
‡ Coculum, i. potius quam cocula & coculus, est vas æneum coctionibus aptum. *Por la olla en que se cozе la vianda.*
Cocus, i. *Por el cozinero.* qui. g. mageiros per ei.
‡ Coculus diminutiuum, à cocus.
Coda, æ. pro eo quod cauda, æ. *Por la cola.*
Codeonas, Lucianus appellat, quę Apul. tintinabula
Codex, icis, siue Caudex, icis. *Por el tronco del arbol.*
Codex, icis. *Por el volumen o libro escripto.*
‡ Codex, icis. *Mano de papel.*
‡ Codex excerptorius. *El cartapacio, donde se escriue lo que saca delos libros.*
Codia, as. per. ω. & ei. diph. interp. papaueris caput.
Codicaria nauis, dicitur quæ ex codicibus est compacta.
Codetum, i. est arboretum à codicibus dictum.
Codicillus, i. dimin. est ab eo quod est codex.
Codicilli, orum. in plurali. *El codicillo del testamento.*
Codicilli pugillares dñr, qui Græce hemeropinaces.
Codicilli pugillares pluraliter dici indicat iurisc.
Codicilli, à Catullo quoq; dicuntur pugillarij vsus.
Codion, Græce pellis villosa interp.
Codon pro tegumẽto pelliceo accipitur apud Sillĩ
Cœlos, per. œ. diph. interp. cauus siue inanis.
Cœlia, as. interp. venter siue vterus mulieris.
Cœliacus, dicitur qui fluxu ventris laborat.
Cœliolysia, Græce alui profluuium interp.
Cœlostomia, as. est cum vox in oris concauo sonat.

Cœlotes

Addendum 3: Page 11 from *Le plus ancien dictionnaire bantu/Het oudste Bantu-woordenboek* (Van Wing and Penders 1928)



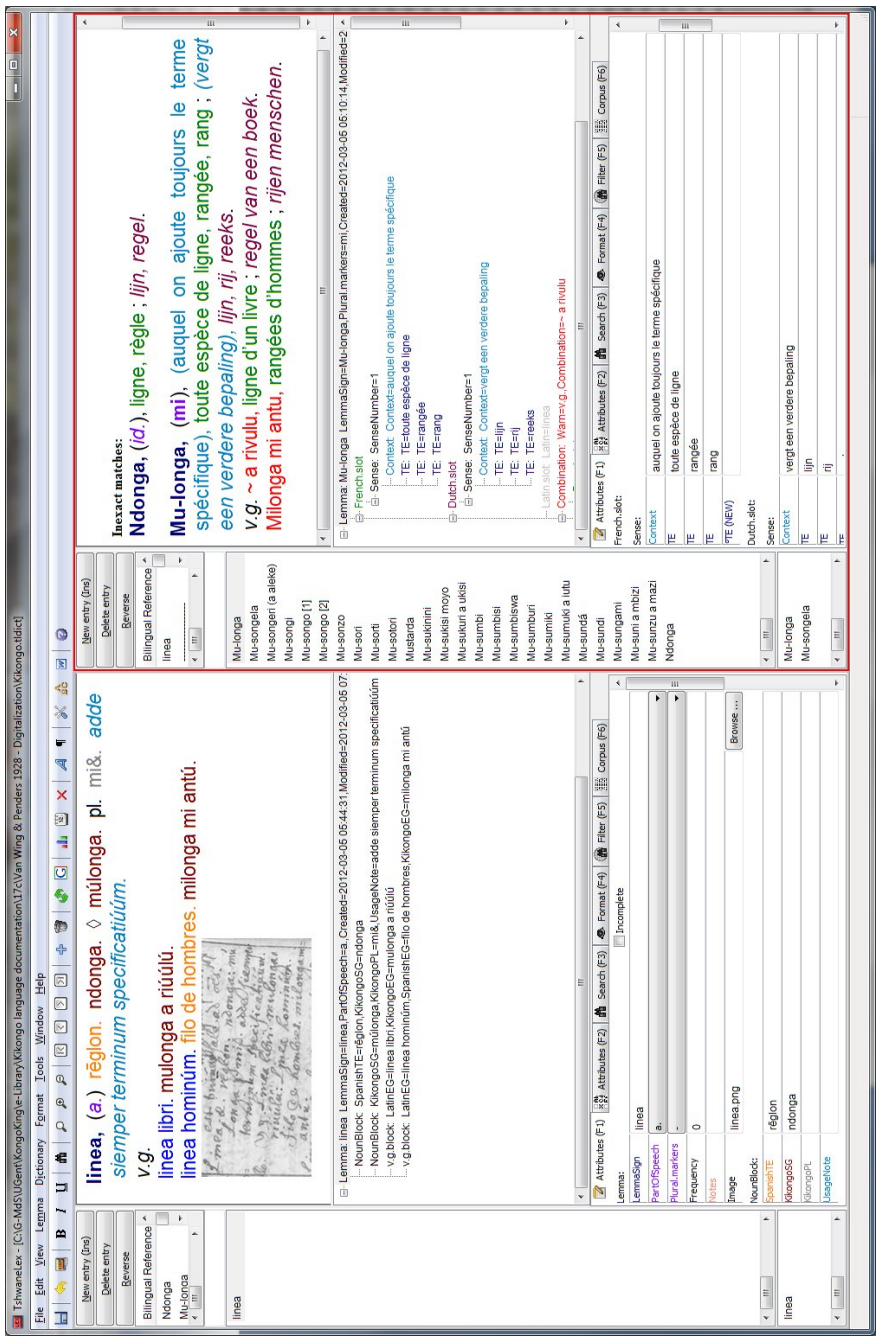
— 11 —

Butamena, *pr :* **mbutamene,** couver, fomenter ; *broeien, stoken.*
Bwa, *pr :* **mbwiri,** tomber, choir, succomber, se précipiter, se pencher en avant, s'incliner, s'écrouler ; *vallen, neerzijgen, bezwijken, voorover-hangen, ineenstorten.*
— **nkuku,** rouiller ; *roesten.*
— **tola-tola,** tomber goutte à goutte, distiller ; *druppel voor druppel vallen, druipen, druppelen.*
Ki—, kina ki—, (i), penché, caduc ; *gebukt, bouwvallig.*
— **riaka, vutuka ku—,** retomber, tomber de nouveau ; *hervallen, op-nieuw vallen.*
Ka— riaka, kavutuka ku—, (ia), qui retombe ; *die hervalt.*
Songa kisinsu kia ku—, chanceler, se balancer ; *schudden, waggelen.*
Bwakuluka, *pr :* **mbwakulukiri,** rougir, être rouge ; *blozen, rood worden, rood zijn.*
Bwakulukisa, *pr :* **mbwakulukisi,** faire rougir, rendre rouge ; *doen blozen, doen rood worden.*
Bweza, *pr :* **mbwezele,** *(vieux),* ajouter, mettre sur ; *(oud), bijvoegen.*
Bwila, *pr :* **mbwiriri,** saisir, prendre, surprendre, capturer ; *vatten, pakken, grijpen, verrassen, betrappen, gevangen nemen.*
Bwilakana, *pr :* **mbwilakene,** lutter, lutter contre, combattre ; *worstelen, strijden, bevechten.*
Ya ku—, ya muanu a ku—, en lutteur, à la manière des lutteurs ; *gelijk de kampvechters.*
Bwilwa, être pris ; *gevangen zijn.*
Kia—, (ia), prisonnier ; *gevangene.*
Bwisa, *pr :* **mbwisisi,** démolir ; *afbreken.*
Bwita, *pr :* **mbwitiri,** commencer à faire nuit, devenir sombre, (on ajoute toujours **kuma ku**) ; *beginnen donker te worden, deemsteren, (er wordt altijd* **kuma ku** *bijgevoegd).*
v. g. **Kuma kwa ku—,** il commence à faire nuit ; *'t begint donker te worden, de nacht valt in.*
Bwulula, *pr :* **mbwulwiri,** retomber ; *hervallen.*
Ka—, (ia), qui retombe, récidive ; *die hervalt, die dezelfde misdaad begaat.*

D

Deus ankentu a luvunu, déesse ; *godin.*
— **tukanga,** [ave], Dieu te garde ! *God beware U!*
Mu ke — ko, mu ke Nzambi mpungu ko, athée, sans Dieu ; *goddeloos, zonder God.*
Dia, *cfr.* **ria.**
Diakono, diacre ; *diaken.*
Diamante, dans : *in :*

Addendum 4: Van Gheel's and Van Wing and Penders' dictionaries in linked-view mode in the TLex *KongoKing Database* (2012)



Why One and Two Do Not Make Three: Dictionary Form Revisited

Anna Dziemianko, *Department of Lexicology and Lexicography, Faculty of English, Adam Mickiewicz University, Poznań, Poland (danna@wa.amu.edu.pl)*

Abstract: The primary aim of the article is to compare the usefulness of paper and electronic versions of OALDCE7 (Wehmeier 2005) for language encoding, decoding and learning. It is explained why, in contrast to Dziemianko's (2010) findings concerning COBUILD6 (Sinclair 2008), but in keeping with her observations (Dziemianko 2011) with regard to LDOCE5 (Mayor 2009), the e-version of OALDCE7 proved to be no better for language reception, production and learning than the dictionary in book form.[1] An attempt is made to pinpoint the micro- and macrostructural design features which make e-COBUILD6 a better learning tool than e-OALDCE7 and e-LDOCE5. Recommendations concerning further research into the significance of the medium (paper vs. electronic) in the process of dictionary use conclude the study. The secondary aim which the paper attempts to achieve is to present the status of replication as a scientific research method and justify its use in lexicography.

Keywords: PAPER DICTIONARIES, ELECTRONIC DICTIONARIES, DICTIONARY USE, ENCODING, DECODING, RETENTION, RESEARCH METHODS, REPLICATION, MENUS, HIGHLIGHTING, NOISE, ACCESS, ENTRY LENGTH

Opsomming: Waarom een en twee nie gelyk is aan drie nie — woordeboekvorm herbeskou. Die primêre doel van die artikel is om die bruikbaarheid van papier- en elektroniese weergawes van OALDCE7 (Wehmeier 2005) te vergelyk vir taalenkodering en -dekodering en die aanleer van taal. Daar word verduidelik waarom, in teenstelling met Dziemianko (2010) se bevindinge betreffende COBUILD6 (Sinclair 2008), maar in ooreenstemming met haar waarnemings (Dziemianko 2011) met betrekking tot LDOCE5 (Mayor 2009), die e-weergawe van OALDCE7 geblyk het nie beter te wees vir taalresepsie en -produksie en die aanleer van taal as die woordeboek in boekvorm nie.[1] 'n Poging word aangewend om die mikro- en makrostrukturele ontwerpkenmerke aan te stip wat e-COBUILD6 'n beter onderrighulpmiddel maak as e-OALDCE7 en e-LDOCE5. Aanbevelings betreffende verdere navorsing oor die belangrikheid van die medium (papier vs. elektronies) in die proses van woordeboekgebruik sluit die studie af. Die sekondêre doel wat die artikel probeer bereik, is om die status van replisering as 'n wetenskaplike navorsingsmetode aan te bied en die gebruik daarvan in die leksikografie te regverdig.

Sleutelwoorde: PAPIERWOORDEBOEKE, ELEKTRONIESE WOORDEBOEKE, WOORDEBOEKGEBRUIK, ENKODERING, DEKODERING, RETENSIE, NAVORSINGSMETODES, REPLISERING, KIESLYSTE, BELIGTING, GERAAS, TOEGANG, LENGTE VAN INSKRYWING

Lexikos 22 (AFRILEX-reeks/series 22: 2012): 195-216

1. Introduction

1.1 The usefulness of paper and electronic dictionaries

Electronic counterparts of printed monolingual English learners' dictionaries, available on CD-ROMs, online, or — increasingly often — on portable electronic devices, are taken for granted these days. Some of them appear to be quite close to their predecessors in book form (Rogers 1996, Nesi 1999). However, it is strongly stressed that e-dictionaries should not be just electronic remakes of existing printed dictionaries, but should rather be compiled from scratch as genuine electronic tools and take advantage of the wide array of technological possibilities (Nielsen and Mourier 2005: 110). Although contemporary electronic dictionaries, also those based on paper ones, do employ various functionalities offered by the electronic medium and/or the Web technology, further improvements are suggested (Müller-Spitzer et al. 2011, Prinsloo et al. 2011, Lew: In press, Kwary 2012). While the compilation of electronic dictionaries for foreign learners of English independently of (or in place of) paper dictionaries might be just a matter of time, the coexistence of the two media at present raises an obvious question of their relative usefulness in different linguistic tasks.

There is a vast body of studies where the effectiveness of paper and electronic dictionary use is compared.[2] Unfortunately, the results do not permit easy generalisation due to the wide range of user- and task-variables as well as different functionalities and lexicographic data available in the diverse electronic dictionaries used in research. Worse yet, even when the design, dictionary and user differences are neglected, hardly any general picture emerges, either.

First, as regards decoding, no effect of paper and electronic dictionary conditions was found by Nesi (2000), Kobayashi (2007), Koyama and Takeuchi (2007) and Chen (2010, 2012). Electronic dictionaries were however observed to significantly facilitate language reception by Osaki et al. (2003), Osaki and Nakayama (2004) or Dziemianko (2010). In the first two of the abovementioned studies, they also proved to significantly help in identifying contextually appropriate meanings.

Second, different conclusions follow also from the few studies where the influence of paper and electronic dictionaries on language production was tested. In the study by Chen (2010), the subjects were requested to formulate sentences with low-frequency words on the basis of the information found in dictionaries available on hand-held electronic devices and on paper. The results obtained in the encoding task did not depend on the dictionary used. In the study by Dziemianko (2010), in turn, the results from the production task, which consisted in supplying prepositions missing from sentences, were significantly better in the group working with the online version of COBUILD6 than in the one consulting COBUILD6 on paper.

Third, conclusions from studies concerned with the role of paper and electronic dictionaries in vocabulary retention are no less confusing. On the one hand, there are investigations which point to no significant effect of the medium on retention (Koyama and Takeuchi 2003, Osaki et al. 2003, Osaki and Nakayama 2004, Kobayashi 2007, Xu 2010, Chen 2010, 2012). There are also those where the medium proved consequential in this respect. The research conducted by Koyama and Takeuchi (2004) revealed that paper dictionary use resulted in better retention than reference to a portable electronic dictionary. Dziemianko (2010), by contrast, concluded that the consultation of COBUILD6 online resulted in better retention of meaning and collocations than the use of the dictionary in book form. Interestingly enough, the authors of both studies refer to the Involvement Load Hypothesis to account for their findings. Koyama and Takeuchi (2004) suppose that the more demanding process of paper dictionary search is beneficial to retention, in line with the assumption that greater effort means deeper processing, which stimulates retention. Dziemianko (2010), in turn, presumes that the saliency of a dictionary entry on the computer screen as well as the lack of distractions in the form of entries irrelevant to the task at hand, which are bound to be seen on the page of a paper dictionary, induce the cognitive involvement which enhances retention.

Finally, even the replication of a study on the usefulness of paper and electronic dictionaries yields results divergent from those obtained in the original investigation. Dziemianko (2011) adopted the same conditions as those in her pervious study (Dziemianko 2010), except for the dictionary. Instead of COBUILD6, the paper and free online versions of LDOCE5 were offered for consultation. Importantly, the subjects who comprised the other sample were as proficient in English and familiar with paper and electronic dictionaries as those who used COBUILD6 (B2-C1 in CEFR). Despite the same tasks in both experiments, the results from the replication do not confirm previous conclusions. Whereas in the 2010 study it was found that the electronic medium enhanced reception, production and the retention of meaning and collocations, in the more recent investigation dictionary format proved to be inconsequential to the scores on the very same language tasks. In other words, success rates in encoding, decoding and retention were comparable across the two dictionary conditions, i.e., LDOCE5 on paper and online.

To account for the results, Dziemianko (2011) points out that in the free online version of LDOCE5 excessive noise in the form of colourful widgets or animated tower advertisements dwarfs lexicographic data. Such unsolicited (promotional) information in loud colours and different shapes must have diverted the subjects' attention away from dictionary information, which became less prominent and quite inconspicuous. Possibly, then, discerning lexicographic information and extracting it from the glutted website became no less difficult than locating it in a paper dictionary. Unfortunately, neither p-LDOCE5 search nor e-LDOCE5 noise contributed to strengthening the memory trace in a way which could positively influence retention. E-COBUILD6, by

contrast, is much clearer and more neatly organised. In particular, there are no advertisements on its website, and dictionary information looks salient on the screen. Possibly that is why it was more useful than COBUILD6 in book form.

The above brief overview of selected recent studies on paper and electronic dictionary use reveals no obvious conclusions concerning the relative usefulness of these media for language reception, production and retention. As already pointed out above, the investigations differ in tasks, subjects, sampling methods, monitoring dictionary use or quantification, which naturally raises serious comparability issues. Unfortunately, the role of dictionary form in other respects, not discussed in the present paper, such as the speed of dictionary consultation, entry navigation, access paths or even dictionary appreciation is no clearer, either (Dziemianko: In press).

1.2 The role of replication

The wide variety (and inconclusiveness) of research into the relative usefulness of paper and electronic dictionaries highlights the need for systematic replication. Commonly seen as merely repeating a study to see if the same results can be obtained (Lindsay and Ehrenberg 1993: 217, Abbuhl 2012: 296), replication constitutes a crucial scientific method. If carefully designed and conducted, it leads to results that can be generalised, rather than just isolated findings (Lindsay and Ehrenberg 1993: 216). It also increases confidence in the results and helps to establish the reliability of research (Seidlhofer 2003: 215, Gass et al. 2011: 210-211). It is even claimed that "the soundest empirical test of the reliability of data is provided by replicating" (Sidman 1960: 70) and "an isolated study remains virtually meaningless and useless in itself" (Lindsay and Ehrenberg 1993: 218).

Gast (2009: 112) gives three reasons why it is worthwhile to replicate previous studies: to assess the reliability of findings (i.e., internal validity), to assess the generality of findings (i.e., external validity) and to look for exceptions (i.e., conditions under which the original findings do not apply). It is thanks to replication that the margin of error is reduced and confidence that findings are not accidental is strengthened. Systematic replication (whereby a researcher carries out a planned series of studies with systematic changes from one study to another and identifies them as a series) is particularly valuable as it makes it possible to establish the generality of findings, or see how broadly the results can generalise beyond the original experiment (Gast 2009: 111-112, 116, 121). Currently, statistical significance is taken for the ultimate objective of a study, rather than just the first step. A statistically significant result means that it is unlikely to be a product of the sampling error and that it is probably real inasmuch as it is likely to be achieved if the whole population is tested. Yet, "[s]ignificance cannot and does not tell us whether the same result would hold again in a different population or under different conditions. To establish that would require much explicit replication" (Lindsay and Ehrenberg 1993:

218). Put differently, "one statistically significant finding cannot be accepted as 'the truth'; only when results are repeated in other studies can we have greater confidence that our decision to accept or reject a hypothesis is correct" (Abbuhl 2012: 306).

Apart from justifying the need for replications, it is necessary to reflect on how research can be replicated. Replications can be plotted along a continuum which extends from exact, through approximate, to conceptual replications, depending on how closely they resemble the original study (Abbuhl 2012: 297-300). Exact replications (also known as literal, strict or virtual), which consist in repeating the original study exactly or as exactly as possible, are mostly unreal, since no groups of subjects with all their idiosyncratic characteristics and experiences can be duplicated (Lindsay and Ehrenberg 1993: 200, Macaulay 2003: 78). In the case of approximate replications, also known as replications with changes (Abbuhl 2012: 298), the original study is repeated, but some (typically non-major) variables are modified, e.g., population, setting or task, yet comparability is not lost. The aim of such replications is to verify the generalisability of the results from the original study to a new population, setting or modality. In fact, the differences in the conditions of the consecutive studies are of the essence; it is they that make it possible to see whether results hold nevertheless (Lindsay and Ehrenberg 1993: 217).[3] Finally, conceptual or constructive replications diverge from the original study to the largest extent; the same research question is investigated, but a different design is followed. In other words, the findings from an existing study supply the starting point, but researchers develop their own methodology. Such replications make it possible to distinguish between method-specific results and those which can be generalised, but the more variables are changed, the less comparable the original study and its conceptual replication become (Abbuhl 2012: 304).

Unfortunately, replication is held in relatively low esteem; it is considered to be inferior to original research (Umapathy 1987: 170) and lacking in prestige (Campbell 1986: 122). The "pressure to be original" (Park 2004: 194) and the mistaken view that any replication boils down to merely repeating an existing study exactly (Lindsay and Ehrenberg 1993: 220) contribute to the low regard for replication as a scientific method. Although its role in theory development cannot be overestimated, irrespective of whether it supports the tested theory or, perhaps even more importantly — not, replication is seldom undertaken.

As regards research into dictionary use, the value of replication seems to be recognised; the method is claimed to be helpful for improving dictionaries and their usability for language learners (McCreary 2002: 182). However, there are relatively few studies openly acknowledged to be replications of some previous investigations, conducted with different degrees of modification (e.g., Greenbaum et al. 1984, Nesi and Meara 1991, Horst 1995, McCreary and Dolezal 1999, McCreary 2002, McCreary and Amacker 2006, Lew and Doroszewska 2009, Lew and Dziemianko 2006, Lew 2010b, Dziemianko 2011, Chen 2012).[4] Admittedly, the study by Greenbaum et al. (1984), which replicates the

survey by Quirk (1974), shows that the method has been employed in user-centred research for at least three decades. Yet, the small number of replications cannot be unmotivated. It might result from the fact that many studies on dictionary use are simply non-replicable (Hartmann 1987: 27). The low esteem which replication has is probably another factor which discourages researchers. Besides, it is by no means easy to ensure that the original study and its replication are closely comparable. Although replications are considered advisable when the researcher's aim is to make a new study parallel to an existing one (Lew 2002), direct comparisons can still be quite difficult to perform. For one thing, as pointed out above, exact replications are virtually nonexistent. For another, approximate replications, where the conditions whose influence is of particular interest are purposely varied, obviously give a chance for systematic comparison, provided that the other conditions remain unchanged. Yet, it takes time and effort to control the latter, which makes approximate replications difficult to accomplish successfully. Finally, the fact that not many researchers openly wish their investigations could be replicated in the future (McCreary and Dolezal 1999, Al-Ajmi 2002, Dziemianko 2006, Lew and Dziemianko 2006, Koyama and Takeushi 2007, Tono 2011) suggests that, in fact, the awareness of the benefits which can be derived from replication might need to be raised. It is tacitly assumed that replication "carries more risk than potential reward for both the replicator and the originator of the research" (Park 2004: 194). After all, failure to obtain the same result might be seen as a proof that the latter was wrong, or that the former is incompetent (Lindsay and Ehrenberg 1993: 218).

Indeed, although replications are said to be crucial "to distinguish the spurious from the real" (Abbuhl 2012: 306), there is a strong bias against negative findings. The file-drawer syndrome prevents the publication of many replications which do not support previous findings (Lindsay 1990, Park 2004: 194). Admittedly, confirming replications (whose results agree with those from the original investigation) are valuable inasmuch as they make the corroborated findings more credible. Yet, disconfirming replications are by no means worthless. Assuming that research is conversation, they prove that there is still a need to discuss the issue which turns out to be more complex than it seemed (Lindsay and Ehrenberg 1993: 218, Abbuhl 2012: 306). Besides, accounting for the divergent results provides ample scope for originality.

In an attempt to meet the need for systematic replication in research into dictionary use, the next part of the paper describes the second approximate replication of the study by Dziemianko (2010) and the obtained results.

2. The replication

2.1 Aim

As mentioned above, Dziemianko (2010) found that e-COBUILD6 was more useful in L2 reception, production and learning (retention of meaning and col-

locations) than COBUILD6 on paper. The results were not confirmed by the first approximate replication carried out by the author herself, where the paper and free online versions of LDOCE5 were employed. No statistically significant differences between the results obtained in the paper and electronic dictionary conditions were noted then in any task (Dziemianko 2011).

The aim of the present study is twofold. First, an attempt is made to investigate the usefulness of OALDCE7 in paper and electronic form for language reception, production and learning. Second, Dziemianko's (2010) findings concerning COBUILD6 are compared with those obtained from both replications.

The following research questions are answered:

1. Which version, paper or electronic, of OALDCE7 is more useful for L2 reception, production and learning (retention of meaning and collocations)?
2. Which dictionary (OALDCE7, LDOCE5 or COBUILD6) and in which form is most helpful in dealing with receptive and productive tasks, and which is the best learning tool?

The CD-ROM and regular printed versions of OALDCE7 were used. The choice of the seventh edition of the dictionary, rather than the latest one, was motivated by the number of copies of the dictionary in book form available in the experimental setting as well as by the functionalities of the electronic version. For one thing, there were enough paper copies of OALDCE7 to go around in the groups in which the study was conducted. For another, the CD-ROM version of OALDCE7 made it possible to see whether some search facilities which it offers (such as automatic scrolling or highlighting the entry for the looked up word, not available in the online versions of LDOCE5 and COBUILD6) matter to dictionary users.

2.2 Materials and subjects

The materials used by Dziemianko (2010), i.e., the pretest, questionnaire, test and unexpected delayed post-test, were employed. The subjects did the same receptive and productive tasks as in the original study. In the receptive task, they explained the meaning of nine nouns and phrases (*backgammon, booby prize, clampdown, collateral damage, down under, dream ticket, flapjack, onus, outcrop*). The productive part consisted in completing sentences with prepositions removed from nine collocations (*on the blink, in cahoots with, up the creek, at gunpoint, wreak havoc on, in the offing, in the pipeline, under sedation, on the trot*). Both tasks featured in the pretest, test proper and retention test. The pretest served to sift out the cases where the subjects knew correct answers. It was accompanied by a questionnaire to gain an insight into the subjects' familiarity with dictionary formats. Once the pretest and the questionnaire had been completed, the test was administered. In the test, the subjects did the same tasks as in the pretest, but with access to either paper or electronic OALDCE7. In the

delayed retention test conducted two weeks later, the sequence of the target structures was reshuffled and no access to dictionaries was allowed. The study was carried out in regular class time (45 minutes).

Great care was taken to ensure that the subjects were as proficient as those in the original research. Overall, 86 students of English (B2-C1 in CEFR) at Poznań University took part in the study; 42 of them consulted p-OALDCE7 and the other 44 the e-OALDCE7. The subjects' proficiency was determined on the basis of the grammar test in the practical English exam taken on a yearly basis. Importantly, the information obtained from the questionnaire indicates that in both experimental conditions the proportions of subjects consulting paper and electronic dictionaries as a matter of routine were comparable (the p-OALDCE7 group: students using paper dictionaries 66.7%, students using electronic dictionaries 69.0%, p=0.83; the e-OALDCE7 group: students using paper dictionaries 63.6%, students using electronic dictionaries 68.2%, p=0.68; Z test for dependent samples, non-significant, alpha-level=0,05).

2.3 Results

2.3.1 Research question one (the usefulness of OALDCE7)

The mean proportions of correct answers in the main and retention tests are illustrated in Figure 1. The results of the repeated-measures ANOVAs for both tests are given in Table 1.[5]

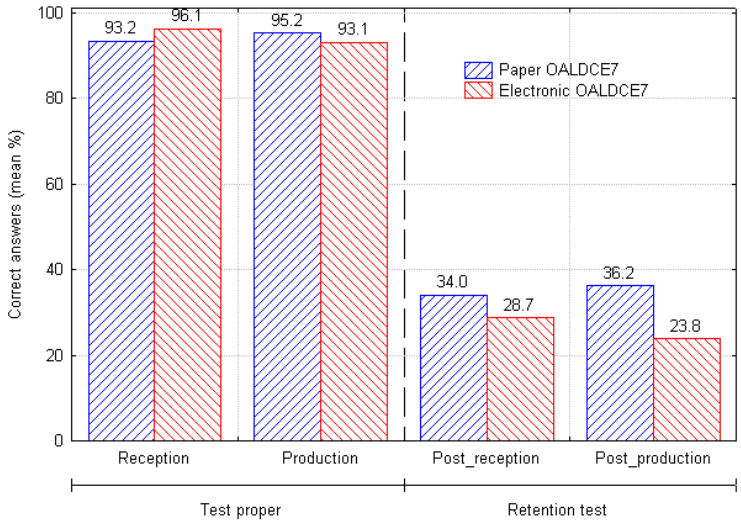


Figure 1: Results obtained in the main and retention tests (OALDCE7)

Factors	Test proper			Retention		
	F	p	partial η^2	F	p	partial η^2
FORM	0.0	0.880	0.001	1.3	0.272	0.075
TASK	0.0	0.880	0.001	0.1	0.790	0.005
TASK*FORM	0.7	0.422	0.041	0.5	0.499	0.029

Table 1: ANOVA summary results (main and retention tests): OALDCE7

In each test, the scores on each task were comparable among the users of paper and electronic versions of OALDCE7 at the accepted level of significance (alpha=0.05). In the main test, the subjects provided over 90 percent of correct answers in each task. The differences in the main test scores between the paper and electronic conditions approximated 3 percent for reception (paper dictionary (PD): 93.2%, electronic dictionary (ED): 96.1%) and production (PD: 95.2%, ED: 93.1%). In the retention test, in turn, active recall in the paper dictionary group (PD: 36.2%) was about half as good again as in the electronic dictionary group (ED: 23.8%). For passive recall, the difference, still in favour of the paper dictionary, amounted to 18 percent (PD: 34% vs. ED: 28.7%). While the differences were statistically insignificant in the light of the ANOVA, their scale seems to suggest that if the sample had been bigger, they might have gained significance. Yet, the low values of the estimate of effect size (partial η^2) computed for the retention test show that the size of each investigated main and interaction effect was very small, which means that only a modest proportion of the respective variance can be accounted for by a given (main or interaction) effect. In particular, only 7.5% of the between subjects variance in retention scores can be attributed to dictionary form (FORM, partial η^2=0.075).

2.3.2 Research question two (comparative usefulness of OALDCE7, LDOCE5 and COBUILD6)

2.3.2.1 Test proper

ANOVA results for the main test scores achieved by OALDCE7, LDOCE5 and COBUILD6 users are given in Table 2.

Factor	F	p	partial η^2
DICTIONARY	0.3	0.728	0.013
FORM	2.8	0.099	0.056
DICTIONARY*FORM	3.5	0.039*	0.127
TASK	0.0	0.999	0.000
TASK*DICTIONARY	0.0	0.965	0.001
TASK*FORM	0.4	0.553	0.007
TASK*DICTIONARY*FORM	0.4	0.690	0.015

Table 2: ANOVA (main test): OALDCE7, LDOCE5 and COBUILD6

The data show that only one interaction (DICTIONARY x FORM) was statistically significant (p=0.039, alpha=0.05; partial η^2=0.127). To explore it in more depth, Table 3 shows the results of the Tukey Honest Significant Difference test.[6] Figure 2 illustrates the interaction graphically.

DICTIONARY	FORM	Mean %		
COBUILD6	paper	92.1	****	
OALDCE7	paper	94.2	****	****
OALDCE7	electronic	94.6	****	****
LDOCE5	electronic	95.1	****	****
LDOCE5	paper	95.9	****	****
COBUILD6	electronic	98.6		****

Table 3: Tukey HSD test: DICTIONARY x FORM (main test)

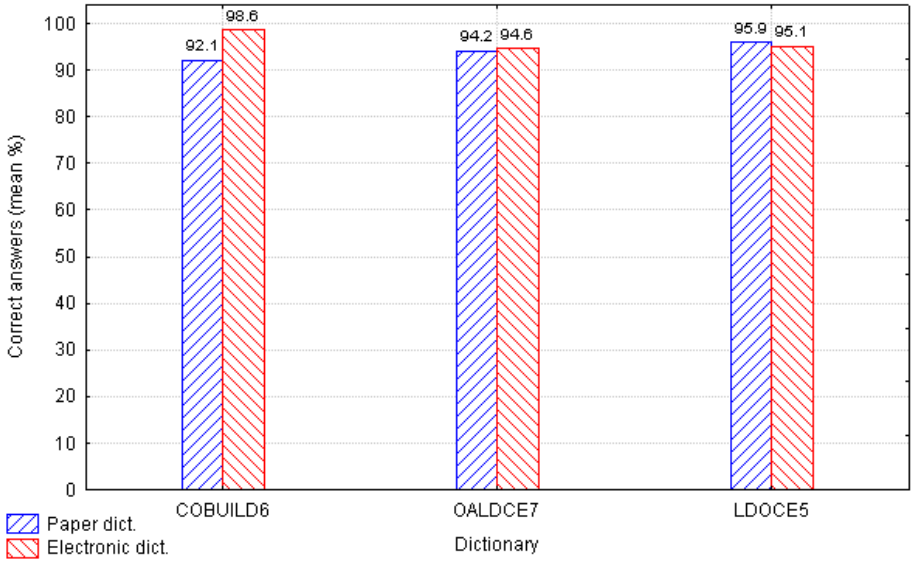


Figure 2: DICTIONARY x FORM: Correct answers (mean %) in the main test

The results of the Tukey HSD test reveal that in the main test, e-COBUILD6 (98.6%) was more useful than COBUILD6 on paper (92.1%, cf. Dziemianko 2010). However, both versions of LDOCE5 and OALDCE7 were comparably helpful.

2.3.2.2 Retention test

Summary ANOVA results for the retention test are collated in Table 4.

Factor	F	p	partial η^2
DICTIONARY	20.9	0.000**	0.465
FORM	0.3	0.564	0.007
DICTIONARY*FORM	3.1	0.054	0.115
TASK	8.8	0.005**	0.155
TASK*DICTIONARY	1.8	0.184	0.068
TASK*FORM	0.1	0.717	0.003
TASK*DICTIONARY*FORM	0.3	0.745	0.012

Table 4: ANOVA (retention test): OALDCE7, LDOCE5 and COBUILD6

The data indicate that the main effects produced by DICTIONARY (p=0.000) and TASK (p=0.005) were statistically highly significant at alpha=0.05. Also, the effect sizes associated with these factors were large and medium, respectively (DICTIONARY: partial η^2=0.465, TASK: partial η^2=0.155). Table 5 gives the results of the Tukey HSD test for the two significant effects, illustrated graphically in Figure 3.

DICTIONARY	Mean %			TASK	Mean %	
OALDCE7	30.7	****		Post_production	39.0	****
LDOCE5	37.5	****		Post_reception	47.9	
COBUILD6	62.2		****			

Table 5: Tukey HSD test: DICTIONARY and TASK (retention test)

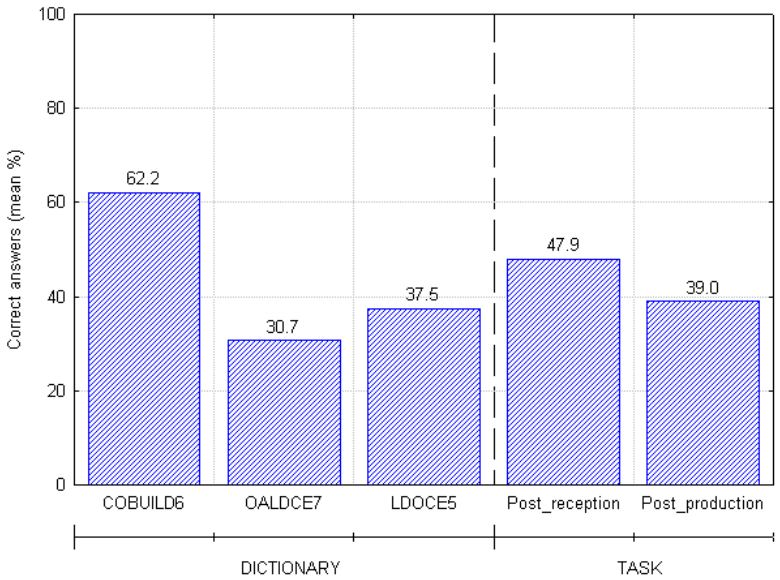


Figure 3: DICTIONARY and TASK: Correct answers (mean %) in the retention test

First, the best retention was observed in the COBUILD6 group, where it exceeded 62% and was significantly better than in the other dictionary conditions. The retention results obtained after reference to OALDCE7 (30.7%) and LDOCE5 (37.5%), only about half as good as among COBUILD6 users, were comparable. Second, meaning (47.9%) proved much easier to remember than collocations (39.0%); passive recall was over one fourth more successful than active recall, and the difference was statistically significant at alpha=0.05.

The interaction Dictionary x Form, which is not quite statistically significant but approaches significance (p=0.054, alpha=0.05; partial η^2=0.115, Table 4), also merits further investigation. Results of the Tukey HSD test for the interaction in question are collated in Table 6. The relevant mean proportions are illustrated in Figure 4.

DICTIONARY	FORM	Mean %			
OALDCE7	electronic	26.3	****		
OALDCE7	paper	35.1	****	****	
LDOCE5	electronic	37.4	****	****	
LDOCE5	paper	37.6	****	****	
COBUILD6	paper	54.0		****	****
COBUILD6	electronic	70.3			****

Table 6: Tukey HSD test: DICTIONARY x FORM (retention test)

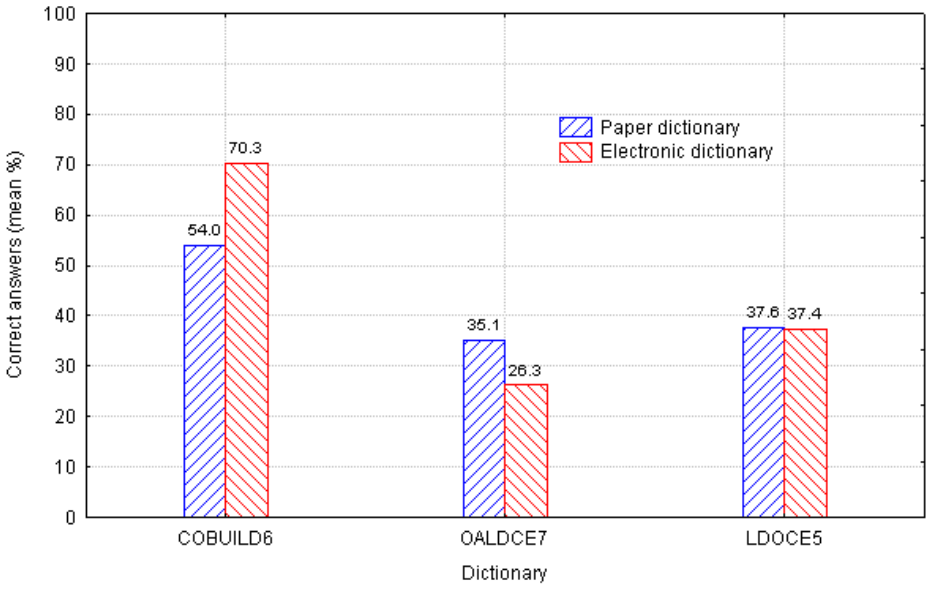


Figure 4: DICTIONARY x FORM: Correct answers (mean %) in the retention test

Three main conclusions follow from the data. First, it transpires that there were no significant differences in retention between the users of paper and electronic versions of LDOCE5 and OALDCE7. Second, reference to e-COBUILD6 yielded significantly better retention results than reliance on the other e-dictionaries; e-COBUILD users (70.3%) remembered about 90 and 170 percent more than the subjects who referred to e-LDOCE5 (37.4%) and e-OALDCE7 (26.3%), respectively. Third, retention among the users of LDOCE5, OALDCE7 and COBUILD6 on paper was comparable. Even though reference to p-COBUILD6 (54.0%) yielded retention results which were about half as good again as those obtained after the consultation of p-OALDCE7 (35.1%) and p-LDOCE5 (37.6%), on the Tukey HSD test, the difference was not statistically significant at alpha=0.05.

3. Discussion

Obviously, the replications led to conclusions different from those obtained in the original study. First of all, in contrast to Dziemianko's (2010) findings concerning COBUILD6, the e-versions of OALDCE7 and LDOCE5 proved to be no better for language reception, production and learning than the dictionaries in book form. Second, e-COBUILD6 was found to be a better learning tool than e-OALDCE7 and e-LDOCE5. It is thus necessary to reflect on the micro- or macrostructural features and factors not intrinsic to any dictionary structures which contributed to the success achieved with the help of e-COBUILD6 and prevented e-OALDCE7 and e-LDOCE5 from being likewise useful.

First of all, it is worth noting that the e-COBUILD6 website is quite crude; it is made up of the search window followed by the entry for the looked up word and a few buttons on the right (to be clicked if users wish to expand their vocabulary, customise the dictionary or get help). In e-OALDCE7, in turn, the entry for the looked up word, if short enough, is displayed along with the entries which follow it. This form of presentation resembles the paper dictionary and diverges from the approach adopted by e-COBUILD6, where only the entry for the looked-up word can be seen on the screen. Undoubtedly, the view of entries in a sequence must have naturally dispersed the subjects' attention and disturbed concentration. Such interface dissimilarities might be a reason why the retention scores of e-COBUILD6 users were better than those of the e-OALDCE7 group. The same factor might also account for the lack of any statistically significant difference between the results obtained with the help of the electronic and paper versions of OALDCE7 in the main and retention tests. In e-LDOCE5, by contrast, the entries for the headwords which follow the looked up word are not displayed, but the website overflows with noise, thereby deflecting users from the dictionary itself and making lexicographic data much less salient and distinct (cf. Dziemianko 2011 and section 1.1). This could be a possible reason why e-LDOCE5 was no more helpful in any experimental task than p-LDOCE5.

Apart from the examination of interfaces, item analysis was conducted with a view to explaining the observed results. Looking at the data for individual target items, Dziemianko (2011) drew interesting conclusions about the role of clickable menus in e-LDOCE5, i.e., vertical menus which consist of several matches, each of which is hyperlinked to an entry or subentry. Figure 5 shows such a menu for *blink*.[7]

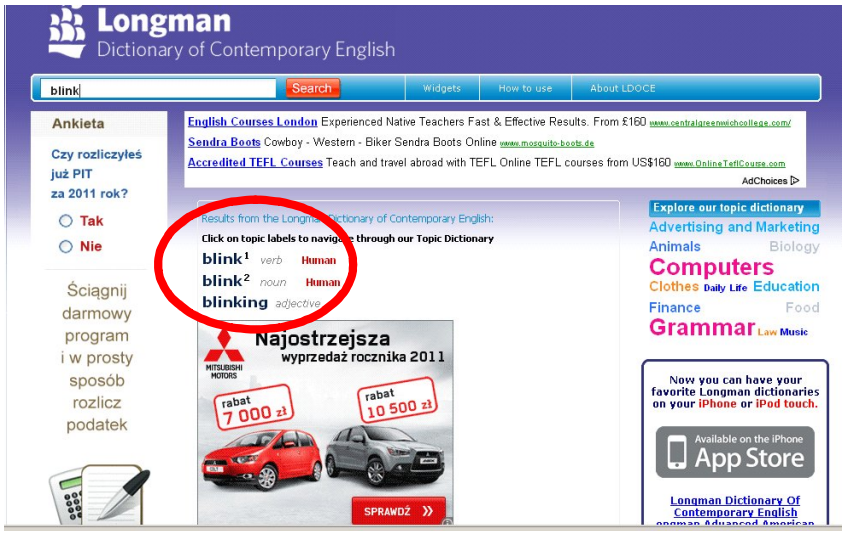


Figure 5: The menu for *blink* in e-LDOCE5 (circled)

Accessing noun phrases through clickable menus in e-LDOCE5 was found to severely impede reception in comparison with p-LDOCE5. No similar effect of clickable menus was identified on production. However, they proved seriously detrimental to passive and active recall (in comparison with the menu-less access paths in e-COBUILD6). Dziemianko (2011) hypothesised that the mechanical rather than cognitive effort invested into coping with the hierarchical, step-wise outer access structure in e-LDOCE5, at which stage *relevant* semantic information is not processed yet, did not strengthen the memory trace, but actually prevented successful reception and retention.

OALDCE7 does not feature clickable menus similar to those in e-LDOCE5, but it offers a different functionality — automatic scrolling whereby the looked up compound, phrase or idiom not given the headword status is immediately shown at the top of the screen. It is worth remembering that the results obtained by e-OALDCE7 users in the receptive task in the test proper were on average 3 percent better than in the group consulting p-OALDCE7 (cf. Figure 1). The largest difference in decoding scores between the experimental conditions was observed for *down under*, which in the paper version is given as the

sixth of the seven idioms explained at the end of the 12-sense entry for *down* (adv). In the electronic version, in turn, *down under* is immediately shown at the top of the computer screen, its identical placement in the entry for *down* notwithstanding. Automatic scrolling to the phrase resulted in 29 percent better score. Even though not quite statistically significant (p=0.080, Z test for independent samples, two-tailed, alpha=0.05), the difference was much beyond the aforementioned average (3 percent).

Interesting observations can be made about active recall, which was on average four times better among e-COBUILD6 users than among the subjects consulting e-OALDCE7. Item analysis reveals that this difference owes most to the collocation *up the creek*, retained over 13 times more often by the e-COBUILD6 group. This tremendous and statistically significant difference (p=0.000, Z test for independent samples, two-tailed, alpha=0.05) results most probably from the fact that the search for *creek* in e-OALDCE7 yields two matches. The first of them, a proper name irrelevant to the task at hand (*Creek — a member of a Native American people, many of whom now live in the US state of Oklahoma*), is highlighted, as shown in Figure 6. In e-COBUILD6, by contrast, *up the creek* constitutes the third subentry of *creek*, none of which is highlighted. The tentative conclusion which can be drawn from the data is that highlighting the entry for the searched word by default does not pay off when its homograph, treated in a separate entry (which is not highlighted), happens to be what dictionary users need. In such a case, default highlighting can result in immensely poorer retention.

Figure 6: The highlighted entry for *Creek* in e-OALDCE7

The second largest difference in active recall between the groups using e-COBUILD6 and e-OALDCE7 was observed for *on the trot*. The subjects who consulted the latter dictionary found the phrase in the section devoted to idioms, located at the end of the entry which consists of four verb senses, a subentry for the phrasal verb *trot (sth) out* and two noun senses. In e-COBUILD6, in turn, *on the trot* constitutes the third (final) subentry, but the two preceding verb subentries are quite short. Judging by the number of senses which separate the headword from the target phrase, the search path in e-OALDCE7 is three times longer than in e-COBUILD6. This might be a reason why e-COBUILD6 users were about 7 times more successful in active recall than the subjects who referred to e-OALDCE7. Apparently, then, the effort exerted to locate the phrase, as measured by entry length, is inversely related to active recall. In other words, the longer the entry is, the lower the chances of successful retention becomes. Yet, this hypothesis needs to be verified in further studies. It is worth noting that the results obtained in the main test for *on the trot* indicate that the phrase was extracted with comparable success from both dictionaries (97% in e-COBUILD6 and 97.7% e-OALDCE7, p=0.840 Z test for independent samples, alpha=0.05). Such an observation supports the surprising findings by Nesi and Tan (2011), who noted that the senses at the end of the entry are identified with the greatest speed and accuracy by dictionary users, followed by those which are given first. The regularity observed in the entry for *trot* not only confirms the saliency of the entry-final position, but also suggests that the effect persists regardless of entry length.[8] Nonetheless, it transpires that the saliency of entry-final positions has widely different consequences for entry navigation (i.e., finding the needed information) on the one hand, and retention on the other.

The foregoing discussion makes it possible to formulate a few suggestions for further research into e-dictionary use. First, it appears that the role of noise on dictionary websites is worth looking into. It goes without saying that advertisements make online dictionaries accessible to anyone free of charge. No wonder, then, that ad-supported online dictionaries are enjoying considerable popularity.[9] Nonetheless, it is open to question whether dictionary websites with and without advertisements are comparably useful. The tentative conclusion following from the present investigation is that unsolicited promotional material diverts users' attention from lexicographic data and actually deprives an online dictionary of much of its usefulness. Second, the effect of the hierarchical nature of data display in electronic dictionaries on retention is another promising area of research. The above assessment of the possible influence of clickable menus on retention, and active recall in particular, is quite pessimistic, but systematic manipulation of fabricated microstructures is necessary to get a deeper insight into the actual significance of clickable menus in electronic dictionaries. Admittedly, research into clickable menus as access facilitators was taken up by Lew and Tokarek (2010), who concluded that such tools help lower-level students navigate a dictionary entry and get to the right

sense, but are of no real benefit to advanced users. Apart from regular clickable menus, the authors looked into the usefulness of clickable menus where the target sense was automatically highlighted. Such menus proved comparably useful at both proficiency levels. However, no attempt has yet been made to investigate the effect of menus in paper or electronic dictionaries on retention (cf. Nesi and Tan 2011, Tono 2011, Lew 2010b). Third, it might be useful to explore the influence which highlighting entries in electronic dictionaries exerts on active and passive recall in the case of homographs treated in different entries, only one of which is highlighted. While highlighting entries by default seems attractive, it transpires that bringing out the entry which does not feature the information that a user wishes to find has a negative impact on retention. At this stage it is worth distinguishing between highlighting entries and highlighting specific senses. The latter was found a welcome navigation enhancement in polysemous microstructures, where it assists users in reaching the relevant sense more quickly and accurately (Lew and Tokarek 2010).

Unfortunately, the present study is not free from limitations. First, a number of subject variables were not controlled. Only the subjects' familiarity with dictionary formats and proficiency in English were taken into consideration, since they were considered most likely to immediately affect dictionary use and language skills. Besides, it needs to be remembered that real dictionaries rather than systematically manipulated microstructures were employed. Such an approach resulted in a naturalistic task, but it made it difficult to pin down specific factors responsible for the observed effects. To establish the role of selected factors, entries need to be fabricated and systematically manipulated, which no doubt creates more tightly controlled, albeit more artificial, conditions. The use of actual paper and electronic dictionaries also means that dictionary form alone may not be the key factor which determines the effectiveness of dictionary consultation. Specific solutions adopted and form-independent typographical structural indicators (Gouws 2003), such as font size and colour, line spacing or layout, which remained beyond control in the studies discussed above, can play an important role in dictionary use (cf. Lew 2010a: 294, Nesi: In press). To reduce their influence, printouts of the electronic dictionary screen display could be used instead of a real paper dictionary. Such task operationalisation could help to isolate the factor of dictionary form (on-screen vs. paper) and free it of the effect produced by typographical structural indicators (cf. Chen 2012). Nonetheless, in this way the paper dictionary user is also largely helped inasmuch as only mini-dictionaries covering the key items rather than complete paper dictionaries are typically produced from printouts, which seriously limits and simplifies outer access (Bergenholtz and Gouws 2007: 243).[10]

All in all, whereas the present study proved to be quite exploratory in nature at the stage of item analysis, it made it possible to develop a few testable hypotheses which merit further attention. In this way it hopefully confirmed that replication as a research method does not entail lack of originality. Impor-

tantly, it also showed that approximate replication helps to validate theories and substantiate generalisations. Ultimately, it is replications that contribute to making research a truly accretive process whereby knowledge is accumulated and consolidated over time, and, by the same token, prevent a discipline from being composed of scattered hypotheses and observations (cf. Santos 1989).

Endnotes

1. In the article, where differences between the dictionaries and their forms are of the utmost importance, the aforementioned, generally accepted acronyms are used for the sake of convenience. In the list of references, full bibliographic information is provided under the names of the respective dictionary editors, not repeated below: Mayor (2009) — LDOCE5, Sinclair (2008) — COBUILD6 and Wehmeier (2005) — OALDCE7.
2. For an overview, see Dziemianko (In press).
3. Naturally, the greater the differences are, the higher the risk that the effect will not be replicated. Yet, if it is confirmed, its generality increases (Gast 2009: 111). By the same token, "failure to replicate or follow up on studies with different populations and in different contexts may lead to de facto generalisation" (Duff 2006: 71).
4. Compare similar remark made by Chi (2009: 14), who also notes the paucity of replications in the field of dictionary use.
5. In any ANOVA discussed below, TASK was the repeated-measures factor.
6. All the means connected by (****) in one column are not different from each other at p=0.05.
7. The screenshot also gives an insight into the amount of noise on the e-LDOCE5 website.
8. Only five-sense entries were employed in the study by Nesi and Tan (2011).
9. See also Lew (2011).
10. Proponents of the Involvement Load Hypothesis would no doubt claim that simplified outer access can affect retention results, the assumption being that any effort invested in word search, including mechanical page turning and scanning running heads, can increase the chances of successful retention. On the other hand, it is suggested that not any involvement, but only semantic involvement affects vocabulary retention in the process of dictionary use. The aforementioned, largely automatic stages of paper dictionary look-up, might not yet evoke adequate semantic or cognitive involvement to influence vocabulary retention (cf. Craik and Lockhart 1972, Dziemianko: In press). Besides, printouts of an electronic dictionary prevent users from scanning entries close to the target ones, which might also affect retention (Chen 2012).

References

Abbuhl, R. 2012. Why, When, and How to Replicate Research. Mackey, A and S.M. Gass (Eds.). 2012. *Research Methods in Second Language Acquisition: A Practical Guide*: 296-312. Oxford: Blackwell Publishing.

Al-Ajmi, H. 2002. Which Microstructural Features of Bilingual Dictionaries Affect Users' Look-Up Performance? *International Journal of Lexicography* 15(2): 119-131.

Bergenholtz, H. and R.H. Gouws. 2007. The Access Process in Dictionaries for Fixed Expressions. *Lexicographica* 23: 237-260.

Campbell, D.T. 1986. Science's Social System of Validity-Enhancing Collective Belief Change and the Problems of the Social Sciences. Fiske, D.W. and R.A. Shweder (Eds.). 1986. *Metatheory in Social Science: Pluralisms and Subjectivities*: 108-135. Chicago: University of Chicago Press.

Chen, Y. 2010. Dictionary Use and EFL Learning. A Contrastive Study of Pocket Electronic Dictionaries and Paper Dictionaries. *International Journal of Lexicography* 23(3): 275-306.

Chen, Y. 2012. Dictionary Use and Vocabulary Learning in the Context of Reading. *International Journal of Lexicography* 25(2): 216-247.

Chi, A. 2009. *Empirical Study of the Efficacy of Integrating the Teaching of Dictionary Use into a Tertiary English Curriculum in Hong Kong*. Hong Kong: Hong Kong University Press.

Craik, F.I.M. and R.S. Lockhart. 1972. Levels of Processing: A Framework for Memory Research. *Journal of Verbal Learning and Verbal Behavior* 11: 671-684.

Duff, P. 2006. Beyond Generalizability: Contextualization, Complexity, and Credibility in Applied Linguistics Research. Chalhoub-Deville, M., C. Chapelle and P. Duff (Eds.). 2006. *Inference and Generalizability in Applied Linguistics: Multiple Research Perspectives*: 65-95. Dordrecht: John Benjamins.

Dziemianko, A. 2006. *User-Friendliness of Verb Syntax in Pedagogical Dictionaries of English*. Tübingen: Max Niemeyer Verlag.

Dziemianko, A. 2010. Paper or Electronic? The Role of Dictionary Form in Language Reception, Production and the Retention of Meaning and Collocations. *International Journal of Lexicography* 23(3): 257-273.

Dziemianko, A. 2011. Does Dictionary Form Really Matter? Akasu, K. and U. Satoru, (Eds.). 2011. *ASIALEX 2011 Proceedings. Lexicography: Theoretical and Practical Perspectives*: 92-101. Kyoto: Asian Association for Lexicography.

Dziemianko, A. In press. On the Use(fulness) of Paper and Electronic Dictionaries. Granger, S. and M. Paquot (Eds.). *Electronic Lexicography*. Oxford: Oxford University Press.

Gass, S., A. Mackey and L. Ross-Feldman. 2011. Task-Based Interactions in Classroom and Laboratory Settings. *Language Learning* 61(Suppl. 1): 189-220.

Gast, D.L. (Ed.). 2009. *Single Subject Research Methodology in Behavioral Sciences*. New York: Routledge.

Gouws, R. 2003. Types of Articles, their Structure and Different Types of Lemmata. Van Sterkenburg, P. (Ed.). 2003. *A Practical Guide to Lexicography*: 34-43. Amsterdam: John Benjamins.

Greenbaum, S., C.F. Meyer and J. Taylor. 1984. The Image of the Dictionary for American College Students. *Dictionaries*: *The Journal of the Dictionary Society of North America* 6: 31-52.

Hartmann, R.R.K. 1987. Four Perspectives on Dictionary Use: A Critical Review of Research Methods. Cowie, A.P. (Ed.). 1987. *The Dictionary and the Language Learner: Papers from the EURALEX Seminar at the University of Leeds, 1-3 April 1985*: 11-28. Tübingen: Max Niemeyer Verlag.

Horst, M.E. 1995. Who Get Summit Mark in the Class? Productive Use of Definitions and Examples in Beginning Learners of English. *Paper presented at TESOL Arabia, Al-Ayn, United Arab Emirates*.

Kobayashi, C. 2007. Comparing Electronic and Printed Dictionaries: Their Effects on Lexical Processing Strategy Use, Word Retention and Reading Comprehension. Bradford-Watts, K. (Ed.).

2007. *JALT2006 Conference Proceedings*: 657-671. Tokyo: JALT.

Koyama, T. and O. Takeuchi. 2003. Printed Dictionaries vs. Electronic Dictionaries: A Pilot Study on How Japanese EFL Learners Differ in Using Dictionaries. *Language Education and Technology* 40: 61-79.

Koyama, T. and O. Takeuchi. 2004. Comparing Electronic and Printed Dictionaries: How the Difference Affected EFL Learning. *JACET Bulletin* 38: 33-46.

Koyama, T. and O. Takeuchi. 2007. Does Look-Up Frequency Help Reading Comprehension of EFL Learners? Two Empirical Studies of Electronic Dictionaries. *CALICO Journal* 25(1): 110-125.

Kwary, D.A. 2012. Adaptive Hypermedia and User-Oriented Data for Online Dictionaries: A Case Study on an English Dictionary of Finance for Indonesian Students. *International Journal of Lexicography* 25(1): 30-49.

Lew, R. 2002. Questionnaires in Dictionary Use Research: A Reexamination. Braasch, A. and C. Povlsen (Eds.). 2002. *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 12-17, 2002, Vol.1*: 267-271. Copenhagen: Centre for Sprogteknologi, Copenhagen University.

Lew, R. 2010a. Multimodal Lexicography: The Representation of Meaning in Electronic Dictionaries. *Lexikos* 20: 290-306.

Lew, R. 2010b. Users Take Shortcuts: Navigating Dictionary Entries. Dykstra, A. and T. Schoonheim (Eds.). 2010. *Proceedings of the XIV Euralex International Congress*: 1121-1132. Ljouwert: Afûk.

Lew, R. 2011. Online Dictionaries of English. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011. *E-Lexicography: The Internet, Digital Initiatives and Lexicography*: 230-250. London and New York: Continuum.

Lew, R. In press. How Can We Make Electronic Dictionaries More Effective? Granger, S. and M. Paquot (Eds.). In press. *Electronic Lexicography*. Oxford: Oxford University Press.

Lew, R. and J. Doroszewska. 2009. Electronic Dictionary Entries with Animated Pictures: Lookup Preferences and Word Retention. *International Journal of Lexicography* 22(3): 239-257.

Lew, R. and A. Dziemianko. 2006. A New Type of Folk-Inspired Definition in English Monolingual Learners' Dictionaries and its Usefulness for Conveying Syntactic Information. *International Journal of Lexicography* 19(3): 225-242.

Lew, R. and P. Tokarek. 2010. Entry Menus in Bilingual Electronic Dictionaries. Granger, S. and M. Paquot (Eds.). 2010. *Elexicography in the 21st Century: New Challenges, New Applications*: 193-202. Louvain-la-Neuve: Cahiers du CENTAL.

Lindsay, R.M. 1990. An Examination of the 'Negative Result' Bias: Some Empirical Evidence and a Statistical Redress. *Paper read at Canadian Academic Accounting Conference, Victoria*.

Lindsay, R.M. and A.S.C. Ehrenberg. 1993. The Design of Replicated Studies. *The American Statistician* 47(3): 217-228.

Macaulay, R. 2003. Repeat after Me: The Value of Replication. *International Journal of English Studies* 3(1): 77-92.

Mayor, M. (Ed.). 2009[5]. *Longman Dictionary of Contemporary English*. Harlow: Longman. *http://www.ldoceonline.com*. [LDOCE5]

McCreary, D.R. 2002. American Freshmen and English Dictionaries: I Had Aspersions of Becoming an English Teacher. *International Journal of Lexicography* 15(3): 181-205.

McCreary, D.R. and E. Amacker. 2006. Experimental Research on College Students' Usage of Two Dictionaries: A Comparison of the *Merriam-Webster Collegiate Dictionary* and the *Macmillan English Dictionary for Advanced Learners*. Corino, E., C. Marello and C. Onesti (Eds.). 2006. *Atti del XII Internazionale Congresso di Lessicografia, Torino, 6-9 Settembre 2006*. 871-885. Allessandria: Edizioni dell'Orso.

McCreary, D.R. and F.T. Dolezal. 1999. A Study of Dictionary Use by ESL Students in an American University. *International Journal of Lexicography* 12(2): 107-146.

Müller-Spitzer, C., A. Koplenig and A. Töpel. 2011. What Makes a Good Online Dictionary? Empirical Insights from an Interdisciplinary Research Project. Kosem, I. and K. Kosem (Eds.). 2011. *Electronic Lexicography in the 21st Century: New Applications for New Users. Proceedings of eLex 2011*: 203-208. Ljubljana: Trojina, Institute for Applied Slovene Studies.

Nesi, H. 1999. A User's Guide to Electronic Dictionaries for Language Learners. *International Journal of Lexicography* 12(1): 55-66.

Nesi, H. 2000. On Screen or in Print? Students' Use of a Learner's Dictionary on CD-ROM and in Book Form. Howarth, P. and R. Herington (Eds.). 2000. *EAP Learning Technologies*: 106-114. Leeds: University Press.

Nesi, H. In press. The Effect of e-Dictionary Font on Vocabulary Retention. *Paper read at eLEX2011: Electronic Lexicography in the 21st Century: New Applications for New Users, Bled, Slovenia.*

Nesi, H. and P. Meara. 1991. How Using Dictionaries Affects Performance in Multiple Choice EFL Tests. *Reading in a Foreign Language* 8(1): 631-643.

Nesi, H. and K.H. Tan. 2011. The Effect of Menus and Signposting on the Speed and Accuracy of Sense Selection. *International Journal of Lexicography* 24(1): 79-96.

Nielsen S. and L. Mourier. 2005. Internet Accounting Dictionaries: Present Solutions and Future Opportunities. *Hermes, Journal of Linguistics* 34: 83-116.

Osaki, S. and N. Nakayama. 2004. Denshijisyo vs. Insatujisyo: Yuyosei to Jikkousei no Chigainituiteno Kosatu [Electronic Dictionary vs. Paper Dictionary: A Comparison of the Two Dictionary Studies]. *Bulletin of Tokyo Denki University, Arts and Sciences* 2: 77-83.

Osaki, S., N. Ochiai, T. Iso and K. Aizawa. 2003. Electronic Dictionary vs. Printed Dictionary: Accessing the Appropriate Meaning, Reading Comprehension and Retention. Murata, M., S. Yamada and Y. Tono (Eds.). 2003. *Proceedings of ASIALEX 2003, Dictionaries and Language Learning: How can Dictionaries Help Human and Machine Learning?*: 205-212. Tokyo: Asialex.

Park, C. 2004. What Is the Value of Replicating Other Studies? *Research Evaluation* 13(3): 189-195.

Prinsloo, D., U. Heid, T. Bothma and G. Faaß. 2011. Interactive, Dynamic Electronic Dictionaries for Text Production. Kosem, I. and K. Kosem (Eds.). 2011. *Electronic Lexicography in the 21st Century: New Applications for New Users. Proceedings of eLex 2011*: 215-220. Ljubljana: Trojina, Institute for Applied Slovene Studies.

Quirk, R. 1974. The Image of the Dictionary. Quirk, R. (Ed.). 1974. *The Linguist and the English Language*: 148-163. London: Edward Arnold.

Rogers, M. 1996. Beyond the Dictionary: The Translator, the L2 Learner and the Computer. Anderman, G. and M. Rogers (Eds.). 1996. *Words, Words, Words. The Translator and the Language Learner*: 69-95. Clevedon: Multilingual Matters.

Santos, T. 1989. Replication in Applied Linguistics Research. *TESOL Quarterly* 23(4): 699-702.

Seidlhofer, B. 2003. *Controversies in Applied Linguistics*. Oxford: Oxford University Press.

Sidman, M. 1960. *Tactics of Scientific Research: Evaluating Experimental Data in Psychology*. New York: Basic Books.

Sinclair, J.M. (Ed.). 2008[6]. *Collins COBUILD Advanced English Dictionary*. Boston: Heinle Cengage Learning / Glasgow: HarperCollins Publishers. *http://www.myCOBUILD.com*. [COBUILD6]

Tono, Y. 2011. Application of Eye-Tracking in EFL Learners' Dictionary Look-Up Process Research. *International Journal of Lexicography*, 24(1): 124-153.

Umapathy, S. 1987. Unfavorable Variances in Budgeting: Analysis and Recommendations. Ferris, K.R. and J.L. Livingstone (Eds.). 1987. *Management Planning and Control*: 163-176. Beavercreek: Century VII.

Wehmeier, S. (Ed.). 2005[7]. *Oxford Advanced Learner's Dictionary of Current English*. Oxford: Oxford University Press. [OALDCE7]

Xu, X. 2010. Study on the Effect of Dictionary Use on Second Language Incidental Vocabulary Acquisition: An Empirical Study of College English Vocabulary Learning Strategy. *Journal of Language Teaching and Research* 1(4): 519-523.

Who can Really be Called a Lexicographer?

Rufus H. Gouws, *Department of Afrikaans and Dutch, University of Stellenbosch, Stellenbosch, South Africa (rhg@sun.ac.za)*

Abstract: Lexicographers define words but still lack a clear and unambiguous understanding of the word *lexicographer*. This paper gives a brief discussion of the problems experienced in trying to determine exactly what a lexicographer is. The distinction between theoretical and practical lexicographers is quite clear but within both these categories there are grey areas where it is not so clear whether a specific participant in lexicographic activities qualifies to be called a lexicographer. The lack of formal professional lexicographic qualifications impedes an unambiguous understanding of the word. The emergence of lexicography as an independent discipline compels lexicographers to take a closer look at criteria relevant to identifying someone as being a lexicographer.

Keywords: EMLEX, LEXICOGRAPHER, LEXICOGRAPHIC PRACTICE, LEXICOGRAPHIC PROCESS, LEXICOGRAPHIC RESEARCH, LEXICOGRAPHIC TRAINING, METALEXICOGRAPHER, PRACTICAL LEXICOGRAPHER, PRIMARY LEXICOGRAPHER, SECONDARY LEXICOGRAPHER, THEORETICAL LEXICOGRAPHER

Opsomming: Wie kan werklik 'n leksikograaf genoem word? Leksikograwe definieer woorde, maar dit ontbreek hulle steeds aan 'n ondubbelsinnige verstaan van die woord *leksikograaf*. Hierdie artikel bied 'n bondige bespreking van probleme wat ervaar word wanneer gepoog word om werklik te begryp wat 'n leksikograaf is. Die onderskeid tussen teoretiese en praktiese leksikograwe is redelik duidelik, maar binne albei hierdie kategorieë is daar grys gebiede waar dit nie sonder meer gesê kan word of 'n bepaalde deelnemer aan leksikografiese aktiwiteite wel daarvoor kwalifiseer om 'n leksikograaf genoem te word nie. Op internasionale vlak bemoeilik die gebrek aan formele professionele leksikografiese kwalifikasies 'n ondubbelsinnige begrip van die woord. Die ontluiking van leksikografie as 'n onafhanklike dissipline dwing leksikograwe om nouer ondersoek in te stel na daardie kriteria wat ter sake is om iemand as 'n leksikograaf te kan identifiseer.

Sleutelwoorde: EMLEX, LEKSIKOGRAAF, LEKSIKOGRAFIESE NAVORSING, LEKSIKOGRAFIESE OPLEIDING, LEKSIKOGRAFIESE PRAKTYK, LEKSIKOGRAFIESE PROSES, METALEKSIKOGRAAF, PRAKTIESE LEKSIKOGRAAF, PRIMÊRE LEKSIKOGRAAF, SEKONDÊRE LEKSIKOGRAAF, TEORETIESE LEKSIKOGRAAF

1. Introduction

"Who can really be called a lexicographer?" can be regarded as one of the most unlikely and even most inappropriate questions to be put to the readers of a

journal focusing exclusively on lexicographic matters. However, although the range of answers to this question will have a lot in common there will also be some differences. These differences may lead to interesting discussions.

In dictionaries a variety of definition types are used. One of these definition types is the circular definition — a type that is not applauded elsewhere but utilised in dictionaries due to its space-saving value. A typical example of the use of a circular definition in a dictionary would be to define the word *lexicographer* as "someone working within the field of lexicography". Such a circular definition is only permissible if the word *lexicography* has been included as a lemma of which the article contains a clarifying definition. This paper will not attempt to answer the question "What is lexicography?" However, in working with the premise that a lexicographer is someone involved in the field of lexicography some aspects regarding lexicography and lexicographers will be discussed.

2. The problem

A question like "Who is an engineer?"or "Who is an architect?" can easily be answered by referring to the academic and professional qualifications needed for someone to be called an engineer or an architect. The answer to a question like "Who is a gardener?" is not to be answered that unambiguously. Is it someone exclusively working in gardens, someone doing it as a hobby, someone doing it under protest but because he/she has a garden that needs to be taken care of they have to do some gardening work? Fortunately or unfortunately there are no formal internationally recognised and accepted professional qualifications that can be used to uniquely identify someone as being a lexicographer. Fortunately there are people involved in lexicographic activities: either on a fulltime basis, or as a hobby or even under protest, e.g. when a publishing house needs to produce a dictionary and someone has to do the job. Peoples' involvement in lexicography covers a wide-ranging variety of lexicographic activities. Some talk about dictionaries without ever having compiled one. Some people are involved in lexicographic activities because they develop computer programmes for dictionaries, or they do the marketing of a dictionary, or they proofread a dictionary manuscript, or they design the lay-out of a dictionary or they review dictionaries for popular or academic publications. But does this involvement in lexicographic activities qualify all of them to be called lexicographers?

In recent decades it has been the case that many lexicographers have been people with a formal training in linguistics, cf. Atkins (2002: 25) saying:

> The most significant difference, I believe, between the 1967 lexicography and that of today is that in the interval my approach to lexicography has benefited from the insights of linguistics. … Linguistic theory, particularly recent work in lexical semantics, can light the way to better lexicography.

But very seldom did this training in linguistics include any direct reference to lexicography or to dictionaries. Although training in linguistics can be a valuable aid when compiling certain types of dictionaries it is no prerequisite for being a lexicographer, and when planning and compiling e.g. a dictionary of medical and health terms very little linguistic training is put to use. It is only in the last ten to fifteen years that formal academic programmes like the MPhil/MA and DLitt/PhD in Lexicography have been offered, e.g. at the University of Stellenbosch, South Africa, and it is only since 2011 that the European Masters in Lexicography (EMLex) has been an option at a number of European universities. Students completing these courses have a formal academic qualification in lexicography — just as students completing a degree in e.g. engineering have a formal academic qualification in engineering. Do these students with an engineering qualification have to embark on a specific project to be called an engineer? Do students completing the academic programmes in lexicography have to work on a dictionary project to be called lexicographers?

A recent advertisement for the post of "Senior Editor, Dictionaries" at a South African publishing house states that the position includes "compiling, editing and proof checking dictionary products". These are lexicographic activities, albeit that the advertisement does not call this editor a lexicographer! When it comes to the requirements for the post the advertisement guides potential candidates by means of a hierarchical set of criteria, distinguishing in descending order between "requirements", "highly recommended" and "recommended". *Requirements* are e.g. indicated as "highly computer literate" and "a good understanding of editorial and production processes", *Highly recommended* is indicated as e.g. "professional experience developing or producing dictionaries" and "interest in how South Africans use dictionaries" whilst *Recommended* is indicated as e.g. "driver's licence and willingness to undertake occasional travel" and "linguistic or lexicographic training". For this position of lexicographer, lexicographic training is regarded on par with having a driver's licence and of an inferior value to being highly computer literate.

The implications of this advertisement should compel people involved in the field of lexicographic training to a process of self-reflection regarding the quality and contents of their training programmes and the needs and requirements of the market for their products, but it also gives us a good idea of what a publishing house regards as important criteria for their future lexicographers. Which combination and ordering of these criteria will be needed to ensure that the successful candidate can be called a lexicographer?

3. The factual situation

Lexicography has two major components, i.e. a theoretical component and a practical component. When being "involved in lexicographic activities" this involvement could refer to either or to both of these components. Gouws (2006) has identified four types of lexicographers:

— those working in the lexicographic practice without any theoretical experience;

— those working in theoretical lexicography without any practical experience;

— those working primarily as practical lexicographers but with some theoretical experience;

— those working primarily as theoretical lexicographers but with some practical experience.

However, such a distinction does not go beyond the initial definition of "someone working within the field of lexicography". But how are lexicographers defined by their fellow lexicographers?

4. Existing dictionary definitions

Arguably the best known definition of a lexicographer is that by Samuel Johnson (1755):

> A writer of dictionaries; a harmless drudge that busies himself in tracing the original, and detailing the signification of words.

Along with numerous other definitions the core of the definition given by Johnson lies in the fact that a lexicographer is primarily regarded as a writer of dictionaries. A much broader approach is found in the Wikipedia definition:

> A person devoted to lexicography is called a lexicographer. (*Wikipedia, the Free Encyclopedia*)

This definition implies that the writing of dictionaries is not the only criterion to qualify someone for being a lexicographer. Bergenholtz and Gouws (2012) have argued that a lexicographer could be a person with a practical or a theoretical involvement in lexicography. The question to be asked in this regard goes to the nature and the extent of this involvement.

Looking at lexicographers and their practical work it is also important to distinguish between the different types of dictionaries compiled by these lexicographers. Different sets of opposites could be taken into account. One such distinction is that between linguistic and encyclopedic dictionaries and a second distinction is that between general language dictionaries and specialised dictionaries. In one of the early typological classifications of lexicographic works, Zgusta (1971) already made provision for encyclopedias as belonging to the family of dictionaries. The need to include this type of reference work within the lexicographic family was also emphasised by Ilson (1988) when stating his policy as first editor of the *International Journal of Lexicography*:

> Our primary concern is with reference works that give lexically relevant information about lexically relevant items. But we realise that the problems facing compilers and users of dictionaries and thesauruses are similar to those facing compilers and users of indexes, encyclopaedias, atlases, and other types of reference work, and our pages are open to the discussion of their problems, too.

This statement of Ilson does not imply that the compiler of an atlas is a lexicographer but it does broaden the domain of lexicographic work, i.e. the domain of labour of the lexicographer. The definition of a lexicographer could therefore even bear reference to someone involved in the compilation of other lexical reference works comparable to dictionaries. This issue will not be discussed here. However, it is relevant to note that suggestions have already been made that the application of lexicographic theory should be extended and applied to a broader category of reference works, cf. Gouws (2011).

The second distinction, i.e. between LGP and LSP dictionaries, is also important for the interpretation of the term *lexicographer*. If it is argued that a lexicographer is someone involved in the compilation of a dictionary the reference to "dictionary" does not imply only general language dictionaries. One of the results of acknowledging lexicography as an independent discipline is that dictionaries constitute the subject matter of lexicography. Whereas linguists would be interested in the linguistic contents of general language dictionaries they have little linguistic interest in e.g. the structure of dictionaries or the contents of dictionaries dealing with languages for special purposes like dictionaries of chemistry, gene technology, statistics or psychology. Albeit true that LSP dictionaries are lexicographic products, their compilers do not have to be linguists or people with any linguistic training. Not being a linguist does not influence the status of the compilers of these dictionaries as being lexicographers.

5. Theoretical and practical lexicographers

With lexicography consisting of two components and lexicographers defined as people involved in lexicographic activities it would be a logical conclusion to say that lexicographers also fall into two groups, i.e. theoretical and practical lexicographers.

5.1 Theoretical lexicographers

People working in the field of theoretical lexicography are also lexicographers, often called metalexicographers. They are those scholars dedicating time and research endeavours to the formulation of lexicographic models, discussions of dictionary structures, lexicographic functions, the contents of dictionaries, dictionary typology, dictionary use, dictionary criticism, etc. In this regard it is important to look at the different components of theoretical lexicography, e.g. as indicated by Wiegand (1984), and to refer to participants in these different

components of theoretical lexicography as metalexicographers.

But how does the scope of the term *lexicographer* compare to that of *theoretical lexicography*? The theoretical component of the work of lexicographers can be divided into at least two major categories: doing research in and writing about lexicography and, secondly, devising course material for lexicographic programmes and teaching lexicography; in short the research and the training components of this endeavour.

There will be little doubt that people writing text books with lexicography as topic, writing papers on lexicographic topics for scientific journals, writing papers for lexicography conferences, doing research for advanced degrees in lexicography and people developing theoretical models for practical dictionary projects can be called lexicographers, more specifically metalexicographers. A question that comes to the fore is the degree and level of involvement here that is needed to qualify someone as being called a lexicographer. What kind of research outputs are required? When it comes to teaching and training, lexicography is taught as an academic subject in a number of university programmes. In the majority of cases it is done as part of a programme in language or linguistics; in only a few instances as part of a formal qualification in lexicography. In very few instances universities have a staff member appointed specifically to teach lexicography. Where this is the case and it is expected from such a staff member to teach lexicography and to supervise advanced postgraduate research but also to do research in lexicography and publish in scientific journals such a professor in lexicography will necessarily qualify to be called a metalexicographer. But when it comes to teaching there is a question as to what can be regarded as the minimum of teaching and the lowest level of teaching done to qualify the presenter as a metalexicographer? Does the teaching of a single undergraduate module dealing with a few aspects of lexicography qualify its presenter to be called a lexicographer? Is it sufficient if the teaching is restricted to dictionary using skills or should it be a more comprehensive treatment of the field of lexicography? Is teaching and training sufficient or should research and research outputs be prerequisites for someone to be called a lexicographer? These questions have not yet been answered sufficiently.

Metalexicographers can be seen as people devoting time and effort to research in the field of lexicography and also people applying this knowledge in teaching and training programmes in order to equip potential practical lexicographers with the necessary expertise to do their work. But in both the main components of research and training it is unclear what the minimum criteria are to qualify someone as a metalexicographer. For the further development of lexicography as a fully-fledged field it is necessary to formulate a clear set of criteria that can distinguish lexicographers from non-lexicographers.

5.2 Practical lexicographers

The biggest grey area when it comes to determining who a lexicographer is, lies in the domain of the lexicographic practice. Without any question someone

writing a dictionary, no matter what type of dictionary, can be called a lexicographer. In this regard it is important to have a clear indication of what a dictionary is. This is especially relevant within the field of LSP lexicography. Terminological lists and glossaries are often called dictionaries. If the compilation of such a dictionary merely demands from someone to select the relevant items to be included as lemmata and to order them alphabetically, a real question can be put as to whether he/she can be called a lexicographer. Here it is important to realise that any lexicographic process consists of different phases, cf. Wiegand (1998), and people involved in these different phases are needed to ensure the eventual production and success of the envisaged dictionary. In many dictionary projects different members of the team participate in different phases of the process. One of these phases is the selection of items to be included as lemmata in a dictionary. This is an extremely important phase and the nature and extent of the treatment allocated to these items do not make the selection less vital as a phase in the lexicographic process. The person responsible for this selection surely has to be regarded as a lexicographer. So even if the eventual dictionary is only a word list, the person responsible for the selection of the items to be included as lemmata and who compiles that lexicographic product qualifies to be called a lexicographer, even if he/she merely executes a dictionary plan compiled by a theoretical lexicographer.

One of the implications of lexicography being regarded as an independent discipline and not a subdiscipline of linguistics is the fact that LSP dictionaries fall within the scope of lexicography, albeit that they do not fall within the scope of linguistics. In recent years LSP dictionaries have often been the result of a team effort in which one or more subject field experts combine with one or more established lexicographers, e.g. the recently published *Woordeboek vir die Gesondheidswetenskappe/Dictionary for the Health Sciences* (Lochner 2011). In these cases the established lexicographer will devise the plan for the dictionary. The subject field expert will supply the subject specific data to be included in the dictionary. Such a subject field expert primarily functions within his/her field and is not primarily a lexicographer. While working together with the established lexicographer to compile a dictionary he/she supplies the relevant subject matter for the specific dictionary and by doing so also becomes a lexicographer — but in a different way compared to the established lexicographer. As a lexicographer he/she contributes in terms of dictionary contents selected for a given target group but usually does not contribute to the working out of the plan of the dictionary or in presenting the data in order to satisfy the envisaged functions of the dictionary.

In some dictionary projects individual team members are only responsible for specific data types. One person would take the responsibility for data regarding pronunciation, another for etymological data, a third for morphological data, etc. Each member of the team contributes to the writing of the dictionary but can each one be called a lexicographer? Someone exclusively responsible for etymology gets the assignment on account of his/her expertise as an etymologist. He/she is not responsible for the lexicographic treatment of

a given word but only to give its etymology — so that it can be presented as an item in a dictionary article. Is he/she an etymologist, a lexicographer, an etymological lexicographer or a lexicographical etymologist?

In this regard it seems as if a distinction can be made between primary and secondary lexicographers. This might imply that there is a hierarchy of some being "more" or "less" lexicographers than others.

The editorial team of many dictionaries also include members that are not directly involved in the process of writing the dictionary. One such member is the computer specialist who designs the programme for the dictionary. Such a programme is specifically designed for a given dictionary. It forms an integral part of the lexicographic process. This team member should also be regarded as lexicographer due to his/her involvement in the lexicographic process but he/she could be named more clearly as a computational lexicographer or even a lexicographical computer specialist.

Where dictionaries include pictorial illustrations in the articles artists are usually commissioned to make the drawings or to design the pictures. These artists do not need any lexicographic skills in order to do their duty. They are asked to make drawings of e.g. a tortoise, cell phone or space shuttle. Whether it is for a dictionary or a comic does not really matter to them. The lexicographer will give them directions regarding the focus, dimensions, etc. of the pictures to be drawn. Although it is important that the picture should satisfy the criteria set by the lexicographer in order to fulfil its function as a dictionary entry, the artist merely executes the orders without bringing lexicographic criteria into play. The artist plays a role in the preparation of the data to be included in the dictionary but does not qualify to be called a lexicographer because their work is not of a lexicographic nature although it is done for inclusion in a dictionary.

The proofreaders of a dictionary also play an important role in the success of the final outcome. They need to be familiar with the editorial system of the dictionary in order to execute their proofreading successfully. However, the work they do remains proofreading and not lexicographic work. Therefore the nature and extent of their role in the lexicographic process do not qualify them to be called lexicographers.

The grey area in the field of lexicographic practice is not as grey as it might seem. Those participants in the lexicographic process whose work demands a form of lexicographic training are lexicographers; those who perform work for the dictionary but not work that qualifies as a lexicographic endeavour should not be called lexicographers. However, lexicographers need to formulate criteria according to which one can unambiguously identify someone as being a lexicographer or not.

6. In conclusion

Those members of homo sapiens who carry the title "lexicographer" constitute

a small and intriguing category. However, even for a lexicographer it is not always as easy to say where exactly the boundaries of this group of people are. There is a core of people involved in lexicographic activities that qualify in the minds of most people to be called lexicographers. This core, including meta-lexicographers and practical lexicographers (some primary and others secondary lexicographers) is surrounded by grey areas growing darker and darker and it is not always clear who can still be called a lexicographer and who not. And it does not help to look in a dictionary. Lexicographers need to pay more attention to the formulation of criteria that can help with a clear identification of lexicographers.

Will all real lexicographers please stand up?

References

Atkins, B.T.S. 2002. Then and Now: Competence and Performance in 35 Years of Lexicography. Braasch, A. and C. Povlsen (Eds.). 2002. *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 13–17, 2002*: 1-28. Copenhagen: Københavns Universitet, Center for Sprogteknologi. (18:2008)

Bergenholtz, H. and R.H. Gouws. 2012. Who is a Lexicographer? *Lexicon* 42: 68-78.

Gouws, R.H. 2006. Leerdergedrewe leksikografie. *Tydskrif vir Taalonderrig* 40(1): 72-84.

Gouws, R.H. 2011. Learning, Unlearning and Innovation in the Planning of Electronic Dictionaries. Fuertes-Olivera, Pedro A. and H. Bergenholtz (Eds.). *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 17-29. London/New York: Continuum.

Ilson, R. 1988. Introduction. *International Journal of Lexicography* 1(1): ii.

Johnson, Samuel. 1755. *A Dictionary of the English Language*. London: J. & P. Knapton/T. & T. Longman et al.

Lochner, J. de V. 2011. *Woordeboek vir die Gesondheidswetenskappe/Dictionary for the Health Sciences.* Cape Town: Pharos.

Wiegand, H.E. 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983*: 13-30. Lexicographica. Series Maior 1. Tübingen: Max Niemeyer Verlag.

Wiegand, H.E. 1998. *Wörterbuchforschung.* Berlin: Walter de Gruyter.

Wikipedia. The Free Encyclopedia. www.yourdictionaries.com.

Zgusta, L. 1971. *Manual of Lexicography.* The Hague: Mouton.

Challenges of Predictability and Consistency in the First Comprehensive Sotho Dictionary

Inge Kosch, *Department of African Languages, University of South Africa, Pretoria, South Africa (koschim@unisa.ac.za)*

Abstract: Predictability and consistency are requirements that should run like a golden thread through the macro-, medio- as well as the microstructure of dictionary articles. Adherence to these requirements is one of the marks of a user-friendly reference work that will allow for easy access and trouble-free retrieval of required information. This article aims to investigate some features of Endemann's (1911) *Wörterbuch der Sotho-Sprache* (Dictionary of the Sotho language) with the focus on challenges of predictability and consistency in the lemmatization approach, the access alphabet, cross references and article treatments. The dictionary has hitherto remained at the outskirts of scholarly investigation, one of the reasons being the fact that the target language is German and that, as such, it is not readily accessible to every scholar of the Bantu languages. A further reason is that it is aimed at subject specialists rather than the general public and places a high demand on the lexicographic skill of the user. Some interesting insights can be gained from this example of an early attempt at dictionary compilation and its shortcomings can serve as a springboard for continual improvement of access routes and user-friendliness of dictionaries in the Sotho languages.

Keywords: PREDICTABILITY, CONSISTENCY, COMPREHENSIVE DICTIONARY, SOTHO LANGUAGES, DICTIONARY USER, LEMMATIZATION STRATEGY, CROSS REFERENCES, ALPHABETIZATION, WORD LEMMATIZATION, STEM LEMMATIZATION, DATA CATEGORIES

Opsomming: Uitdagings van Voorspelbaarheid en Konsekwentheid in die Eerste Omvattende Sothowoordeboek. Voorspelbaarheid en konsekwentheid is vereistes wat soos 'n goue draad deur die makro-, medio-, sowel as die mikrostruktuur van woordeboekartikels behoort te loop. Nakoming van hierdie vereistes is een van die kenmerke van 'n gebruikersvriendelike naslaanbron wat maklike en probleemvrye toegang tot verlangde inligting sal verleen. Die doelwit van hierdie artikel is om enkele eienskappe van Endemann (1911) se *Wörterbuch der Sotho-Sprache* (Woordeboek van die Sothotaal) te ondersoek met die klem op uitdagings betreffende voorspelbaarheid en konsekwentheid in die lemmatiseringsbenadering, die toegangsalfabet, kruisverwysings en beskrywing van artikels. Die woordeboek lê buite die kollig van wetenskaplike navorsing, een van die redes synde die feit dat die teikentaal Duits is en dat dit as sulks nie geredelik toeganklik is vir elke Bantoetaalkundige nie. 'n Verdere rede is dat dit gemik is op vakkundiges eerder as die gewone publiek en hoë vereistes stel aan die

naslaanvaardighede van die gebruiker. Hierdie vroeë poging tot woordeboeksamestelling bied 'n aantal interessante insigte en die tekortkominge kan op hul beurt weer dien as vertrekpunt vir voortdurende verbetering van toegangroetes en gebruikersvriendelikheid van woordeboeke in die Sothotale.

Sleutelwoorde: VOORSPELBAARHEID, KONSEKWENTHEID, OMVATTENDE WOORDEBOEK, SOTHOTALE, WOORDEBOEKGEBRUIKER, LEMMATISERINGSTRATEGIE, KRUISVERWYSINGS, ALFABETISERING, WOORDLEMMATISERING, STAMLEMMATISERING, DATAKATEGORIEË

1. Introduction

The aim of this article is to focus on some problems of predictability and consistency with reference to the *Wörterbuch der Sothosprache*, compiled by the Berlin missionary Karl Endemann (1911). The work can be regarded as the first attempt at a comprehensive dictionary, because the term 'Sotho' in the title does not refer to one language only, but to three, namely Northern Sotho (also known as Sesotho sa Leboa or Sepedi), Tswana (Setswana) and Southern Sotho (Sesotho), including their dialects where such are distinguished (that is, in Northern Sotho and Tswana). These languages constitute the Sotho group of languages and are spoken in South Africa and parts of southern Africa. The dictionary is mono-directional with Sotho (as defined above) as the source language and German as the target language. The examples chosen for discussion have been extracted randomly from the dictionary, but with the express purpose of raising some interesting points that can lead to a continual improvement of an optimal retrieval experience for the user.

Defining a dictionary simply as a book containing the words of a language arranged in alphabetical order and describing their meanings, underestimates the complexities associated with dictionary compilation in the Bantu languages. In the first instance, the so-called words are not all words in the sense of autonomous words with independent meanings — in fact a large number of the dictionary entries (or lemmata) are items which are smaller than words and only acquire a 'meaning' or rather 'function' by virtue of their combination with other linguistic elements. Secondly the access alphabet may differ from the traditional alphabet and either display gaps or a proliferation of more article stretches than the 26 letters provided for in the traditional alphabet. Where clusters of two or three letters serve as article stretches, a user may not find a particular word in its expected alphabetic slot. In their *Comprehensive Northern Sotho Dictionary*, Ziervogel and Mokgokong (1975) make extensive use of such clusters, but to the dictionary user, as rightly pointed out by Prinsloo and De Schryver (1999: 261), "it is nothing more than sheer frustration to eventually find for example a word commencing on *bj* alphabetically after *bu* in the dictionary". The reason for this apparent anomaly lies in the fact that the cluster "BJ" occurs as an alphabetical stretch after the letter "B". These are just some of

the unexpected features a first-time user of a Bantu language dictionary will have to contend with.

The dictionary under investigation, though rich in information, did not impact lexicographic practices in South Africa to any noteworthy degree. This situation can in part be ascribed to the fact that it was accessible only to a select number of target users, namely native speakers of German. It also had limited success as a learner's dictionary as it propagated a new, and, in the author's opinion, more scientific orthography and was intended rather for the subject specialist than for the beginner.

Predictability relates to the user being able to find the information where he or she expects to find it. **Consistency** implies that a predetermined pattern is followed in the way the data is presented. These two elements are essential parts of the dictionary conceptualisation plan or style manual. As pointed out by Gouws and Prinsloo (2005b: 16), the result of the lack of a properly planned microstructural programme leads to "the situation where the lexicographers decide in a haphazard way to include a certain data category in a specific article and omit it from the next". They also state that "one of the most annoying experiences in the process of dictionary consultation is to be confronted with an inconsistent presentation of data" (Gouws and Prinsloo 2005b: 9). A random approach to the microstructure with an arbitrary presentation of data negatively impacts the quality and status of the dictionary as an authoritative and accessible source of information (Gouws 2003: 109).

2. Lemmatization strategies

The agglutinative nature of the Bantu languages has led lexicographers to adopt either a *stem* approach or a *word* approach to the lemmatization of words. In *stem lemmatization* words are alphabetized on the first letter of the stem, which means the prefix, if there is an overt prefix, is disregarded. In *word lemmatization* words are alphabetized on the first letter of the word, which would be the first letter of the prefix, if the word has a prefix. Some dictionaries exclusively make use of stem lemmatization for all word categories (e.g. Endemann 1911, Ziervogel and Mokgokong 1975), whereas others make use of both approaches, depending on the word category, e.g. the stem approach for verbs and the word approach for nouns (e.g. Kriel 1983, de Schryver 2007). For a detailed discussion of the difference between stem lemmatization and word lemmatization and the implications they have on user-friendliness, the reader is referred to Prinsloo (2009).

Endemann followed a stem approach like Ziervogel and Mokgokong (1975), but with two major differences (Kosch 2011: 115). Firstly, in the case of nouns, the prefix is not indicated to the right of the stem, but to the left, separated from it by means of a hyphen, for example, under the letter "A" we find the following derivations from *aga* 'build' (Endemann 1911: 55):

(1) *mo-àχi*, Pl. *va-*,[1] "Erbauer, Ansiedler, Bewohner".

↓

se-àχi, Pl. *li-*, "Maurer, der viel und gut baut".

↓

aχiŏa, Pass. v. *aχa*.

↓

aχiša, Kaus. v. *aχa*, "bauen helfen, siedeln helfen, benachbart sein".

The entries are arranged in a way that is appealing to the eye, because the first letter of the stems ("A" in this case) is aligned vertically (as indicated by arrows for the purpose of this discussion). This type of presentation has been referred to as the left-expanded article structure. The benefits of this approach are expressed as follows by Gouws and Prinsloo (2005a: 38):

> Giving the complete orthographic words, in spite of a procedure of stem lemmatisation, by including the prefixes in the article-initial position is user-friendly and reassuring, especially to the inexperienced user [knowing] that (s)he is dealing with the right word.

Secondly, the majority of entries are presented as main lemmata. Lemmata which belong together semantically and structurally are thus not presented as sub-lemmata under one lemma sign. For example, *agela* 'build for', *agelela* 'fence in, build up' and *agelelana* 'fence in for each other', are all accorded their own lemma status, just like *aga* 'build', from which they are derived.

The left-expanded article arrangement may have its advantages in that it allows the user to recognize full orthographic words at one glance, but it also presents a challenge. It presupposes knowledge of the prefix-truncation rule, whereby the user needs to be able to identify the prefix of a word (e.g. *mo-* in *moagi* 'builder'), then to remove it successfully and to recognise that the letter under which *moagi* will have to be looked up is "A". Besides this, the user is required to know the morphophonological rules of the language. For example, should he/she want to look up the word *mmelegi* (Endemann: *mmeleχi*) 'child bearer/burden bearer' (from *-belega* 'give birth, bear, carry on back'), he/she would have to know that the prefix *mo-* followed by a verb stem commencing on *b*, changes to *mm* (*mo-b* > *mm*), hence *mmelegi* (< *mo-belegi*, Endemann: *mo-veleχi*). The word would hence have to be looked up under the first letter of the stem, in this case "B" for *-belegi* (or "V" for *-veleχi* according to Endemann's access alphabet as will be explained in section 4). It is clear that lemmatization according to the first letter of the stem is not user-friendly in all cases and requires a higher level of access skill, since "such stems hardly come naturally into the minds of the users, when they look up the words in a dictionary" (Kiango 2005: 269).

3. Cross references

In order to indicate the relationship between lemmata, Endemann used various mechanisms to link the entries either to previous or following entries. Space does not allow for an in-depth discussion of these, but example (2) should serve as an illustration: The relationship of *agela* to *aga* is indicated by means of the description 'Direkt. v. *aχa*' (Directive/Applicative of *aga*). In other cases the derivational relationship of one entry to the next is indicated by means of the marker 'Davon…' (i.e. from whence …), e.g. (Endemann 1911: 55):

(2) *aχa*, V. tr., "bauen; Gebäude, Gehege aufführen; aufmauern; sich ansiedeln; horsten". …

aχela, **Direkt. v. *aχa***, "bauen für …, bebauen". **Davon**

aχelela, Direkt., "einfriedigen (Garten, Hof usw.), umhegen" … **Davon**

aχelelana, Rezipr., "im freundlichen Verkehr mit einander sein" (wörtl. "einander umbauen, umhegen").

Space should always be used judiciously in the layout of a dictionary. There is no denying the fact that the many cross references in the dictionary under review take up quite some space. However, given the lemmatization strategy that is adopted here, cross referencing cannot be done away with and is in the interest of users as it alerts them to grammatical and semantic relationships between entries. Endemann made extensive use of cross references of the type illustrated in example (2) above, especially in the case of verbs. In the case of deverbatives, however, the relationship with the basic lemma is not always indicated, e.g. for *moagi* 'builder', *seagi* 'expert builder' and *moagiši* 'assistant builder'[2] there are no cross references to *aga.* One could argue that Endemann may have regarded cross references in cases like *moagi*, *seagi* and *moagiši* to *aga* as obsolete, because these lemmata appear in close proximity and in the same article stretch as *aga*. Nevertheless, it would have been good pedagogical practice to be consistent with the cross references, especially in cases where the related words are somewhat removed from each other by unrelated lemmata and occupy another slot in the article stretch, where they belong alphabetically, but not semantically.

Where a related entry is accommodated in a different article stretch altogether as a result of morphophonological changes, the user could expect a cross-reference to the word it derives from, but this is not always the case. In *kago* 'building', for example, which appears under the letter "K" and *ikagela* 'build for oneself' which appears under "I", cross references have been given to *aga* and *agela* respectively. However, for *tefo* 'payment' under "T", there is no cross reference to *lefa* 'pay'. Gouws and Prinsloo (2010) rightly observe that alphabetical ordering for the African languages has serious detrimental conse-

quences for grammatical relations. Without cross references, users will be the poorer for their look-up activity.

4. Access alphabet

Regrettably, users very seldom read (still less memorize) the instructions for use (Svensén 1993: 230), when consulting a dictionary. Instead, they go straight to the central list, where they anticipate finding items in strict alphabetical order. It is only when the search proves unsuccessful — either because there are inexplicable gaps in the alphabet or because the words are not found where the users expected to find them — that they might reluctantly resort to the front matter of the dictionary for specific conventions and an explanation regarding the ordering of lemmata. Users of Endemann's dictionary will only grasp the format of the access alphabet once they understand that the orthography was designed to be as close to a phonetic orthography as possible. The reason why the letters 'B', 'D' and 'G', for example, are absent from the access alphabet, is because Endemann did not regard the sounds these orthographic symbols are used for in a phonetic alphabet, as part of the Sotho sound inventory, namely the bilabial plosive stop [b], the apico-alveolar plosive stop [d] and the dorso-velar plosive stop [g] respectively. He ascribed the incorrect usage of these symbols to ignorance of phonetics, e.g. with regard to the letter B, he declares: "Die bisher übliche Schreibung *b* statt ʋ im Sotho und Tžŏana beruht auf Unkenntnis bezüglich der Phonetik" ("The customary spelling of *b* instead of ʋ in Sotho and Tswana used up to now is based on ignorance regarding the phonetics") (Endemann 1911: 3). He proposed other symbols to reflect the correct pronunciation of these three sounds respectively, namely the Greek letter ν (V), an l with a dot as a diacritic under the symbol ɬ (Ḷ) and the Greek letter χ (X).

Standard alphabet:	Endemann:
B, b [b]: bilabial plosive stop	V, ν [β]: bilabial fricative continuant
D, d [d]: apico-alveolar plosive stop	Ḷ, ɬ [ɻ]: palato-alveolar flap (retroflex)
G, g [g]: dorso-velar plosive stop	X, χ [ɣ, x]: dorso-velar fricative continuant[3]

From a scientific point of view Endemann's propagation of new symbols may have been a laudable attempt to enforce correct pronunciations, but it also diminished the appeal of the dictionary as a useful learning tool to its potential users. Especially the beginner, oblivious to finer nuances of pronunciation, would probably not have searched for an entry like *-belega* 'give birth' under the letter 'V' but rather under 'B', only to find this letter missing completely from the access alphabet.

Table 1 presents Endemann's access alphabet (lower case letters have been added, because they differ in type in some cases from the capital letters), com-

pared to the traditional alphabet. The rightmost representations in phonetic script indicate all the initial sounds of lemmata accommodated under each letter of the access alphabet (in broad phonetic script, i.e. without indication of ejection, vowel raising, etc.).

Table 1: Access alphabet in Endemann (1911)

Traditional alphabet	**Endemann's access alphabet**	**Initial sounds of lemmata accommodated under letter of access alphabet**
A	A/ɑ	[a]
B	–	
C	–	
D	–	
E	E/e	[e, ɛ, j[4]]
F	F/ƒ	[ɸ, ɸs, ɸʃ]
G	–	
H	H/h[5]	[ɦ, h]
I	I/i	[i]
J	J/j	[dʒ]
K	K/k	[k, kh, kxh]
L	L/l L̩y/l̩y	[l, ɹ] [ʒ]
M	M/m	[m]
N	N/n N̰/n̰ Ṅ/ṅ	[n, ɲ] [ŋ‖] [ŋ]
O	O/o	[o, ɔ, w[6]]
P	P/p	[p, ph, ps, psh, pʃ, pʃh]
Q	–	
R	R/r	[r]
S	S/s Š/š S̩/s̩	[s] [ʃ] [ɬ]
T	T, Th/t, th T̩, T̩h/t̩, t̩h	[t, th, ts, tsh, tʃ, tʃh] [tl, tlh]
U	U/u	[u, w[7]]

V	V/v	[β, βʒ]
W	–	
X	X/χ X̌/χ̌	[x, ɣ] [ɦ]
Y	–	
Z[8]	Z/z Ž/ž	(dial.: only in one example) (dial.: only in one example)

Endemann only made limited use of clusters of sounds as article stretches, unlike Ziervogel and Mokgokong (1975) who used many digraphs (combinations of two symbols) and trigraphs (combinations of three symbols) in their dictionary. To amplify the symbols Endemann added diacritics like a tilde or a dot. These peculiar combinations as well as the absence of certain letters render the access alphabet unpredictable. Endemann compiled the dictionary according to what he thought would be the most suitable for his intended users, but as shown in Table 1, the access alphabet is not necessarily logical to the user. In compiling dictionaries, lexicographers should therefore be wary of serving technical and scientific considerations above the users' interests.

In two cases Endemann gives two symbols as an article stretch, namely 'T', 'Th' and ' T̲', 'T̲h'. This is a confusing notation, leaving the user to anticipate one of two possibilities: either that the lemmata with the unaspirated initial consonants will be treated exhaustively first, followed by the lemmata beginning with the aspirated consonants, or that the stems with aspirated plosives as initial sounds are integrated alphabetically with the rest of the stems. The latter is indeed the case, which means that an entry such as *taelo̲* 'command' occurs before *thatano̲* 'mutual love', which in turn will precede *timelo̲* 'loss'. One wonders why two symbols for an article stretch are only given in the case of 'T', 'TH' and ' T̲', 'T̲H' and not in some of the other article stretches where the same principle would apply.

5. Data categories

In each language the data dictate the microstructure, but for each article there should at least be an obligatory microstructure and the ordering of the data types should be fixed (e.g. morphological data, part of speech, translation, usage in cotext, etc.). Endemann did not seem to adhere to a consistent pattern as far as the microstructural components are concerned. The article treatments rather suggest that he improvised as he compiled the dictionary, dictated by the nature of the material that presented itself to him. He was led by intuition in his treatment of each article as to what information to include and what to omit — at least one can accept that he did not record non-attested forms under the urge of completing paradigms. No phonetic transcription is given — the very purpose of the orthography that Endemann proposed, was that the pronunciation would be correctly derivable from the

practical orthography, making a phonetic transcription obsolete.

The grid in Table 2 below captures the data categories encountered in the dictionary for the treatment of the lemmata. Not all slots are consistently filled for every entry: some entries of which the meaning is transparent, have limited treatment and some articles only supply the minimum information such as a translation equivalent, for example:

(3) p. 53 *vo-a̰t̰à* "Torheit" (i.e. negligence, stupidity)

In many cases the relevant sense of a lemma sign is not explained by means of its usage in a cotext; in other cases lengthy explanations, personal experiences, opinions or a word of caution regarding correct pronunciation are recorded as first item under an article stretch, e.g. under the letter "R":

Endemann (1911: 418):

> Das *r* des Sotho darf nur alveolar gesprochen werden. Der Europäer der sein *r* velar spricht, muss sich hier Mühe geben, die richtige Aussprache zu lernen.
>
> The *r* of Sotho may only be pronounced as an alveolar. The European who pronounces his *r* in a velar manner, must make an effort to acquire the correct pronunciation. (own translation)

The numbering in the following table corresponds with the paragraphs below the table where each data category is discussed in more detail.

Table 2: Basic indicators of articles

5.1 Origin/source of words
5.2 Part of speech and grammatical data (may include notes on spelling or pronunciation)
5.3 Semantic data (Translation equivalent, paraphrase on meaning, meaning relations); etymological data
5.4 Cotext (Application in cotext/proverbs/idioms); data on cultural value; pragmatic data; usage notes
5.5 Elaboration of articles (i.a. reference to other languages, personal note, reprimand or word of caution)
5.6 Sketches

5.1 Origin/source of words

The first entry after the lemma sign is invariably an indication of where the

item was attested, either in a language, dialect or particular region. This category is marked by bracketed abbreviations, which are explained in the front matter (Endemann 1911: 36, 37). For example, in (4) the word was attested among Tswana-speakers (Tž.). Sometimes more than one origin for an item is included between brackets, e.g. (Tž, Kχ for 'Tswana, Kgatla') or an "etc." is added which leaves the user somewhat perplexed, as he/she would not know what the "etc." refers to, compare example (5).

(4) p. 232 *mo-kχo̲kχo̲ma* (Tž.) "Luftröhre"
(*mokgokgoma*, attested in Tswana: "air pipe")

(5) p. 49 *aŏa̋!* (P. etc.), Interj., "nein!", *aŏa s̲e̲!* "nein doch!" (erfreut, auch spöttisch). (*aowa*, attested in Pedi, etc. Interjection "no!", *aowa hle* "no, please!" pleased, also mocking)

The author explains that where such bracketed labels occur, this does not necessarily mean that the items are not in use elsewhere. Where the lemma is deemed to occur across the board in the Sotho languages and dialects under investigation, this data category is absent as in example (6).

(6) p. 233 *vo-kχo̱le̲*, "Ferne, Entfernung" (*bokgole* i.e. distance)

The label marking the origin is thus not an essential part of every article treatment.

5.2 Part of speech and grammatical data

The indication of the part of speech is not altogether transparent or consistent. The part of speech is omitted in the case of nouns with overt class prefixes. This is a space-saving device and probably adopted because the user is expected to know that the presence of a class prefix (cf. *vo-* and *mo-* in examples (7) and (8)) automatically implies a noun. If the noun is countable, the plural class prefix will follow the lemma, compare *va-* in example (8):

(7) p. 55 *vo-aχišo̲* "Nachbarschaft". (*boagišo* 'assistance'[9])

(8) p. 55 *mo-aχiši*, Pl. ***va-***, "Nachbar". (*moagiši* 'assistant builder'[10])

In contrast to these examples, there seems to be no pattern when the class prefix is not overt, as in the case of nouns in class 9. In some cases the nominal category of the noun is marked by a capital letter N, compare example (9), but in other examples the noun is left unidentified for part of speech, compare example (10):

(9) p. 561 *thọta*, **N**., Pl. *ḭi-*, "Bug, Erdhügel, Erdwall, kleiner Hügel, steigender Boden, Hochland". Siehe *marota, morota.* (i.e. rising land, hunch or hump)

(10) p. 561 *thope*, Pl. *ɟi-*, "kurzer, platter, breiter Speer" (i.e. short, flat, broad spear)

In irregular nouns, the plural form is written in full, e.g. *vānana* (*banana* 'little girls') in (11) and *nyŏákọ* (*nywako* 'houses') in (12). Examples like these constitute a challenge to users, because they would not think of looking up *ngwanana* 'little girl' or *ngwako* 'house' under the letter "A", but under the letter "N" for *ng* (Endemann *ṅ*). In (12) the user is given cases of similar sound changes as examples, e.g. *ngwaga* from **moaga* 'year'. This was meant as an educational exercise, but it sidetracks the user and increases his/her look-up time.

(11) p. 48 *mo-anana,* gespr. *ṅŏanana*, Pl. ***vānana***, "Mädchen, weibliches Kind".

(12) p. 41 *mo-á kọ*, gespr. *ṅŏákọ*, Pl. ***nyŏákọ***, "Haus". Der Pl. *ny*. ist = *myako* aus *me-ako*; vergl. *moaχa* = *ṅŏaχa, anyŏa* f. *amya*, *enyŏa* f. *emya*.

5.3 Semantic data

A translation or explanation of the meaning is present in all article treatments or alternatively the user is referred to another form with a different spelling under which the lemma can be looked up. The meaning explanation follows the word category indicator (where present) and is indicated between double quotation marks, but not where the lemma needs to be circumscribed, e.g.

(13) p. 253 *molàlạ* (N.). Pl. *me-*, Name einer Baumart, deren Bast zu Stricken geflochten wird und von deren Saft Bier gemacht wird. (Name of a type of tree, of which the bark is plaited into cords and of which beer is made from its sap.)

There are a fair number of entries about which uncertainty exists regarding the precise meaning equivalent. This negatively affects the user's confidence because his or her consultation effort does not reach a satisfactory conclusion. The semantic gaps are probably due to the fact that Endemann prepared the dictionary for publication after his return to Germany, where he was not able to obtain missing information in the field among the Sotho speakers themselves, but had to rely only on his notes and information from some informants who were in Germany at the time, cf. example (14):

(14) p. 163 *mo-kikiana* (Tž), Pl. *me-*, Name einer hohen Baumart (näheres bietet die Quelle nicht). (Name of a tall type of tree — the source does not provide more detail).

It is evident that the final version of the dictionary should have been completed in the field for increased accuracy and completeness of information.

5.4 Cotext

Information that may be encountered in the treatment of an article is its application in a cotext (such as in idiomatic expressions or proverbs), e.g. (15).

(15) p. 54 *ma-ávola* "Anteile". *Leǰ́u ke maávola, le ávotžŏe* "Tod ist Anteil (d.h. das bestimmte, zuerteilte Los), er ist zuerteilt". (Redensart beim Tode eines Kindes).

("Shares". "Death is a share, i.e. the designated, allocated destiny), it is dished out." (Expression at the death of a child).

Such examples are regrettably in the minority, although the meaning equivalents of most lemmata are explained as they were applied in the traditional milieu. Applications in cotexts would have been a great learning tool.

5.5 Elaboration of article

Even though Endemann provides users' guidelines in the front matter, he relegated many observations to the individual entries themselves, rendering some dictionary articles disproportionately long. The dictionary articles thus often provide more than the user bargains for, and this interferes with the user's information retrieval process. As pointed out by Gouws and Prinsloo (2005b: 170): "… the lexicographer should realise that users very seldom want to read through an article. The typical dictionary consultation procedure is aimed at finding one specific item or data type". Descriptions should be short and compact. Endemann particularly falls prey to digression. In the case of the verbal particle *ke* (p. 150) "it is", for example, he takes pains to point out its wrong application based on a wrong translation in German, taking up more than one column of space.

The treatments of some lemma-signs are used as a forum to give lengthy explanations of phonological rules, rules which were evidently felt not to have been addressed adequately in the front matter. Furthermore, contrary to good lexicographic practice, the author expresses personal preferences and judgments regarding wrong perceptions and pronunciations, often as the first component of an article.

In a number of articles the compiler demonstrates his wide knowledge of other languages outside the Sotho group and one wonders how much the user gains through comparisons with other languages, compare the reference to the languages Nyamwezi, Yao and Zulu in example (16):

(16) p. 167 *kŏàla* (S.-S.), V. tr. = *tzŏàla*, "verschliessen". Vergl. Nyamw.: *lugala*, Yao: *ugala*, Zu.: *vala*.

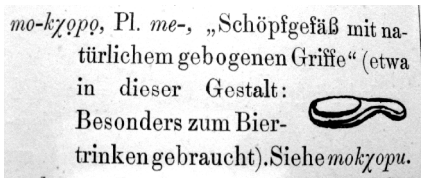
Over-completeness where too much extra-linguistic information is given, can

be counterproductive.

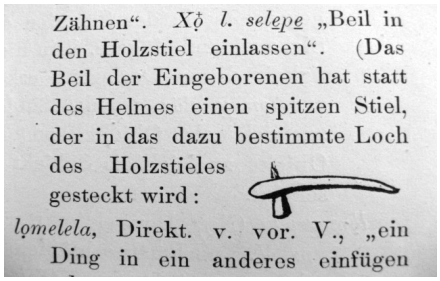
5.6 Sketches

If one decides to support certain data with illustrations, this approach should be followed consistently throughout a dictionary and not in an ad hoc way. In Endemann's dictionary, only two instances were found where a sketch formed part of the article. The sketches depict cultural objects or tools.

Sketch 1: mo-kχǫpǫ (Endemann 1911: 241)


mo-kχǫpǫ, Pl. *me-*, „Schöpfgefäß mit natürlichem gebogenen Griffe“ (etwa in dieser Gestalt: Besonders zum Biertrinken gebraucht). Siehe *mokχǫpu*.

Sketch 2: selepe (Endemann 1911: 281)


Zähnen“. *Xǫ̊ l. selepe* „Beil in den Holzstiel einlassen“. (Das Beil der Eingeborenen hat statt des Helmes einen spitzen Stiel, der in das dazu bestimmte Loch des Holzstieles gesteckt wird:
lǫmelela, Direkt. v. vor. V., „ein Ding in ein anderes einfügen

Obviously space is a restricting factor, but then again one wonders why the sketches were limited to these particular two entries. Were they just entered as an afterthought or because the description was felt to be inadequate? What about the many other lemmata in the dictionary referring to cultural objects, unknown in the culture of the target user, which could equally have benefited from a pictorial illustration for extra-linguistic information? Landau (1989: 112) mentions that "Zgusta sees the primary purpose of illustrations as that of depicting unusual or unfamiliar things".

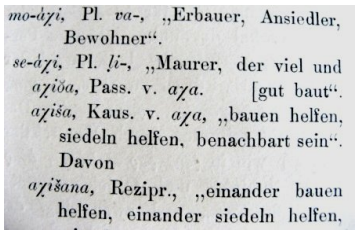
6. Space saving strategies/Visual presentation

Limitations of space and cost restrictions will always remain a major con-

sideration while lexicographers are bound to the paper dictionary. The lemmata in Endemann's dictionary are displayed in alphabetical order in two columns on a page, the columns being separated by a vertical line. In some instances the treatment of a particular lemma exceeds its allotted space within the given column and spills over into the line of the previous or next lemma. In such a case it is separated from the treatment of the other lemma by a square bracket, as can be seen in Sketch 3a, where the phrase "Maurer, der viel und gut baut" (Builder that builds much and well) spills over onto the next line. In Sketch 3b, the translation of *mo-kŏalava* "steiniger Rücken" (stony ridge) is completed in the preceding line. This may be a convenient space saving strategy on the side of the publishers, but to the user it may convey the message of poor dictionary planning and slow down his or her look-up time.

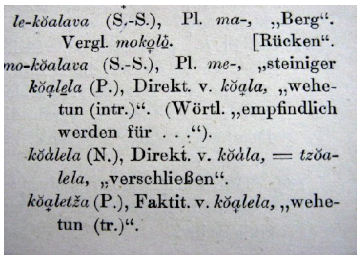
Sketch 3: Overlapping of lines (Endemann 1911: 55)

(a)



mo-áχi, Pl. *va-*, „Erbauer, Ansiedler, Bewohner".
se-áχi, Pl. *ḷi-*, „Maurer, der viel und [gut baut".
aχiŏa, Pass. v. *aχa*.
aχiša, Kaus. v. *aχa*, „bauen helfen, siedeln helfen, benachbart sein". Davon
aχišana, Rezipr., „einander bauen helfen, einander siedeln helfen,

(b)



le-kŏalava (S.-S.), Pl. *ma-*, „Berg". Vergl. *mokọlŏ*. [Rücken".
mo-kŏalava (S.-S.), Pl. *me-*, „steiniger
kŏạlẹla (P.), Direkt. v. *kŏạla*, „wehetun (intr.)". (Wörtl. „empfindlich werden für . . .").
kŏàlela (N.), Direkt. v. *kŏàla*, = *tzŏalela*, „verschließen".
kŏạletža (P.), Faktit. v. *kŏạlela*, „wehetun (tr.)".

7. Conclusion

Predictability and consistency play a crucial role in optimising the user's effective and trouble-free retrieval of information. The word approach is the most user-friendly, because it functions in a predictable way as far as the access alphabet is concerned. However, it is also the most space-consuming in a paper dictionary. At face value, Endemann's approach looks like a word approach and does not create unnatural forms, because entries can be recognised as full orthographic words with prefixes preceding their stems. However, all the entries are still lemmatized on their stems. It may be true that the left-expanded article structure boosts the confidence of users as they are able to recognise full

words at one glance, but this approach requires more than just basic reference skills. It expects the user to have an understanding of the word structure and phonological changes operating in the particular language. For any scholar, especially someone with no prior knowledge of the structure of a Bantu language, the look-up time will be slowed down significantly. Before being able to look up the meaning of words, the user would first have to consult the explanatory introduction which should form part of any good dictionary (Lombard, 1970: 12).

Using a dictionary as a forum to introduce a new orthography based on pronunciation, proved to be a deterrent to users as it complicated access to the central list. Already the first encounter with the access alphabet diminishes a user's appreciation of the work. Landau (1989: 76) aptly observes that "before a dictionary can be written for a language, the language must have developed more or less standard spellings". Differences in the data categories accommodated in the treatment of articles were observed, with some data categories absent, or discussed at length, leading to imbalanced entries.

The solution to the problem of user-friendly dictionaries for the Bantu languages, and one which dispels the greatest obstacles of the lexicographer, obviously lies in the electronic dictionary, where lemmata can be listed according to the first letter of the word in the case of nouns without space saving concerns. Illustrations, especially of cultural objects, of which sadly only two appear in Endemann's dictionary, can be of great historical value for posterity and with today's technology, these could be facilitated by pop-up functionalities from within an entry, without affecting the layout of the article in any way.

Gouws and Prinsloo (2005a: 34) mention examples of Bantu language dictionaries that employed a lemmatisation strategy resembling left-expanded article structures and that went unnoticed in South African Bantu language lexicography (e.g. the *Concise SiSwati Dictionary* by Rycroft, 1982 and the *Southern Sotho-English Dictionary* by Mabille and Dieterlin, 1988). Endemann's work could just as well have been included here, as it followed the same approach and "went unnoticed". However, it deserves recognition, despite its shortcomings as far as predictability and consistency in certain respects are concerned, because it was the very first dictionary compiled for the Sotho languages that used the left-expanded approach for lemmatisation. The shortcomings in the dictionary that were pointed out during the discussion can hopefully serve as a springboard for continual improvement of access routes and user-friendliness of dictionaries in the Sotho languages.

Endnotes

1. Examples extracted from Endemann are given in the author's orthography.
2. Endemann's meaning equivalent of *moagiši* in German is *Nachbar* "neighbour". The correct word for neighbour is *moagišani*, but the latter is not given by Endemann.

3. Regarding the voiced and voiceless variants of χ ([ɣ] and [x] respectively) Endemann did not see the need to distinguish them in writing in the practical orthography (1911:8).
4. For the palatal glide [j] Endemann uses the symbol ĕ.
5. Endemann compromises here for space saving reasons and uses the letter "H" for stems which are also encountered elsewhere in the Sotho languages with an initial "F" ([f]) or "X" ([x] or [ɣ]).
6. For the labio-velar glide [w] Endemann uses the symbol ŏ.
7. Only one example where [w] is represented as ŭ.
8. The sounds represented by "Z" result from the omission of a preceding υ or ¦ according to Endemann. The precise phonetic representation cannot be established from the information given.
9. The German translation equivalent given for *boagišo* is 'neighbourhood' instead of 'assistance'.
10. The translation equivalent in German is given as *Nachbar* which means 'neighbour'. See also endnote 2.

References

De Schryver, G.-M. (Editor-in-chief). 2007. *Oxford Bilingual School Dictionary: Northern Sotho and English/ Pukuntšu ya Polelopedi ya Sekolo. Sesotho sa Leboa le Seisimane. E gatišitšwe ke Oxford*. Cape Town: Oxford University Press Southern Africa.

Endemann, K. 1911. *Wörterbuch der Sothosprache*. Abhandlungen des Hamburgischen Kolonialinstituts, Band VII. (Reihe B. Völkerkunde, Kulturgeschichte und Sprachen Band 4). Hamburg: L. Friedrichsen & Co.

Gouws, R.H. 2003. Aspekte van mikrostrukturele verskeidenheid en inkonsekwentheid in woordeboeke. *Lexikos* 13: 92-110.

Gouws, R.H. and D.J. Prinsloo. 2005a. Left-Expanded Article Structures in Bantu with Special Reference to IsiZulu and Sepedi. *International Journal of Lexicography* 18(1): 25-46.

Gouws, R.H. and D.J. Prinsloo. 2005b. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.

Gouws, R.H. and D.J. Prinsloo. 2010. Lexicography Workshop Notes: Dictionaries — the Theory and Practice of the Lexicographic Process. Workshop held at the University of Pretoria, 10–11 May 2010.

Kiango, J.G. 2005. Problems of Citation Forms in Dictionaries of Bantu Languages. *Nordic Journal of African Studies* 14(3): 255-273.

Kosch, I. 2011. Innovation and Compromise in K. Endemann's Dictionary of the Sotho Language (1911). *South African Journal of African Languages* 31(1): 110–120.

Kriel, T.J. 1983. *Pukuntšu Woordeboek (Noord-Sotho–Afrikaans, Afrikaans–Noord-Sotho)*. Pretoria: J.L. van Schaik.

Landau, S.I. 1989. *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.

Lombard, D.P. 1970. Bantoetaalstudie: Dictionaries in Northern Sotho. *Bantu Education Journal* 16(3): 12-13.

Prinsloo, D.J. 2009. Current Lexicography Practice in Bantu with Specific Reference to the Oxford Northern Sotho School Dictionary. *International Journal of Lexicography* 22(2): 151-178.

Prinsloo, D.J. and G.-M. De Schryver. 1999. The Lemmatization of Nouns in African Languages with Special Reference to Sepedi and Cilubà. *South African Journal of African Languages* 19(4): 258-275.

Svensén, B. 1993. *Practical Lexicography: Principles and Methods of Dictionary-Making*. Oxford/New York: Oxford University Press.

Ziervogel, D. and P.C. Mokgokong. 1975. *Groot Noord-Sotho Woordeboek/Comprehensive Northern Sotho Dictionary*. Pretoria: J.L. van Schaik/Unisa.

The Role of Syntactic Class, Frequency, and Word Order in Looking up English Multi-Word Expressions

Robert Lew, *Department of Lexicography and Lexicology, Faculty of English, Adam Mickiewicz University in Poznań, Poland (rlew@amu.edu.pl)*

Abstract: Multi-word lexical units, such as compounds and idioms, are often problematic for lexicographers. Dictionaries are traditionally organized around single orthographic words, and so the question arises of where to place such complex lexical units. The user-friendly answer would be to include them primarily under the word which users are most likely to look up. But how do we know which words are likely to be looked up? The present study addresses this question by examining the roles of part of speech, word frequency, and word position in guiding the decisions of Polish learners of English as to which component word of a multi-word expression to look up in the dictionary. The degree of word frequency is found to be the strongest predictor, with less frequent words having a significantly greater chance of being selected for consultation. Then there is an independent part of speech-related preference for nouns, with adjectives being second, followed by verbs in third place. Words belonging to the remaining syntactic categories (adverbs, prepositions, conjunctions, determiners, and pronouns) are hardly looked up at all. However, word placement within the multi-word expression does not seem to matter much. This study has implications for dictionary makers in considering how to list multi-word-expressions.

Keywords: MULTI-WORD EXPRESSION, MWE, MULTI-WORD UNIT, MULTI-WORD ITEM, DICTIONARY, DICTIONARY CONSULTATION, DICTIONARY ACCESS, LEXICOGRAPHY, ENGLISH LANGUAGE, POLISH LEARNERS

Opsomming: Die rol wat sintaktiese kategorie, frekwensie en woordorde speel in die naslaan van Engelse meerwoordige uitdrukkings. Meerwoordige leksikale items, soos samestellings en idiome, is dikwels problematies vir leksikograwe. Woordeboeke word tradisioneel gerangskik om enkele ortografiese woorde, en dus ontstaan die vraag waar sulke komplekse leksikale items geplaas moet word. Die gebruikersvriendelike antwoord sou wees om hulle primêr in te sluit onder die woord wat gebruikers gewoonlik eerste sou naslaan. Maar hoe weet ons watter woorde gewoonlik nageslaan word? Die huidige studie spreek hierdie vraag aan deur die rol te ondersoek wat die woordsoort, woordfrekwensie, en woordposisie speel om Poolse aanleerders van Engels se besluite te rig oor watter woord om binne 'n meerwoordige uitdrukking in die woordeboek na te slaan. Daar is gevind dat die woordfrekwensie die sterkste voorspeller is, met minder gebruiklike woorde wat 'n beduidend groter kans het om vir naslaan gekies te word. Dan is daar 'n onafhanklike voorkeur wat verband hou met die woordsoort vir selfstandige naamwoorde, met byvoeglike naamwoorde in die tweede plek, gevolg deur werkwoorde in die derde

plek. Woorde wat behoort tot die oorblywende sintaktiese kategorieë (bywoorde, voorsetsels, voegwoorde, bepalers en voornaamwoorde) word skaars opgesoek. Woordplasing binne die meerwoordige uitdrukking skyn egter nie baie saak te maak nie. Hierdie studie het implikasies vir woordeboekmakers wanneer hulle oorweeg hoe om meerwoordige uitdrukkings te lys.

Sleutelwoorde: MEERWOORDIGE UITDRUKKING, MWE, MEERWOORDIGE EENHEID, MEERWOORDIGE ITEM, WOORDEBOEK, WOORDEBOEKRAADPLEGING, WOORDEBOEK-TOEGANG, LEKSIKOGRAFIE, ENGELSE TAAL, POOLSE AANLEERDERS

1. Introduction

At first sight it seems that dictionaries treat headwords as if users' look-up strategy is based only on single words. This implies a simplistic view of lexical items as single words, which agrees with the naive view of language, and also with the Chomskyan 'slot-and-filler' model of language[1], which itself may owe much to the impact of the structure of the (printed) dictionary on the linguist as a naive dictionary user, cf. Nowakowski (1990).

However, for describing lexical phenomena, a Sinclairian view of language may be more fitting. It emphasizes the *idiom principle*, whereby words tend to cluster into more or less fixed chunks, and such chunks often express relatively unitary meanings. On this view, the lexicographer would owe it to the user to offer a fair treatment of such *multi-word expressions* (MWEs; also *multi-word items*, *units*, or just *multi-words*) in a dictionary, giving such complex lexical items the same status as has so far been the privilege of items lexicalized in orthographically simplex words. In English, common formations of this type include noun compounds, phrasal verbs and 'idioms' in a narrower sense (the broader sense including all of the above). Multi-word sequences of the less fixed type are usually classified as collocation, which (when defined more narrowly) differs from the previously given types of multi-words in terms of (1) semantics, in that it does not typically denote a unitary concept, but rather a complex one; and (2) structure, in that it tends to be less deterministic and more flexible. A broader, distributional view of collocation might encompass all of the preceding types of word chunks, and so this view is not as helpful for lexicographers, who usually prefer working with finer categories.

The topic of the present study is multi-word expressions with more or less unitary meaning. Specifically, the question is where to place (the lexicographer's perspective) and find (the user's perspective) such items in the dictionary. If we accept that the prevailing lexicographic tradition for languages with alphabetic writing systems is to arrange mostly single-word headwords alphabetically, then we need to decide under which orthographic word one should place multi-words, assuming that the full treatment cannot be given under every single constituent word. A related question is under which lemmas a restricted (brief) treatment, usually in the form of a cross-reference, should be offered.

Dictionary compilers may choose to adopt a variety of approaches, taking into account word order (under the first content word is a frequent solution), word class (usually prioritizing nouns, sometimes verbs) and word frequency (listing multi-words under the less frequent components). Often, though, no uniform strategy is declared in the preface and none can be generalized from a mere inspection of the entries.

1.1 Previous studies on how users look up multi-words

Before user studies became mainstream, metalexicographers offered guidelines based on intuition. Careful attention is given to the issue of MWE placement in Zgusta et al. (1971: 269-270) in the form of four principles. First, multi-words should not be included in the entries for articles, prepositions and *be* as a copulative verb. Further, Zgusta et al. claim that preference should be given to component words which are semantically least clear in the context of the MWE. This principle may be hard to apply in practice, being rather subjective. The third guideline warns against prioritizing attributive elements. The final recommendation is to use the (linearly) first content word within the multi-word expression. Of these guidelines, the first and fourth have become rather popular in English lexicographic practice. A combination of the two results in a decision, sometimes mentioned in the front matter, to list multi-word expressions under the first content word.[2]

Actual user preferences in looking up multi-words have been studied by Béjoint (1981), Tono (1987), Bogaards (1990, 1991, 1992), Atkins and Knowles (1990), and Atkins and Varantola (1998). These studies will be summarized briefly below.

Béjoint (1981) investigated user look-up preferences of French students of English using a list of eight English multi-word expressions (*artificial insemination*, *boil down to*, *false alarm*, *magnetic tape*, *come down with*, *lose sight of*, *rid of*, *fountain pen*). He found that (1) learners would prefer *not* to have separate entries for compounds; (2) in nominal compounds the noun is preferred; and (3) in what Béjoint terms *verbal compounds*, French students preferred verbs over adverbs and prepositions, but in the one case of *lose sight of*, which also included a noun, there was a slight preference for the noun.

Tono (1987) investigated the headword choices of 129 Japanese learners of English looking at 62 idioms in specific syntactic patterns. Overall, Tono found a preference for (1) content words over function words; (2) less familiar words; and (3) words with more restricted combinability. As far as I am able to tell, familiarity and combinability were assessed impressionistically and only after the fact.

Bogaards (1990) compared the look-up preferences of a large sample of speakers of French and Dutch in 52 multi-words, and found fairly consistent but L1-dependent look-up strategies. French speakers appeared to have been guided by word frequency, going for the less frequent words, and then by syntactic structure, preferring superordinate (independent) to subordinate (depend-

ent) elements. In contrast, speakers of Dutch seemed to have looked primarily at part of speech, choosing nouns, and then adjectives and verbs, in this order. Bogaards (1991) and Bogaards (1992) explored in more detail the role of frequency differences in determining the choices of native speakers of French.

The EURALEX/AILA Research Project on Dictionary Use (Atkins and Knowles 1990; Atkins and Varantola 1998) does not turn up much useful data on looking up MWEs, primarily because it reports on only three items (a fourth item was found problematic and had to be discarded). We do learn, however, that look-up behaviour does not seem to vary by the L1 of the learner (French, Italian, German, and Spanish), and that the words that learners select for lookup are often not the ones at which the multi-words in question are listed in the leading monolingual learners' dictionaries.

1.2 Possible factors affecting the look-up of multi-words

Bogaards (1991: 204) lists seven factors that might potentially affect the look-up behaviour of dictionary users. These are:

1. grammatical vs content words
2. order of the words
3. word frequency
4. grammatical category
5. syntactic structure
6. semantic value
7. idiomatic character

The general picture that emerges from past studies on looking up multi-word expressions is that users appear to be guided by word frequency, grammatical category and syntactic structure or word order.

In terms of word frequency, users tend to prefer less frequent words. It is quite likely that frequency is an indirect factor, acting through the mediation of word familiarity, but the latter is harder to measure and is a personal (subjective) attribute of limited use in dictionary design. In contrast, corpus frequency is relatively easy to measure and is collective rather than subjective.

When it comes to word class, users tend not to look up closed-class words such as articles or prepositions, and prefer content words. Amongst the content words, there may be some preference for nouns.

As far as word order is concerned, strangely enough, there is not much in the way of direct reports, and Bogaards (1991: 204) dismisses it as 'fairly uninteresting', but this factor may be hard to distinguish from syntactic role. For instance, Bogaards (1990) found that in French nouns modified by adjectives,

nouns tend to be looked up, and accounted for this in terms of a preference for syntactically superordinate elements. But, in fact, since in French adjectives typically follow nouns, it is hard to judge whether the choices made are not in fact a consequence of simple linear order — that is, users picking the first content word they come across — rather than an awareness of syntactic status. Matters are complicated even further by the same choices being explainable also in terms of a preference for nouns vis-à-vis adjectives. All in all, the role of word position seems an interesting one to examine, if only because it is taken so seriously by dictionary publishers.

Thus, in the present study an attempt will be made to investigate the role of three factors: part of speech, word order, and frequency in attracting users' attention as potential candidates for dictionary lookup.

2. The study

2.1 Aim

The aim of the study is to assess the effect of part of speech, word position (within the MWE), and lexical frequency on the users' selection of elements in multi-word expressions that they would most readily look up.

2.2 Participants

Participants in the study were 40 Polish secondary school students aged 17 and 18, with males and females roughly equally represented. As learners of English, participants were at the B1 proficiency level as per the Common European Framework of Reference for Languages.[3]

2.3 Instrument

The principal instrument used was the Headword Choice Test designed specifically for this study. The test consisted of 36 English multi-word expressions which were, in equal measure, noun compounds (e.g. *life jacket*) and sentence idioms (e.g. *have a heart of gold*; *still waters run deep*). The items were presented on a single page laid out in two columns, 18 items in each, with instructions in Polish written across the top. There were four versions of the Headword Choice Test (labelled A, B, C, and D) differing only in the order or items, in an effort to counterbalance any order effects.

The selection of items for the Headword Choice Test was guided by the goal to have a balanced representation of words in terms of the combination of the three design factors: lexical frequency, part of speech, and word position within the MWE. And so, it was important to include both frequent and rare nouns, placed initially or otherwise within the MWE. In doing so, we were con-

strained by what is possible in the language. Function words, being closed-class items such as articles or prepositions, tend to be very frequent, and their position relative to lexical words is subject to language-specific syntactic constraints. For this reason, it was not possible to obtain data with all theoretical combinations of frequency, part of speech (POS), and word position.

For word frequency, the Corpus of Contemporary American English (COCA, Davies 2008–) was consulted. Lemmatized frequency counts were used (checked in May 2009). Raw frequency counts were subsequently categorized into three frequency bands: frequent (over 48,000 occurrences in COCA), medium (between 10,000 and 48,000 tokens), and rare (below 10,000). As a result, the 83 content words (tokens) in the MWEs included 31 frequent items, 29 medium-frequency words, and 23 rare items.

In terms of part of speech, items were selected so that at least the three major classes of content words (nouns, verbs, and adjectives) would feature in a variety of word positions and represent a range of frequencies.

When it comes to word order, the literature suggests a special role for the first content word in a multi-word. For this reason, and because the target multi-words varied in length between two and five words, word position was treated as a two-level factor: initial versus non-initial.

The materials and procedure were piloted on a small group of eight subjects similar to our participants in terms of educational level and English proficiency. No problems with the instruction, items, or procedure were noted during the pilot study. All participants in the pilot study completed the task in less than ten minutes.

2.4 Procedure

Participants were provided with printouts of the Headword Choice Test described above. They were instructed by the experimenter in their native language (Polish) to underline, for each item on the list, the one word which they would look up in a dictionary if they wished to find out the meaning of the complete expression. The same instruction was included in writing at the top of the test sheet.

Participants worked individually with no access to additional materials. Based on the results of the pilot study, they were allowed 15 minutes to complete the task. All students started at the same time and when finished, the experimenter collected the sheets. The session proceeded smoothly and all participants managed to complete the task on time.

3. Results

All word selection data were entered into a database for further processing. Then, for each individual word token, the number of times it had been under-

lined was computed. This number corresponded to the number of subjects, out of the total of 40, who indicated by underlining that they would look up the multi-word item under this specific headword. Such headword selection counts were then analyzed in terms of how they were affected by word position in the MWE, part of speech, and lexical frequency. The measures presented in sections 3.2 to 3.4 below express the mean number of participants who indicated that they would have chosen a given word over other components of the MWE, further averaged for all words with a particular level of a design variable (e.g. initial, verb, medium frequency, etc.). This manner of computing lookup preference measures is unaffected by raw counts of particular categories and so the numbers are directly comparable within each factor.

In what follows, selection counts per item are tabulated (3.1). Further on, the roles of the three factors of interest are presented descriptively in turn (3.2-3.4). Finally, a multivariate GLM analysis is computed to assess the strength of the influence of each of the three factors and portion of variability they explain (3.5).

3.1 Headword selection data

Table 1 below gives complete data on headword selection for the 36 multi-word expressions tested. Each potential headword is followed in parentheses by the number of participants (out of 40) who underlined this particular word. For example, in item 1., *artificial insemination*, 17 participants underlined the adjective *artificial*, while 23 underlined *insemination*. Item 7. is slightly irregular: while most participants went for *red-handed*, two participants underlined just the second portion of this hyphenated word, *handed*. Similarly, in item 29., six participants underlined just the *self* portion of *self-made*. Although this does not cause major problems, hyphenated words are probably best avoided in such designs.

ID	multi-word (underline counts)
1.	artificial (17) insemination (23)
2.	at (0) daggers (37) drawn (3)
3.	back (11) door (29)
4.	be (1) a (0) wet (6) blanket (33)
5.	blind (26) trust (14)
6.	bury (17) the (0) hatchet (23)
7.	catch (10) sb (0) red-handed (28) [handed (2)]
8.	clean (6) slate (34)
9.	couch (28) potato (12)
10.	detached (34) house (6)
11.	estate (30) car (10)
12.	everyday (18) life (22)

13.	first (12) come (3) first (7) served (18)
14.	frame (35) of (0) mind (5)
15.	go (2) to (0) pieces (38)
16.	haste (27) makes (2) waste (11)
17.	have (1) a (0) clue (39)
18.	have (2) a (0) heart (33) of (0) gold (5)
19.	have (1) the (0) last (9) word (30)
20.	hide (27) and (0) seek (13)
21.	in (0) the (0) nick (31) of (1) time (8)
22.	life (10) jacket (30)
23.	like (1) headless (33) chicken (6)
24.	miss (21) the (1) point (18)
25.	new (0) broom (20) sweeps (19) clean (1)
26.	once (2) bitten (21) twice (4) shy (13)
27.	out (19) of (1) the (0) blue (20)
28.	rock (27) the (0) boat (13)
29.	[self (6)] self-made (32) man (2)
30.	sliding (30) doors (10)
31.	still (5) waters (19) run (3) deep (13)
32.	teeter (30) on (0) the (0) brink (10)
33.	time (20) is (0) money (20)
34.	train (17) of (0) thoughts (23)
35.	with (0) open (10) arms (30)
36.	world (13) cup (27)

Table 1: Target multi-word expressions with selection counts.

An examination of the selection counts suggests that, as in most previous studies, Polish learners of English tend to ignore function words and very frequent words. This becomes even clearer if we focus on the items that all participants ignored (i.e. they were never underlined) in looking up the target multi-word expressions (**Table 2** on the next page).

Those items tend to be frequent function words or relatively delexicalized verbs (*is*, *made*). Other such semantically shallow verbs (*have*, *go*) were underlined only once or twice. The item *sb* is something of a special case, being an abbreviation for *somebody* that is most often used in dictionary metalanguage and other language-teaching materials, but its status as a regular word is questionable. The article *the* is not on the list: while most instances of it were ignored, it was underlined by a single participant in *miss the point*. Such cases emphasize the point that user behaviour is to some degree erratic, and no uniform policy on its own will ensure that *all* users will fully benefit from the entries, however well structured.

word	POS	frequency band	position
a	article	frequent	noninitial
and	conjunction	frequent	noninitial
at	preposition	frequent	initial
in	preposition	frequent	initial
is	verb	frequent	noninitial
made	verb	frequent	noninitial
on	preposition	frequent	noninitial
(sb)	(pronoun/metalng)	(rare)	(noninitial)
to	preposition	frequent	noninitial
with	preposition	frequent	initial

Table 2: Words never underlined by participants.

Beyond the above observations, it is hard to make reliable generalizations by just scrutinizing tabulated count data. Therefore, we will now attempt to examine how headword selections depend on the three design factors: word position, part of speech, and word frequency.

3.2 Word position

The position of the word within the MWE did not appear to make much difference to our participants. Across all word tokens in the MWEs, the average multi-word-initial word was selected by 15.4 subjects, compared with 14.7 for the noninitial word. This is an unremarkable difference that would probably have little practical significance even if found to be statistically significant (detailed inferential statistics follow in section 3.5 below).

Thus, perhaps somewhat surprisingly, our Polish learners did not exhibit a marked preference for looking up initial components of multi-words. This would indicate that the frequent practice of dictionary makers to list multi-words under the first (content) word is of limited utility, at least for Polish learners.

3.3 Part of speech

Unlike word position, part of speech appears to have had a non-trivial impact on users' decision as to which word to look up (see **Figure 1**). Nouns come out at the top, with a mean of 21.1 selections falling on the noun. Adjectives are the second most preferred word class (16.8), ahead of verbs (10.7). The least often looked up word classes are adverbs (5.0) and prepositions (3.0) (this line-up excludes articles, conjunctions and pronouns, which were not underlined at all, and for which there is little data).

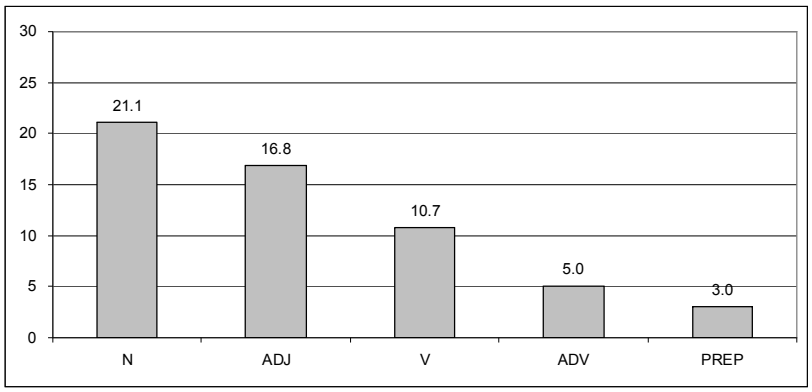


Figure 1: Lookup preference (in mean selection counts) by part of speech.

The rather poor standing of verbs compared to adjectives is perhaps somewhat surprising. Possibly, this may be related to the relative semantic vagueness of verbs in multi-word expressions.

3.4 Word frequency

Word frequency as expressed in frequency bands again appears to have played a role in guiding the participants' decisions as to which words to look up (see **Figure 2**). Words in the *rare* category registered the highest mean selection count (25.1). Medium frequency words received an average of 17.5 selections, with 11.4 being the figure for frequent words. We will revisit the role of frequency in more detail in section 3.5.1 below.

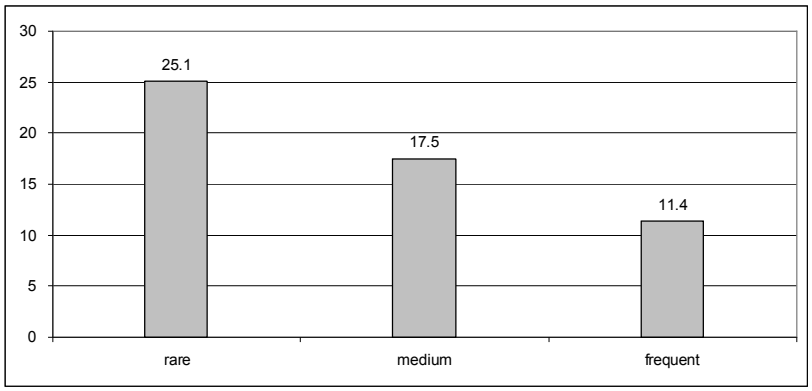


Figure 2: Lookup preference (in mean selection counts) for rare, medium, and frequent words.

3.5 A factorial analysis

To assess more systematically the degree of influence that word position, part of speech and word frequency have on the likelihood of the word being selected when looking up MWEs, a factorial General Linear Model (GLM) analysis was conducted on word selection counts as the dependent variable, with the three design factors as predictor variables. This analysis was conducted with the help of the Statistica 8 software suite.

For those unfamiliar with General Linear Modelling, for practical purposes it can be thought of as a generalization of Analysis of Variance (ANOVA) which allows continuous factors, not just categorical ones. Looking at our data, in a conventional ANOVA we would have been forced to use discrete frequency bands as levels of the frequency factor, much as in **Figure 2**. In contrast, the GLM approach has made it possible to utilize the full frequency information and thus obtain a more complete mathematical model of reality. To make frequency figures independent of corpus size, raw frequencies were converted to items per million (ipm, a customary measure in corpus statistics). Further, to reflect the fact that psycholinguistically meaningful differences in word frequency tend to be exponential rather than linear, a common logarithm of ipm was computed.

The data for the less central syntactic categories were not complete in terms of the availability of all combinations of word frequency and word position, so could not be analyzed due to numerous empty cells in the design. For this reason, the GLM analysis was restricted to nouns, verbs and adjectives (these, however, cover 87% of the data; besides, some previous studies also ignored function words).

The results of the GLM analysis are given in **Table 3**. Readers familiar with ANOVA tables should have no problems understanding the results. The table also includes partial η^2 ('eta-squared'), a measure of effect size commonly used in similar designs, as well as observed test power, assuming an alpha level of 0.05.

	SS	**df**	**MS**	**F**	**p**	**partial η^2**	**observed power (α=0.05)**
Intercept	6446.9	1	6446.9	82.2	**0.000**	**0.540**	1.00
Frequency	2680.0	1	2680.0	34.2	**0.000**	**0.328**	1.00
POS	557.4	2	278.7	3.6	**0.034**	**0.092**	0.64
Position	204.7	1	204.7	2.6	0.111	0.036	0.36
POS*Position	233.4	2	116.7	1.5	0.233	0.041	0.31
Error	5488.0	70	78.4				

Table 3: A three-way GLM analysis of word lookup preference with word frequency, word position and part of speech as factors. Factors in bold are statistically significant.

Table 3 indicates that apart from the intercept (a constant non-zero component, as it were), the two design factors that reach significance are frequency and part of speech. However, the effect size for the latter is much smaller than for the former, which roughly means that frequency predicts a greater portion of the participants' lookup preference. The role of part of speech has already received sufficient coverage in 3.3 above, so let us now turn our attention to frequency.

3.5.1 Frequency

Figure 3 plots word selection counts for individual words against their corpus (COCA) frequency data. Frequency is expressed as a common logarithm of items-per-million, a relative frequency measure often preferred in corpus statistics because of its independence of corpus size.

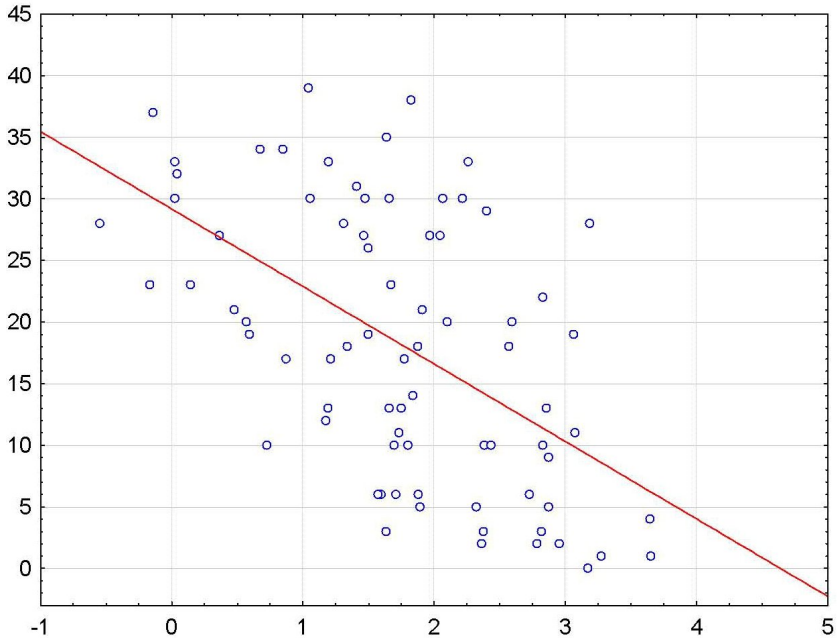


Figure 3: Scatterplot of word selection counts (a measure of lookup preference) against the common logarithm of relative word frequency, with a regression line fitted (count = 29.2 – 6.3 × log(ipm)).

It can be seen that, in broad outline, the lower the frequency, the greater the tendency for the word to attract attention. To formalize this tendency, a regression line was fitted, and it predicts the word selection count as the intercept of 29.2 minus 6.3 times the logarithm of normalized frequency (formulaically,

count = 29.2 – 66.3 × log(ipm)). While the data points appear to cluster along the regression line, it is also true that they do so rather loosely. This means that lexical frequency only predicts a relatively modest portion of the look-up decisions. There are other factors at play, including of course part of speech. We should also bear in mind that corpus frequency is only a general indicator of word familiarity. Learners are likely to be more directly guided by how familiar a lexical item appears to them, and while the number of times they have encountered a word certainly plays an important role, everyone's experience with words is different. Finally, learners of a language are probably exposed to types of texts in proportions different from those reflected in a general corpus.

3.5.2 Part of speech by word position interaction

The interaction of part of speech by word does not reach significance ($F_{(2, 70)}$=1.5, p=0.2), so one can only speak of tendencies here. The graph (**Figure 4**) patterns into what is often referred to as a crossed interaction. For nouns, it does not matter if the noun is phrase-initial or not. For adjectives and verbs, however, there does seem to be some (albeit not significant) preference in the sample for the initial position.

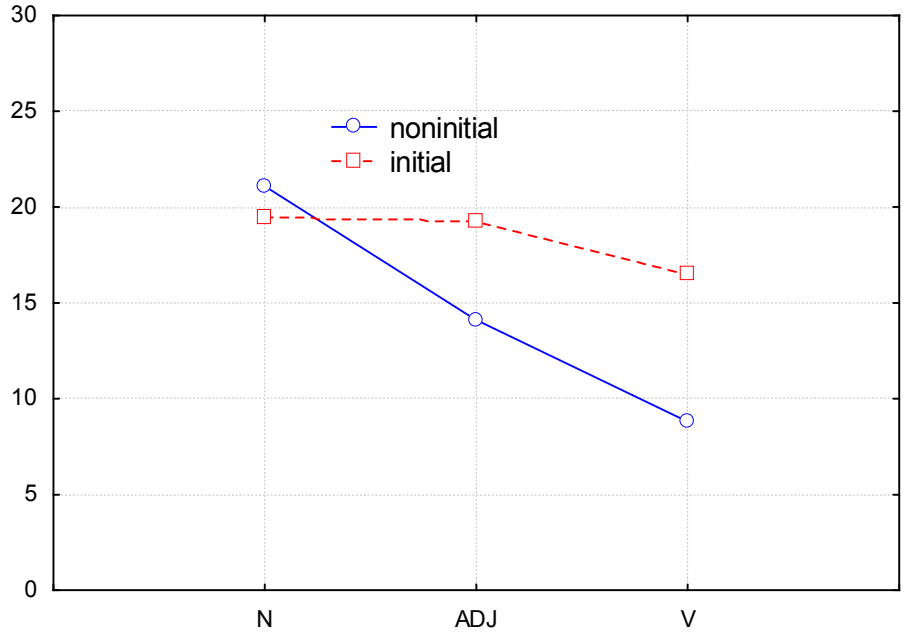


Figure 4: Interaction plot of part of speech and word position.

Perhaps this pattern means — though at present this is little more than a guess — that participants mostly looked for unfamiliar words and then nouns, but if

these strategies did not yield a clear winner, initial word position may have come into play.

4. Discussion

When faced with a known multi-word expression, Polish learners of English prefer to look up low-frequency words found in the MWE, probably because those are the words they tend to be less familiar with, and/or because they realize that common words often have very long entries where it is easy to miss something. Apart from the frequency, learners are guided by part of speech, preferring nouns, and then adjectives and verbs, in this order. They tend to ignore function words (articles, prepositions, pronouns) and adverbs, as well as verbs in their delexicalized uses.

Our findings on the whole concur with those obtained in previous studies for native speakers of other languages. The role of frequency features in all investigations, with the possible exception of native speakers of Dutch in Bogaards (1990), and it is telling that in our study frequency stands out as the most robust predictor of headword selection (partial $\eta^2 = 0.328$, **Table 3.**). The noun > adjective > verb hierarchy tallies with that noted by Bogaards (1990) for Dutch speakers. The potential POS-dependent role of word position has not been noted before, but this effect was not significant in our study.

Not all the findings overlap, though. On a detailed level, one of the items included in the present study, *artificial insemination*, was also tested by Béjoint (1981). He found a very clear preference (93%) for *insemination*, but in the present study the preference for this word was only marginal (58%). The disparity could be due to the different L1 (French versus Polish), or to divergent dictionary cultures (regular users more or less consciously adapt to what they encounter in dictionaries), or else — perhaps most likely — to a difference in the level of participants (secondary school students versus English majors at university).

This study suffers from a number of limitations. Most obviously, it is limited to Polish learners of English at a specific level.

The task does not exactly mimic an actual look-up situation. As in all previous studies, participants were asked to mark words rather than actually look them up in a dictionary. The advantage of the underlining task is that it is much quicker than actually looking words up, and thus it frees up the time in which to test a greater number of items, but there is no guarantee that learners operate in exactly the same way in the two situations.

Finally, MWEs are presented out of context, which is not how users would encounter them in real texts. In a broader context, learners may not realize they are dealing with MWEs and, instead, believe that they have a problem understanding some sense of a simplex word. It is, however, possible that in such a case they would follow similar strategies in selecting the word to look up.

5. Implications for lexicographers

The present findings suggest that lexicographers, in deciding where to treat an MWE in full, should be guided primarily by word frequency, going for the least frequent constituent. Doing so should not pose much of a practical problem since in this day and age dictionary compilation is already heavily corpus-based. There may even be potential for a degree of automation here (Kilgarriff et al. 2010). Where there is no clear[4] lowest-frequency word, nouns should be given priority, but in those instances it might be wise to duplicate the full treatment under the second least-frequent item. Cross-references should be given at all nouns, adjectives, and verbs except extremely frequent ones such as *be* or *have*.

All these decisions on the treatment of multi-word-expressions should be described in the front matter of the dictionary. Even if the average user will not make good use of that information, there is a chance that their teacher might.

6. MWEs in paper and electronic dictionaries

The issue of where to place multi-word expressions is a particularly relevant one for paper dictionaries, where restrictions of space make it rather impractical to present such items under many headwords at the same time. If one has to pick one lemma under which to embed the MWE, it is important that it is a lemma that most users would expect the expression to be placed under. Other lemmas can, and often do, include cross-references to the headword with the full treatment, giving users access to the expression, even if through an indirect route.

An unorthodox solution was adopted in *Cambridge International Dictionary of English* (CIDE, Procter 1995): this dictionary included a complete index of multi-word expressions in a separate section. Later editions did not retain this feature, and such an index is probably not an effective solution.

In electronic dictionaries it is perfectly possible to store an MWE in a single place, but present the full treatment under multiple lemmas. While this is not a huge technical problem, it is not at all obvious that this is indeed the best option, as doing so would significantly inflate entries, making them harder to navigate. This is especially important on devices with small displays, such as mobile phones, where presentation space is radically limited (Lew 2010: 299, in press). Thus, the issue of which component word of an MWE is the one users would most readily look up remains at least partially relevant for electronic dictionaries. It will become less of a problem once the dictionary can reliably recognize multi-word items typed directly into the search box. In fact, such a capability is slowly becoming a reality (Lew 2011, 2012), though progress is hampered by the fact that multi-word expressions often exhibit significant variation in form.

Still, success in the above case would be contingent on the dictionary users realizing that they are dealing with a multi-word item in the first place. There is no doubt that MWEs sometimes go unrecognized, and yet users may still choose to look up one of their components when faced with a comprehension problem they see as being due to a particular problem word within the scope of the MWE. In such a case, they may still chance upon the MWE within the entry, provided it is salient enough. Thus, felicitous placement of MWEs remains important even in those electronic dictionaries which are capable of finding them independently of headwords.

7. Educating dictionary users

Dictionary users in formal educational settings should be given training in dictionary (reference) skills (Lew and Galas 2008; Bae 2011; Ronald and Ozawa 2011). As part of that training, they should be made aware of the importance of multi-word expressions and taught to identify them in texts. They should receive hands-on practice on how to effectively find MWEs in dictionaries. Further, users should become aware that a good candidate to start the search is the word that looks the least familiar, but if this fails, they should try the noun. Regular users of a specific dictionary should make an effort to find out what its MWE placement strategy is, if there is one (of course, explicit advice in the front matter will help, see 5 above). For electronic dictionaries, they should check if multi-word expressions may be typed directly into the search box, and if so, follow this strategy. If this does not work, they might consider switching to a dictionary that does offer this functionality.

Acknowledgements

I wish to thank my student assistant Ewelina Łukasik for help with data collection, Adam Kilgarriff for suggestions regarding the presentation of results, and Paul Meara for his thoughts on the perception of word frequency. I am grateful to Arleta Adamska-Sałaciak for reading through the manuscript. I have also been fortunate in receiving very helpful suggestions from the anonymous reviewers.

Dedication

This article is dedicated to the memory of Paul Bogaards.

Notes

1. In this model individual words neatly fill the terminal nodes of a syntactic structure, with word combinability mostly restricted to syntax.

2. For example, *Macmillan English Dictionary* (Rundell 2007: ix) instructs users to '[l]ook for fixed expressions at the entry for the first main word in the expression'.
3. The CEFR is a system proposed by the Council of Europe which aims at harmonizing standards of attainment in foreign language learning.
4. Bogaards (1992) proposes a minimum difference in rank of about 2500 for French, but I have doubts whether a difference so expressed is a useful measure across a broad range of frequency. For example, a difference between frequency ranks of 10 and 2510 is dramatic, but one between 50,000 and 52,500 will be rather hard to notice, if not somewhat arbitrary.

References

Atkins, B.T.S. and F.E. Knowles. 1990. Interim Report on the EURALEX/AILA Research Project into Dictionary Use. Magay, T. and J. Zigány (Eds.). 1990. *BudaLEX '88 Proceedings. Papers from the 3rd International EURALEX Congress, Budapest 4–9 September 1988:* 381-392. Budapest: Akadémiai Kiadó.

Atkins, B.T.S. and K. Varantola. 1998. Language Learners Using Dictionaries: The Final Report on the EURALEX/AILA Research Project on Dictionary Use. Atkins, B.T.S. (Ed.). 1998. *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators:* 21-81. Lexicographica Series Maior 88. Tübingen: Max Niemeyer.

Bae, S. 2011. Teacher-Training in Dictionary Use: Voices from Korean Teachers of English. Akasu, K. and S. Uchida (Eds.). 2011. *ASIALEX2011 Proceedings, LEXICOGRAPHY: Theoretical and Practical Perspectives:* 45-55. Kyoto: Asian Association for Lexicography.

Béjoint, H. 1981. The Foreign Student's Use of Monolingual English Dictionaries: A Study of Language Needs and Reference Skills. *Applied Linguistics* 2(3): 207-222.

Bogaards, P. 1990. Où cherche-t-on dans le dictionnaire? *International Journal of Lexicography* 3(2): 79-102.

Bogaards, P. 1991. Word Frequency in the Search Strategies of French Dictionary Users. *Lexicographica* 7: 202-212.

Bogaards, P. 1992. French Dictionary Users and Word Frequency. Tommola, K.H., K. Varantola, T. Salami-Tononen and J. Schopp (Eds.). 1992. *EURALEX '92 Proceedings. Papers submitted to the 5th EURALEX International Congress, Tampere, Finland:* 51-59. Tampere: Department of Translation Studies, University of Tampere.

Davies, M. 2008–. The Corpus of Contemporary American English: 425 Million Words, 1990–Present. *Available online at http://corpus.byu.edu/coca/.*

Kilgarriff, A., V. Kovár and P. Rychlý. 2010. Tickbox Lexicography. Granger, S. and M. Paquot (Eds.). 2010. *eLexicography in the 21st Century: New Challenges, New Applications. Proceedings of ELEX2009:* 411-418. Louvain-la-Neuve: Cahiers du CENTAL.

Lew, R. 2010. Multimodal Lexicography: The Representation of Meaning in Electronic Dictionaries. *Lexikos* 20: 290-306.

Lew, R. 2011. Online Dictionaries of English. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography:* 230-250. London/New York: Continuum.

Lew, R. 2012. How Can We Make Electronic Dictionaries More Effective? Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography:* 343-361. Oxford: Oxford University Press.

Lew, R. In press. Space Restrictions in Paper and Electronic Dictionaries and Their Implications for the Design of Production Dictionaries. Banski, P. and B. Wójtowicz (Eds.). In press. *Issues in Modern Lexicography.* München: Lincom Europa.

Lew, R. and K. Galas. 2008. Can Dictionary Skills Be Taught? The Effectiveness of Lexicographic Training for Primary-School-Level Polish Learners of English. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15–19 July 2008*: 1273-1285. Barcelona: Universitat Pompeu Fabra.

Nowakowski, M. 1990. Metaphysics of the Dictionary Versus the Lexicon. Tomaszczyk, J. and B. Lewandowska-Tomaszczyk (Eds.). 1990. *Meaning and Lexicography:* 3-19. Amsterdam/Philadelphia: John Benjamins.

Procter, P. et al. (Ed.). 1995. *Cambridge International Dictionary of English.* (CIDE.) Cambridge: Cambridge University Press.

Ronald, J. and S. Ozawa. 2011. Electronic Dictionary Use: Identifying and Addressing User Difficulties. Akasu, K. and S. Uchida (Eds.). 2011 *Asialex2011 Proceedings, Lexicography: Theoretical and Practical Perspectives:* 436-446. Kyoto: Asian Association for Lexicography.

Rundell, M. (Ed.). 2007. *Macmillan English Dictionary*. Second Edition. London: Macmillan. (MED2)

Tono, Y. 1987. *Which Word Do You Look up First? A Study of Dictionary Reference Skills.* Unpublished M.Ed. dissertation. Tokyo Gakugei University, Tokyo.

Zgusta, L., V. Cerný, Z. Hermanová-Novotná and D. Heroldová. 1971. *Manual of Lexicography.* Janua Linguarum. Series Maior 39. The Hague: Mouton.

Example Sentences in Bilingual School Dictionaries

Liezl Potgieter, *Department Afrikaans and Dutch, University of Stellenbosch, Stellenbosch (liezlpotgieter@gmail.com)*

Abstract: Dictionaries are written for a specific user group and the content and presentation must therefore be directed/aimed at that specific target group. In the past dictionaries were often compiled without the lexicographers really taking time and researching the intended users and their specific needs. School dictionaries especially were often compiled after very little or no attention being paid to the users and their needs. This article will look at the role of example sentences in dictionary articles, different types of example sentences as well as the criteria for example sentences in bilingual school dictionaries in order to ensure that the final product is natural, typical, informative, intelligible and of use to the user.

Keywords: SCHOOL DICTIONARY, USER, USER NEEDS, EXAMPLE SENTENCE, VOCABULARY, LEMMA, CITATION, CITATION EXAMPLE, COMPETENCE EXAMPLE

Opsomming: Voorbeeldsinne in tweetalige skoolwoordeboeke. Woordeboeke word saamgestel vir 'n spesifieke gebruikersgroep en die inhoud en aanbieding moet daarom op die spesifieke teikengroep gerig wees. In die verlede is woordeboeke dikwels saamgestel sonder dat die leksikograwe werklik tyd daaraan bestee het om navorsing te doen oor die beoogde gebruikers en hulle spesifieke behoeftes. Skoolwoordeboeke is veral saamgestel sonder dat daar veel of enige aandag aan die gebruikers en hulle behoeftes geskenk is. In hierdie artikel sal daar gekyk word na die rol van voorbeeldsinne in woordeboekinskrywings, verskillende tipes voorbeeldsinne sowel as die kriteria vir voorbeeldsinne in tweetalige skoolwoordeboeke om te verseker dat die eindproduk natuurlik, tipies, informatief, verstaanbaar en van nut is vir die gebruiker.

Sleutelwoorde: SKOOLWOORDEBOEK, GEBRUIKER, GEBRUIKERSBEHOEFTES, VOORBEELDSIN, WOORDESKAT, LEMMA, SITAAT, SITAATVOORBEELD, GEBRUIKSVOORBEELD

1. Introduction

Dictionaries are written for a specific user group and the content and presentation must therefore be directed/aimed at that specific target group. In the past dictionaries were often compiled without the lexicographers really taking time and researching the intended users and their specific needs. School dictionaries especially were often compiled after very little or no attention being paid to the users and their needs. Often a number of articles were extracted from an existing bilingual dictionary and then published as a school dictionary, when in reality the so-called school dictionary was nothing more than a condensed ver-

sion of the more comprehensive bilingual dictionary. One such example is Bosman, Van der Merwe and Barnes' *Tweetalige Skoolwoordeboek* which, according to the introduction in the tenth edition is a "shortened version" of the well-known *Tweetalige Woordeboek* by Bosman, Hiemstra and Van der Merwe.

These shortened versions published as school dictionaries seldom meet the needs of the target users and often are more of a hindrance than a help. One of the big shortcomings in several bilingual school dictionaries, including *Tweetalige Skoolwoordeboek* and its successor, *Pharos Afrikaans–Engels Skoolwoordeboek* published in 2004, is that they supply the user with little or no contextual guidance (for example, labels) and/or cotextual guidance (example sentences).

Luckily, during the last few years lexicographers have become more and more aware of the importance of the so-called user perspective — determining who the intended target user is and what his or her specific needs are with regard to the dictionary. According to Tarp (2005: 8) "all theoretical and practical lexicographic work should be based on the dictionary functions which represent the assistance provided by a dictionary — by means of its lexicographic data — to a specific type of user in solving the specific type of problems related to a specific type of user situation". Over and above this, pedagogical lexicography in South Africa has also gained momentum and during the past few years several vastly improved bilingual and hybrid school dictionaries with example sentences as part of the dictionary articles have been released. Examples of such dictionaries are listed in the table below, along with the age group of the learners (the intended target users) the dictionary is aimed at.

Dictionary	Publisher	Year of release	Age group
Nuwe woordeboek sonder grense [NWSG]	Maskew Miller Longman	2004	Grade 4-6
Bilingual Learner's Dictionary [BLD]	Pharos Dictionaries	1998	—
Oxford English–Afrikaans School Dictionary [OEASD]	Oxford University Press	2007	Grade 4-9
Oxford English–Northern Sotho School Dictionary [OENSD]	Oxford University Press	2007	Grade 4-9
Oxford English–Zulu School Dictionary [OEZSD]	Oxford University Press	2010	Grade 4-9
Grondslagfasewoordeboek [GSFW]	Maskew Miller Longman	2010	Grade 1-3

2. The role of example sentences in dictionary articles[1]

Example sentences play an important role in school dictionary articles and should supplement and possibly extend the definitions (Creamer 1987). Accord-

ing to Creamer (1987: 241) an example can "take the burden off a definition by showing various ways the entry can be translated in context, indicate typical modifiers and illustrate points of usage". When specifically focusing on learner dictionaries one finds that examples in learner dictionaries are "intended to illustrate the meanings of words more clearly than is sometimes possible within the definition" (Herbst 1989: 1382). The role of example sentences in dictionary articles (especially in school dictionaries) should therefore not be underestimated.

3. Different types of example sentences

A distinction can be made between three different types of example sentences, namely citations, citation examples and competence examples (Bergenholtz and Tarp 1995: 139). Citations are quotes from recorded language usage (usually recorded in a corpus) and have been taken from authentic texts without subsequent adaptation by the lexicographer (Bergenholtz and Tarp 1995: 139). Although citations have the advantage that the user may always be certain that the example represents reality, the citations may also reflect slightly deviant usage (Bergenholtz and Tarp 1995: 139). Furthermore, even with a large corpus it may be difficult for lexicographers to find examples of the appropriate length, citations often aren't concise enough to be effective examples and it might not always be possible to find an apt example illustrating the specific usage of the lemma (Bergenholtz and Tarp 1995: 139; Gouws 1989: 228). In such cases lexicographers often make use of citation examples which are shortened or adapted citations from which names, meaningless adverbial phrases, data in brackets, et cetera are excluded. Although the adaptation of citation examples can take a lot of time, they are more accessible and user-friendly than citations and can also save a lot of space in the dictionary (Potgieter 2011: 124).

The last type of examples — competence examples — are thought up by the lexicographer and based on the lexicographer's own abilities. It can often take a lexicographer up to half an hour to find a suitable citation from the corpus, but it will take him or her only a few minutes to think up a competence example. Competence examples are usually commonly found in Language for General Purposes (LGP) dictionaries and school dictionaries. But competence examples do hold the danger that they can sometimes be very stereotypical and/or reflect the lexicographer's personal views. Lexicographers are therefore warned to stay away from stereotypes and try to leave their personal views/opinions out of the examples[2].

In the case of school dictionaries, the lexicographers are free to decide whether they want to make use of citations, citation examples or competence examples. They can also choose to use a combination of all three types of example sentences. The important thing is not the source of the example, but whether the example sentence fits the criteria as listed and discussed below.

4. Criteria for example sentences in school dictionaries

When compiling a bilingual school dictionary the following criteria should be kept in mind when writing/choosing example sentences. The more criteria the example sentences meet, the better and more suited they will be for school dictionaries and the users of these dictionaries.

4.1 The example sentences must fit into a context the users will understand

It is important that the example sentences chosen by the lexicographer to illustrate the use of the lemma, should be used within a context that the dictionary user will know and understand. It is therefore important that the examples should fit into the learners' world. If the context or use of the word is too foreign or the example sentence is too complicated, it will end up being more of a hindrance than a help.

The following is an example of an example sentence that is too far removed from the learners' world and therefore too foreign for a learner to understand. Therefore it does not clarify the meaning or use of the lemma to the learner.

> **lemile** ▪ cultivated ♦ E be e le bona bomang ba ka tlago ba sekiša kgoši ge a **lemile** mašemo ao e lego a gagwe? *Who could have prosecuted the chief when he cultivated his lands?*

Article 1: "lemile" in OENSD

As can be seen, although the headword "lemile" (cultivated) is used in the example sentence, the sentence in itself really has nothing to do with the word "cultivated" and doesn't help the learner in understanding the word "cultivated". The following example sentence, taken from the OEZSD is much more successful in explaining and demonstrating the meaning and use of the word "cultivate": "They agreed that they would cultivate one type of crop."

A second example of an example sentence that doesn't help explaining the meaning and/or use of the lemma, is the following.

> **danger** ▪ kotsi ♦ Ask questions to make sure that the learners understand the **danger** of lightning. *Botšiša dipotšišo go kgonthišiša gore baithuti ba kwešiša kotsi ya legadima*.

Article 2: "danger" in OENSD

The following two example sentences from BLD and NWSG are much better illustrations of the meaning and use of the word "danger" and will therefore be of much bigger help to the learners consulting the dictionary. "Mense wat te vinnig ry, is 'n gevaar op die pad./People who speed are a danger on the road."

[from BLD] and "Die matrose op die sinkende skip is in groot gevaar [danger/ peril]." (The sailors on the sinking ship are in great danger) [from NWSG].

Other examples of example sentences that are of little or no help to learners in understanding the meaning or use of the lemma are the following:

> **black person** … Moswana; Mothomoso … Remember, a **black person** speaks the truth when he says that one who plots against others is often his own victim. *Fela elelwa gore Moswana o be a rerešitše ge a re, sekhukhuni se bonwa ke sebataladi.*
>
> Article 3: "black person" from OENSD
>
> **completion** … phethagatšo … The municipality is trying very hard to achieve the **completion** of what they promised the community. *Mmasepala o leka ka maatla go fihlelela phethagatšo yeo ba e tshepišitšego setšhaba.*
>
> Article 4: "completion" from OENSD

4.2 The example sentences must fit the learners' vocabulary

The next important criterion that the lexicographer has to keep in mind when choosing and/or compiling example sentences is that the example sentences must fit the learners' vocabulary. Segler mentions that an oft-cited criterion for the usefulness of examples is that they are comprehensible to the learner. "The more vocabulary items used in the example are unfamiliar to the learner, the less likely it is that the example will be helpful to him" (sic) (Segler 2007: 20). Since the school dictionaries listed above are mainly aimed at language learners between grades 4 and 9, the vocabulary used in the example sentences should also fit the vocabulary used by that age group. According to Drysdale (1987: 213) examples are comprehensible if they use styles, registers and vocabulary that are "both idiomatic and intelligible at the students' level of comprehension". If the majority of the words used in the example sentences are not found in the learners' general or defining vocabulary, the learners won't understand the example sentence and once again it will be more of a hindrance than a help. Article 5 is an example of a sentence where some of the words used (see the highlighted text) do not form part of the learners' active vocabulary, are above the learners' comprehension and therefore unintelligible. As a result the example sentence can cause confusion instead of clarifying the lemma.

> **šeo** … there they are *(close to you)* … Mehlala ya mainamatšo šeo letlakaleng la bobedi. *There are the examples of deverbative nouns, on the second page.*
>
> Article 5: "šeo" from OENSD

By rewriting the sentence as follows "There are the examples of different types of fruit.", it would much better illustrate the lemma as the learners wouldn't then focus on the words they don't know or understand, but on the use of the lemma (underlined).

It is also recommended that lexicographers not make use of unnecessarily difficult phrasing in cases where there are simpler ways available in which to phrase the sentence without it losing its meaning. Article 6 is a good example of a sentence where unnecessarily difficult phrasing was used (see highlighted text).

> **centimetre** … sentimetara … The car came to a standstill 30 centimetres from the post. *Koloi e ile ya ema disentimetara tše masometharo go tloga koteng.*
>
> Article 6: "centimetre" from OENSD

In this particular example the lexicographer could rather have used the word "stopped" instead of the phrase "came to a standstill". None of the meaning would have been lost, but the learners wouldn't have been confronted with an unnecessarily difficult phrase distracting their attention from the word represented by the lemma of which the usage is illustrated in the sentence.

4.3 The word represented by the lemma should be the focus of the example sentence

Another important criterion in the compilation/choosing of example sentences for school dictionaries is that the word represented by the lemma should be the focus of the example sentence. This implies that when the dictionary user reads the example sentence supplied he or she should immediately be able to see which word is the lemma discussed in that particular article. The following example sentence is a good example of a sentence where the word represented by the lemma isn't the focus of the sentence and if the specific word wasn't printed in bold the reader would have had no way of knowing which word in this particular sentence is the word the sentence is supposed to further explicate.

> **him** … **2 yena** Our group made an appointment with the teacher to discuss our project with **him**. *Ithimba lethu linqume isikhathi sokubonana nothisha ukuthi lixoxe naye ngeprojekthi yethu.*
>
> Article 7: "him" from OEZSD

Just from reading the sentence it is difficult to deduce if the lemma of the article is *group*, *appointment, teacher, discuss, project* or *him*. The following example sentences from OEASD provide much better examples of the use of the word "him":

> **him** … hom … Do you know **him**? *Ken jy hom?* … I gave **him** some food. *Ek het hom kos gegee.*
>
> Article 8: "him" from OEASD

The following article (article 9) is another example of a poor example sentence where it is difficult to deduce what the word is that the sentence is supposed to explicate while the one thereafter is a better example of a good example sentence for the word "bad".

> **mpe** … 1 bad … Ke phošo go itefetša go dilo tše mpe tše batho ba go dirilego tšona. *It is a mistake to avenge yourself as a result of bad things that people have done to you.*

Article 9: "mpe" from OENSD

> **bad** … 1 sleg … *Ek het slegte nuus vir jou: 'n motor het jou hond omgery*. I have bad news for you: a car has run over your dog.

Article 10: "bad" from BLD

4.4 The example sentences must clearly illustrate the meaning of the word represented by the lemma

The next important criterion to keep in mind, is that the example sentences should clearly illustrate the meaning of the word represented by the lemma and therefore not be vague or obscure.

As mentioned before, example sentences play an important role in school dictionaries helping to illustrate the meaning of words more clearly than is sometimes possible within the definition (Herbst 1989: 1382) and therefore their value should not be underestimated. It is therefore important that the word that is supposed to be illustrated in the example sentence should be the focal point of the example sentence and that the sentences shouldn't be too vague or obscure for the learner to be able to clearly see and understand the word represented by the lemma's meaning.

In article 11 the example sentence is vague and obscure and it is difficult to deduce the meaning of the word represented by the lemma from the example sentence whereas the example sentence in article 12 illustrates the meaning of the word "end" a lot better.

> **end** … noun … mafelelo; bofelo; pheletšo … If he sees you holding that dress like that, that will be the **end** of you. *Ge a ka go bona o swere roko yeo bjalo, e tla ba mafelelo ka wena.*

Article 11: "end" from OENSD

> **end** … noun … einde … They get paid at the **end** of the month. *Hulle word aan die einde van die maand betaal.*

Article 12: "end" from OEASD

It is also important that when making use of example sentences extracted from corpora the sentence shouldn't lose its illustrative value when removed from the context in which it was originally written.

By ensuring that the word represented by the lemma is the focal point in the example sentence and that the sentence clearly illustrates the use of this word, the lexicographer ensures that the example sentences are of a much higher quality and are of much more use to the learners.

4.5 The example sentences shouldn't be too long and drawn out

Another important criterion when it comes to example sentences in school dictionaries is that the sentences shouldn't be too long and drawn out. By keeping the sentences short and to the point, it is much easier to clearly illustrate the meaning of the word and to keep the word represented by the lemma the focus of the example sentences.

It is often the case that when learners are confronted with long and drawn out sentences that these sentences are citations selected from a corpus. These sentences often are long because they provide "superfluous detail in the name of authenticity" (Cowie 1999: 137). Cowie (1999: 137) argues that stripping away this extralinguistic information may be even more helpful for production than a lengthy example. Jacobsen et al. (1991: 2788) observes that "often an authentic example can be abbreviated or paraphrased without losing any of its illustrative value". Zgusta (1971: 265) also writes that when making use of citations "probably the best thing to do is quote … a reduced part of a passage in a text from which those parts that are inessential are omitted".

When looking at the following examples one can see that they are so long and drawn out that it is difficult to follow the sentence and be able to see which word's meaning is supposed to be illustrated.

> **thelela** … slip … Go kwagala gore ba rile ge ba tshela ye nngwe ya dinoka tša moo moruti Merensky a **thelela** godimo ga leswika gomme a robega letsogo. *It was said that when they crossed one of the rivers, Pastor Merensky slipped on a rock and broke his arm.*
>
> Article 13: "thelela" from OENSD

By shortening the sentence, the meaning and use of the lemma is illustrated much more clearly: Peter slipped on a rock and broke his arm.".

Here are some more examples.

> **son-in-law** … umkhwenyana … Parents who love their daughters dearly would want to have a say about who marries them, about the **son-in-law** to be. *Abazali abathanda amadodakazi abo kakhulu bangafuna ukuba nezwi kulabo ababashadayo, abazoba abakhwenyana babo.*
>
> Article 14: "son-in-law" from OEZSD

The following shorter example sentence would have been a better choice: "My daughter's husband is my son-in-law.".

-khiqiza … produce; manufacture … Inkampani **ekhiqiza** izimpahla zikagesi i-LG Electronics isisayine isivumelwano seminyaka emithathu sokuxhasa iBafana Bafana. *The company that manufactures electrical equipment, LG Electronics, signed an agreement to sponsor Bafana Bafana for three years.*

Article 15: "-khiqiza" from OEZSD

A better example would have been: "LG Electronics is a company that manufactures electrical equipment.".

4.6 The example sentences should help the learners with text production

Another important criterion is that the example sentences should help the learners with text production by showing users which words are typically used with the word represented by the lemma. One of the functions of example sentences is to show the headword "in action" and show points relating to the word represented by the lemma, for example collocations of verbs, use of articles, grammatical behaviour, phrases used to accompany nouns, et cetera. That way, the learners will know which words to use in conjunction with the word represented by the lemma. Here are a few examples:

ruk[1] … while … Die kos sal oor 'n **rukkie** reg wees. *The food will be ready in a little while.*

Article 16: "ruk[1]" from OEASD

promise … ukwethembisa; ukuthembisa … The athlete showed **promise** at an early age. *Umgijimi ukhombisa ukwethembisa nje eseneminyaka emincane.*

Article 17: "promise" from OEZSD

lecture … thuto … The professor gave an interesting **lecture** on South African English. *Profesa o file thuto ya go kgahliša ka ga Seisimane sa ka Afrika-Borwa.*

Article 18: "lecture" from OENSD

loop … Hulle het in die **loop** van [during] die jaar hierheen getrek.

Article 19: "loop" from NWSG

4.7 The example sentences should include names from different cultures and language groups

Lastly it is recommended that in cases where people are featured in the example sentences, the lexicographers should ensure that both male and female names are used (about 50% of each) and that the names also originate from the different cultures and language groups relevant to the target users. Should lexicographers run out of names, it is recommended they invest in name books, for example *The A-Z of Names* (Isabell Gauche) or *Call me by my name* (Lindiwe

Mtembu-Salter) or consult one or more of the many websites with names found on the internet. This will ensure that the final product is culturally and racially inclusive and no racial, language or cultural groups are left out.

5. In closing

The listed criteria are only a few of the things that contribute to creating natural, typical, informative and intelligible example sentences in school dictionaries. It is important that an example needs to be an utterance that the user of a dictionary is likely to hear spoken by speakers of a language, it should complement the definition and help the user understand it better and it should be intelligible (Hiles 2009: 26-27). At all times lexicographers should ensure that the example sentences they choose for school dictionaries should help the user in language learning and usage and not hinder his or her learning process.

Endnotes

1. In this article only example sentences will be looked at and not collocations and/or phrases.
2. For years there has been a debate as to which type of example — citation, citation example or competence examples — is the best. For more on this debate, see Prinsloo and Gouws 2000.

Sources

Bergenholtz, H. and S. Tarp (Eds.). 1995. *Manual of Specialised Lexicography: The Preparation of Specialised Dictionaries*. Amsterdam/Philadelphia: John Benjamins.

Bosman, D.B., I.W. van der Merwe and A.S.V. Barnes. 1998. *Tweetalige Skoolwoordeboek. Afrikaans–Engels/Engels–Afrikaans.* Cape Town: Pharos Woordeboeke.

Bosman, D.B., I.W. van der Merwe and L.W. Hiemstra. 1994. *Tweetalige Woordeboek/Bilingual Dictionary.* Cape Town: Tafelberg.

Cowie, A. 1999. *English Dictionaries for Foreign Learners: A History*. Oxford: Oxford University Press/Clarendon Press.

Creamer, T. 1987. Beyond the Definition: Some Problems with Examples in Recent Chinese–English and English–Chinese Bilingual Dictionaries. Cowie, A.P. (Ed.). 1987. *The Dictionary and the Language Learner. Papers from the EURALEX Seminar at the University of Leeds, 1–3 April 1985*. Lexicographica. Series Maior 17. Tübingen: Max Niemeyer: 238-245.

Drysdale, P. 1987. The Role of Examples in a Learner's Dictionary. Cowie, A.P. (Ed.). 1987. *The Dictionary and the Language Learner. Papers from the EURALEX Seminar at the University of Leeds, 1–3 April 1985*. Lexicographica. Series Maior 17. Tübingen: Max Niemeyer: 213-223.

Du Plessis, M. 1998. *Tweetalige aanleerderswoordeboek/Bilingual Learner's Dictionary.* Cape Town: Pharos Woordeboeke.

Gauche, I. 2005. *The A-Z of names. Discover the Promise your Name Holds*. Cape Town: Struik Christian Books.

Gouws, R., L. Potgieter en S. Burgess. 2010. *Grondslagfasewoordeboek. Afrikaans/English.* Cape Town: Maskew Miller Longman.

Gouws, R.H. 1989. *Leksikografie.* Cape Town: Academica.

Gouws, R.H., M. Stark and L. Gouws. 2004. *Nuwe woordeboek sonder grense*. Cape Town: Maskew Miller Longman.

Herbst, T. 1989. Dictionaries for Foreign Language Teaching: English. Hausmann, F.J. et al. (Eds.). 1989. *Dictionaries: An International Encyclopedia of Lexicography. Handbooks of Linguistics and Communication Science*. 5.2. Berlin: Walter de Gruyter: 1379-1385.

Hiles, L. 2009. Examples in South African School Dictionaries: From Theory to Practice. Unpublished M.Phil. Thesis. Stellenbosch: University of Stellenbosch.

Jacobsen, J., J. Manley, and V. Pedersen. 1991. Examples in the Bilingual Dictionary. Hausmann, F.J. et al. (Eds.). 1991. *Dictionaries: An International Encyclopedia of Lexicography. Handbooks of Linguistics and Communication Science* 5.3. Berlin: Walter de Gruyter: 2782-2789.

Kromhout, J. 2004. *Pharos Skoolwoordeboek. Afrikaans–Engels/Engels–Afrikaans.* Cape Town: Pharos.

Mtembu-Salter, L. 2001. *Call me by my Name. More than 2000 Xhosa, Zulu, Swazi and Ndebele Names.* Cape Town: Kwela Books.

Oxford Afrikaans–Engels/English–Afrikaans Skoolwoordeboek/School Dictionary. 2007. Cape Town: Oxford University Press.

Oxford IsiZulu–InsiNgisi/English–Zulu Isichazamazwi Sesikole/School Dictionary. 2010. Cape Town: Oxford University Press.

Oxford Sesotho sa Leboa–Seisimane/English–Northern Sotho Pukuntšu ya Sekolo/School Dictionary. 2008. Cape Town: Oxford University Press.

Potgieter, L. 2011. Die ontwerp van vertalende vakwoordeboeke met vertalers as teikengebruikers: 'n teoretiese model. Unpublished Doctoral Dissertation. Stellenbosch: University of Stellenbosch.

Prinsloo, D.J. and R.H. Gouws. 2000. The Use of Examples in Polyfunctional Dictionaries. *Lexikos* 10: 138-156.

Segler, T.M. 2007. *Investigating the Selection of Example Sentences for Unknown Target Words in ICALL Reading Texts for L2 German*. Doctoral Dissertation. Edinburgh: University of Edinburgh. [Online]. Available at: http://www.era.lib.ed.ac.uk/bitstream/1842/1750/3/Segler%2520TM%2520thesis%252007.pdf [2011, 8 February].

Tarp, Sven. 2005. The Pedagogical Dimension of the Well-conceived Specialised Dictionary. *Ibérica. Journal of the European Association of Languages for Specific Purposes* 10: 7-21.

Zgusta, Ladislav. 1971. *Manual of Lexicography.* The Hague: Mouton.

Die leksikografiese bewerking van verwantskapsterme in Sepedi

D.J. Prinsloo, Departement Afrikatale, *Universiteit van Pretoria, Pretoria, Suid-Afrika (danie.prinsloo@up.ac.za)*

Opsomming: Die verwantskapsterminologie van Sepedi is omvangryk en vorm 'n komplekse sisteem. In teenstelling met tale soos Afrikaans, Engels en Duits word groter uitdagings aan die Sepedileksikograaf gestel ten opsigte van die identifisering van verwantskapsterme en die lemmatisering en bewerking daarvan. Voorafstudie van die verwantskapsterminologiestelsel is 'n voorvereiste vir suksesvolle gebruikersleiding. Die Sepedi leksikograaf is die tussenganger tussen veral die onervare woordeboekgebruiker en hierdie ingewikkelde verwantskapsterminologiesisteem, en moet derhalwe sorg vir effektiewe lemmatisering en voldoende bewerking van verwantskapsterme. Die aard en omvang van verwantskapsterminologie in Sepedi word ontleed en daar word aangetoon dat verwantskapsterminologie in Sepedi problematies is, veral ten opsigte van die lemmatisering van samestellings, en in besonder, afgeleide enkelwoordvorme en frases, soos byvoorbeeld veelvuldige besitskonstruksies. Ten einde toegang tot verwantskapsterme in die woordeboek te vergemaklik word 'n leksikografiese konvensie vir die lemmatisering van verwantskapsterme voorgehou. Korpusvoorkomste van verwantskapsterme word aangegee en ruimte word afgestaan vir 'n kritiese evaluering van die lemmatisering en bewerking van verwantskapsterme in Sepediwoordeboeke.

Sleutelwoorde: VERWANTSKAPSTERME, SEPEDI, STAMBOOM, PATERNE, LEMMATISERING, WOORDEBOEKKONVENSIE, LEKSIKOGRAFIESE BEWERKING, RESEPTIEWE WOORDEBOEKGEBRUIK, PRODUKTIEWE WOORDEBOEKGEBRUIK, GEBRUIKSFREKWENSIE

Abstract: The lexicographical treatment of kinship terminology in Sepedi. Kinship terminology in Sepedi is extensive and forms a complex system. In contrast to languages such as Afrikaans, English and German the lexicographer faces greater challenges in respect of the identification of kinship terms and the lemmatisation and treatment thereof in Sepedi dictionaries. Preparational studies of the kinship terminology system are a prerequisite for successful user guidance. The Sepedi lexicographer is the mediator, especially between the inexperienced dictionary user and this complicated kinship terminology system, and therefore has to provide for effective lemmatisation and sufficient treatment of kinship terms. The nature and extent of kinship terminology in Sepedi are analysed and it is indicated that kinship terminology in Sepedi is problematic, especially in respect of the lemmatisation of compounds, and in particular, derived single-word forms and phrases, such as for example multiple possessive constructions. In order to ease access to kinship terms in dictionaries, a lexicographic convention for the lemmatisation of kinship terms is suggested. Corpus occurrences of kinship terms are indicated and space is allocated for a critical evaluation of the lemmatisation and treatment of kinship terms in Sepedi dictionaries.

Lexikos 22 (AFRILEX-reeks/series 22: 2012): 272-289

Keywords: KINSHIP TERMS, SEPEDI, FAMILY TREE, PATERNAL, LEMMATISATION, DICTIONARY CONVENTION, LEXICOGRAPHIC TREATMENT, RECEPTIVE DICTIONARY USE, PRODUCTIVE DICTIONARY USE, FREQUENCY OF USE

1. Inleiding

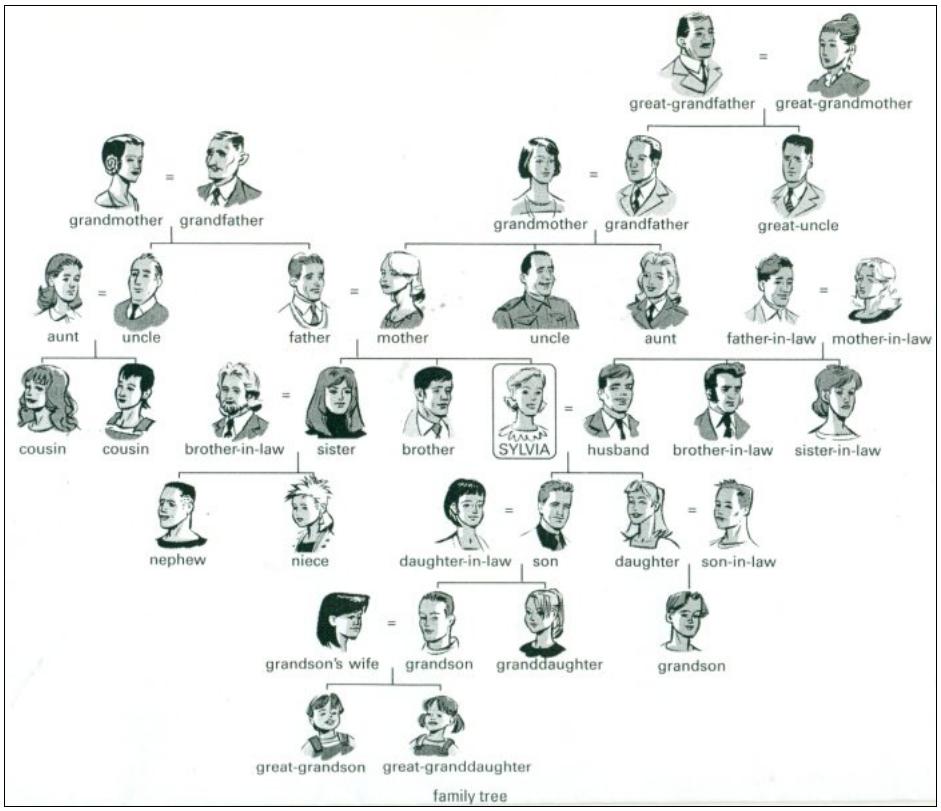
Die verwantskapsterminologie van Sepedi is omvangryk en vorm 'n komplekse sisteem. Prinsloo en Van Wyk (1992) poog om 'n paterne boomstruktuur te beskryf in terme van verskillende generasies, manlike en vroulike verwante, in terme van hoe 'n man (Ego) en sy vrou alle familielede van die man sal aanspreek of na hulle verwys.

Verwantskapsterme kom as enkelwoorde, samestellings en in konstruksies (hoofsaaklik besitskonstruksies) voor en die leksikograaf se dilemma is om lemmatiseringstrategieë te identifiseer wat alle verwantskapsterme in Sepedi in 'n omvattende woordeboek sal ondervang. Bestaande woordeboeke bied bloot selektief lemmas aan wat op grond van gebruiksfrekwensie in 'n korpus, of intuïtief geselekteer is. In hierdie artikel word 'n meer uitgebreide sisteem soos deur Prinsloo en Van Wyk (1992) voorgehou word as vertrekpunt geneem vir die voorgestelde lemmatiseringsproses. Hierdie sisteem, hoewel omvattend, bied ten beste net die terme vir familielede van die man. Die spieëlbeeld, naamlik hoe 'n vrou en haar man die *vrou* se verwante aanspreek of na verwys, kom nie eers hier ter sprake nie.

Die leksikograaf se rol is dié van 'n tussenganger tussen 'n komplekse struktuur aan die een kant en die woordeboekgebruiker aan die ander kant. Vir die doel van hierdie artikel word aanleerders van Sepedi as teikengebruikers geïdentifiseer.

Daar sal eerstens gepoog word om 'n lemmatiseringstrategie vir verwantskapsterme in Sepedi te formuleer uit die aanbod van enkelwoorde, samestellings en konstruksies in Prinsloo en Van Wyk (1992). Tweedens word verwantskapsterme in Prinsloo en Van Wyk (1992) se voorkoms in die *Pretoria Sepedi Corpus* (PSC) bestudeer en die lemmatisering en bewerking van verwantskapsterme in bestaande woordeboeke vir Sepedi onder die loep geneem. Evaluering van Sepediwoordeboeke word in terme van die lemma-aanbod en die kwaliteit van die bewerking gedoen. Derdens word 'n nuwe leksikografiese konvensie voorgestel vir die lemmatisering van verwantskapsterme in omvattende aanleerderswoordeboeke vir Sepedi.

Die lemmatisering van verwantskapsterme in 'n algemene woordeboek is 'n gegewe en geen woordeboek kan bekostig om nie terme soos *oom*, *broer*, *suster*, *oupa* of *ouma* te lemmatiseer en te bewerk nie. Die omvang en kompleksiteit van verwantskapsterminologiestelsels verskil egter van mekaar. So, byvoorbeeld, gee *Macmillan English Dictionary* (MED) die volgende diagram in figuur 1.

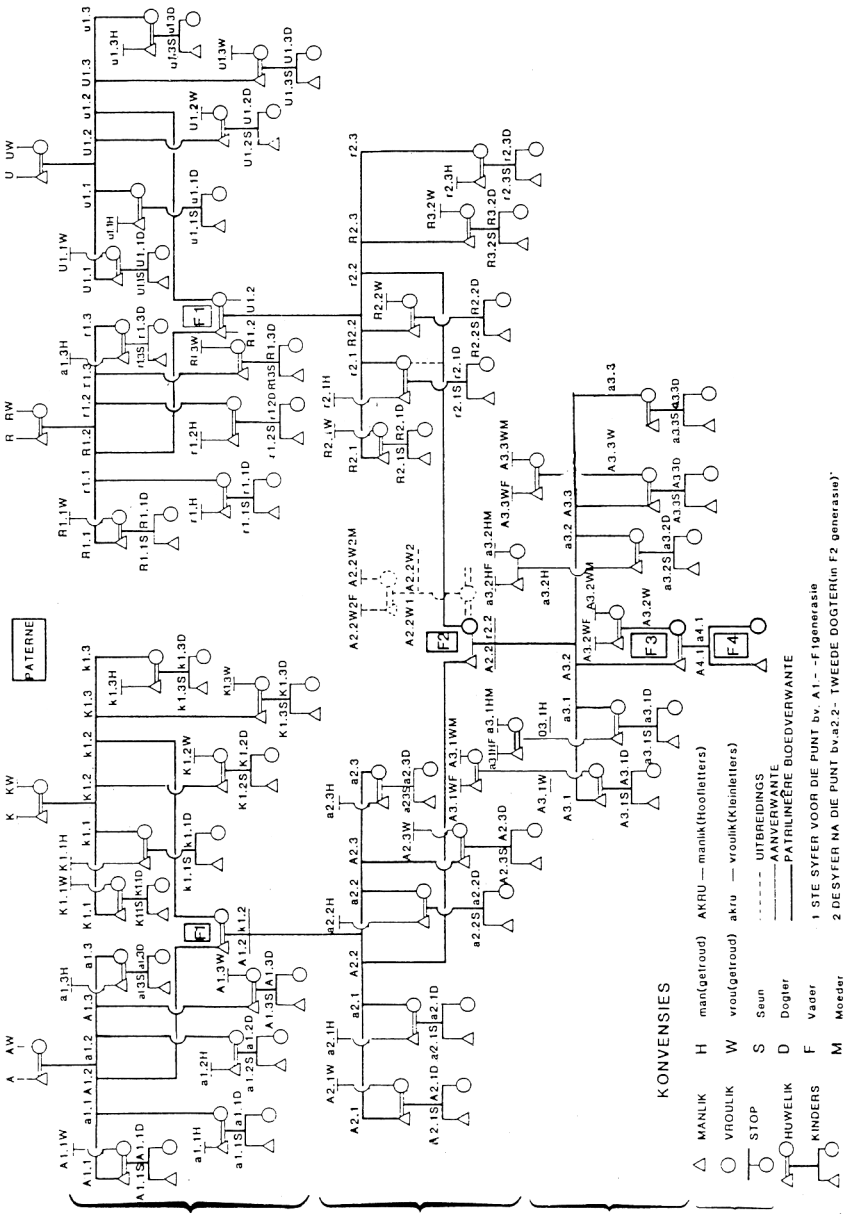


Figuur 1: *Familiestamboom* in MED

Prinsloo en Bosch (2012) verwys na figuur 1 as 'n struktuur van *interpreteerbare kompleksiteit* (comprehensible complexity) waarvan die terme met behulp van standaard leksikografiemetodes bewerk kan word terwyl hulle meen dat verdere gebruikersleiding nodig is in die geval van verwantskapsterminologiestelsels vir isiZulu en Sepedi.

2. Sepedi verwantskapsterminologie as 'n komplekse sisteem

Teenoor figuur 1 staan die verwantskapstelsel van Sepedi in figuur 2 in skrille kontras en kan laasgenoemde nie bloot leksikografies hanteer word deur met die blote bewerking van verwantskapsterminologie lemmas in die sentrale teks van die woordeboek te volstaan nie.

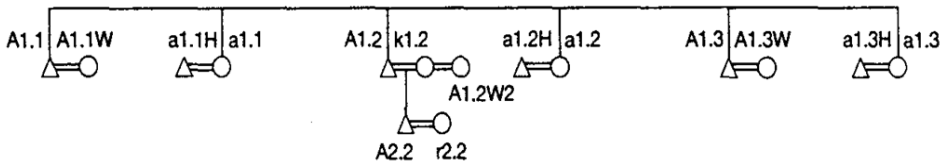


Figuur 2: Paterne diagram vir Sepedi (Prinsloo en van Wyk 1992: 45)

Hierdie figuur is selfs te kompleks vir aanbieding in die agtertekste-afdeling van die woordeboek en ten beste geskik as die verwysingsadres in 'n buiteteks,

soos byvoorbeeld Prinsloo en Van Wyk (1992). Figuur 2 reflekteer verder slegs die paterne relasies (hoe 'n man A2.2 en sy vrou r2.2 die *man* se verwante sal aanspreek of na hulle sal verwys).

Figuur 2 bied nietemin die ideale vertrekpunt vir die leksikograaf in die lemmatiseringsproses van verwantskapsterme. In die eerste plek moet hy/sy die komplekse struktuur kan afbreek in terme van die "boustene", byvoorbeeld die verskillende generasies soos 'n man se vader se broers en susters in figuur 3. So, byvoorbeeld, kan die Sepedi leksikograaf die stelsel in figuur 2 afbreek tot 19 boustene soos deur Prinsloo en Van Wyk (1992) gedoen is. Sekere gedeeltes of die samevoeging van gedeeltes kan dan as verteerbare vereenvoudigde skematiese voorstellings met voorbeelde in die agtertekste-afdeling aangebied word.



Figuur 3: Verwantskapsterme vir 'n man se vader se broers en susters

Aanspreekvorme/variante (Verwysingsterme tussen hakies)		
F2-Generasie	Ego A2.2 (manlik)	Ego r2.2 (vroulik)
Oom	A1.1 Ramogolo (Ramogolo)	
Tante	A1.1W Mmamogolo (Mmamogolo/ Mogatša ramogolo)	Soos vir ego A2.2 met opsionele byvoeging van 'wa mogatšaka' 'van my eggenoot'
Tante	a1.1 Rakgadi (Rakgadi/Rakgadi yo mogolo)	
Oom	a1.1H Mogatša' rakgadi/Rakgadi (Mogatša' rakgadi yo mogolo)	
Vader	A1.2 Tate (Tate)	Tate/Ratswale (Ratswale)
Moeder	k1.2 Mme/Mma (Mme/Mma)	Mme/Mma/Matswale (Mme/ Mma/Matswale)
Vader se vrou	A1.2W2 Mmane/Mmangwane (Mmane/Mmangwane Mogatša' tate wa bobedi)	Mmane/Mmangwane (Mmane/ Mmangwane Mogatša wa ratswale wa bobedi)
Oom	A1.3 Rangwane (Rangwane)	
Tante	A1.3W Mmane/Mmangwane (Mmane/Mmangwane Mogatša rangwane)	Soos vir ego A2.2 met opsionele byvoeging van 'wa mogatšaka' 'van my eggenoot'
Tante	a1.2 Rakgadi (Rakgadi (yo monyane))	

Oom	a1.2H Mogatša' rakgadi/Rakgadi (Mogatša' rakgadi (yo monyane))	
Tante	a1.3 Rakgadi (Rakgadi (yo a latelago rakgadi yo monyane)) (yo monyane)	Soos vir ego A2.2 met opsionele byvoeging van 'wa mogatšaka' 'van my eggenoot'
Oom	a1.3H Mogatša' rakgadi/Rakgadi (Mogatša' rakgadi (yo a latelago rakgadi yo monyane))	

Tabel 1: Verwantskapsterme vir 'n man se vader se broers en susters
(Prinsloo en Van Wyk 1992: 47-48)

In hierdie voorbeeld word die verwantskapstruktuur en verwantskapsterme van 'n man se vader se broers en susters meer gebruikersvriendelik as in figuur 2 aangebied bloot omrede kleiner gedeeltes een vir een aan die orde gestel word. Dit doen nie afbreuk aan die geheelbeeld nie maar bied die ooreenkomste en verskille, veral ten opsigte van hoe die persoon aangespreek word teenoor hoe daar na die persoon verwys moet word in verteerbare eenhede. Die boomdiagram vorm steeds 'n kernelement van die inligtingsaanbod. Die eerste taak in terme van die afbreek van komplekse verwantskapsterminologiestelsels behels dus bestudering van die struktuursamestelling en die identifisering en analise van die boustene, bv. figure 2 en 3.

3. Lemmatisering

Die tweede hoofaktiwiteit lê op grammatikale vlak in terme van die analise, groepering en redusering van sintaktiese strukture en is sonder meer afgestem op lemmatisering. Die benadering hier is gerig op die onervare teikengebruiker wat Sepedi wil leer en 'n omvattende woordeboek vir reseptiewe (dekoderende) sowel as produktiewe (enkoderende) doeleindes wil gebruik. Geen grondige kennis van die grammatika van Sepedi word dus voorveronderstel nie. In terme van die Funksieteorie (Tarp 2008), lê die behoefte hier veral op die vlak van die kognitiewe en kommunikatiewe vaardighede.

3.1 Lemmatisering van enkelwoorde

Die maklikste gevalle is waar die verwantskapsterm deur 'n onafgeleide enkelwoord in die enkelvoud soos *malome* 'oom' en *mokgotse* 'swaer' benoem word. Hierdie verwantskapsterme kan sonder meer gelemmatiseer en leksikografies bewerk word.

Die tweede oorweging raak die lemmatisering van enkelvouds- versus meervoudsvorme van verwantskapsterme. Enkelvouds- en meervoudsvorming word in Sepedi deur prefikse gedoen. Van Wyk (1987) wys daarop dat die

meervoudsprefiks *bo-* van klas 2a, byvoorbeeld *botate* nie sonder meer meervoudigheid aandui nie maar 'n respeksvorm kan wees, of meervoudigheid in die vorm van *vader-hulle* aandui. Tabel 2 illustreer die sisteem vir die eerste agt naamwoordklasse uit die naamwoordklassesisteem van Sepedi.

Klas	***Prefiks***	***Voorbeeld***	***Besitskakel***	***Besitlike voornaamwoord***
1	mo-	monna 'man'	wa	gagwe
2	ba-	banna 'mans'	ba	bona
1a	**ø/N-**	**tate 'vader'**	**wa**	**gagwe**
2b	**bo-**	**botate 'vaders'**	**ba**	**bona**
3	mo-	motse 'stat'	wa	wona
4	me-	metse 'statte'	ya	yona
5	le-	lesogana 'jongman'	la	lona
6	ma-	masogana 'jongmans'	a	ona
7	se-	selepe 'byl'	sa	sona
8	di-	dilepe 'byle'	tša	tšona

Tabel 2: Die eerste agt naamwoordklasse van Sepedi

Enkelvouds- en meervoudsaanduiding deur middel van prefikse is uiteraard problematies deurdat die leksikograaf moet besluit of hy/sy slegs enkelvoude of enkelvoudsvorme sowel as meervoudsvorme wil lemmatiseer. So, byvoorbeeld, moet die leksikograaf besluit of slegs *monna* en *tate*, of beide *monna* en *banna*, en *tate* en *botate* in tabel 2 gelemmatiseer moet word. Laasgenoemde neem uiteraard meer ruimte in die woordeboek in beslag en die lemmas kom in verskillende artikeltrajekte tereg, wat op sy beurt weer kruisverwysings impliseer indien die leksikograaf nie beide die enkelvouds- en meervoudsvorme volledig wil bewerk nie.

Pukuntšu Dictionary (PUKU 1) lemmatiseer beide enkelvouds- en meervoudsvorme. Die gebruiker kan dus *monna* onder *mo-* en *banna* onder *ba-* respektiewelik in die alfabetiese gedeeltes M en B in die woordeboek opsoek. Dieselfde geld vir *mokgotse* 'swaer, skoonsuster' waar ook die meervoudsvorm *bakgotse* 'swaers', en beide *malome* 'oom, ma se broer' en die meervoudsvorm *bomalome* 'ooms, ma se broers' as afsonderlike artikels aangebied word. Hierteenoor word slegs enkelvoudsvorme in *Pukuntšu woordeboek, Noord-Sotho–Afrikaans, Afrikaans–Noord-Sotho* (PUKU 2) gelemmatiseer en word daar van die gebruiker verwag om *banna* onder *monna* op te soek en *botate* onder *tate*. In laasgenoemde geval voorveronderstel die leksikograaf kennis van die naamwoord-

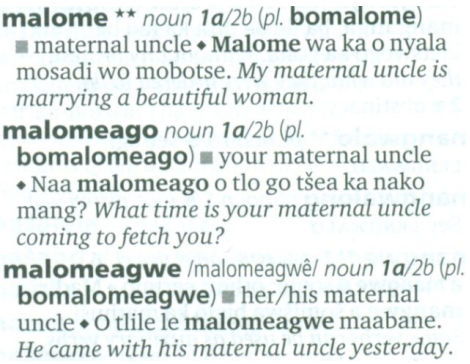
struktuur van die gebruiker maar bied ook sekere hulpmiddels in die voorwerk aan wat die gebruiker kan help om die enkelvoudsvorme te bepaal.

Die derde aspek vir besinning ten opsigte van die lemmatisering van enkelwoorde behels afgeleide of saamgestelde vorme, veral die hoëfrekwensie-afleidings *-ago* 'van jou', *-agwe* ' van hom of haar', *mma-* 'moeder' en *mogatša-* 'eggenoot'. Tabel 3 lys byvoorbeeld die 20 vorme waarin *malome* in die PSC voorkom.

Verwantskapsterm	**Voorkomste in die PSC**	**Vertaling**
bomalome	**20**	ooms
bomalomeabo	1	hulle ooms
bomalomeabona	1	hulle ooms
bomalomeago	**5**	jou ooms
bomalomeagwe	**9**	sy/haar ooms
bomalomeatšona	1	hulle ooms
bommamalome	**1**	die moeders van oom/oom se moeder-hulle
bomogatšamalomeabona	1	eggenotes van hulle oom
gamalome	**5**	by oom se plek
gamalomeagwe	**1**	by sy oom se plek
malomeago	**83**	jou oom
malomeagwe	**201**	sy/haar oom
malomeake	3	my oom
malomealena	1	julle oom
malomeatšona	2	hulle oom
mmamalome	**10**	oom se moeder
mmamalomeagwe	**3**	sy/haar oom se moeder
mogatšamalome	**9**	oom se eggenoot
mogatšamalomeagwe	**9**	sy/haar oom se eggenoot
ngwanamalome	2	oom se kind

Tabel 3: Afleidings van *malome* in die PSC

Die vraag is of hierdie afleidings, byvoorbeeld die paradigma *malomeago* 'jou oom', *malomeagwe* 'sy/haar oom', *mmamalome* 'oom se moeder' en *mogatša-malome* 'oom se vrou' afsonderlik gelemmatiseer moet word. Vergelyk byvoorbeeld *Oxford Bilingual School Dictionary: Northern Sotho and English* (ONSD) ten opsigte van *malomeago* en *malomeagwe* in figuur 4.



malome ** *noun* ***1a***/*2b* (*pl.* **bomalome**)
■ maternal uncle ◆ **Malome** wa ka o nyala mosadi wo mobotse. *My maternal uncle is marrying a beautiful woman.*

malomeago *noun* ***1a***/*2b* (*pl.* **bomalomeago**) ■ your maternal uncle ◆ Naa **malomeago** o tlo go tšea ka nako mang? *What time is your maternal uncle coming to fetch you?*

malomeagwe /malomeagwê/ *noun* ***1a***/*2b* (*pl.* **bomalomeagwe**) ■ her/his maternal uncle ◆ O tlile le **malomeagwe** maabane. *He came with his maternal uncle yesterday.*

Figuur 4: *Malome*, *malomeago* en *malomeagwe* in ONSD

In wese voeg die laaste twee artikels bloot die betekenis 'van jou' en 'van hom/haar' toe. Vir 'n skoolwoordeboek soos ONSD wat op die basis van gebruiksfrekwensie saamgestel is, is dit aanvaarbaar en gebruikersvriendelik. Die leksikograaf kan egter besluit dat hierdie afleidings nie gelemmatiseer moet word nie omdat hulle genoegsaam deursigtig is en ook nie enige morfo-fonologiese veranderinge tot gevolg het nie. Die leksikograaf kan dan die gebruiker tegemoetkom deur die afleidings *-ago* en *-agwe* binne die artikel van die onafgeleide term te bewerk soos wat dit in PUKU 1 gedoen is waar beide die vorme *malomago* en *malomagwe* in die artikel van *malome* bewerk word. Dat die omvang van afleidings van verwantskapsterme met byvoorbeeld *-ago* en *-agwe* baie groot is blyk byvoorbeeld uit die frekwensies van die tien hoogsgebruikte afleidings met *-agwe* in die PSC:

mmagwe (1,483), 'sy/haar moeder'	tatagwe (1,237) 'sy/haar vader'
rragwe (287) 'sy/haar vader'	mogatšagwe (217) 'sy/haar eggenoot'
morwagwe (207) 'sy/haar seun'	malomeagwe (201) 'sy/haar oom'
morwediagwe (163) 'sy/haar dogter'	morwarragwe (146) 'sy/haar broer'
mogolwagwe (109), 'sy/haar neef'	rakgadiagwe (87), 'sy/haar vader se suster'

Die vraag is of daar nie 'n alternatiewe lemmatiseringstrategie vir afgeleide vorme gevind kan word nie en hierdie aspek word in paragraaf 3.3 hieronder verder bespreek.

3.2 Lemmatisering van frases, woordgroepe en grammatikale konstruksies

Wanneer meer ingewikkelde strukture onder die loep geneem word, is dit wenslik om benewens enkelwoordlemmatisering ook ander opsies te oorweeg. Vergelyk die volgende konstruksies in tabel 5 van Prinsloo en Van Wyk (1992: 50) in tabel 4, as basis vir die eersvolgende analise:

Ego A2.2 (male)
A3.1WF Mokgotse (Mokgotse (wa ka) ka morwa yo mogolo)
A3.1WM Mogatša wa Mokgotse (Mogatša wa mokgotse (wa ka) ka morwa yo mogolo)
a3.1HF Mokgotse (Mokgotse (wa ka) ka morwedi yo mogolo)
a3.1HM Mogatša wa mokgotse (Mogatša wa mokgotse (wa ka) ka morwedi yo mogolo)
A3.2WF Mokgotse (Mokgotse (wa ka) ka morwa wa bobedi)
A3.2WM Mogatša wa mokgotse (Mogatša wa mokgotse (wa ka) ka morwa wa bobedi)
a3.2HF Mokgotse (Mokgotse (wa ka) ka morwedi wa bobedi)
a3.2HM Mogatša wa mokgotse (Mogatša wa mokgotse (wa ka) ka morwedi wa bobedi)
A3.3WF Mokgotse (Mokgotse (wa ka) ka morwa wa phejane)
A3.3WM Mogatša wa mokgotse (Mogatša wa mokgotse (wa ka) ka morwa wa phejane)

Tabel 4: Man se eie kinders se skoonouers

In tabel 4 tree *mokgotse* 'swaer' in 'n verskeidenheid konstruksies op waarvan die besitskonstruksie die mees frekwente is byvoorbeeld *mogatša wa mokgotse* (*mogatša* (naamwoord klas 1), *wa* (besitskakel klas 1) en *mokgotse* (naamwoord klas 1)) 'eggenoot van *mokgotse'*. Die leksikograaf het hier te make met die twee sleutelwoorde *eggenoot* en *swaer* en kan wel besluit om *mogatša wa mokgotse* as multiwoord lemma op te neem. Dit sou byvoorbeeld 'n opsie wees in grondslagfase woordeboeke waar daar nie noodwendig aanvaar kan word dat die gebruiker vertroud is met die besitskonstruksie in Sepedi nie.

'n Tweede opsie is om van die standpunt uit te gaan dat die struktuur *eggenoot van iemand* deursigtig is wat nóg die aanbod as multiwoord lemma nóg enige verdere gebruikersleiding vereis. Dit sou byvoorbeeld gepas wees vir meer gevorderde aanleerders van die taal vir wie beide die besitskonstruksie en waarskynlik ook die betekenis van die hoëfrekwensiewoord *mogatša* (228 voorkomste in die PSC) bekend is.

Die leksikograaf word egter gekonfronteer met 'n verskeidenheid van nog meer komplekse strukture, byvoorbeeld a.3.2HM waar verwysing deur 'n manlike persoon na sy tweede dogter se man se moeder ter sprake is:

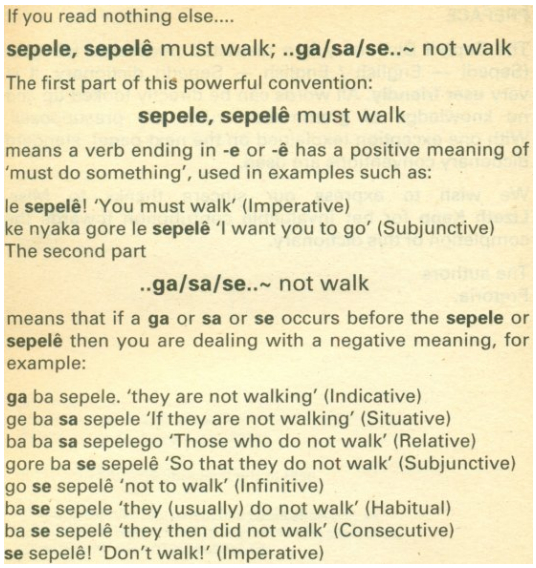
> *Mogatša wa mokgotse (wa ka) ka morwedi wa bobedi*
> (eggenoot van swaer (van my) met dogter van tweede)
> 'my tweede dogter se eggenoot se moeder'

Hier het die leksikograaf met drie besitskonstruksies (die eerste twee in tandem) asook met die gebruik van die instrumentalis (*ka* 'met') en 'n rangtelwoord (*bobedi* 'tweede') te make. Die vraag is hoe so 'n komplekse struktuur gelemmatiseer moet word. Die leksikograaf sal met reg huiwer om hierdie voorbeeld as 'n nege-woord lemma op te neem maar besef wel die behoefte aan sterker gebruikersleiding as die blote lemmatisering van enkelwoorde uit die frase. Dit kan ook geargumenteer word dat die struktuur, hoewel baie kompleks, steeds volkome deursigtig is en dat daar gevolglik op enkelwoordlemmas besluit kan word. Dit sou egter beteken dat die onervare gebruiker tot soveel as nege lemmas sou moes opsoek ten einde die betekenis van *mogatša wa mokgotse (wa ka) ka morwedi wa bobedi* te konstrueer.

Indien al die terme in tabel 4 in die artikel van *mokgotse* bewerk sou word, sal dit tot 'n baie lywige artikel vir *mokgotse* lei. Soortgelyke lemmatisering van die ander komplekse konstruksies in die verwantskapsterminologiesisteem sal nie haalbaar wees in terme van beskikbare spasie in 'n papierwoordeboek nie. Net soos in die geval van afgeleide enkelwoordvorme in 3.1 hierbo is die vraag of daar nie 'n alternatiewe oplossing vir die lemmatisering van frases, woordgroepe en grammatikale konstruksies gevind kan word nie.

3.3 'n Leksikografiese konvensie vir die lemmatisering van verwantskapsterme

'n Oplossing vir die herhalende patrone in 3.2 asook vir die talle enkelwoord-afleidings wat in 3.1 en in tabel 3 ten opsigte van *malome* bespreek is, kan gevind word in die samestelling van 'n leksikografiese konvensie vir verwantskaps-terminologie. Só 'n konvensie is 'n kompakte substitusiestruktuur vir die redusering van strukture soortgelyk aan die sogenaamde *ga/sa/se* konvensie van Prinsloo en Gouws (1996) wat sedertdien in ONSD, *Popular Northern Sotho Dictionary* (POP) en *New Sepedi Dictionary, English–Sepedi* (NSD) gebruik word vir werkwoordstamme wat op 'n *-e* eindig. Die doel is om 'n kragtige konvensie te ontwerp wat vir 'n aantal herhalende partone voorsiening maak. Só 'n konstruksie moet verstaanbaar vir die teikengebruiker wees en in die gebruikersgids van die woordeboek verduidelik word. Vergelyk, byvoorbeeld, die bekendstelling en uitvoerige verduideliking van hierdie konvensie in die voorwerk van NSD, en ook die groot aantal strukture wat deur dié konvensie ondervang word, in figuur 5.



If you read nothing else....

sepele, sepelê must walk; **..ga/sa/se..~** not walk

The first part of this powerful convention:

sepele, sepelê must walk

means a verb ending in **-e** or **-ê** has a positive meaning of 'must do something', used in examples such as:

le **sepelê**! 'You must walk' (Imperative)
ke nyaka gore le **sepelê** 'I want you to go' (Subjunctive)

The second part

..ga/sa/se..~ not walk

means that if a **ga** or **sa** or **se** occurs before the **sepele** or **sepelê** then you are dealing with a negative meaning, for example:

ga ba sepele. 'they are not walking' (Indicative)
ge ba **sa** sepele 'If they are not walking' (Situative)
ba ba **sa** sepelego 'Those who do not walk' (Relative)
gore ba **se** sepelê 'So that they do not walk' (Subjunctive)
go **se** sepelê 'not to walk' (Infinitive)
ba **se** sepele 'they (usually) do not walk' (Habitual)
ba **se** sepelê 'they then did not walk' (Consecutive)
se sepelê! 'Don't walk!' (Imperative)

Figuur 5: Gebruikersleiding en aanduiding van die reikwydte van die *ga/sa/se* konvensie in NSD

Vir die lemmatisering van verwantskapsterme word gepoog om 'n enkele konvensie te ontwerp ten einde 'n komplekse stel enkelwoordafleidings, asook 'n komplekse stel frases sinvol te lemmatiseer. Net soos in die geval van die ontwerp van die *ga/sa/se* konvensie is dit moontlik om 'n konvensie te ontwerp wat voorsiening maak vir alle elemente van deursigtige herhalingspatrone soos *mmago-*, *mmelega-*, *-ake*, *-ago*, *-agwe* en besitskonstruksies soos *mogatša (wa)* ~. Die eerste stap is om die elemente te lys wat kandidate is vir opname in die konvensie.

ga/bo/mma/mogatša (wa)/ngwana (wa)/	Elemente wat voor die lemma optree
~	Die lemma bv. *malome*
/ago/agwe/abo(na)/ake/atšo(na)/alena/ayo(na)/ aona/asona/yo mogolo/wa bobedi/phejane/ mafelelo	Elemente wat na die lemma optree

Tabel 5: Kandidaatelemente vir opname in 'n leksikografiese konvensie vir verwantskapsterme

Die struktuur in tabel 5 maak voorsiening vir 'n wye verskeidenheid enkelwoordafleidings en frases. So, byvoorbeeld, ondervang tabel 5 alle afleidings van *malome* in tabel 3 hierbo asook die grootste gedeelte van strukture in tabel 4 en die volledige paradigma van besitlike voornaamwoorde. Hierdie struktuur is egter te lywig, onprakties en gebruikersonvriendelik vir 'n leksikografiese konvensie. Net soos in die geval van die ontwerp van die *ga/sa/se* konvensie (Prinsloo en Gouws: 1996) is dit nodig om die reeks elemente in tabel 5 te verminder op grond van frekwensie tot 'n voorgestelde:

ga/bo/mma/mogatša/~/ake/ago/agwe/abo(na)/ake/atšo(na)/alena

of 'n verder gereduseerde:

bo/mma/mogatša/~/ago/agwe

Die leksikograaf kan besluit watter konvensie die mees doeltreffendste vir die tipe gebruiker en die spesifieke woordeboek sal wees.

4. Die leksikografiese bewerking van verwantskapsterme in bestaande Sepedi woordeboeke

Uitvoerige besinning ten opsigte van leemtes in die lemmatisering en bewerking van verwantskapsterminologie is nie moontlik binne die bestek van 'n artikel nie en slegs die mees problematiese aspekte word hier onder die loep geneem. Die eerste leemte raak die lemma-aanbod deurdat daar in die meeste woordeboeke oënskynlik geen seleksiestrategie vir die opname van verwant-

skapsterme is nie en terme gevolglik maar op sterkte van die intuïsie van die leksikograaf opgeneem is. In tabel 6 word enkelwoord verwantskapsterme uit Prinsloo en Van Wyk (1992) gelys wat in die PSC voorkom, en hulle opname (√), of weglating (x), uit *Groot Noord-Sotho-woordeboek, Noord-Sotho–Afrikaans/ Engels* (GNSW), *Sesotho sa Leboa/English Pukuntšu Dictionary* (SLEPD), ONSD, *Pukuntšutlhaloši ya Sesotho sa Leboa* (PTLH), PUKU 1 en POP. Terme wat in hakies gegee word, is alternatiewe spellings.

Term	Vertaalekwivalent	PSC	GNSW	SLEPD	ONSD	PTLH	PUKU 1	POP
kgaitšedi/ (kgaetšedi)	broer of suster	38	√ / √	x/√	x/ √	x/ √	√ /√	√/√
koko	ouma	223	√	√	√	√	√	√
malome	oom	558	√	√	√	√	√	√
matswale	skoonmoeder	14	√	x	x	x	x	√
mma	moeder	1060	√	**X**	√	√	√	√
mmakgolo	ouma	3	x	x	x	x	√	x
mmamalome	oom se moeder	11	x	x	x	x	x	x
mmamogolo	moeder se ouer suster	57	√	√	x	√	√	√
mmane	moeder se jonger suster	61	√	√	√	√	√	√
mmangwane	moeder	39	√	√	x	√	√	√
mmatswale	skoonmoeder	75	√	√	√	√	x	√
mme	moeder	5524* >100	√	√	√	**X**	√	√
mogadibo	broer se vrou	82	√	√	√	√	√	√
mogadikana / (mogadikane)	medevrou	3	x/√	√/√	x/x	√/√	x/√	x/√
mogaditšong	medevrou	61	√	√	x	√	√	√
mogatšaka	my eggenoot	188	x	x	√	x	x	√
mogatšake	my eggenoot	22	x	x	x	x	x	x
mogolle	ouer broer/suster	90	√	**X**	√	**X**	√	√
mogolwake	my ouer broer/suster	39	x	x	x	x	x	x
mokgonyana	skoonseun	260	√	√	√	√	√	√
mokgotse	swaer/skoonsuster	189	√	√	√	√	√	√
molamo	swaer	315* >100	√	√	√	√	√	√
monyanana	jonger broer of suster	28	√	x	x	√	√	√
moratho	jonger broer of suster	114	√	√	√	√	√	√
morwa	seun	3803	√	√	√	√	√	√
morwaka	my seun	7	x	x	x	x	x	√
morwake	my seun	45	x	√	x	√	x	√
morwarre	broer	127	√	√	√	√	√	√
morwedi	dogter	1000	√	√	√	√	√	√
morwediaka	my dogter	7	x	x	x	x	x	x
morwediake	my dogter	34	x	√	x	√	x	x
motlogolo	kleinkind	177	√	√	√	√	√	√
motswala	neef/niggie	120	√	√	√	√	√	√
motswalake	my neef/niggie	14	x	x	x	x	x	√
ngwanangwanake/ ngwanangwanaka	my kleinkind	5	x/x	x/√	x/x	x/√	x/x	x/x
ngwanego	jou familielid	3	x	x	x	x	x	x

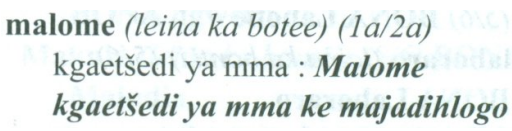
ngwetši	skoondogter	637	√	√	√	√	√	√
nnake	my jonger broer/ suster	8	√	√	x	√	√	x
Rakgadi	vader se suster	348	√	√	√	√	√	√
Rakgolo	grootvader	190	√	√	√	√	√	√
rakgolokhukhu	voorvader	8	√	√	x	√	√	x
ramogolo	vader se ouer broer	114	√	√	√	√	√	√
rangwane	vader se jonger broer	354	√	√	√	√	√	√
ratswale	skoonvader	26	√	√	x	√	√	√
rra	vader	55	x	√	x	x	√	√
Samma	jonger broer of suster	40	√	√	x	√	√	√
tate/tata	vader	1517	√/√	**X**/√	√/x	√/x	√/√	√/√
tatemogolo	oupa	9	x	x	x	x	x	x

* geskatte frekwensie van verwantskapsterm in geval van homonieme word aangedui deur ">"
Vetgedrukte hoofletters '**X**' veronderstel belangrike leemtes ten opsigte van die lemmatisering van verwantskapsterme.

Tabel 6: Frekwensie in die PSC en opname in GNSW, SLEPD, ONSD, PTLH, PUKU 1 en POP van enkelwoord verwantskapsterme in Prinsloo en Van Wyk (1992)

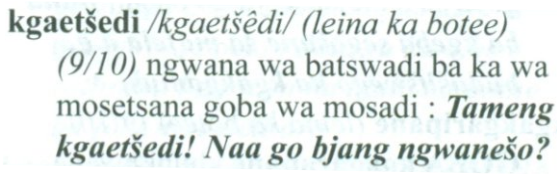
Nie alle weglatings, in vergelyking met die ander woordeboeke moet as leemtes beskou word nie. So, byvoorbeeld, is die weglating van *matswale*, *mmakgolo*, *ratswale*, ens. uit ONSD regverdigbaar in terme van gebruiksfrekwensie wat as kriteria vir opname in dié woordeboek dien. Weglating van *mma* en *tate*, uit die SLEPD, met 'n gebruiksfrekwensie van respektiewelik 1,060 en 1,517 in die PSC, kan egter as 'n leemte of fout beskou word. Net so is *mogolle* opvallend afwesig in SLEPD en PTLH.

Die tweede leemte geld die kwaliteit van bewerking van verwantskapsterme wat wel in die woordeboek opgeneem is. In baie gevalle is die gebruikersleiding wat woordeboeke ten opsigte van verwantskapsterme bied nie duidelik nie of selfs misleidend. Die inligtingsaanbod is gewoon nie genoegsaam kontrasterend, gedetailleerd of diskriminerend nie. So, byvoorbeeld, vind die gebruiker wat die woord *malome* in PTLH opsoek die betekenis *kgaetšedi ya mma* 'moeder se broer'.


malome *(leina ka botee) (1a/2a)*
kgaetšedi ya mma : ***Malome***
kgaetšedi ya mma ke majadihlogo

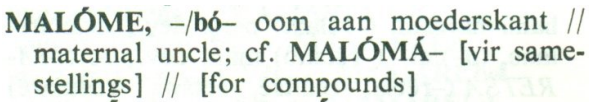
Figuur 6: *Malome* in PTLH

Die gebruiker wat onseker is oor die presiese betekenis van die ander verwantskapsterm in die betekenisparafrase, *kgaetšedi*, soek dit vervolgens in dieselfde woordeboek op:


kgaetšedi */kgaetšêdi/ (leina ka botee)*
(9/10) ngwana wa batswadi ba ka wa mosetsana goba wa mosadi : ***Tameng kgaetšedi! Naa go bjang ngwanešo?***

Figuur 7: *Kgaetšedi* in PTLH

Uit die betekenis "n vroulike kind van my ouers' (*ngwana wa batswadi ba ka wa mosetsana goba wa mosadi*) kan die onervare gebruiker foutiewelik konkludeer dat *malome* verwys na "n vroulike kind van my ouers van my ma' in plaas van 'moeder se broer'. SLEPD vaar nie veel beter nie deur bloot *malome* met *uncle* te vertaal sonder enige verwysing na moederskant. POP en GNSW doen beter met die vertaling van *malome* as *my maternal uncle* 'my oom aan moederskant'. Die GNSW-artikel word in figuur 8 aangegee.


MALÓME, –/**bó**– oom aan moederskant // maternal uncle; cf. **MALÓMÁ**– [vir samestellings] // [for compounds]

Figuur 8: *Malome* in GNSW

In POP en SLEPD word die vertaalekwivalente *malome*, *ramogolo*, *rangwane* vir die lemma *uncle* aangegee. Meer gebruikersleiding soos *moeder se broer*, *vader se ouer/jonger broer* is hier noodsaaklik vir beide teksresepsie en teksproduksie.

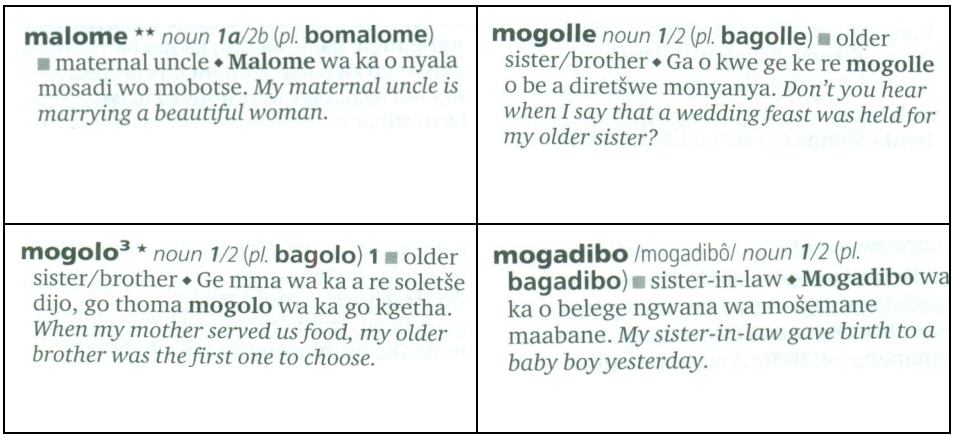
SLEPD gee ook die vertaalekwivalente vir *kgaetšedi* as 'my sister, sister' terwyl PUKU 1 die vertaalekwivalente 'jonger suster van 'n broer' en 'jonger broer van 'n suster' aangee. Die gebruiker is dus nie seker of *kgaetšedi* slegs na 'suster' of na 'suster of broer' kan verwys nie.

Volgens GNSW beteken *mogadibo* 'broer se vrou' maar is dit 'n term wat slegs deur sy suster gebruik word. SLEPD en ONSD gee egter sonder enige beperking die Engelse vertaling *sister in law* 'skoonsuster' aan. Hierteenoor gee PUKU 1 naas 'skoonsuster' ook 'skoondogter' as vertaling aan. Dit kan ook die gebruiker verwar.

Die inligtingsaanbod ten opsigte van 'n spesifieke lemma verskil uiteraard in woordeboeke maar die vraag is of die leksikograaf in hierdie gevalle behoorlike voorafstudie van die terme gedoen het.

Die leksikograaf moet ook seker maak dat die inligtingsaanbod by die meervoudsvorm nie verskil van dié by die enkelvoudsvorm nie. So, byvoorbeeld, gee PUKU 1 by *mokgotse* ook skoonsuster as vertaalekwivalent maar nie in die artikel van die meervoudslemma *bakgotse* nie.

Dit is ten slotte opmerklik dat die gebruikersvoorbeelde dikwels nie bydra tot die semantiese kommentaar nie. Vergelyk die volgende voorbeelde uit ONSD:



Figuur 9: *Malome*, *mogolle*, *mogolo* en *mogadibo* in ONSD

Meer as 50%–80% van die woordeboekspasie vir die artikels van *malome*, *mogolle*, *mogolo* en *mogadibo* dien nie werklik enige doel ter verheldering van die betekenis nie want in al vier gevalle kan die betrokke lemma bloot met enige ander toepaslike naamwoord vervang word. Dit is wat Rundell (Atkins et al. 1997) lys as een van die tipiese kenmerke van swak voorbeelde naamlik "are natural, typical ... but completely pointless (e.g. Sicilian: a quarrel between two Sicilians)". Selfs die MED se bewerking van *uncle* stel in hierdie opsig teleur deurdat die gebruiksvoorbeeld *the business was owned by my uncle* nie tot betekenisverheldering bydra nie:

> **uncle** ˈʌŋk(ə)l noun [countable]
> the brother of one of your parents, or the husband of your AUNT . You are his **niece** or **nephew**
> The business was owned by my uncle.
> a letter from Uncle Richard
> **a** used by children in front of the name of a man who is a close friend of their parents
> How long have you known Uncle Phil and Auntie Carol?

Hier kan *uncle* netsowel met enige ander naamwoord vervang word. MED is wel voortreflik deurdat Ego in die bewerking van die lemma in ag geneem word dit wil sê in watter verhouding *ek* tot *uncle* staan, naamlik sy neef of niggie. Dit dra veel meer by tot die betekenisinligting as die gegewe gebruikersvoorbeeld. Weglating van die gebruikersvoorbeeld in bogenoemde gevalle

doen weinig afbreuk aan die kwaliteit van die artikel en die leksikograaf kan dit selfs oorweeg om die verhouding waarin die persoon tot die lemma staan as die tipiese gebruiksvoorbeeld aan te bied by verwantskapterme, bv. *ek is my oom se niggie* en *my broer is sy neef*.

5. Samevatting

Die Sepedi leksikograaf is terdeë bewus van sy/haar rol as tussenganger tussen komplekse grammatikale strukture in die taal aan die een kant en veral die onervare woordeboekgebruiker aan die ander kant. Sekere strukture in die taal soos verwantskapsterminologie, die modale vorme van die werkwoord en kopulatiewe vereis innoverende lemmatiseringstrategieë ten einde die gebruiker in staat te stel om die verlangde inligting in die woordeboek te vind. Die aard en omvang van verwantskapsterme in Sepedi vra ook meer as die blote lemmatisering op lukrake wyse en op sterkte van die intuïsie van die leksikograaf, of die blote gebruik van 'n frekwensieafsnypunt in 'n Sepedi korpus. Wat die lemmatisering van verwantskapsterminologie betref, is aangetoon dat dit problematies is, veral ten opsigte van die lemmatisering van samestellings, oftewel afgeleide enkelwoordvorme en die lemmatisering van frases soos veelvuldige besitskonstruksies. Die leksikograaf moet in die eerste plek alle leksikografiese bewerkingstrategieë soos vertaalekwivalente, parafrases (definisies), gebruiksvoorbeelde en kruisverwysings na buitetekste en eksterne bronne maksimaal benut. Vir die leksikografiese bewerking van komplekse strukture soos verwantskapsterminologie is dit egter nodig om verdere gebruikersleiding te gee, veral ten opsigte van die toegangstruktuur vir afgeleide vorme en verwantskapsterme wat as multiwoord konstruksies voorkom — derhalwe die aanbod van 'n leksikografiese konvensie vir die lemmatisering van verwantskapsterme in omvattende woordeboeke vir onervare gebruikers wat in hierdie artikel voorgestel word.

6. Literatuurlys

A. Woordeboeke

(GNSW) Ziervogel, D. en P.C.M. Mokgokong. 1975. *Pukuntšu ye kgolo ya Sesotho sa Leboa, Sesotho sa Leboa–Seburu/Seisimane/Groot Noord-Sotho-woordeboek, Noord-Sotho–Afrikaans/Engels/Comprehensive Northern Sotho Dictionary, Northern Sotho–Afrikaans/English.* Pretoria: J.L. van Schaik.

(MED) Rundell, M. (Red.). 2002. *Macmillan English Dictionary for Advanced Learners*. CD-ROM. Londen: Macmillan Education. Tweede uitgawe 2007 op *CD-ROM*. Londen: A&C Black Publishers.

(NSD) Prinsloo, D.J. en B.P. Sathekge. 1996. *New Sepedi Dictionary, English–Sepedi (Northern Sotho), Sepedi (Northern Sotho)–English.* Pietermaritzburg: Shuter & Shooter.

(ONSD) De Schryver, G.-M. et al. (Reds.). 2008. *Pukuntšu ya Polelopedi ya Sekolo. Sesotho sa Leboa le Seisimane. E gatišitšwe ke Oxford/Oxford Bilingual School Dictionary. Northern Sotho and English.* Kaapstad: Oxford.

(POP) Kriel, T.J., D.J. Prinsloo en B.P. Sathekge. 1997. *Popular Northern Sotho Dictionary, Northern Sotho–English, English–Northern Sotho*. Kaapstad: Pharos.

(PTLH) Mojela, M.V. (Red.). 2007. *Pukuntšutlhaloši ya Sesotho sa Leboa*. Pietermaritzburg: Nutrend.

(PUKU 1) Kriel, T.J. 1983. *Pukuntšu Dictionary*. Pretoria: J.L. van Schaik.

(PUKU 2) Kriel, T.J. en E.B. van Wyk. 1989. *Pukuntšu woordeboek, Noord-Sotho–Afrikaans, Afrikaans–Noord-Sotho*. Pretoria: J.L. van Schaik.

(SLEPD) Mojela, M.V., M.C. Mphahlele, M.P. Mogodi en M.R. Selokela. 2006. *Sesotho sa Leboa/ English Pukuntšu Dictionary.* Kaapstad: Phumelela.

B. Ander literatuur

Atkins, B.T. Sue, M. Rundell en E. Weiner. 1997. *Salex'97. A Training Course in the Compilation of Monolingual Dictionaries.* Ongepubliseerde kursusmateriaal van 'n tutoriaal gehou by die Woordeboekeenheid vir Suid-Afrikaanse Engels, Rhodes Universiteit, Grahamstad, 15–26 September 1997.

Prinsloo, D.J. en Sonja E. Bosch. 2012. Kinship Terminology in English–Zulu/Northern Sotho Dictionaries — A Challenge for the Bantu Lexicographer. Fjeld, Ruth Vatvedt en Julie Matilde Torjusen (Reds.). 2012. *Proceedings of the 15th Euralex International Congress*, 7–11 August 2012: 296-303. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.

Prinsloo, D.J. en R.H. Gouws. 1996. Formulating a New Dictionary Convention for the Lemmatization of Verbs in Northern Sotho. *South African Journal of African Languages* 16(3): 100-107.

Prinsloo, D.J. en J.J. van Wyk. 1992. Verwantskapsterminologie van die Noord-Sotho. *South African Journal of Ethnology* 15(2): 43-58.

Tarp, Sven. 2008. *Lexicography in the Borderland Between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Lexicographica. Series Maior 134. Tübingen: Max Niemeyer.

Van Wyk, E.B. 1987. Proclitic *bo* of Northern Sotho. *South African Journal of African Languages* 7(1): 34-42.

Devices for Information Presentation in Electronic Dictionaries*

D.J. Prinsloo, *Department of African Languages, University of Pretoria, Pretoria, South Africa (danie.prinsloo@up.ac.za)*,

Ulrich Heid, *Department of Information Science and Natural Language Processing, Hildesheim University, Hildesheim, Germany and Department of African Languages, University of Pretoria (heid@uni-hildesheim.de)*,

Theo Bothma, *Department of Information Science, University of Pretoria (theo.bothma@up.ac.za)*
and
Gertrud Faaß, *Department of Information Science and Natural Language Processing, Hildesheim University, Hildesheim, Germany and Department of African Languages, University of Pretoria (gertrud.faass@uni-hildesheim.de)*

Abstract: Electronic dictionaries should support dictionary users by giving them guidance in text production and text reception, alongside a user-definable offer of lexicographic data for cognitive purposes. In this article, we sketch the principles of an interactive and dynamic electronic dictionary aimed at text production and text reception guiding users in innovative ways, especially with respect to difficult, complicated or confusing issues. The lexicographer has to do a very careful analysis of the nature of the possible problems to suggest an optimal solution for a specific problem. We are of the opinion that there are numerous complex situations where users need more detailed support than currently available in e-dictionaries, enabling them to make valid and correct choices. For highly complex situations, we suggest guidance through a decision tree-like device. We assume that the solutions proposed here are not specific to one language only but can, after careful analysis, be applied to e-dictionaries in different languages across the world.

Keywords: ELECTRONIC DICTIONARIES; USER GUIDANCE; TEXT PRODUCTION; TEXT RECEPTION; DICTIONARY DESIGN, DECISION TREE STRUCTURE, COPULATIVES, KINSHIP TERMINOLOGY, INFORMATION PRESENTATION DEVICES

Opsomming: Inligtingsaanbiedingsinstrumente in elektroniese woordeboeke. Elektroniese woordeboeke behoort woordeboekgebruikers te ondersteun deur hulle te lei ten opsigte van teksproduksie en teksresepsie volgens 'n gebruikergedefinieerde aanbod van leksi-

* This article is an extended and reworked version of a paper presented at the *Second Conference on Electronic Lexicography, Electronic Lexicography in the 21st Century — New Applications for New Users*, eLex-2011, in Bled, Slovenia, 10–12 November 2011 (cf. Prinsloo et al. (2011)).

kografiese data vir kognitiewe doeleindes. In hierdie artikel skets ons die beginsels waarop 'n interaktiewe en dinamiese elektroniese woordeboek berus, gemik op teksproduksie en teksresepsie wat die gebruikers op innoverende wyse lei, veral ten opsigte van moeilike, gekompliseerde of verwarrende aspekte. Die leksikograaf is genoodsaak om 'n noukeurige analise te doen van die aard van moontlike probleme ten einde 'n optimale oplossing aan te bied vir 'n spesifieke probleem. Ons is van mening dat daar verskeie komplekse gevalle bestaan waar gebruikers meer gedetailleerde ondersteuning benodig as wat tans in e-woordeboeke beskikbaar is ten einde korrekte keuses te kan maak. Vir hoogs problematiese situasies stel ons leiding deur middel van 'n keuseboom-tipe instrument voor. Ons veronderstel dat die oplossings wat hier voorgestel word nie tot 'n enkele taal beperk is nie, maar na versigtige analise op verkillende tale van die wêreld toegepas kan word.

Sleutelwoorde: ELEKTRONIESE WOORDEBOEKE, GEBRUIKERSLEIDING, TEKSPRODUKSIE, TEKSRESEPSIE, WOORDEBOEKONTWERP, KEUSE-BOOMSTRUKTUUR, KOPULATIEWE, VERWANTSKAPSTERMINOLOGIE; INLIGTINGSAANBIEDINGSINSTRUMENTE

1. Introduction

In lexicography, the electronic era was met with great enthusiasm and expectations. Early publications on electronic dictionaries were all about the potential of the new medium and the expected revolution it would bring along, thereby antiquating the paper dictionary in a decade or two. De Schryver (2009), however, rightfully expresses disappointment in respect of the pace of development of electronic dictionaries. More exciting was the introduction of what could be called "true electronic features" such as pop-up boxes, audible pronunciation and sophisticated search features. Some electronic dictionaries also solve lemmatisation problems, which cannot be resolved in paper dictionaries. Typical examples are *isiZulu.net*[1] where Zulu words can be looked up without prior stem identification by simply typing in the word, or commercial products such as Amazon Kindle e-books[2] that link inflected word forms to lemmatised forms, e.g. *went* to *go* or German *ging* to *gehen*.

Electronic dictionaries of today, however, could enter a more advanced dimension in fulfilling more sophisticated needs of the users, e.g. through multiple access routes. Rundell (2009) refers to "game changing" developments that have "expanded the scope of what dictionaries can do and (in some respects) changed our view of what dictionaries are for". De Schryver (2009) calls in this context for an adaptive and intelligent dictionary (aiLEX) that will be able to "study and understand its user" and consequently "present itself to that user". "Intelligent" dictionaries in this sense are currently outside the scope of this article and of our approach: "intelligent" probably assumes a greater or lesser use of artificial or computational intelligence in the underlying programming of the dictionary, which is a realistic expansion of e-dictionaries but not one we are currently planning.

Our focus is more on supporting dictionary users by giving them guidance in text production and text reception, alongside a user-definable offer of lexicographic data for cognitive purposes. In most cases, what is currently offered in dictionaries

claiming that they give guidance in text production, is in fact still more on the level of text reception, or it leads to an overload of information, or the information provided is not appropriate for a given situation. We thus intend to sketch the principles of an interactive and dynamic electronic dictionary aimed at text production and text reception, which guides the user in innovative ways, especially with respect to difficult, complicated or confusing issues. What is at stake in the medium term are improvements in access possibilities of electronic dictionaries, including (but not discussed in this paper) access to more information presented by linking the dictionary with other online sources, cf. Heid, Prinsloo and Bothma (2012).

Our approach will result in dictionaries which are different from existing ones, and where the individualised lexical offer for a specific user in a specific usage situation and with a specific information need is paramount. In our view, it is important that a dictionary provides features that allow an individual user to get access, in the most intuitive way, to exactly the bit of information he/she needs in a given situation. This does not mean that predefined typical user needs as stated in the Function Theory of lexicography (cf. e.g. Tarp 2008a) are not used — our approach simply goes one step further. The process of individualising the lexical offer which we sketch in this article follows the principles of the Function Theory and adds at the same time the concept of dynamic interaction, as Tarp (2009a: 292) states, "in order to conceive dictionaries capable of meeting all the users' needs in specific types of situations". The improvements envisaged are not limited to adding more sophisticated search options or to the increased utilisation of pop-up windows — it will, in fact, utilise totally different concepts such as a step-by-step text production guidance and interactive selection processes.

These processes will be illustrated by means of examples ranging from fairly simple to highly complex from different languages, viz. Germanic, Romance and African, where the user can be guided to the correct text production or text reception choices by means of devices ranging from simple examples to complex decision trees.

We do not assume that the type of solutions we offer below are applicable to all lexical items in all text production or text reception information needs — the lexicographer will have to do a very careful analysis of the nature of the possible problems to suggest an optimal solution for a specific problem, and in many cases users may not need complex solutions. However, we are of the opinion that there are numerous complex situations where users need more detailed support than currently available in e-dictionaries, to make valid and correct choices. We assume that the solutions proposed here are not specific to one language only but can, after careful analysis, be applied to e-dictionaries in many different languages across the world.

There are two possible approaches to provide lexicographic support in the above-mentioned situations, viz. a "stand-alone" dictionary that will be consulted as an information tool in its own right, or the integration of the lexicographic tool into a text production environment, for example into the user's word processor. Depending on the point of departure, the user could either

find the information he/she is looking for by direct/intended dictionary-lookup or during the actual process of typing text, for example in a situation where the user can move the cursor in the text he/she is creating from word to word to get basic information.

To set up a database that could provide access to the data needed for the processes described briefly above requires a very careful planning and design of the database. The first step, however, would be to select a number of examples and to describe these in detail, deciding on the nature of the solution to be offered in each case, for example, in a text production or in a text reception situation. Once this is described in detail, it will have to be formalised according to an XML schema, a database structure or another data representation format which will need to be designed in such a way that it will make provision for different levels of complexity and detail. Careful attention will have to be paid to the granularity of the data, to enable the user to easily navigate complex steps, and in the process not to present the user with unnecessary or confusing data — only the data required at each step in the decision process are to be presented to the user at any given time. Designing such a database and the schemas associated with it is a complex process which will not be addressed in this paper. After a brief reference to the Function Theory of lexicography (section 2), we will provide a short and preliminary survey and classification of devices for information access in electronic dictionaries (section 3). We will then address new devices for information access which we conceive of as being appropriate for the purpose of providing adequate information for different use situations, viz. tables, diagrams, guidance paths and interactive decision trees. All these devices will be discussed in some depth (section 4), before a detailed example of decision support for a complex text production problem will be given (section 5). We conclude in section 6 and point to intended future work.

2. Dictionaries as language information tools: the views of Function Theory

The work on information presentation in electronic dictionaries presented in this article will make use of elements of the Function Theory of lexicography as proposed, inter alia, by Bergenholtz and Tarp (cf. Bergenholtz and Tarp 2002, 2003, 2004, 2005, Tarp 2008 and 2008a, 2009 and 2009b, 2011, 2012), which claims that "dictionaries and other lexicographical work are above all *utility tools* conceived and produced with the genuine purpose of satisfying specific types of human needs, i.e. *information needs*, existing in one or several individuals in society" (Bothma and Tarp 2012: 89). The Function Theory currently works with four types of lexicographically relevant situations (cf. Tarp 2008a), viz. communicative, cognitive, operative and interpretive situations.

In this article we refer only to communicative and cognitive situations. Communicative situations may be further subdivided into a number of situations such as text production and text reception, text translation, text revision etc. Cognitive situations may also be divided into various sub-situations and refer to situations where the user may need to acquire specific knowledge to

perform a task. Operative and interpretive situations refer respectively to situations where a user needs instructions on how to perform a physical or mental action and to situations where a need exists to interpret and understand a sign, signal, symbol etc. For a brief summary of the Function Theory, see Bothma and Tarp (2012, section 2).

3. A classification of devices for information access in electronic dictionaries

In this section, we make an attempt at provisionally classifying the devices available to lexicographers who design electronic dictionaries, for giving e-dictionary users access to lexicographic data. In our view, three types of devices need to be distinguished: contents-related devices, navigational devices and presentational devices. In section 4, we will propose a few presentational devices in addition to the standard ones briefly mentioned in section 3.3, which in our view may considerably improve the effectiveness and efficiency of electronic dictionaries, at least for certain types of phenomena.

Contrary to the situation for printed dictionaries, where both contents-related and navigational devices (i.e. items ("Angaben") and text structure markers ("Strukturanzeiger")) have been discussed in detail (Wiegand 1989), we are not aware, as yet, of an inventory or a state of the art description of the devices commonly used in electronic dictionaries. Unfortunately, we will not be able, in this paper, to even outline such a state of the art; our overview will thus have to remain sketchy. There are neither any standards nor guidelines that would suggest which devices to use for what purpose, for which users or which types of phenomena to be explained in the dictionary.

3.1 Contents-related lexicographic devices

Electronic dictionaries have inherited from printed dictionaries a number of well-known lexicographic devices such as paraphrase of meaning (definitions), examples of usage, pronunciation guidance, part of speech, etc. Such devices are commonly used in many printed and electronic dictionaries, and will not be discussed any further.

3.2 Navigation devices

In addition, electronic dictionaries make use of general design principles of web design, for example with respect to navigation devices: this includes features such as scrolling facilities, clickable headlines, site maps etc. *Wiktionary*[3], the online dictionary that belongs to *Wikipedia*, as well as the *Oxford English Dictionary Online*[4] provide examples of site maps for dictionary entries: they offer an overview of their articles in terms of clickable headlines (cf. Bothma 2011: 84). Again, these devices are also commonly in use, and will hence not be discussed any further.

3.3 Devices for information presentation

There are several devices for on-demand data provision, such as expandable or collapsible indications, cross-referencing by means of links, mouse-sensitive areas which display data when being pointed at with the mouse cursor, or pop-up boxes with lexicographic data that can be opened by pointing the mouse cursor at them and clicking. All of these provide information to the user if he/she activates them. Such devices allow for the personalisation of the lexical offer, in so far as they allow the user (at least to some extent) to decide himself or herself, how much he/she wants to see.

3.3.1 Mouse-over

An example of data provision on demand based on the mouse-over feature in an existing e-dictionary is the mouse-sensitive behaviour of the subcategorisation indications in ELDIT, the electronic learners' dictionary for German and Italian (cf. Abel 2002) realised at EURAC in Bolzano[5]: it displays subcategorisation patterns in a terminology which deliberately avoids linguistic terminology (assuming that users are not familiar with terms like "subject", "controlled infinitival" etc.) as standard practice. In figure 1, we reproduce a screen from ELDIT which shows the entry for IT *dare* ("[to] give"), with a valency formula saying "someone gives something to someone" (*qualcuno dà qualcosa a qualcuno*). If one of the fields of the formula is being pointed at with the mouse cursor (cf. figure 1), the linguistic term of the respective element of the valency formula is displayed (in figure 1, this is *complemento oggetto*), and the corresponding word or phrase in the example sentence is highlighted. The same highlighting happens if a relevant part of the example sentence is pointed to with the mouse cursor: in this case, the respective element of the valency frame and the corresponding linguistic term are indicated.

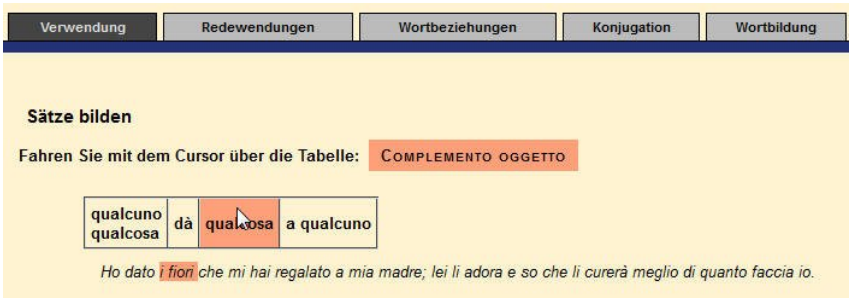


Figure 1: Screenshot from ELDIT: mouse-sensitive elements of valency frames, as an example of data provision on demand

3.3.2 Pop-up boxes

Similar to the above example, other elements of the data display can be made mouse-sensitive. The following example concerns an illustration, where lexicographic data about objects belonging to a theme or being part of a larger object are shown on demand. This device has also been prototyped, for example, in ELDIT[6]; an example is the illustration in the entry for *Haus* ("house") with its pop-up entry for *Fenster* ("window") in figure 2. The cursor changes to a hand when active items are encountered and upon clicking on the item, additional data are displayed in the pop-up box.

Figure 2: Screenshot from ELDIT of a mouse-sensitive illustration: a pop-up box as an example of data provision on demand (user clicking on the resp. item)

See also the *English Dictionary for South Africa* (2012, CD-ROM version) where items in illustrations are hyperlinked to the dictionary article.

This device could very well be used in technical dictionaries to illustrate terms that denote elements of complex technical objects, e.g. in explosion diagrams. To the best of our knowledge, it is, however, not much used yet in specialised dictionaries.

4. New devices for information presentation in electronic dictionaries

This section will be devoted to a discussion of different presentational devices that have so far not often been used in electronic dictionaries; they seem, how-

ever, to be relatively effective and efficient for a number of cognitive as well as communicative dictionary functions, especially on lexical items that belong to (potentially rather complex) structured systems.

4.1 Phenomena belonging to structured systems

To act as language information tools, (general language) dictionaries should cover phenomena from all levels of linguistic description: pronunciation, morphology, syntax, semantics, and pragmatics. Many such phenomena are systematic in one way or another: for example, many of the morphological and syntactic properties of words and word combinations must be seen in the context of the grammatical system of the language or of a given subsystem. They can only (or at least better) be interpreted with reference to this grammatical (sub) system. This holds for many function words, for words belonging to inflectional paradigms, for complementation patterns or for complex words built according to specific morphological word formation principles. We say that such items belong to "structured systems".

Typically, the particularised presentation of lexical data in semasiological dictionaries, i.e. the individualised access to each lemma entry, does not bring the systematic nature of such phenomena to the fore, but rather obscures it by distributing the members of the set across the whole macrostructure. For some dictionary use situations, this is not a major issue, and some lexicographers counterbalance this effect by including systematic morphological or syntactic overview tables (inflection paradigms, inventories of closed class items, subcategorisation tables, etc.) into their dictionaries, for example as outer texts, in an appendix or in a dictionary grammar (cf. Gouws 2009, 2010).

The same property of being part of a structured system is also present in certain lexical semantic phenomena, especially in those which can be structured in terms of taxonomies (e.g. animals, plants), of ordered lists (e.g. names of the days of the week or military ranks) or in terms of relational networks (e.g. kinship terms, cf. Prinsloo and Bosch 2012). In specialised languages, often the meaning of a term is not independent from a given system of related terms; well-known examples are technical taxonomies or juridical terms which denote concepts that are part of a legal system.

The above examples all concern lexical items that are related with other lexical items in one way or another. For a user, knowing about such relationships may in some cases just be part of his or her grammatical knowledge, in other cases it may be a necessary precondition for successful use of the respective items in communication. A prominent example of the latter case is the translation of legal terms which belong to legal systems that are not isomorphic between source and target language (cf. Mayer 1998).

To present items belonging to a structured system, we suggest the use of one of the different presentational devices which we intend to discuss in the following: tables, diagrams, guidance paths and decision trees. While the first two are mainly oriented towards cognitive functions, the latter two are primar-

ily relevant for communicative functions.

4.2 Devices for presenting structured systems: tables

In the first place, presenting the structured system which a given item belongs to, only serves a cognitive purpose: the user of the dictionary may learn about the wider context of the item he or she is looking up. In the case of translation, for example of legal documentation, it may, however, also have a communicative function.

Tables, being two-dimensional, tend to provide good possibilities for presenting items that denote objects which belong to lists or taxonomies, i.e. to systems with up to two classification criteria. A table is thus a recommended lexicographic device for restoring contextualisation where the relation between a list of items that have a certain number of characteristics in common has to be illustrated. Tables may be linked (in the sense of data on demand) to each individual entry they contain, and they can thus also be made accessible from each item.

Table 1 is an example of a table that summarises morphosemantically related lexical items. It contains a summary of eight moods as distinguished by Lombard (1985) for Sepedi (also known as Sesotho sa Leboa or Northern Sotho). This table gives a bird's eye view of the forms used for the different moods and distinguishes between imperfect (present tense) and perfect (past tense) forms, moods (1-3 versus 4-8) and gives examples of positive and negated forms in each case. Finally, a concise summary of the negation strategies and a simplistic indication of the meaning of each mood is presented in the rightmost column.

1. INDICATIVE	pres	pos	*monna o reka puku* the man buys a book	*…-a*
		neg	*monna ga a reke puku* the man does not buy a book	*ga ...-e* STATE-MENTS
	past	pos	*monna o rekile puku* the man bought a book	
		neg	*monna ga se a reka puku* the man did not buy a book	*ga se*+cons SC+pres
2. SITUATIVE	pres	pos	*ge monna a reka puku* if the man buys a book	
		neg	*ge monna a sa reke puku* if the man does not buy a book	*sa ...-e* IF/WHILE
	past	pos	*ge monna a rekile puku* if the man bought a book	
		neg	*ge monna a sa reka puku* if the man did not buy a book	*sa+pres*

3. RELATIVE	pres	pos	*monna yo a rekago puku* the man who is buying a book	*…ago*
		neg	*monna yo a sa rekego puku* the man who is not buying a book	*sa...-ego* THAT (IS DOING)
	past	pos	*monna yo a rekilego puku* the man who bought a book	
		neg	*monna yo a sa rekago puku* the man who did not buy a book	*sa…-ago*
4. SUBJUNCTIVE		pos	*... (gore) monna a reke puku* … (so that) the man buys a book	MUST (DO SOMETHING)
		neg	*... (gore) monna a se reke puku* … (so that) the man does not buy a book	*se ...-e*
5. CONSECUTIVE		pos	*... monna a reka puku* … (then) the man bought a book	AND THEN (DO/DID SOMETHING)
		neg	*... monna a se reke puku* … (then) the man did not buy a book	*se ...-e*
6. INFINITIVE		pos	*... go reka puku* … to buy a book	*go …-a* TO (DO SOMETHING)
		neg	*... go se reke puku* … not to buy a book	*go+se ...-e*
7. IMPERATIVE		pos	*reka puku!* buy a book!	DO (SOME-THING)!
		neg	*se reke puku!* do not buy a book!	*se …-e*
8. HABITUAL		pos	*... monna a reke puku* the man (usually) buys a book	*…-e* USUALLY (DO SOME-THING)
		neg	*... monna a se reke puku* the man (usually) does not buy a book	*se ...-e*

Table 1: The modal system in Sepedi: table presentation

We will come back to possibilities of making use of table 1 for communicative purposes below, in section 4.4.1.

4.3 Devices for presenting structured systems with multiple relations: diagrams

As is the case with tables, diagrams also fulfil a contextualisation function, but in addition they answer to the requirements of visually presented relations. A diagram of the wiring system of a car not only brings together all wires, globes and fuses, but also presents them schematically in relation to each other. The same is true, for example, for kinship terminology, where not only a set of terms that belong together is brought together, but also the exact relations in terms of e.g. older and younger generations, male and female, husbands and wives, are illustrated (cf. the diagram extracted from the *Macmillan English Dictionary for Advanced Learners* (2007: 502), figure 3). Diagrams allow the lexicographer to keep track of more than two types of ordering criteria, thus being adequate for presenting items that denote objects related through several different types of relations. Kinship terminology is a good example of this type of sets of lexical items; this is illustrated in figure 3.

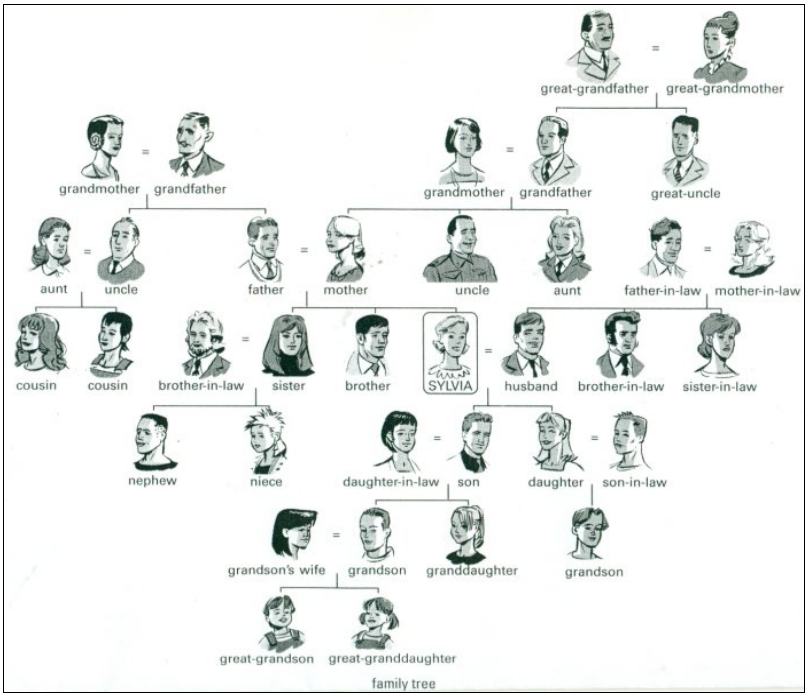


Figure 3: Family tree in MED (2007: 502)

4.4 Devices for guiding users through structured systems: guidance paths

Tables as well as diagrams provide an overview of a structured system, a sort of bird's eye view. Both do not necessarily presuppose any reading direction or any specific entry point. In this sense, these devices rather serve cognitive, overview-related functions, much more than communicative ones.

In our view, situations where the user needs to precisely search for a given item from a structured system are best served by what we call "guidance paths", i.e. paths through tables or diagrams that can be activated by the user.

4.4.1 Guidance paths through tables

Table 2 contains a small extract from table 1 (cf. section 4.1), for the lexical material used in Sepedi to express the situative. It is similar in layout to table 1, but converted into a set of choice options, represented by the arrows. By following the appropriate path, the user will be guided to a valid expression, e.g. by following the path "situative → past → positive" to *-ile*.

2. SITUATIVE	pres.	pos	ge monna a reka puku if the man buys a book	
		neg.	ge monna a sa reke puku if the man does not buy a book	sa...-e IF/WHILE
	past	pos	ge monna a rekile puku if the man bought a book	
		neg	ge monna a sa rekapuku if the man did not buy a book	sa+pres

Table 2: Extract from table A (modal system of Sepedi), with choice options for text production

Table 3 reflects another extract from the modal system of Sepedi (cf. table 1) and suggests a guidance path for expressing the negative form of the moods that do not distinguish time. The path links the relevant moods with the negative forms, and the vertical bars and framed boxes on the right hand side emphasise the fact that a single negation strategy is used in all cases. This layout is aimed at production guidance in the formation of negated sentences.

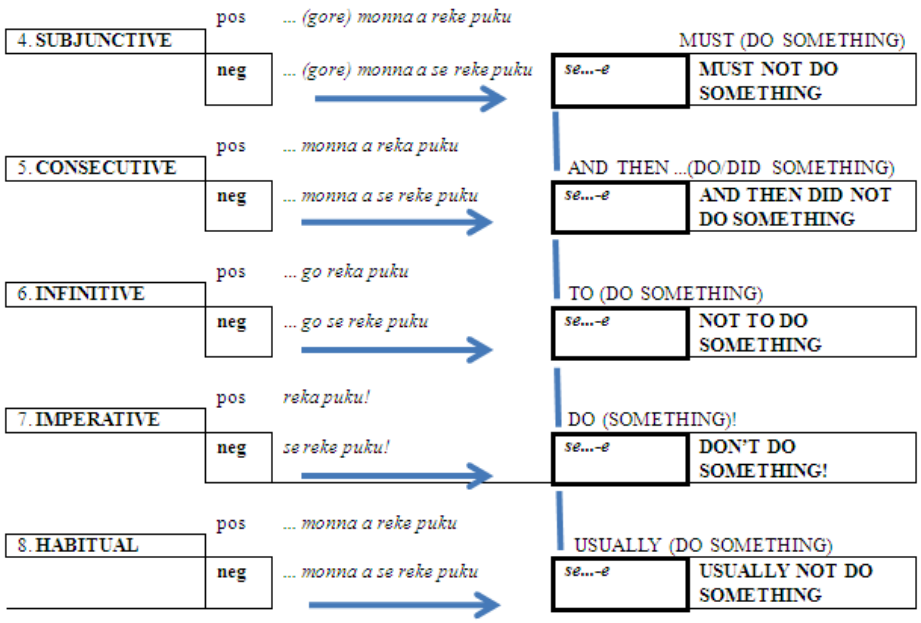


Table 3: A guidance path for expressing negated forms of moods in Sepedi

Table 4 is a guidance path highlighting the tripartite relation between (i) the grammatical term, used to denote the mood, (ii) a basic semantic indication of its meaning, given in context with a positive and negated example and a simplistic summary of the negation strategy used. The emphasis here is on reception guidance in terms of the meaning of a given modal expression (cf. the arrows in the schema), to be read from the middle either to the right hand side (paraphrase) or to the left hand side (grammatical terminology). The schema can also be used for text production, either starting from meaning paraphrases (rightmost column), or from the grammatical terminology (leftmost column), with the guidance path leading in either case to the lexical expressions in the middle column.

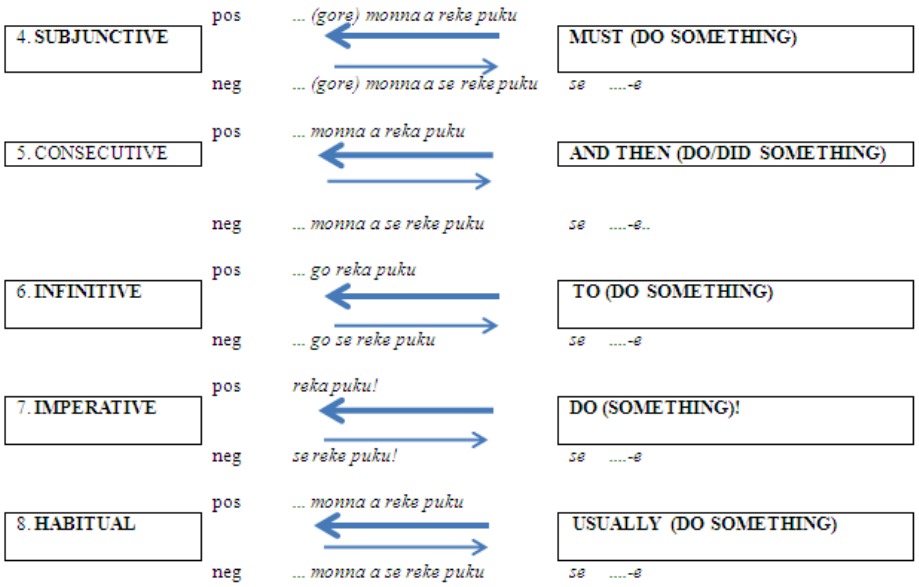


Table 4: Possible guidance paths for highlighting the relation between a Sepedi mood and its meaning

4.4.2 Guidance paths through diagrams

If we take up the example of kinship terminology discussed above, in section 4.3, we may recall that the table shown in figure 3 serves mainly cognitive purposes, as it allows the user to understand which names of family members exist in English. In a text reception situation, the user who learns English as a foreign language may wish to understand what a term like *cousin* means: ideally, he or she will just have to enter the item into a search box of the dictionary and get all those relatives highlighted and linked to the "central" person in the diagram ("Sylvia") who can be called *cousin*. So, the dictionary proposes a path through the relational diagram. Along this path, the distinctive properties of the items related (in this case the distinctive properties of *cousin*) can be collected, cf. figure 4.

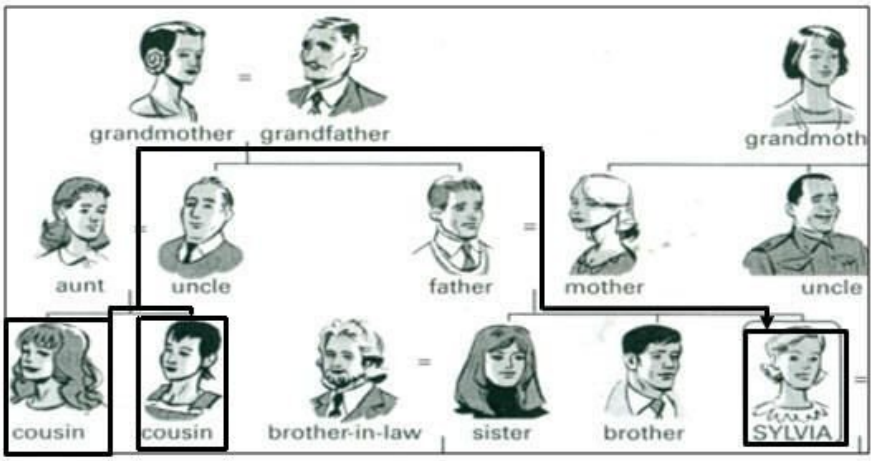


Figure 4: Guidance path for text reception, within (an extract from) the kinship diagram from figure 3

In a text production situation, likewise, a path through the diagram should be constructed, but this time step by step, on the basis of the user's decisions and moves, e.g. from "Sylvia" to her father (step 1 in figure 5), further on to the parents of her father (step 2), to the uncle (step 3), and finally to the children of the uncle (step 4). When the path stops, the dictionary should provide the lexical items searched for. Obviously, the paths for text reception and text production are the same, but used in different ways. For a more complicated system of kinship terms than that of English, namely that of isiZulu and Sepedi, see Prinsloo and Bosch 2012 and our discussion in section 4.5.3, below.

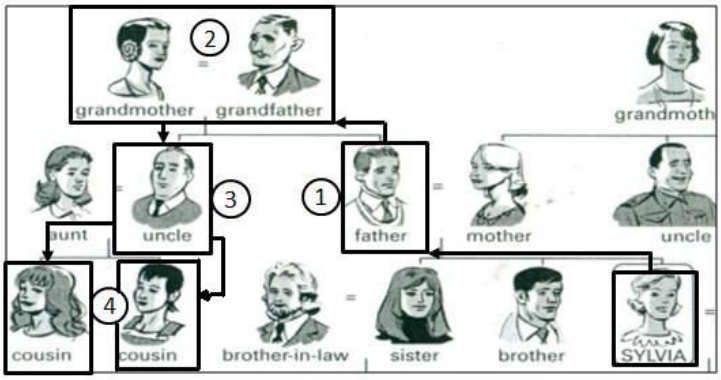


Figure 5: Guidance path for text production, within the kinship diagram from figure 3

The procedures sketched above provide guidance for communicative situations to the user, for text understanding or for text production. They could be implemented by means of tagged static data, including overlays to diagrams or tables.

4.5 Interactive decision trees

The example of kinship terminology shows the guidance an electronic dictionary could give to the user by means of simple highlighting of paths that lead through a structured system. For issues of low and medium complexity, within (morpho) syntax, ontologies and lexical semantic networks, the decisions to be taken at each step in a text production scenario, as well as the respective meaning components derived from the full path in a text reception scenario, may be relatively intuitive and easy to follow for a user.

An alternative, and maybe even more efficient way of supporting the user, especially in text production and translation towards a foreign language, is what we call *interactive decision trees*.

This device is an attempt to model lexical choice, for production-oriented communicative functions, as a decision process. Text production decisions are made, under constraints from different levels of linguistic description. At each choice point, a few options are open, and the full decision algorithm thus reminds of a tree where the nodes are the choice points and the arcs are the options. This is why we speak of a decision tree, as it is customary in computer science.

Where the interaction of the constraints leading to the right lexical expression is rather complex, the user may thus profit more from stepwise guidance through the tree than from a full table-based or diagram-based overview of the underlying system.

4.5.1 A simple example of decision trees

A trivial example of a decision tree, which we discuss here to show the logic of the device rather than its linguistic contents, is the translation of English possessive determiners to French. The morphosyntactic systems of the two languages are not isomorphic. While English possessives have for some persons different forms depending on the natural gender of the possessor (*his* (masc.) vs. *her* (fem.)), French possessives agree with the grammatical gender and number of the possessed object (*mon livre* (masc. object) vs. *ma bouteille* (fem. object) vs. *mes livres*, *mes bouteilles* (plural object)), but don't mark the natural gender of the possessor.

To correctly translate English possessives to French, the user has to decide on a number of properties of the source and target items involved. Obviously, the relevant linguistic attributes (e.g. number, gender, …) and their values (e.g.

singular, plural, …) are indicated in tables that give the full picture of the possessive determiners in the two languages. But for the actual task of translating, say, *your bottle* to French, the user only needs to be aware of a limited number of decisions to be taken: the first choice point concerns the number of the English item (cf. step 1 in figure 3); let us assume the user selects "singular"; the next decision has to do with the number of the possessed object (step 2, user decides again for "singular"); this leads to the third choice point, concerning the grammatical gender of French *bouteille*, namely "feminine" (step 3). The decision tree ends at the French form *ta*, which is appropriate in this context. Figure 6 schematises the process: solid lines indicate the choices made by the user, and broken lines point to alternative options at each choice point.

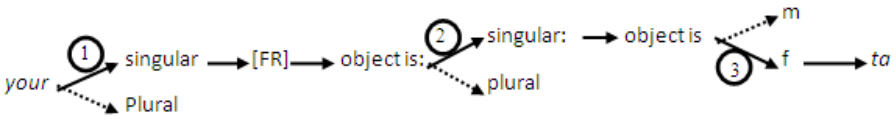


Figure 6: Schema of a decision tree for the translation of English *your* to French, in the context *your bottle*

Obviously, the above made-up example is oversimplified and redundant with most learners' grammatical knowledge. It was merely used to show the logic and the type of guidance we intend to provide through interactive decision trees.

There are, however, other cases that are more complex and may thus indeed require the use of decision trees to provide users with exactly the information needed in a given situation. Sepedi copulatives are a prime example of such a case; we will thus elaborate on the facts pertaining to this construction in the remainder of this section, giving a fuller picture of possible user guidance with respect to this phenomenon in section 5.

4.5.2 Decision trees for complex phenomena: the case of Sepedi copulatives

The system of Sepedi copulatives could be presented in a bird's eye view as a table (cf. Appendix 1) but by virtue of its sheer size, this solution would be suboptimal for both text production and text reception. Standard particularised dictionary entries would lose the information about the systematicity of the phenomena; we thus see the interactive decision trees as the most effective and efficient device to convey exactly the data a user would need with respect to these phenomena in a given type of communicative situation.

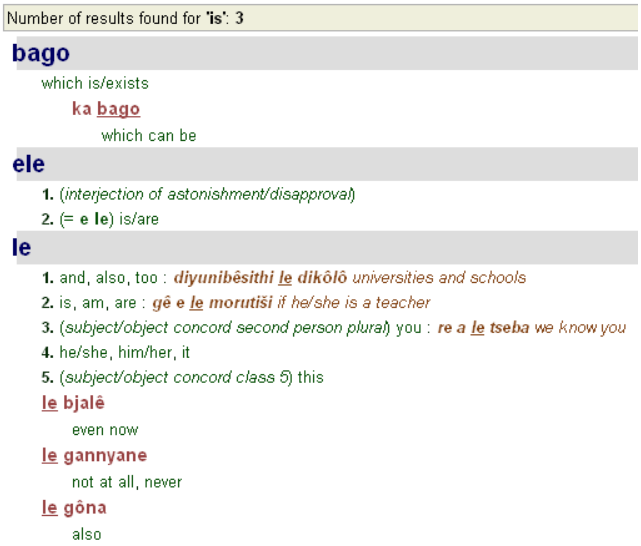
In a text production situation, the decision algorithm for the selection of copulatives entails distinguishing between an *identifying* vs. a *descriptive* vs. an

associative relation existing between the subject and its complement, see also sections 1 to 3 in Appendix 1:

(1)
is
[identifying. copulative], ke lengwalo (it *is* a letter)
[descriptive. copulative], mosadi o bohlale (the woman *is* clever)
[associative copulative], Satsope o na le Sara (Satsope *is* with Sara)

Learners of Sepedi who want to use copulatives in speech or text production have at best to do an intensive study of the copulatives from dictionaries and grammar books. Dictionaries typically provide basic and sometimes even inadequate information (cf. figure 7, below). Grammar books such as Poulos and Louwrens (1994) on the other hand, provide an overload of grammatical information (37 pages), in an effort to cover all the relevant and possible copulatives. Such details may be useful in a cognitive situation where the user would like to learn everything about the copulative, but they are hardly useful in a text production situation where the user simply wants guidance on which form to use. Such information overload could easily lead to "information death" (cf. Bergenholtz and Bothma (2011)).

Dictionaries, and especially electronic dictionaries, fail to give even basic communicative guidance or to treat all three main copulative relations shown in (1). Consider the article for the lemma *is* in the *Sesotho sa Leboa (Northern Sotho)–English Dictionary* (2003) in figure 7.


Number of results found for 'is': 3

bago

which is/exists

ka bago

which can be

ele

1. (*interjection of astonishment/disapproval*)

2. (= **e le**) is/are

le

1. and, also, too : ***diyunibêsithi le dikôlô*** *universities and schools*

2. is, am, are : ***gê e le morutiši*** *if he/she is a teacher*

3. (*subject/object concord second person plural*) you : ***re a le tseba*** *we know you*

4. he/she, him/her, it

5. (*subject/object concord class 5*) this

le bjalê

even now

le gannyane

not at all, never

le gôna

also

Figure 7: The lemma is in the *Sesotho sa Leboa (Northern Sotho)–English Dictionary* (2003)

In this example, two of the three copulative categories, i.e., the identifying and the associative copulatives, have not been treated, not to mention giving proper receptive or productive guidance. Paper dictionaries for Sepedi reflect the same deficiencies.

In the e-environment it is, however, possible to provide the user with the required guidance on which form is the correct one for a given situation, and to provide exactly the amount of information that is needed for each of the possible choices. Compared with a table, a decision tree will in such a case reduce the amount of presented information considerably, and the user can, at any stage, decide that his/her information need has been met and return to his/her primary task, namely to write a text.

For example, when the user wants to write *the woman is clever* in Sepedi he/she should be guided to *mosadi o bohlale* and be guarded from the typical error **mosadi ke bohlale*. The user can then be guided to subsequent levels of decisions, e.g. concerning person and noun class of the subject, tenses and moods, as well as a number of lexicalised exceptions, cf. Appendix 1 for a tabular overview of the full set of items to be considered, and section 5 for a detailed example of how decision-tree-based guidance can be conceived of.

4.5.3 Presenting decision trees to the dictionary user

The phenomena sketched above may usefully be presented to the user in terms of subsequent choices that could take the form of check boxes or other selection devices, in a graphical user interface (GUI). The visual appearance of the interface should make clear that the selections are the result of a decision process involving several steps. Instead of complex tables giving all options, a path through sub-tables should be shown, but together with links to synoptic tables which indeed allow the user to see the full picture if he/she wishes to. For a set of function words of the same category, the basic decision tree is constant. Users will only follow different paths through this tree, depending on their actual needs.

The choice points and options may equally be presented in the form of interlinked questions of a questionnaire. These are based on exactly the same logic and internal representation and thus formally equivalent to the decision trees. An example of this device has been proposed by Prinsloo and Bosch (2012), for Zulu and Sepedi kinship terminology, the system of which is considerably more complex (i.e. contains more attributes) than that of English discussed above. Without going into the details of this system, we reproduce, in figure 8, the questionnaire provided by Prinsloo and Bosch (2012), to exemplify this option of presentation. The authors use five levels of selection boxes in their questionnaire; these correspond to five choice points, cf. figure 8.

The example in figure 8 gives step-by-step guidance to correctly address a male speaker's father's older sister supplemented by additional information on the target term.

Level 1

I am a boy ☑	I am a girl ☐

Level 2

I want to speak **to** my aunt ☐	I want to speak **about** my aunt ☑

Level 3

My aunt is:	
My father's:	**My mother's:**
Older sister ☑	Older sister ☐
Younger sister ☐	Younger sister ☐
Older brother's wife ☐	Older brother's wife ☐
Younger brother's wife ☐	Younger brother's wife ☐

Level 4

MY FATHER'S OLDER SISTER
Rakgadi
Rakgadi yo mogolo
Click here for:
Additional information on a man's father's brothers and sisters ☐
Additional information on a man's wife's father's brothers and sisters ☐
Additional information: on kinship relation trees ☐
Other relations using *Rakgadi* ☑

Level 5

OTHER RELATIONS USING RAKGADI
My father's older brother's **daughter**:
Rakgadi e ka ba morwedi wa ramogolo
"Rakgadi can be the daughter of my father's elder brother"

My father's younger brother's **daughter**:
Rakgadi e ka ba morwedi wa rangwane
"Rakgadi can be the daughter of my father's younger brother"

Figure 8: Section of decision tree guidance for Sepedi kinship terms, presented as questionnaire, from Prinsloo and Bosch (2012)

4.5.4 Technical options for realising decision trees

The internal representation of lexical and grammatical data should be adapted to the decision-tree-like access to the data. For this, not only synoptic tables of

the respective words, but also a representation of the selection rules is needed.

We list in the following a number of interface solutions, by order of decreasing specialisation, which should be considered:

— Solve the problem, suggest the correct solution, and give a visual presentation and link to *Read mor*e sections such as FAQs or outer texts.

— Supply a link to *Read more* information where distinctions on a cognitive level are made.

— Supply a link to guidance on the basis of e.g. *Frequently made errors*.

— Give good, typical examples of use throughout.

All envisaged forms of presentation should be based upon a grammatical description of the construction to be tackled. One could argue that most of these issues have been sufficiently described in standard grammars. However, one should not assume that the format of these descriptions is such that they are ready to use for building decision trees. A reorganisation of the data and a fine-grained structuring (e.g. through markup) will be necessary.

In a text production or a text reception situation, a user can consult the (stand-alone) dictionary to obtain the required information. However, it is also possible that the support tool is integrated into a word processor the user is using to construct his/her text. In such a case, the user may require feedback on his/her own text production efforts based on his/her grammatical knowledge without specifically consulting the dictionary. The e-dictionary should then be integrated into the word processor as a grammar checker, similar to the features currently available in popular word processing software.

The following example is intended to show how we envisage practical text production work with a version of the decision tree dictionary integrated with the user's text processing environment.

Let us depart from a most common error scenario in Sepedi, for example, the user typing **selepe ke bogale* in order to express *the axe is sharp*. Learners usually know that *ke* means *it is* and that no distinction is made between *he is*, *she is*, *it is* and *they are* in Sepedi: all convert to *it is*, e.g. *(mosadi) ke mooki* "she is (it is) a nurse" so they incorrectly use *ke* with *bogale* "sharp". As a second example consider **mosadi o mooki* instead of *mosadi ke mooki* "the woman is a nurse". Learners are accustomed to using the subject concord *o* with class 1 nouns in sentence construction and it is the correct form in two out of the three copulative relations (descriptive and associative copulatives; so attempting to use it also in the identifying copulative is a common error).

The student types **selepe ke bogale* in a word processor linked to the electronic dictionary and either *ke* alone or all three words together are flagged as incorrect. A quick solution is offered by means of a suggestion box, in this case offering three possibilities, namely *se*, *se le* "is" and *se lego* "who/what is". The user who has basic knowledge of the modal system will know which one to

select.

Most users, however, would need further guidance, and this is offered by a decision process guiding him/her through the three possible moods (see Appendix 1, sections 1–3) and the Indicative *se*, Situative *se le* or Relative s*e lego* of the decision tree examples below in section 5 for the descriptive copulative with sub-decisions. The process for **mosadi o mooki* is similar, i.e. a decision process guiding the user through the three possible moods (Indicative *ke*, Situative *e le* or Relative *e lego*) of the decision tree for the identifying copulative respectively, with sub-decisions.

4.5.5 Lexicographic processes towards the preparation of decision trees

The process to produce a dictionary article that provides a decision tree requires at least three sequential steps, building on one another:

— Step 1 would be to acquire comprehensive and accurate data for the set of rules etc. to be described. This includes the grammatical rules as well as relevant examples, common errors, etc.

— In Step 2 the lexicographer, in collaboration with a database expert, needs to reorganise the data so that it will be possible for a programmer to implement a decision tree. This requires at least two sub-processes:

  — The logic of the decision process needs to be worked out very carefully, i.e., what is the logical sequence of the decisions, how much information is required to make and/or support the decisions, when are what type of examples needed, when are links to outer texts required, etc.

  — The data need to be marked up in such a way that each of the data elements defined in the analysis of a specific complex problem can be identified at the required level of granularity. This implies that the database should make provision for such extensions, either by using an extensible XML schema or additional tables and fields in a relational database, depending on the original design of the system (cf. Bothma (2011)).

— In Step 3 the programmer takes the flow diagram of the decision tree together with all the explanations, examples and linked data, and implements this. The programmer should also design a "user-friendly" interface that is intuitive for the average user and guides him/her to follow the correct trail through the decision tree for the given information need.

It may be feasible to use multiple devices for the same phenomenon, as is clear from the preceding examples. For example, in the case of the kinship terms, one can provide a diagram only (see section 4.3 above), or a diagram with guidance

paths by means of different overlays (see section 4.4.2 above), or even a decision tree (see section 4.5.3 above). As indicated earlier, not all articles in a dictionary will necessarily be enhanced by means of any of these devices. Similarly, if one of these devices is applicable in a given case, it would evidently not be optimal to use multiple devices for the same phenomenon in the same article. The lexicographer will therefore have to do a very careful analysis of all lemmas to identify those lemmas and structured systems that can optimally benefit from any enhancement through the use of one of the above-mentioned information presentation devices, and he or she will subsequently have to do an equally careful analysis to decide which of the proposed technologies will be the best option in any given situation.

5. Exemplification: complex cases of copulative selection

In this section, we come back to the structured system of the Sepedi copulative, which was mentioned in section 4.5.2, above. We now explain possible guidance scenarios in detail.

5.1 Different levels of user guidance

Figure 9 provides a schematic illustration of a pop-up guidance screen sequence for **selepe ke bogale*.

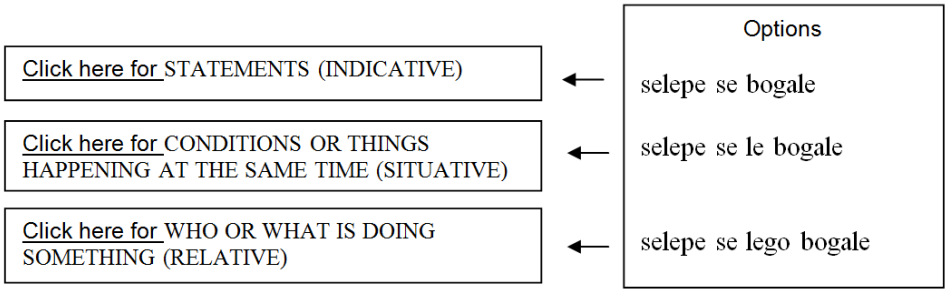


Figure 9: Dictionary feedback for **selepe ke bogale*

The user is informed that the only three correct options are *selepe se bogale*, *selepe se le bogale* and *selepe se lego bogale*. In many circumstances, observing the three correct options will enable the student from his/her knowledge of the language to select the correct one. If more guidance in respect of the descriptive relations in the Indicative, Situative and Relative is required, the user can click the buttons on the left hand side in figure 9 to display the information given in Figures 10, 11 and 12. The first option *selepe* ***se*** *bogale* is in the Indicative Mood, *selepe* ***se le*** *bogale* in the Situative Mood and *selepe* ***se lego*** *bogale* in the Relative Mood.

These grammatical terms are, however, avoided on the first level of guidance, and even on the second level, the distinction in meaning, i.e. statements, versus conditions etc. is given priority to the grammatical terms Indicative, Situative and Relative, which are given in brackets. These terms could be perceived as user-unfriendly by those users who do not know the grammar. In figures 3–5 they are, however, mentioned for the benefit of users who do have knowledge of the grammar.

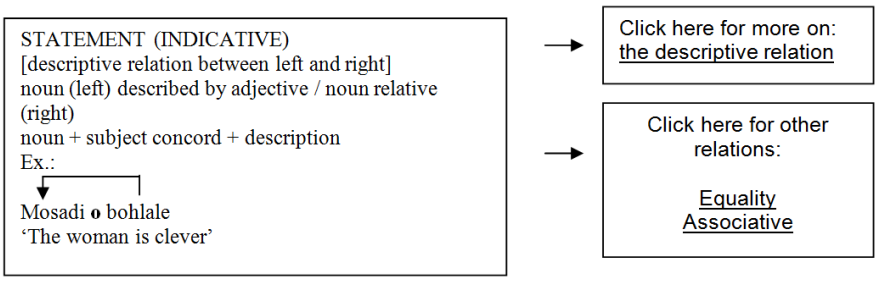


Figure 10: Pop-up information boxes for *selepe se bogale*

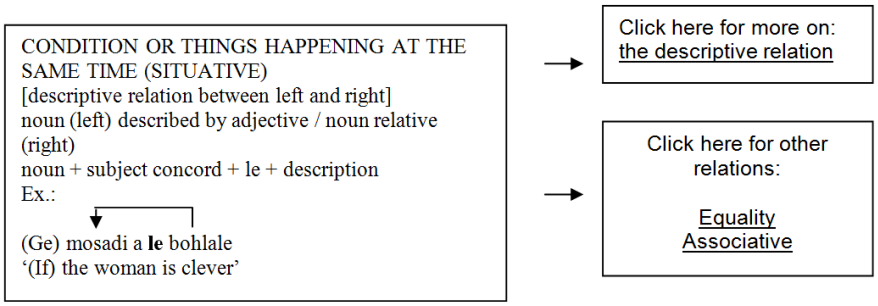


Figure 11: Pop-up information boxes for *selepe se le bogale*

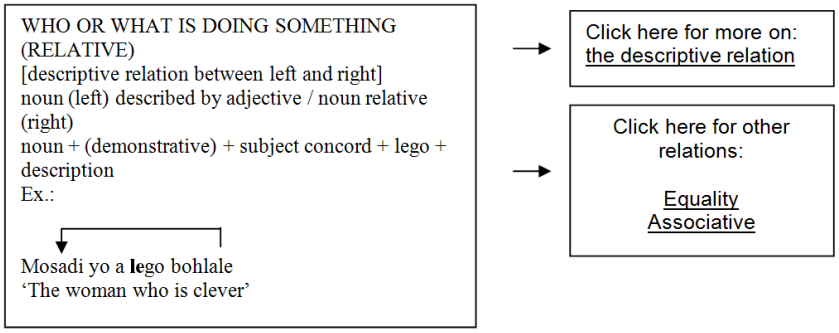


Figure 12: Pop-up information boxes for *selepe se lego bogale*

In each case, the panel given in the left part of the mock-up provides the information needed for text production. Users with more (cognitive) needs can access a fuller picture via the buttons on the right hand side. Clicking the top buttons on the right hand side in figures 10–12 results in the presentation of more detailed information on the descriptive copulative, as shown on the left hand side of these figures. The buttons at the bottom on the right hand side of these figures result in pop-up boxes giving more information on the other two relations, i.e. equality and associative for each specific mood. So, for example, clicking the "Equality" option in figure 10 will provide more information on equality relations in the Indicative mood, the "Associative" option in figure 11 gives more information on associative relations in the Situative mood, etc.

5.2 From text production guidance to full grammatical guidance

Pop-up boxes giving more information and typical examples of the descriptive relations can be provided on a third level. These pop-up boxes will cover the entire sections 1–3 in the appendix. Compare, for example, guidance towards section 2, as in figure 13.

STATEMENT (INDICATIVE)

Positive	Negative
(Nna) *ke* bohlale	(Nna) *ga ke* bohlale
'I am clever'	'I am not clever'
(Rena) *re* bohlale	(Rena) *ga re* bohlale
'We are clever'	'We are not clever'
Monna *o* bohlale	Monna ga *a* bohlale
'The man is clever'	'The man is not clever'

Figure 13: Pop-up information box for persons and classes in the Indicative Mood of the Descriptive Copulative

A second scenario is where comprehensive guidance is required, e.g. when the user wants to know how to say *is* in Sepedi. In this case a combination of decision processes is required. These processes are enriched with data from processed corpora linked with the dictionary.

6. Conclusions

The project described in this article is driven by two underlying motivations, namely the urge to compile electronic dictionaries that can do better than current ones through maximal utilisation of advanced information technologies,

and the need for dynamic dictionaries guiding the user in innovative ways. We believe that step-by-step guidance, mainly through sequences of choices, the provision of additional relevant information on request as well as protection against incorrect conclusions users may draw from the data available to them, are the cornerstones of the design of such dynamic dictionaries.

As indicated in section 2, this project makes use of certain aspects of the Function Theory of lexicography, viz. we discuss the information needs from the perspectives of text reception, text production and cognitive situations. The examples we have discussed show that the same basic data can be used for providing the required level of guidance for each of the envisaged situations. Through the use of presentation devices of the kind illustrated in this article, the data are filtered to show only those items that are relevant for the specific information need. It is therefore not necessary to create different underlying data sets to guide the user in different situations. It is only necessary to provide the appropriate filters and presentation devices to extract precisely the data needed to solve the user's specific information need.

Careful analysis of the data needed to solve the user's information need in any given situation is therefore required: if the user has a text reception information need, data required for text production should not be included in addition in the material he/she gets presented, and *vice versa*. This selection task should be performed by a lexicographer with an excellent understanding of the language structures of the specific language, as well as of pitfalls of text reception and text production in the language. However, it also requires very serious input from an expert in database technologies and/or knowledge representation languages to organise and code the data in such a way that all relevant and indeed only relevant data required in terms of the specific type of information need can be extracted from the database. Data markup and data granularity are therefore extremely important issues to be considered (cf. also Bothma 2011: 90, 100).

Our current work only describes different devices that can be used to present information to users. How this can be implemented in principle is shown in the examples that we have provided. Practical implementation issues have clearly not yet been addressed in this article. This constitutes future work we intend to do, which includes a careful study of different database technologies and knowledge representation languages, to decide how practical implementation can best be done. Part of such a study is to do rapid prototyping of the same phenomenon by means of different technologies to find out what would, in practice, be the most effective and efficient way of realising the theory implicit in our discussions. And, obviously, once this is done, user evaluation would be required in terms of both the functionalities that are provided and also the specific interfaces that can be designed for each type of device.

As a matter of principle, data will be provided to the user only on demand. Therefore we don't envisage that each and every user will automatically be confronted with, for example, decision tree guidance in the case of

looking up the Sepedi copulatives in a text production situation. If the user finds sufficient data in the "standard" description of the copulative for text production and his/her information need is satisfied, the search ends and the user can carry on with his/her work task. If, however, the user needs further guidance because his/her knowledge of the language is insufficient to interpret the "standard" description correctly, additional information would be available on demand by means of the decision tree. The user is therefore in control of the information searching process and the amount of information displayed will not lead to information overload for the user who does not need extra information. Inherently such a choice of being presented with more (or less) information results in adaptive electronic dictionaries, as envisaged by Tarp (e.g. Tarp 2011), Bothma (2011) and many others. Such presentation devices will evidently not be added for all lemmas in a dictionary since not all of them are equally complex. Again, the lexicographer will be required to make a very careful analysis to decide which lemmas require such an extensive additional treatment.

We believe that continued research along the lines we have indicated in this article, together with further enhancements such as linking to corpora and other external data (cf. e.g. Heid, Prinsloo and Bothma 2012) and enhanced adaptive and interactive features will indeed lead to innovative electronic dictionaries that can better address the specific, individualised information needs of users in different situations.

7. Endnotes

1. http://www.isizulu.net
2. http://www.amazon.com
3. http://www.wiktionary.org
4. http://www.oed.com
5. http://dev.eurac.edu:8081/MakeEldit1/Eldit.html
6. http://dev.eurac.edu:8081/MakeEldit1/Eldit.html

8. References

Abel, Andrea. 2002. Darstellung der Verbvalenz in einem elektronischen Lernerwörterbuch Deutsch–Italienisch (ELDIT). Neue Medien, neue Ansätze. Braasch, A. and A. and C. Povlsen (Eds.). 2002. *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 13–17, 2002*: 413-418. Copenhagen: Center for Sprogteknologi, University of Copenhagen.

Amazon Kindle. [Online.] Available http://www.amazon.com.

Bergenholtz, Henning and Theo J.D. Bothma. 2011. Needs-adapted Data Presentation in e-Information Tools. *Lexikos* 21: 53-77.

Bergenholtz, Henning and Sven Tarp. 2002. Die moderne lexikographische Funktionslehre. Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen. *Lexicographica. International Annual for Lexicography* 18: 253-263.

Bergenholtz, Henning and Sven Tarp. 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31:171-196.

Bergenholtz, Henning and Sven Tarp. 2004. The Concept of "Dictionary Usage". Dollerup, Cai (Ed.). 2004. *Worlds of Words. A Tribute to Arne Zettersten. Nordic Journal of English Studies* 3(1): 23-36.

Bergenholtz, Henning and Sven Tarp. 2005. Wörterbuchfunktionen. Barz, Irmhild, Henning Bergenholtz and Jarmo Korhonen (Eds.). 2005. *Schreiben, Verstehen, Übersetzen, Lernen. Zu ein- und zweisprachigen Wörterbüchern mit Deutsch*: 11-25. Frankfurt/Bern/New York/Paris: Peter Lang.

Bothma, Theo J.D. 2011. Filtering and Adapting Data and Information in the Online Environment in Response to User Needs. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 71-102. London/New York: Continuum.

Bothma, Theo J.D. and Sven Tarp. 2012**.** Lexicography and the Relevance Criterion. *Lexikos* 22: 86-108.

De Schryver, Gilles-Maurice. 2009. State-of-the-Art Software to Support Intelligent Lexicography. Zhu, R. (Ed.). 2009. *Proceedings of the International Seminar on Kangxi Dictionary & Lexicology:* 565-580. Beijing: Beijing Normal University. Also: http://www.hcxf.cn/read.asp?id=570.

ELDIT. http://dev.eurac.edu:8081/MakeEldit1/Eldit.html.

English Dictionary for South Africa. 2012. Cape Town: Pharos and KDictionaries.

Faaß, Gertrud. 2010. *A Morphosyntactic Description of Northern Sotho as a Basis for an Automated Translation from Northern Sotho to English.* PhD Dissertation, University of Pretoria. http://upetd.up.ac.za/thesis/available/etd-10092010-134539.

Heid, Ulrich, Danie J. Prinsloo and Theo J.D. Bothma. 2012. Dictionary and Corpus Data in a Common Portal: State of the Art and Requirements for the Future. *Lexikographica* 28. To appear.

isiZulu.net. [Online.] Available http://www.isizulu.net.

Lombard, Daniel P. 1985. *Introduction to the Grammar of Northern Sotho*. Pretoria: J.L. van Schaik.

Macmillan English Dictionary for Advanced Learners. Rundell, M. and G. Fox. (Eds.). 2002. *Macmillan English Dictionary for Advanced Learners.* London: Macmillan.

Mayer, Felix. 1998. Eintragsmodelle für terminologische Datenbanken: ein Beitrag zur übersetzungsorientierten Terminographie. [Forum für Fachsprachen-Forschung, Bd. 44.] Tübingen: Narr.

Oxford English Dictionary Online (OED). [Online.] Available http://www.oed.com.

Poulos, George and Louis J. Louwrens. 1994. *A Linguistic Analysis of Northern Sotho.* Pretoria: Via Afrika.

Prinsloo, Danie J. and Sonja E. Bosch. 2012. Kinship Terminology in English–Zulu/Northern Sotho Dictionaries — A Challenge for the Bantu Lexicographer. Fjeld, Ruth Vatvedt and Julie Matilde Torjusen (Eds.). 2012. *Proceedings of the 15th Euralex International Congress,* 7–11 August 2012: 296-303. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.

Prinsloo, Danie J., Ulrich Heid, Theo J.D. Bothma and Gertrud Faaß. 2011. Interactive, Dynamic Electronic Dictionaries for Text Production. Kosem, I. and K. Kosem. 2011. *Electronic Lexicography in the 21st Century. New Applications for New Users. Proceedings of eLex 2011,* Bled, 10–12 November 2011: 215-220. Bled: Trojína.

Rundell, Mike. 2009. *The Road to Automated Lexicography: First Banish the Drudgery ... Then the Drudges?* Keynote speech at eLex, Louvain-la-Neuve, 22–24 October 2009. http://uclouvain.be/en-271028.html.

Sesotho sa Leboa (Northern Sotho)–English Dictionary. http://africanlanguages.com/sdp/.

Tarp, Sven. 2008. Revival of a Dusty Old Profession. *Hermes. Journal of Linguistics* 41: 175-188.

Tarp, Sven. 2008a. Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography. Tübingen: Niemeyer.

Tarp, Sven. 2009. Reflections on Data Access in Lexicographic Works. Nielsen, Sandro and Sven Tarp (Eds.). 2009. *Lexicography in the 21st Century. In Honour of Henning Bergenholtz:* 43-62. Amsterdam/Philadelphia: John Benjamins.

Tarp, Sven. 2009a. Reflections on Lexicographical User Research. *Lexikos* 19: 275-296.

Tarp, Sven. 2009b. Beyond Lexicography: New Visions and Challenges in the Information Age. Bergenholtz, Henning, Sandro Nielsen and Sven Tarp (Eds.). 2009. *Lexicography at a Crossroads. Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow.* Linguistic Insights. Studies in Language and Communication 90: 17-32. Bern: Peter Lang.

Tarp, Sven. 2011. Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011. e-*Lexicography: The Internet, Digital Initiatives and Lexicography*: 55-70. London/New York: Continuum.

Tarp, Sven. 2012. Online Dictionaries: Today and Tomorrow. *Lexicographica 28.* To appear.

Wiegand, Herbert E. 1989. Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven. Hausmann, F.J. et al. (Eds.). 1989. *Wörterbücher. Dictionaries. Dictionnaires.* Volume 1: 409-462 Berlin: De Gruyter.

Wiktionary. [Online.] Available http://www.wiktionary.org.

APPENDIX: THE COPULATIVE IN SEPEDI

Overview of the copulative constellations (Faaß 2010: Table 3.30, p.128)

Copulative	*Identifying (1)*		*Descriptive (2)*		*Associative (3)*	
Category	stative	dynamic	stative	dynamic	stative	dynamic
Tense						
Present	X	X	X	X	X	X
Perfect	X	X	X	X	X	X
Future		X		X		X
Mood						
Indicative (pos/neg)	X	X	X	X	X	X
Situative (pos/neg)	X	X	X	X	X	X
Relative (pos/neg)	X	X	X	X	X	X
Consecutive (pos/neg)		X		X		X
Subjunctive (pos/neg)		X		X		X
Habitual (pos/neg)		X		X		X
Infinitive (pos/neg)		X		X		X
Imperative (pos/neg)		X		X		X

STATIVE COPULATIVES

Section (1): Identifying Copulative (Equality between noun and complement)

Indicative ***The man*** ***is a teacher***			***Situative*** ***If/while the man*** ***is a teacher***		***Relative*** ***The man who*** ***is a teacher***	
Pers/Num/Cl.	**pos**	**neg**	**pos**	**neg**	**pos**	**neg**
1PS	*ke*	*ga ke*	*ke le*	*ke se*	*ke lego*	*ke sego*
1PP	*re*	*ga re*	*re le*	*re se*	*re lego*	*re sego*
2PS	*o*	*ga o*	*o le*	*o se*	*o lego*	*o sego*
2PP	*le*	*ga le*	*le le*	*le se*	*le lego*	*le sego*
CL1 – CL18	*ke*	*ga se*	*e le*	*e se*	*e lego*	*e sego*

Section (2): Descriptive Copulative (Complement describes noun)

(1) Indicative The man is clever			*(2) Situative If/while the man is clever*		*(3) Relative The man who is clever*	
Pers/Num/Cl.	**pos**	**neg**	**pos**	**neg**	**pos**	**neg**
1PS	*ke*	*ga ke*	*ke le*	*ke se*	*ke lego*	*ke sego*
1PP	*re*	*ga re*	*re le*	*re se*	*re lego*	*re sego*
2PS	*o*	*ga o*	*o le*	*o se*	*o lego*	*o sego*
2PP	*le*	*ga le*	*le le*	*le se*	*le lego*	*le sego*
CL1	*o*	*ga a*	*a le*	*a se*	*a lego*	*a sego*
CL2	*ba*	*ga ba*	*ba le*	*ba se*	*ba lego*	*ba sego*
CL 3	*o*	*ga o*	*o le*	*o se*	*o lego*	*o sego*
CL 4	*e*	*ga e*	*e le*	*e se*	*e lego*	*e sego*
CL 5	*le*	*ga le*	*le le*	*le se*	*le lego*	*le sego*
CL 6	*a*	*ga a*	*a le*	*a se*	*a lego*	*a sego*
CL 7	*se*	*ga se*	*se le*	*se se*	*se lego*	*se sego*
CL 8	*di*	*ga di*	*di le*	*di se*	*di lego*	*di sego*
CL 9	*e*	*ga e*	*e le*	*e se*	*e lego*	*e sego*
CL 10	*di*	*ga di*	*di le*	*di se*	*di lego*	*di sego*
CL 14	*bo*	*ga bo*	*bo le*	*bo se*	*bo lego*	*bo sego*
CL 15-18	*go*	*ga go*	*go le*	*go se*	*go lego*	*go sego*

Section (3): Associative Copulative (Complement and noun associated)

(1) Indicative The man is with his dog			*(2) Situative If/while the man is with his dog*		*(3) Relative The man who is with his dog*	
Pers/Num/ Class	**pos**	**neg**	**pos**	**neg**	**pos**	**neg**
1PS	*ke na le*	*ga ke na (le)*	*ke na le*	*ke se na (le)*	*ke nago le*	*ke se nago (le)*
1PP	*re na le*	*ga re na (le)*	*re na le*	*re se na (le)*	*re nago le*	*re se nago (le)*
2 PS	*o na le*	*ga o na (le)*	*o na le*	*o se na (le)*	*o nago le*	*o se nago (le)*
2PP	*le na le*	*ga le na (le)*	*le na le*	*le se na (le)*	*le nago le*	*le se nago (le)*
CL1	*o na le*	*ga a na (le)*	*a na le*	*a se na (le)*	*a nago le*	*a se nago (le)*
CL2	*ba na le*	*ga ba na (le)*	*ba na le*	*ba se na (le)*	*ba nago le*	*ba se nago (le)*
CL 3	*o na le*	*ga o na (le)*	*o na le*	*o se na (le)*	*o nago le*	*o se nago (le)*
CL 4	*e na le*	*ga e na (le)*	*e na le*	*e se na (le)*	*e nago le*	*e se nago (le)*
CL 5	*le na le*	*ga le na (le)*	*le na le*	*le se na (le)*	*le nago le*	*le se nago (le)*
CL 6	*a na le*	*ga a na (le)*	*a na le*	*a se na (le)*	*a nago le*	*a se nago (le)*
CL 7	*se na le*	*ga se na (le)*	*se na le*	*se se na (le)*	*se nago le*	*se se nago (le)*
CL 8	*di na le*	*ga di na (le)*	*di na le*	*di se na (le)*	*di nago le*	*di se nago (le)*
CL 9	*e na le*	*ga e na (le)*	*e na le*	*e se na (le)*	*e nago le*	*e se nago (le)*
CL 10	*di na le*	*ga di na (le)*	*di na le*	*di se na (le)*	*di nago le*	*di se nago (le)*
CL 14	*bo na le*	*ga bo na (le)*	*bo na le*	*bo se na (le)*	*bo nago le*	*bo se nago (le)*
CL 15-18	*go na le*	*ga go na (le)*	*go na le*	*go se na (le)*	*go nago le*	*go se nago (le)*

Do We Need a (New) Theory of Lexicography?

Sven Tarp, *Department of Afrikaans and Dutch, University of Stellenbosch, South Africa* and *Centre for Lexicography, Aarhus University, Denmark (st@asb.dk)*

Abstract: In the current transition from printed to electronic dictionaries the question has been raised whether we need a new theory of lexicography that may guide the conception and production of lexicographical e-tools or if we can use the theories already developed in the era of printed works. In order to answer this question the article first of all discusses the question whether a lexicographical theory exists, is possible at all or even wanted. It shows that the various approaches to this question are mainly due to the fact that the very concept of theory is widely disputed within lexicographical circles. In this respect, it briefly discusses the Anglo-Saxon academic tradition according to which science is only related to natural phenomena and where all other phenomena are referred to the spheres of art and craft, and it shows that this tradition is widely opposed by the traditions in other parts of the world. Upon this basis, the article shows that a lexicographical theory is not only possible but that various highly useful theories already exist. Finally, it claims that these theories, especially the ones that are not only focusing on the printed dictionary form, should not be rejected but should be further enhanced and improved in close interaction between lexicography and other consultation disciplines within the broad area of information science.

Keywords: LEXICOGRAPHY, LEXICOGRAPHICAL THEORY, CONCEPT OF LEXICOGRAPHICAL THEORY, FUNCTION THEORY, ACADEMIC STATUS OF LEXICOGRAPHY, INTERDISCIPLINARY VOCATION OF LEXICOGRAPHY, DICTIONARIES, CONSULTATION TOOLS, INFORMATION SCIENCE

Opsomming: Bestaan daar 'n behoefte aan 'n (nuwe) leksikografieteorie? Tydens die huidige oorgang van gedrukte na elektroniese woordeboeke is die vraag geopper of ons 'n nuwe leksikografieteorie benodig wat die ontwerp en produksie van leksikografiese e-gereedskap kan rig en of ons die teorieë wat reeds ontwikkel is tydens die era van gedrukte werke kan gebruik. Om hierdie vraag te beantwoord bespreek die artikel eerstens die vraag of 'n leksikografieteorie bestaan en of dit enigsins moontlik of selfs wenslik is. Dit toon dat die verskillende benaderings tot hierdie kwessie hoofsaaklik te danke is aan die feit dat die hele konsep van teorie wyd binne leksikografiese kringe betwis word. Kortliks word die Angel-Saksiese akademiese tradisie bespreek waarvolgens die wetenskap slegs betrekking het op natuurlike verskynsels en waar alle ander verskynsels binne die sfeer van die kunste en vaardighede val. Dit toon ook dat hierdie tradisie in groot teenstelling is met die tradisies in ander dele van die wêreld. Vanuit hierdie vertrekpunt toon die artikel dat 'n leksikografieteorie nie alleen moontlik is nie, maar dat verskeie hoogs nuttige teorieë reeds bestaan. Ten slotte word beweer dat hierdie teorieë, veral dié wat nie net fokus op die gedrukte woordeboekvorm nie, nie verwerp behoort te word nie, maar verder uitgebrei en verbeter moet word in noue wisselwerking tussen die leksikografie en ander raadple-

gingsvakrigtings binne die breë veld van die inligtingswetenskap.

Sleutelwoorde: LEKSIKOGRAFIE, LEKSIKOGRAFIETEORIE, KONSEP VAN LEKSIKOGRAFIETEORIE, FUNKSIETEORIE, AKADEMIESE STATUS VAN DIE LEKSIKOGRAFIE, INTERDISSIPLINÊRE LEKSIKOGRAFIEBEROEP, WOORDEBOEKE, RAADPLEGINGSHULPMIDDELS, INLIGTINGSWETENSKAP

1. Introduction

Lexicography is in the middle of a challenging but also troublesome transition from printed to electronic dictionaries, cf. Fuertes-Olivera and Bergenholtz (2011a), Granger and Paquot (2012). In this process the question has been raised whether we need a new theory that may guide the conception and production of lexicographical e-tools or if we can use the theories already developed in the era of printed works. In order to answer this question it is first of all necessary to give answer to another question, i.e. whether a lexicographical theory exists, is possible at all or even wanted. This is due to the fact that the very concept of theory is widely disputed within lexicographical circles; cf. Bergenholtz et al. (2008).

In a recent publication where he discusses the theory of lexicographical functions, Yukio Tono (2010: 2) asks: "*Do we really need a 'theory'*?" Tono's final answer to his own question is affirmative. At a high level of abstraction, i.e. independent of the specific theory in question, this view is shared by the Russian tradition (e.g. Scerba 1940, Sorokoletov 1978), the German lexicographical tradition (e.g. Duda et al. 1986, Wiegand 1989, Zöfgen 1994, Weber 1996), Danish lexicographers (e.g. Kromann et al. 1984, Bergenholtz and Tarp 2003), various South African lexicographers (e.g. Gouws 2011), the Czech-American lexicographer Zgusta (1992), and at least part of the Chinese tradition (e.g. Yong and Peng 2008), among many others. However, other lexicographers, especially some of those belonging to the Anglo-Saxon tradition, do not support this vision of lexicography. For instance, Atkins and Rundell (2008: 4) "*do not believe that such a thing exists*" (i.e. a theory of lexicography). And in a recent book, Henri Béjoint (2010: 381) writes:

> I simply do not believe that there exists a theory of lexicography, and I very much doubt that there can be one. Those who have proposed a general theory have not been found convincing by the community, and for good reasons. A theory is a system of ideas put forward to explain phenomena that are not otherwise explainable. A science has a theory, a craft does not. All natural phenomena need a theory, but how can there be a theory of the production of artefacts? There are theories of language, there may be theories of lexicology, but there is no theory of lexicography. Lexicography is about all a craft, the craft of preparing dictionaries, as well as an art, as Landau (2001) says. It may be becoming more scientific, but it has not become a science.

This point of view is strongly embedded in an Anglo-Saxon academic tradition according to which science is only related to natural phenomena and where all other phenomena are referred to the sphere of art and craft. As mentioned, this tradition is widely opposed by the traditions in other parts of the world. For instance, in Danish universities the students can study *cultural science* (kulturvidenskab), *art science* (kunstvidenskab), *literature science* (litteraturvidenskab), *social science* (samfundsvidenskab), *information science* (informationsvidenskab), among many other disciplines. More or less the same disciplines may be studied as sciences with academic degrees at German and South African universities.

Whether lexicography should or should not be considered a science is a question of dispute that basically has to do with the concept of science used by the particular scholar. As mentioned, Béjoint (2011) and the Anglo-Saxon academic tradition reject it. Wiegand (1998), who bases his views on a concept of science formulated by Posner (1988), considers lexicography to be a scientific research area but still not a science as it does not yet fulfil two of fifteen required criteria formulated according to this concept. By contrast, the Russian lexicographer V. Dorosevskij (quoted in Sorokoletov 1978: 79), who bases himself on another concept of science, defined lexicography as "*the science of the classification processes of word material and its presentation in dictionaries*". Although with another definition, this view is shared by Tarp (2008) who relates to the following concept of science taken from *Philosophisches Wörterbuch* (cf. Buhr and Klaus 1971: 1083, 1169):

> A science is a system of knowledge growing out of social practice and developing on an on-going basis, comprising the acknowledgement of the most important properties, causal connections and legal considerations of nature, society and philosophy; rooted in the form of concepts, categories, defined goals, laws, theories and hypotheses, and constituting the basis of Man's growing mastery of his natural and social environment. A science also consists of its own history, pre-theoretical ideas, contributions to methodology, directions for practical action etc.

If this concept of science is related to existing lexicographical theory and practice, it can be concluded that lexicography satisfies all the demands made on any branch of human activity that claims to be a separate science or area of academic study, cf. Tarp (2008: 6):

— It has its own object of study: dictionaries, or to be more precise the planning, production, design and usage of lexicographical works and the close relationship between these works and specific types of social need;

— it is rooted in the form of concepts, categories, theories and hypotheses;

— it comprises both the history of dictionaries and its own history, including pre-theoretical ideas;

— it contains independent contributions to methodology;

— it includes directions for practical action.

As already mentioned, this vision of the academic status of lexicography basically depends on the underlying concept of science and in order to avoid unnecessary misunderstandings it should therefore be recommended that any future discussion of this status is always explicitly related to a specific concept of science and not just to science in general. In this respect, it is noteworthy that even a well-established discipline as information science which has the term *science* included in its very name comprises an internal tendency belonging to the Anglo-Saxon tradition which rejects or at least raises doubts over its status as a science, e.g. Machlup (1983), Brooks (1989) and Buckland (2012), a debate which other scholars regard as "not particularly rewarding", cf. Norton (2010: 8).

2. Lexicographical theory

It is impossible to read the academic literature on lexicography without coming up with the word *theory*. It is used as reference both to *general theories* that cover the whole subject field of lexicography as a discipline and to *specific theories* that cover any subarea of this enormous discipline. Some of the general theories are well-known: [Towards a] *General Theory of Lexicography*, cf. Scerba (1940), the *General Theory of Lexicography*, cf. Wiegand (1998), and the *Function Theory*, cf. Bergenholtz and Tarp (2002, 2003). In the above quoted article, Sorokoletov (1978: 79) writes:

> In the Soviet period lexicography developed into an *independent discipline with its own theory*, own tasks and own methods for their solution.

In criticizing previous studies on the history of Chinese lexicography, among other things because "*dictionary compilation is viewed as a purely linguistic activity*", Yong and Peng (2008: 5) also stress the need for a theory to support this type of research:

> It is frequently apparent in their research [the previous one] that more emphasis is laid on the parts than on the whole, that more attention is paid to the isolated analysis of cases than to *theoretical generalizations*, and that more consideration is given to accumulation of practical experiences than to formulation of *lexicographical theories*.

Even a scholar like Piotrowski (2009: 485) who is strongly critical of the function theory does not deny the need for a "*new theory of lexicography*". By analogue, Bogaards (2010) who has looked for the word *theory* in a corpus of texts from the *International Journal of Lexicography* also confirms that many scholars refer to theory in their reflections on dictionaries although he himself doubts that such a theory already exists.

Apart from these references to the existence of and need for more or less general theories, in the academic literature quite a lot of references to more specific lexicographical theories can also be found, e.g. a *theory of bilingual lexicog-*

raphy (Duda et al. 1986), a *theory of the lexicographical example* (Hausmann 1985 and Zöfgen 1987), a *theory of lexicographical language description* (Wiegand 1983), a *theory of the dictionary form* (Wiegand and Morán 2009), etc. It is therefore simply wrong when it is claimed that no lexicographical theory exists although everybody is welcome to accept, critize or reject the existing theories but not to deny or ignore their existence. It goes without saying that the various lexicographical theories proposed are of a highly varying character as some of them are strongly embedded in linguistics whereas others are based upon the idea that lexicography is an independent discipline (and even science) with a strong interdisciplinary vocation. Besides, it must be admitted that reference to theory is frequently made without clarifying what is meant by theory. Just as was the case with the concept of science, there are also various "competing" concepts of theory to be found in the academic and philosophical literature. In this respect, the function theory is explicitly based upon the following concept of theory, once more taken from *Philosophisches Wörterbuch* (cf. Buhr and Klaus 1971: 155, 1083):

> A theory is a systematically organised set of statements about an area of objective reality or consciousness, i.e. logical structures reflecting the fact that certain things have certain properties, or that certain relationships exist between these things.

In this veil, Tarp (2008: 9-11) defines a *lexicographical theory* in general as a systematically organised set of statements about dictionaries and other lexicographical works and their relationship with specific types of social need, and at the same time stresses that it is necessary to distinguish between different types of theory: *general* and *specific theories*, *integrated* and *non-integrated theories* (i.e. specific theories integrated or not in a general theory of lexicography), and *contemplative* and *transformative theories*.

It is noteworthy that Bogaards (2010: 316) — one of the few lexicographers who has published a specific criticism of the function theory — does not accept that this theory actually is a theory, most of all because it "lacks any form of empirically verifiable or falsifiable hypotheses". A similar comment was sent to the author of this article (Bogaards: Personal communication):

> You finally define a theory (of lexicography) as 'a systematically structured set of statements'. Without going into long epistemological debates, I would say that a theory should first of all consist of a number of hypotheses and axioms, as well as of a set of methods to falsify the hypotheses.

Here it becomes clear that the discrepancy as to the validity of this specific theory is due to the underlying concept of theory used by each researcher. The function theory does not consider hypotheses to be part of the concept of theory, but as something pre-theoretical that may eventually lead to theory (or improved theory) and which belongs to the broader concept of science, cf. Buhr and Klaus (1971: 1169). Falsification is a highly useful method to validate theo-

ries within many disciplines but, as Bergenholtz et al. (2008: 157-159) has argued, the idea that every theory has to be falsifiable is an idea taken over from natural sciences and the Popperian tradition which, as a rule, is not valid inside various social disciplines where a lot of *empirically grounded theories* have been formulated that are absolutely necessary in order to understand practice and make progress within these disciplines although they cannot be falsified in the same manner and with the same rigid criteria as theories related to natural sciences. The same can be said about lexicographical theories.

3. A lexicographical theory is possible

As mentioned above, Béjoint (2010: 381) bases himself upon the idea that "*there is no theory of lexicography*", "*lexicography is about all a craft, the craft of preparing dictionaries, as well as an art*" and that "*a science has a theory, a craft does not*". He then asks: "*How can there be a theory of the production of artefacts?*"

What should we think about these statements? First of all, it must be admitted that the practical production of dictionaries is neither a science nor a theory but a thousand-year-old cultural practice that can reasonably be defined or viewed as a craft which, as all other crafts, has developed historically with a view to satisfying certain needs detected in society. But this does in no way exclude that this craft — as well as the needs that give rise to it, its practical products (dictionaries and other lexicographical tools) and the use made of these — can be subjected to observation, empirical studies and theoretical generalizations. If one takes the point of departure in the above definition of the concept of theory it is perfectly possible:

(1) to *observe and study* this craft in all its dimensions,

(2) to isolate relevant *phenomena with certain properties*,

(3) to establish the *relations* existing between them,

(4) to make *statements* about these phenomena and relations, and

(5) to *systematize* these statements.

This is all it takes to *formulate a coherent theory of lexicography* if the underlying concept of theory is the one discussed above. It is also worth noting that this is actually what is done in the bulk of the academic literature on lexicography, at least in terms of the first four of the above steps, as it is still relatively rare to meet the necessary theoretical generalizations, cf. Yong and Peng (2008: 5), and the systematization of the statements made in the form of logical structures. The inevitable conclusion is therefore that a theory of lexicography *is possible*, and that various — frequently competing — general and specific theories of lexicography *do exist*.

It was such a theory capable of explaining, guiding and even renovating

the existing practice that Scerba (1940) intended to formulate in his groundbreaking contribution to lexicography, this was what Wiegand did with his "general theory of lexicography", and what has been done with the "theory of lexicographical functions". It may be that these theories "have not been found convincing" by the Anglo-Saxon lexicographical community, but this does not mean that they do not exist, are not possible and even highly needed by those who try to solve the complex problems within present-day lexicography.

Whereas specific theories may deal with whichever area or subarea of the discipline, a general theory of lexicography should necessarily include *all aspects* of lexicography (including *all types* of dictionaries and other lexicographical works) and *cannot be restricted only to a subset of dictionaries*, e.g. dictionaries where special linguistic knowledge have been required. *Oxford University Press*, for instance, is not only the publisher of a range of advanced learners', school and concise dictionaries of English, bilingual as well as monolingual, but also of a whole series of dictionaries of archeology, classical civilization, military history, philosophy, law, economics, sociology, art, music, chemistry, biology, nursing, and so on. Other publishers have produced lexicographical works dealing with thousands of other topics, the confection of which requires highly specialized knowledge of these disciplines, but not of linguistics as a separate academic discipline.

A general theory of lexicography cannot ignore that lexicographical works are multi-faceted cultural artifacts and utility tools which, during the millenniums, have met a wide range of different needs detected in society and covered almost all spheres of human activity and knowledge, nor should it depart from the differences that separate all these works in terms of their specific content, structure etc., but from the aspects and elements that unite them and are common to all of them. These uniting aspects and elements can be considered *the core of lexicography* which, according to the function theory, is:

— the design of utility tools

— that can be quickly and easily consulted

— with a view to meeting punctual information needs

— occurring for specific types of users

— in specific types of extra-lexicographic situations.

If this core activity is studied and generalized in the form of a systematic set of statements, the theory formulated in this way will be able to guide the design not only of an entirely new generation of lexicographic works of the kind we already know, but also of consultation tools covering areas that have still not been treated lexicographically. No theory is born fully armed as Athena from Zeus' forehead and existing lexicographical theory, including the function theory, necessarily has to pass through a continuous process of validation and perfection as a result of the fruitful interaction with its *alter ego*, the lexicographical

practice, which is also in a process of continuous progress. As the Danish lexicographer Henning Bergenholtz repeats time after time, cf. Nielsen and Tarp (2009: ix): "*Nothing is more practical than a good theory*".

Such a good theory of lexicography does not only exist today (although it should be continuously perfected), it is also highly needed in order to produce high-quality lexicographical tools in the on-going and troublesome transition from printed to electronic dictionaries.

4. A new theory?

The title of this contribution also raises the question whether we need a *new theory* of lexicography? The answer to this question is both *yes* and *no*, cf. Tarp (2011: 54). On the one hand it is evident that theories only dealing with printed dictionaries, e.g. the one developed by Wiegand and Morán (2009) regarding the dictionary form, cannot give answers to the challenges of electronic dictionaries. On the other hand, the use of the computer and information sciences as well as the increased focus on information in present-day society have made it clear that lexicography is above all an information discipline, a vision shared by a growing number of scholars, e.g. Bothma (2011), Fuertes-Olivera and Bergenholtz (2011b), Heid (2011), Leroyer (2011), among others. When an abstraction is made from the concrete and specific content of the needs that lexicographical works have intended to meet during its whole existence, what is left is their common character of *information needs*. Besides, if a distinction — as it was already done by Hausmann (1977: 144) — is made between global information needs, i.e. the needs related to a more profound study of a specific subject field (or part of it), and punctual information needs related to a single and limited topic within a larger subject field, or to the solution of specific tasks or problems, then it becomes clear that lexicographical works and tools are by excellence *artefacts designed to be consulted* in order to meet punctual information needs.

A short panoramic overview shows that lexicographical works are not the only types of artefacts produced with a view to satisfying punctual information needs. Manuals, how-to's, user guides, telephone books, and indexes in books are all text types totally or partially designed to be consulted in order to retrieve punctual information for one purpose or another. And the list can easily be extended, including even train, bus and flight plans, and other similar reference tools.

It is obvious that these tools have not been designed and produced according to lexicographical principles but it is nevertheless evident that they all have something fundamental in common with dictionaries and other lexicographical tools although they have developed from different traditions. One may in fact speak about the contours of one big discipline including all types of consultation tools designed to meet punctual information needs, a discipline which necessarily should develop its own general theory as part of information

science. In the development of such a theory, lexicography has both a lot to contribute to other theories dealing with punctual consultation tools and to information science in general and a lot to learn from these theories and this science. In this respect, it is worth quoting the visionary words by Gouws (2011: 29):

> Looking back at the development of the theory and practice of lexicography it is clear that for too long the practice of printed dictionaries had to go without a sound theory, for too long lexicography did not establish itself as an independent discipline, for too long the pool of lexicographers had been restricted to experts from a single field, for too long innovation in the lexicographic practice was impeded by its theory being a follower and not a leader, for too long lexicographic theory was exclusively directed at being implemented in the production of dictionaries. Looking at the future, the planning and compilation of electronic dictionaries and the further development of a coherent and medium-unspecific theory we need to unlearn a lot, we need to learn a lot so that we can be innovative and produce better reference tools, including even dictionaries.

The final conclusion is that lexicography will not cease to be an independent discipline with its own specific subject field as well as its own theory and practice but that it will tend to relate more and more to and interact with similar disciplines within the broad area of information science. Consequently, existing theory should not be rejected but should continuously be enhanced and improved through this hopefully fruitful interaction with sister disciplines. In this way, lexicographical theory will place itself in an even better position to assist and guide the present transition from printed to electronic dictionaries.

The gauntlet is down. Who will pick it up?

Literature

Atkins, B.T. Sue and Michael Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

Béjoint, Henri. 2010. *The Lexicography of English.* Oxford: Oxford University Press.

Bergenholtz, Carsten, Henning Bergenholtz and Sven Tarp. 2008. Leksikografi i videnskabsteoretisk perspektiv: sand, falsk eller irrelevant. *LexicoNordica* 15: 155-168.

Bergenholtz, Henning and Sven Tarp. 2002. Die moderne lexikographische Funktionslehre. Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen. *Lexicographica* 18: 253-263.

Bergenholtz, Henning and Sven Tarp. 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31:171-196.

Bogaards, Paul. 2010. Lexicography: Science without Theory? De Schryver, Gilles-Maurice (Ed.). 2010. *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks*: 313-322. Kampala: Menha Publishers.

Bothma, Theo J.D. 2011. Filtering and Adapting Data and Information in the Online Environment in Response to User Needs. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011.

e-Lexicography: The Internet, Digital Initiatives and Lexicography: 71-102. London/New York: Continuum.

Brooks, Terrence A. 1989. The Model of Science and Scientific Models in Librarianship. *Library Trends* 38(2): 237-249.

Buckland, Michael. 2012. What Kind of Science *Can* Information Science Be? *Journal of the American Society for Information Science and Technology* 63(1): 1-7.

Buhr, Manfred and Georg Klaus (Eds.). 1971. *Philosophisches Wörterbuch. Band 1–2*. Berlin: Das europäische Buch.

Duda, W., M. Frenzel, E. Wöller and T. Zimmermann. 1986. *Zu einer Theorie der zweisprachigen Lexikographie. Überlegungen zu einen neuen russisch–deutschen Wörterbuch.* Berlin: Akademie-Verlag.

Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011a. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. London/New York: Continuum.

Fuertes-Olivera, Pedro A. and Henning Bergenholtz. 2011b. Introduction: The Construction of Internet Dictionaries. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 1-16. London/New York: Continuum.

Gouws, Rufus H. 2011. Learning, Unlearning and Innovation in the Planning of Electronic Dictionaries. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 17-29. London/New York: Continuum.

Granger, Sylviane and Magali Paquot (Eds.). 2012. *Electronic Lexicography*. Oxford: Oxford University Press.

Hausmann, Franz Josef. 1977. *Einführung in die Benutzung der neufranzösischen Wörterbücher*. Tübingen: Max Niemeyer.

Hausmann, Franz Josef. 1985. Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels. Bergenholtz, Henning and Joachim Mugdan (Eds.). 1985. *Lexikographie und Grammatik: Akten des Essener Kolloquiums zur Grammatik im Wörterbuch 28–30.6.1984*: 118-129. Lexicographica: Series Maior 3. Tübingen: MaxNiemeyer.

Heid, Ulrich. 2011. Electronic Dictionaries as Tools: Towards an Assessment of Usability. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 287-304. London/New York: Continuum.

Kromann, Hans-Peder, Theis Riiber and Poul Rosbach. 1984. Überlegungen zu Grundfragen der zweisprachigen Lexikographie. Wiegand, Herbert Ernst (Ed.). 1984. *Studien zur neuhochdeutschen Lexikographie V. Germanistische Linguistik* 3-6: 159-238. Hildesheim/New York: Georg Olms.

Landau, Sidney. 2001. *Dictionaries: The Art and Craft of Lexicography*. Second Edition. New York/ Cambridge: Cam-bridge University Press.

Leroyer, Patrick. 2011. Change of Paradigm: From Linguistics to Information Science and from Dictionaries to Lexicographic Information Tools. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 121-140. London/New York: Continuum.

Machlup, F. 1983. Semantic Quirks in Studies of Information. Machlup, F. and U. Mansfield (Eds.). 1983. *The Study of Information: Interdisciplinary Messages*: 641-671. New York: John Wiley & Sons.

Nielsen, Sandro and Sven Tarp. 2009: Nothing is More Practical Than a Good Theory. Nielsen, Sandro and Sven Tarp (Eds.). 2009. *Lexicography in the 21st Century. In honour of Henning Bergenholtz*: ix-xi. Amsterdam/Philadelphia: John Benjamins.

Norton, Melanie J. 2010. Information and Information Science. Norton, Melanie J. (Ed.). 2010. *Introductory Concepts in Information Science*: 3-18. Second Edition. New Jersey: Information Today.

Piotrowski, Tadeusz. 2009. [Review of] Sven Tarp. Lexicography in the Borderland Between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography. *International Journal of Lexicography* 22(4): 480-486.

Posner, Roland. 1988. What is an Academic Discipline? Claussen, Regina and Roland Daube-Schackat (Eds.). 1988. *Gedankenzeichen. Festschrift für Klaus Oehler zum 60. Geburtstag*: 165-185. Tübingen: Stauffenburg Verlag.

Scerba, L.V. 1940. Towards a General Theory of Lexicography. *International Journal of Lexicography* 8(4), 1995: 315-350.

Sorokoletov, F.P. 1978. Traditionen der sowjetrussischen Lexikographie. Wolski, Werner (Ed.). 1982. *Aspekte der sowjetrussischen Lexikcographie. Übersetzungen, Abstracts, bibliographische Angaben*: 63-88. Tübingen: Max Niemeyer.

Tarp, Sven. 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Lexicographica. Series Maior 134. Tübingen: Max Niemeyer.

Tarp, Sven. 2011. Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 54-70. London/ New York: Continuum.

Tono, Yukio. 2010. A Critical Review of the Theory of Lexicographical Functions. *Lexicon* 40: 1-16.

Weber, Nico. (Ed.). 1996. *Theorie der Semantik und Theorie der Lexikographie*. Tübingen: Max Niemeyer.

Wiegand, Herbert Ernst. 1983. Was ist eigentlich ein Lemma? Ein Beitrag zur Theorie der lexikographischen Sprachbeschreibung. Wiegand, Herbert Ernst (Ed.). 1983. *Studien zur neuhochdeutschen Lexicographie* III: 401-474. Germanistische Linguistik 1–4/1982. Hildesheim/New York: Georg Olms Verlag.

Wiegand, Herbert Ernst. 1989. Der gegenwärtige Status des Lexikographie und ihr Verhältnis zu anderen Disziplinen. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta (Eds.). 1989-1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie*: 246-280. Berlin/New York: Walter de Gruyter.

Wiegand, Herbert Ernst. 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie.* 1. Teilband. Mit 159 Abbildungen im Text. Berlin/New York: Walter de Gruyter.

Wiegand, Herbert Ernst and María Teresa Fuentes Morán. 2009. *Estructuras lexicográficas. Aspectos centrales de una teoría de la forma del diccionarío*. Granada: Edicíones Tragacanto.

Yong, Heming and Jing Peng. 2008. *Chinese Lexicography. A History from 1046 BC to AD 1911.* Oxford: Oxford University Press.

Zgusta, Ladislav. 1992. The Czech–Chinese Dictionary and the Theory of Lexicography. *International Journal of Lexicography* 5(2): 85-128.

Zöfgen, Ekkehard. 1987. Kollokation — Kontextualisierung — (Beleg-)Satz. Anmerkungen zur Theorie und Praxis des lexikographischen Beispiels. Barrera-Vidal, Alberto, Hartmut Kleineidam, Manfred Raupach (Eds.). 1987. *Französische Sprachlehre und bon usage. Festschrift für Hans Wilhelm Klein zum 75. Geburststag*: 219-238. München: Max Hueber.

Zöfgen, Ekkehard. 1994. *Lernerwörterbuch in Theorie und Praxis. Ein Beitrag zur Metalexikographie mit besonderer Berücksichtung des Französischen*. Lexicographica. Series Maior 59. Tübingen: Max Niemeyer.

School Dictionaries for First-Language Learners

Sven Tarp, *Department of Afrikaans and Dutch, University of Stellenbosch, South Africa, and Centre for Lexicography, Aarhus School of Business and Social Sciences, University of Aarhus, Denmark (st@asb.dk)*
and
Rufus H. Gouws, *Department of Afrikaans and Dutch, University of Stellenbosch, Stellenbosch, South Africa (rhg@sun.ac.za)*

Abstract: Acknowledging the ambiguity of terms like *school dictionary, children's dictionary, first language, mother-tongue* this article motivates a specific use of *school dictionary, first language* and *learner* and focuses on various problems in these dictionaries. The typical functions of these dictionaries are discussed with reference to the lexicographic needs of first-language learners. Looking at a few existing dictionaries, suggestions are made for the inclusion and presentation of certain data types. The importance of the use of natural language in the paraphrases of meaning is discussed. It is emphasised that lexicographers should consult teachers and curriculum experts when planning school dictionaries and that the grade and age of the target user needs to be taken into account. The aim of this article is not to give final solutions to the questions raised but merely to recommend that a number of factors — or variables — are taken into account when planning future school dictionaries. In this respect, a number of questions are formulated that need to be answered when planning the compilation of school dictionaries.

Keywords: CHILDREN'S DICTIONARY, DICTIONARY CULTURE, FIRST LANGUAGE, LEARNER, LEXICOGRAPHIC FUNCTIONS, LEXICOGRAPHICAL NEEDS, LEXICOGRAPHY, MOTHER-LANGUAGE, PARAPHRASE OF MEANING, SCHOOL DICTIONARY

Opsomming: Skoolwoordeboeke vir eerstetaalleerders. Met inagneming van die dubbelsinnigheid van terme soos *skoolwoordeboek, kinderwoordeboek, eerste taal, moedertaal* kies hierdie artikel vir 'n spesifieke gebruik van die terme *skoolwoordeboek, eerste taal* en *leerder.* Die fokus is op verskeie probleme in hierdie woordeboeke. Die tipiese funksies van hierdie woordeboeke word bespreek met verwysing na die leksikografiese behoeftes van eerstetaalleerders. Na aanleiding van 'n paar bestaande woordeboeke word voorstelle gemaak vir die opname en aanbieding van sekere datatipes. Die belang van natuurlike taal in die betekenisparafrases word bespreek. Dit word beklemtoon dat leksikograwe ook onderwysers en kurrikulumdeskundiges moet raadpleeg wanneer skoolwoordeboeke beplan word en dat die graad en ouderdom van die teikengebruiker in ag geneem moet word. Die doel van hierdie artikel is nie om finale oplossings te gee vir die bestaande probleme nie maar eerder om aan te beveel dat sekere faktore — of veranderlikes — in ag geneem moet word wanneer skoolwoordeboeke beplan word. In hierdie verband word 'n paar vrae geformuleer waarop antwoorde gevind moet word wanneer beplanning gedoen word vir die daarstel van skoolwoordeboeke.

Lexikos 22 (AFRILEX-reeks/series 22: 2012): 333-351

Sleutelwoorde: BETEKENISPARAFRASE, EERSTE TAAL, KINDERWOORDEBOEK, LEERDER, LEKSIKOGRAFIE, LEKSIKOGRAFIESE BEHOEFTES, LEKSIKOGRAFIESE FUNKSIES, MOEDERTAAL, SKOOLWOORDEBOEK, WOORDEBOEKKULTUUR

1. Introduction

In discussing the development of reference sources, including dictionaries, McArthur (1986) classifies these tools as "containers of knowledge". Today the average member of a literate speech community still regards a dictionary as one of the most important sources from which information regarding language and a variety of other matters can be retrieved. Dictionaries are practical tools and as such the nature and extent of their use should never be underestimated. One of the many environments where a user relies on "the" dictionary is within the school system where teachers ever so often refer learners to dictionaries for solutions regarding a wide-ranging series of problems. Unfortunately very often the dictionary consultation does not help to solve the problems of these users. This is due to different reasons, e.g. a school dictionary that is not really directed at the needs of its intended target users, the teacher referring the user to a dictionary used in school that is not really a school dictionary or even, and quite often, the lack of dictionary using skills.

A limited knowledge regarding school dictionaries is often the result of the lack of a dictionary culture. When referring to the notion of dictionary culture, cf. Hausmann (1989), a distinction can be made between a societal and an individual dictionary culture, cf. Gouws (2012). A societal (also known as a collective) dictionary culture implies that a general and wide-spread dictionary culture prevails within a given speech community. In contrast, an individual (also known as ideolectal) dictionary culture prevails within the individual member of a speech community, in spite of the lack of a societal dictionary culture. Where a given speech community lacks a societal dictionary culture the notion of *school dictionaries* is likely to be treated in an insufficient or haphazard way.

One way to enhance a societal dictionary culture and ensure a more scientific approach to the notion of school dictionaries is to introduce both dictionary using skills and intensive dictionary using opportunities in the early and later school years. This can only be done if school children have sufficient access to well-devised school dictionaries. In this regard government intervention may be necessary. A practical example of such an intervention is the education system in Brazil where policy demands that each student in school must have his/her own monolingual dictionary of Brazilian Portuguese.

Wiegand (1989: 251) maintains that lexicography is a scientific practice, aimed at the production of dictionaries, so that a further practice, i.e. the cultural practice of dictionary use can be initiated. One of the problems regarding school dictionaries is that this scientific practice has not always been done in a sufficiently scientific way. The planning and compilation of school dictionaries

should be seen as a team effort, combining lexicographic, curriculum as well as pedagogical expertise. The lack of any one of these fields of expertise will lead to a dictionary of a lesser quality.

Different typological classifications of dictionaries exist, cf. Malkiel (1967), Zgusta (1971), Geeraerts (1984) and Gouws (1989), to name but a few. One of the problems experienced in the use of dictionaries at school level lies in the fact that in many classifications the category of "school dictionary" is such a vague and ambiguously defined typological category. The main problem with this category resides in its general and unspecific nature. Where the notion of school dictionary is defined as a dictionary used in school it can be interpreted that each and every dictionary that is used in a school, and not only those specifically compiled for use in schools, can be regarded as a school dictionary. This leads to the above-mentioned problem where school learners consult a dictionary that is not a school dictionary and the consultation does not lead to the required results.

A competing term to *school dictionary* is *children's dictionary,* cf. Tarp (2011). This term is often used for dictionaries primarily compiled for use by pre-school children but they are also used during the foundation phase. A dictionary like the *Oxford Children's Thesaurus* clearly states on its back cover that it is directed at users of 8 years of age and older, i.e. school-going users. In comparison the *Collins Junior Illustrated Dictionary,* with a title that seems to be identifying a slightly more advanced user group, indicated on its back cover that the target users are of the age of six years plus.

In this paper the focus will be on school dictionaries for first language learners. However, in the title of this paper there are at least three problematic concepts, i.e. *school dictionary, first language* and *learner.* These terms need to be disambiguated for the purpose of this paper.

It is currently a widely accepted criterion that one of the core components in the planning of any dictionary has to be a clear identification of the intended target user of the envisaged dictionary. Quite often one finds an indication of the target user, e.g. *learners*, in the title of a dictionary. The word *school* often prevails in the titles of dictionaries but when looking at the functions, structures and contents of the dictionary it is not clear at all why this word occurs in the title. In some instances it merely indicates that the dictionary is of a more restricted extent. This typically prevails when the so-called school dictionary is a reduced version of a more comprehensive dictionary. This reduction is typically done by deleting some articles or by omitting some entries in certain articles. This leads to a version that often has a higher degree of textual condensation than its more comprehensive counterpart and is even more difficult to use. Such a dictionary is not a school dictionary. The term *school dictionary* refers to those dictionaries specifically compiled for use in schools, albeit that some of them are also used in a pre-school environment, cf. Tarp (2011). Whether they are titled *school dictionary* or *children's dictionary* or even have a title with no reference to the school environment, if they are planned and compiled to be used

in schools they are regarded as school dictionaries. A school dictionary therefore is a dictionary with the genuine purpose to assist users that are school learners to find the solution for problems related to their school work. It is important that school dictionaries should clearly be distinguished from other dictionaries used in school but not specifically planned and compiled with school learners as their envisaged target user group. Dictionaries of the latter type are not regarded in this paper as school dictionaries.

The term *first language* is also problematic because different terms like *mother tongue, native language, home language* and *primary language* are often used to refer to the same thing. All these terms are problematic. Within South African schools the term *home language* is used in opposition to *additional language*, with the first term implying a language subject with a higher academic level than the second. Yet, this term does not imply that it is the best language of the student. A student could have language X as the language used at home and the language in which he/she is the most comfortable. Due to a variety of reasons this student may opt for or be compelled to take language Y as the higher level language at school. In this paper the intricacies of these terms will not be discussed. The term *first language* is used here to refer to the higher grade language, with the implication that in the majority of cases it will be the language in which the student is the most comfortable, which often will be the language he/she uses at home and typically will be the language used as his/her medium of instruction at school. However, it is not necessarily the student's home/mother/native/primary language. The focus on dictionaries for first language learners motivates the decision to exclude bilingual dictionaries from the discussion.

The term *learner* is problematic because in different environments different interpretations are attached to this term. In the typological classification of dictionaries the *learner* referred to in the category *learners' dictionary* indicates a specific category of users, i.e. those users, mostly adults, who study a foreign language. Within the South African educational environment a learner refers to a student attending a school, i.e. from the first to the last school year. Yet again, this paper will not endeavour to judge the decisions by either lexicographers or educationalists regarding the selection of a given term. In this paper the term *learner* refers to a student attending school, i.e. a student more or less from the age of five up to the age of eighteen.

This paper should not be seen as a once-off or isolated investigation. It links directly with an earlier paper, i.e. Tarp and Gouws (2010), and is partially based on presentations by both Tarp and Gouws in two workshops focusing on school dictionaries, one in Stellenbosch (2010) and one in Pretoria (2011), as well as a workshop in Namibia (2011).

This paper primarily refers to printed dictionaries and uses examples from these dictionaries. However, the underlying theoretical principles are not only directed at printed dictionaries but could also be applied to electronic dictionaries. This links with an important approach in lexicography that electronic

dictionaries do not need a separate theory. A general theory of lexicography should rather be formulated in such a way that it can form the basis for the planning and compilation of both printed and electronic dictionaries, albeit it that provision needs to be made for certain medium-specific aspects.

The world of school dictionaries for first language learners is full of problems. This paper will not endeavour to give solutions to these problems but will rather identify and discuss some of the problems and make a few suggestions that could play a role in working towards a better dispensation. The paper will not take a contemplative view by only looking at existing dictionaries, but will also have a transformative approach with proposals for an improved lexicographic practice.

2. Functions of school dictionaries for first-language learners

Existing school dictionaries for first-language learners vary a lot in form and content, not only from one country or language community to another, but even within the same country and language community. This variety may not only be explained by the different traditions but also by the fact that school dictionaries by their very nature may have a big number of different functions in terms of the foreseen user group which has to be categorised according to age (grade) and the corresponding intellectual, linguistic, cultural, and encyclopedic development of the school children, as well as in terms of the various types of learning situations where the children may need to consult or use a dictionary. The publishing houses inevitably have to adapt to this reality and focus on one or a few aspects as no single dictionary will be able to cover the whole spectrum of needs of a highly heterogeneous user group in all the relevant situations. The inevitable result is the existing variety of school dictionaries of which many claim to cover a much bigger group of users and situations than they actually do, a fact that contributes to the lowering of the prestige and quality of the dictionaries.

In order to evaluate existing school dictionaries for first-language learners and come up with recommendations for future dictionaries it is first and foremost necessary to establish the real needs of the potential users of these dictionaries, i.e. the school children studying to improve their first language. These needs are not only intimately connected with the personal characteristics of the user group itself but also with the various learning situations where lexicographically relevant needs may occur for this group.

Most publishers of school dictionaries refer to reception and production of written and oral texts when they explain in which situations their dictionaries may be used, and it is a fact that these two communicative situations are the fundamental situations that have to be covered by school dictionaries although they may be further subdivided into "normal" text reception and production and exercise-related reception and production.

To these communicative situations should be added two fundamental

cognitive situations, i.e. vocabulary learning and grammar learning, where the children — frequently together with the teacher — use the material provided by the dictionary to study and assimilate these two basic components of the language. However, it is important to stress that vocabulary learning most often — and especially for the younger children — goes together with encyclopedic and cultural learning as you cannot learn a word without knowing what it refers to. In this respect, school dictionaries may also assist the school children in a third cognitive situation, i.e. learning about the world, getting world knowledge as a basis for vocabulary learning and interwoven with this. Dictionaries directly conceived to assist vocabulary building contain, as a rule, thematic sections — mostly with illustrations, and sometimes even with illustrations as "lemmata" — whereas dictionaries assisting grammar learning contain special sections — in printed dictionaries frequently placed in the front or back matter — where inflection, word formation, punctuation and other grammatical phenomena are treated in a systematic way. In these cases, the school dictionaries are not primarily used as consultation tools but as mini-text books which can be studied section by section instead of using "normal" text books.

Apart from the two communicative and the three cognitive situations mentioned, all of which are directly related to the learning of a language, school dictionaries in some countries may also provide assistance in a another type of cognitive situation where the children need to know something about their language. This is the case when the national curriculum, for instance in South Africa, requires that the school children in specific grades should learn about the origin and history of their language and its words. However, it is important to note that the corresponding etymological function displayed by various South African school dictionaries has nothing to do with the learning of the language, but only with acquiring a learned knowledge about the language.

Finally, some school dictionaries, especially the so-called children's dictionaries, may have an additional underlying function of a operational character, i.e. to assist the school children in getting into the habit of using dictionaries and developing dictionary skills (for more about operational situations, cf. Tarp 2008a). In fact, Martínez de Sousa (1995: 158), in his Spanish dictionary of lexicography, exaggerates this function and defines a "children's dictionary" solely as a "dictionary especially conceived to initiate the children in the use of this type of work".

To sum up, school dictionaries for first-language learners may have the following seven fundamental communicative, cognitive, and operational functions, of which only five are directly related to the learning of the first language:

Communicative functions directly related to language learning

1. To assist school children with text reception (written or oral)
2. To assist school children with text production (written or oral)

Cognitive functions directly related to language learning

3. To assist school children with the learning of the grammar
4. To assist school children with vocabulary learning
5. To assist school children with learning about the world

Cognitive function not directly related to language learning

6. To assist school children with learning about etymology

Operational function not directly related to language learning

7. To assist school children in developing dictionary skills.

It goes without saying that several of these fundamental functions are restricted only to school children of a certain age (grade) and that the individual school dictionary — even when being a high quality product — does not have to display the functions that are not relevant to the age and grade of the foreseen user group. Apart from that it should be noted that each of these fundamental functions may be further subdivided as a result of the necessary categorisation and subdivision of the highly heterogeneous user group composed of school children of different grades and ages.

As already mentioned, we believe that the most important situations where dictionaries may provide assistance in the learning of the first language are text reception and text production, because the cognitive situations in the first place give rise to increased *knowledge* about this language whereas communication is the mediating element through which the information provided by dictionaries may be transformed into language *skills* which is the main objective of first-language (mother-tongue) learning. Tarp (2008b: 134-135) writes:

> If a person at a certain language level has difficulty in understanding or formulating a mother-tongue text, the solution may be to consult a reception or production dictionary, since the successful communication resulting from this consultation (which is its direct purpose) can reflect on and increase the mother-tongue skills which are always the basic precondition for any successful communication.

For this reason, in the following we will concentrate on the two communicative situations without any disrespect for the other situations where dictionaries may also provide assistance.

3. First-language learners' lexicographical needs

If an abstraction is made from the specific characteristics of the heterogeneous

group of school children in terms of their variation in age and corresponding intellectual, linguistic, cultural, and encyclopedic development, it is possible to make a list of lexicographically relevant needs that this potential user group may experience in relation to text reception and text production.

When school children irrespective of age or grade experience problems in understanding written or oral text, they may need information about the *meaning of individual words*, *idioms* or *proverbs*, and when these reception problems lead to a lexicographical consultation they may furthermore need an *adequate access system* as well as information about *orthography*, *part of speech* and *irregular inflection forms* in order to confirm that they have actually found the right article and the corresponding lexicographical data from which they can retrieve the information needed to solve their original reception problem.

When the same children experience problems in relation to text production they may need information about *orthography*, *pronunciation*, *inflection*, *pragmatic restrictions*, *word formation*, *syntactic properties*, *collocations*, *synonyms*, *antonyms* etc. In order to access the dictionary and confirm that they have found the right article they may need an *adequate access system* as well as information about *meaning* and *part of speech*.

However, when one focuses on the specific characteristics of school children in terms of their intellectual, linguistic, cultural, and encyclopedic development, then a relatively big differentiation has to be made not only regarding their lexicographically relevant needs but also with respect to the type of access system, the amount and types of lemmata, the amount and types of lexicographical data included, and the way these data are presented. The problem is not only the evident fact that the needs vary and change as a function of the school children's increase in age and mental development but also that too few and too simple lexicographical data may not satisfy the needs of upper grade learners, whereas too many and too complicated data may hamper or even obstruct the consultation process for the lower-grade learners and prevent them from accessing the relevant data and retrieving the needed information. Two examples from existing school dictionaries will illustrate this problem:

length lengths
NOUN **1** The **length** of something is the distance that it measures from one end to the other.
NOUN **2** The **length** of something like a holiday is the period of time that it lasts.

Article from the *Collins Junior Illustrated Dictionary*

length /…/ noun **1** [C/U] **MATHS, SCIENCE** a measurement of the distance from one end of something to the other. In a **two-dimensional** object, length is the greatest **dimension**: *The boat was 16 feet* ***in length****.* * *He ran half* ***the length of*** *the pitch with the ball.* * *Measure* ***the length of*** *the line.* **2** [C/U] a measurement of how long something takes to do or of how long it lasts: ***The length*** *of you talk must be at least 10 minutes.* **3** [C/U] a measurement of how long a book or piece

of writing is: *His latest novel is twice* ***the length of*** *his previous one*. **4** [C] a piece of something that is long and thin: ***a length of*** *rope*
PHRASES **at (great/some) length** for a long time and with a lot of detail: *Austin was questioned at length by detectives*.
go to great/extreme/any etc lengths to try in a very determined way to achieve something: *They have gone to great lengths to make us feel welcome*.
the length and breadth of sth every part of a large area

Article from the *Macmillan School Dictionary*

A first-language learner at grade 10 or 12 having text reception or production problems will hardly get any help in the (relatively) over-simplified article designed for school children of much lower grades, whereas the latter will probably get completely lost in the (relatively) complex data included in the article conceived for the former.

4. Giving a paraphrase of meaning

Over many years different research projects focusing on the needs of dictionary users have indicated that the explanation of meaning is usually regarded as the type of data for which monolingual dictionaries are the most frequently consulted. Lexicographers have different opinions regarding the term with which to classify the item giving the explanation of meaning. In the majority of dictionaries the term *definition* is used to refer to this item. Well-founded criticism against this term can be found in Wiegand (1985). Where a definition, as used within the field of logic, presents the meaning of a given word, the explanation of meaning in a dictionary conveys that part of the meaning relevant to the target user. Gouws (2011: 62) shows the varying extent of the explanation of the meaning of the word *bridge* in five different dictionaries. The item giving the meaning should actually be referred to as the *paraphrase of meaning*, cf. Wiegand (1985, 1989a), because it paraphrases the meaning as deemed appropriate for a given user group and user situation. Important is not only the extent of the explanation but also the way in which it is given.

When looking at the data presentation in school dictionaries it is therefore important to pay particular attention to the contents and presentation of the paraphrases of meaning used in a given dictionary.

One of the problems users often have when consulting a dictionary is to understand and correctly interpret the entries in the dictionary article. This is due to high levels of textual condensation and the use of unnatural language in conveying the data. For many members of a speech community a school dictionary is their first introduction to the world of reference tools. It is extremely important that school dictionaries should present data in such a way that the intended target user can achieve an optimal retrieval of information without being impeded by strange codes, abbreviated entries and syntactically reduced paraphrases of meaning or a mere presentation of a synonym as the explana-

tion of meaning. Tarp and Gouws (2010: 479) already referred to the importance of the use of natural language to improve the comprehensibility of the paraphrase of meaning. Albeit that they are directed at learners of different age group the difference in ease of comprehension between the following two paraphrases of meaning should convince lexicographers rather to opt for the use of natural language in the paraphrase of meaning:

> **magnet magnets**
> NOUN A **magnet** is a special piece of metal. It pulls or attracts iron or steel towards it. Magnets can also push other magnets away.
>
> Article from the *Collins Junior Illustrated Dictionary*

> **magnet** (say **mag**-nuht) *noun* (*plural* **magnets**)
> a piece of metal, rock or other substance that can make metal objects move towards it.
>
> Article from the *Oxford South African School Dictionary*

In these paraphrases of meaning the difference between the two articles does not only reside in the one using natural language and the other not. It also comes to the fore in the fact that the *Collins Junior Illustrated Dictionary* employs more than one sentence to explain the meaning, instead of trying to put a full explanation into a single non-sentential phrase.

Lexicographers of school dictionaries should consult teachers to find out whether learners are able to understand the presentation of data in their first-language dictionaries. The dictionary using skills of the learners need to have an influence on the way in which the paraphrase of meaning is presented. Proposals to enhance the quality of the paraphrase of meaning in order to ensure that the target users of school dictionaries will achieve an optimal retrieval of information need to be preceded by user studies that involve school learners. Lexicographers need to utilise the results of these studies and need to join hands with teachers and curriculum designers to plan and compile new school dictionaries.

5. Indicating the learners' age and school grade

In an overview article on children's dictionaries Hausmann (1990: 1365) provides the following detailed description of existing dictionaries of this type:

(1) The layout is especially clear. Space is not saved. The letters are bigger than in general dictionaries. Colours are generally used. The dictionaries often are of a big format.

(2) All the lemmata, or a considerable part thereof, are illustrated.

(3) There are no definitions; or when there are, they are not conventional.

(4) Narrative texts (lexicographic story-telling) substitute the traditional microstructure.

(5) There is no information about the lemma, or when it is given, it is only very little.

(6) Abbreviations are not used.

(7) Exercises are given.

(8) The macrostructure is very selective, never with more than 5 000 lemmata. Generally it is between 200 and 2 000 lemmata.

(9) In most cases, the lemmata refer to concrete things.

(10) The users are children below 10 years.

Hausmann — in the same vein as Bergenholtz et al. (1997) and Hartmann and James (1998) — also notes that there is no clear dividing line between children's dictionaries and school dictionaries. This difficulty in establishing a coherent typology seems to be based on the fact that all these authors mainly look at the features — and not at the purpose or functions — of the two "types" of dictionaries. In an attempt to correct this focus, Tarp (2011: 227) writes:

> There seems to be a problem with the *logical linguistic relation* between the terms used and their conceptual content. In most countries, children start in school between the age of 5 and 7, are considered children at least up to the age of 12 or 14, and continue in school up to the age of 15 or 16. This means that they, for a long period, are school children. Consequently, if a school dictionary is defined as a dictionary conceived to be used by pupils in school, most school dictionaries are at the same time "children's dictionaries".

If this logic is followed, the features described above by Hausmann (1990) are actually the features that characterise — or should characterise — school dictionaries adapted to the linguistic and mental development of first-language learners in the first grades. However, although it is evident *that* the specific lexicographical needs of school children vary according to their age and grade, it is far from evident *how* it varies and *which* should be the lexicographical consequences. A comparison between two South African school dictionaries from the same publishing house but designed for learners of grades 3-7 and grades 8-12, respectively, illustrates the problem:

> **record**[1] (*say* **rek**-ord) *noun* (*plural* records)
> **1** a written list of things that you have done, seen, or found out. *He keeps a* ***record*** *of the money he spends*.
> **2** the best that has been done so far. *Tamara's time for the race was a* ***record***. *Zweli broke the* ***record*** *for high-jump.*
>
> Article from the *Oxford South African Illustrated School Dictionary* (grades 3-7)

record[1] **/rek**-ord/
noun* **1 information kept in a permanent form, especially in writing. **2** (*Computing*) a number of related pieces of information dealt with as a unit.
3 a disc on which sound has been recorded; a piece of music recorded on a disk. **4** facts known about a person's past life, performance, or career **He has a superb record at Wimbledon *a criminal record*. **5** the best performance or most remarkable event etc. of its kind that is known **she holds the world record for the100 metres*.

Article from the *South African Oxford Secondary School Dictionary* (grades 8-12)

These two articles — and the two dictionaries as such — share the following data categories:

— Lemma
— Pronunciation
— Part of speech
— Inflection
— Definition
— Examples
— Usage notes
— Etymology
— Word formation

As to the differences, the most important ones are that the *Oxford South African Illustrated School Dictionary* includes:

— Bigger letters
— More space
— Search fields in terms of senses
— Illustrations (few)
— More data on inflection
— The alphabet repeated on each page
— Exercises in dictionary use

whereas the *South African Oxford Secondary School Dictionary* is characterised by:

— Many more lemmata
— Many more senses

— Exercises in writing

— Tables with word groups

Generally, the shared data categories are presented in a more or less similar way in the two dictionaries. The explanation of meaning, for example, are in both cases provided in the form of small "unnatural-language" paraphrases of meaning that vary very little in terms of abstraction level. Four questions arise immediately from this analysis:

1. Can first-language learners at grade 3 really make use of the *Oxford South African Illustrated School Dictionary*?
2. Do the data included in the *South African Oxford Secondary School Dictionary* really satisfy the lexicographical needs of first-language learners in grade 12?
3. Are the differences between the two dictionaries justified by school children's transition from grade 7 to grade 8?
4. Is it possible to design school dictionaries for first-language learners that cover five or more grades each?

The complexity of these questions is underlined by the fact that the same publishing house has also produced a third school dictionary — the *Oxford South African School Dictionary* — which according to the front page is designed for first-language learners from grades 4 to 9, i.e. covering a total of 6 grades. This dictionary seems to place itself somewhere in between the two former ones, following the same principles and with a little more lemmata and senses than the one designed for grades 3 to 7 and a little less than the one designed for grades 8 to 12. The following article will illustrate the similarities to the corresponding one in the two other dictionaries:

> **record**[1] (*say* **rek**-awd) *noun* (*plural* records)
> **1** notes about things that have happened: *Keep a **record** of all the books you read.*
> **2** a thin round piece of plastic that makes music when you play it on a special machine: *a **record** company * Put another **record** on.*
> **3** the best, fastest, highest or lowest that has been done in a sport: *She holds the school **record** for long jump. * He did it in **record** time (= very fast). * She's hoping to break the **record** for the 100 metres* (= to do it faster than anyone has done before).
>
> Article from the *Oxford South African School Dictionary* (grade 4-9)

As a rule, the four questions above cannot be answered by lexicographers alone but require expert knowledge also from language teachers and designers of national curriculums. However, when it comes to printed dictionaries Tarp and Gouws (2010) have strongly suggested that there should be elaborated at least four different dictionaries for different grade clusters for first-language learners

in a 12-grades school system. In this respect, school dictionaries for the upper grades should contain the data really needed to assist not only text reception but also text production, meaning that they should also include data categories such as *syntactic properties* and *collocations*, i.e. data absolutely necessary to back up the fluent text production expected from upper first-language learners but seldom found explicitly and to the necessary extent in English school dictionaries for first-language learners, even when also designed for the upper grades.

6. Some challenges

Many publishing houses are perfectly aware of the fact that high-quality school dictionaries can only be produced when integrating expert knowledge from various disciplines. In his *Introduction* to the *Macmillan School Dictionary*, the editor-in-chief Michael Rundell for instance writes that in planning the dictionary, one of its two "very valuable sources" used was:

> expert advice: at every stage, experienced teachers, textbook writers, and syllabus designers have contributed their expertise, giving us a clear idea of what the dictionary's users really need to know.

The *Macmillan School Dictionary* is according to its own presentation a school dictionary conceived "for students learning through the medium of English" without specifying the age or grade of the intended user group. There is little doubt that "experienced teachers, textbook writers, and syllabus designers" are necessary not only to give an idea of what the "users really need to know" but also of what they are still not prepared for to assimilate. Although the quoted expert knowledge may have contributed to the high-quality lexicographical product that the *Macmillan School Dictionary* surely constitutes in terms of *upper-grade* "students learning through the medium of English", many *lower-grade* and even *intermediate-grade* school children learning through the same medium may most probably find many of its articles — e.g. the article *length* shown above — too complex and too over-loaded when consulting the dictionary for assistance in text reception or production. That was at least the comments from some of the experts participating in the workshop in Pretoria (2011) where this specific article was discussed. In this respect, it is important that non-lexicographical expert knowledge is used to its full extent, i.e. also to define the limits of the group of school children that can really make use of a particular dictionary.

The second "very valuable source" that Rundell refers to in his *Introduction* is a 20-million word corpus "containing hundreds of school textbooks and exam syllabuses, for every subject from agriculture to zoology". Such a well-composed corpus is surely a very valuable source, especially for the selection of lemmata and additional data to be included in the articles, and many school dictionaries published today are in fact corpus-based, e.g. the three South African school dictionaries from Oxford quoted above. Rundell rightly states:

> Using state-of-the-art software to analyse this corpus, we have built up a detailed picture of the terms and concepts that are vital for the study of the main school subjects. We know, for example, which words are used most frequently in textbooks about plant science, religious studies, the environment, or information technology. This gives us a reliable scientific basis for selecting the words to include in the dictionary and for deciding how much information is needed about each word.

What could be added here is that a well-composed corpus of text books and exam syllabuses may also give a reliable scientific basis for determining the amount and complexity of the lexicographical data which lower and intermediate learners are actually able to handle and assimilate without getting lost in data primarily destined to satisfy the needs of upper-grade students. This challenge is related to another comment which Rundell makes in his *Introduction*:

> But the corpus helps us in other ways too. It shows us how concepts are explained in the textbooks that students actually use in the classroom, and this gives us a model for our own definitions — ensuring that they are always relevant and easy to follow.

We have received information from quite a number of lexicographers and teachers in South Africa and Namibia according to which school children of a specific grade frequently prefer to consult dictionaries designed for learners of a lower grade because they have certain difficulties in extracting the needed information from the dictionaries that are supposed to assist learners at their level. This does not only apply to the South African and Namibian situation, cf. De Schryver and Prinsloo and (2011). There may be several linguistic, cultural, regional and social reasons for this, basically related to the learners' first-language proficiency level and dictionary culture. We are not aware of any statistically reliable user research in this respect, and it cannot be excluded that the problem also extends to the explanations provided in text books which for this reason should maybe not be considered models for lexicographical definitions, especially not when the school dictionary in question is designed for users of various grades. The *Macmillan School Dictionary* informs that its definitions are "written in simple English" and it must be admitted that this seems to be the case. However, the problem — at least in South Africa and Namibia — seems to be that there is a certain contradiction between the requirements formulated in the official curriculum and the way the corresponding dictionaries live up to these requirements in terms of user-friendliness. Although learners should be able to retrieve the information defined by the curriculum and corresponding to their specific grade from a dictionary covering this grade, it does not exclude that the dictionaries should be more user-friendly, e.g. with the paraphrases of meaning written in an even simpler language, the appropriate lexicographical data simplified and presented in a more didactic way, and the access system improved.

The proper understanding of the fact that school children pass through a vigorous linguistic, intellectual, cultural, and encyclopedic development during their years in school is paramount to the production of high-quality lexicographical tools adapted to this very heterogeneous group of dictionary users. The aim of this article is not to give final solutions but merely to recommend that a number of factors — or variables — are taken into account when planning future school dictionaries. In this respect, we think that theoretical lexicographers, publishing houses and curriculum designers should consider and find answers to the following questions:

(1) In which of the following learning situations are — or could — dictionaries be used by first-language learners of the different ages and grades:

   a. Text reception — normal or in combination with special exercises?

   b. Text production — normal or in combination with special exercises?

   c. Vocabulary learning and training?

   d. Grammar learning?

   e. Encyclopedic and scientific learning?

   f. Learning etymology?

   g. Learning dictionary skills?

(2) Which amount and types of lemmata and other lexicographical data categories do these situations require in order to satisfy the learners' information needs in the different grades sufficiently? And in which grades should the respective data categories be introduced?

(3) Is it possible to meet the needs occurring in all these learning situations in the framework of one school dictionary or should the future vision be to design dictionaries of which each only assists the users in one — or a few — of these situations?

(4) Which data categories introduced to benefit learners at an upper level do actually disturb data access and information retrieval for learners at a lower level because the latter may find it difficult navigating in articles with too many and too complex data categories included?

(5) How should explanations (paraphrases of meaning) be written and provided in the various grades in order to avoid being too difficult for learners of a specific grade, compelling them to resort to dictionaries designed for lower grades?

(6) Is it possible to design printed school dictionaries for first-language learners that cover 4, 5 or even more grades each? Or should there be a further sub-classification, e.g. as proposed by Gouws and Tarp (2010) for

school dictionaries for first-language learners of Afrikaans: grades 1-3, grades 4-7, grades 8-10, and grades 11-12?

(7) Are there local or regional differences in school children's first-language proficiency that make it almost impossible to design school dictionaries for a specific grade? And should publishing houses instead give very detailed information about the content of their respective school dictionaries in order to empower the individual schools and first-language teachers to evaluate and decide which dictionaries are most apt for their specific group of school children?

(8) Which consequences may it have for future school dictionaries that South African school classes most often incorporate children from various linguistic backgrounds?

(9) Is it an option in the multilingual South Africa — now or in the near future — to design electronic school dictionaries that could be adapted to each type of school child in terms of first-language proficiency and dictionary skills, and to each type of learning situation where dictionaries are or could be used in school?

7. Conclusions

From the discussions that we have had with publishers and other experts in the workshops organised in Stellenbosch (2010) and Pretoria (2011) it has become clear that a major obstacle in the production of a range of high-quality school dictionaries adapted to the needs of first-language learners in different grades is the relatively limited sales and the publishing houses' corresponding lack of interest in making investments in the necessary (but also risky) product development. In this respect we consider that the South African school system could learn from two other "third-world" countries, i.e. Brazil and Cuba. As already mentioned above, during the last decade the Brazilian government has provided each school child with a Brazilian-Portuguese first-language dictionary, a fact that has not only raised commercial sales but also given a strong impetus to the whole dictionary culture and the corresponding theoretical reflections, cf. Welker (2008). In Cuba, a country with relatively few resources and a population a quarter of that of South Africa, the 1 200 pages, 2-volume *Diccionario Básico Escolar*, now in its third edition, has until now been printed and distributed (some of it freely) in more than 100 thousand copies and at a ridiculously low price of less than one dollar per volume due to both subventions and the use of very cheap paper that may not be that fancy and last so long but nevertheless serves its purpose, cf. Tarp (2012).

We do not necessarily recommend that these two examples are copied in South Africa or any other country, but they could and should be used as inspiration in order to find solutions to a very serious problem, i.e. a low societal

dictionary culture and the alarming lack of the required high-quality lexicographical tools that can meet the growing information needs in society and thus contribute to its development. We are convinced that the problem starts in early school and therefore should be addressed here through the promotion, in one way or another, of school dictionaries adapted to the real needs of first-language learners of the various grades.

8. References

8.1 Dictionaries

Collins Junior Illustrated Dictionary. Goldsmith, Evelyn (Ed.). 2005. *Collins Junior Illustrated Dictionary* London: Collins Educational.

Diccionario Básico Escolar. Miyares Bermúdez, Eloína (Ed.). 2009. *Diccionario Básico Escolar*. Santiago de Cuba: Editorial Oriente.

Macmillan School Dictionary. Rundell, Michael (Ed.). 2004. *Macmillan School Dictionary*. Oxford: Macmillan.

Oxford Children's Thesaurus. Allen, Robert (Ed.). 2003. *Oxford Children's Thesaurus.* Oxford: Oxford University Press.

Oxford South African Illustrated School Dictionary. Reynolds, Mary (Ed.). 2011. *Oxford South African Illustrated School Dictionary*. Cape Town: Oxford University Press Southern Africa.

Oxford South African School Dictionary. Reynolds, Mary (Ed.). 2010. *Oxford South African School Dictionary*. Cape Town: Oxford University Press Southern Africa.

South African Oxford Secondary School Dictionary. Reynolds, Mary (Ed.). 2006. *South African Oxford Secondary School Dictionary*. Cape Town: Oxford University Press Southern Africa.

8.2 Other literature

Bergenholtz, Henning, Ilse Cantell, Ruth Vatvedt Fjeld, Dag Gundersen, Jón Hilmar Jónsson and Bo Svensén. 1997. *Nordisk leksikografisk ordbok.* Oslo: Universitetsforlaget.

De Schryver, Gilles-Maurice and Daniël J. Prinsloo. 2011. Do Dictionaries Define on the Level of their Target Users? A Case Study for Three Dutch Dictionaries. *International Journal of Lexicography* 24(1): 5-28.

Geeraerts, Dirk. 1984. Ontwikkeling in de lexicografische theorievorming. *Jaarboek van de Stichting Instituut voor Nederlandse Lexicologie* 1981-1982: 56-65.

Gouws, Rufus H. 1989. *Leksikografie*. Cape Town: Academica.

Gouws, Rufus H. 2011. Expanding the Scope of the Application of Enrichment in Dictionary Articles. Jesenšek, Marko (Ed.). 2011. *Izzivi sodobnega slovenskega slovaropisja*: 59-74. Maribor: Zora.

Gouws, Rufus H. 2012. (To appear.) Establishing and Developing a Dictionary Culture for Specialised Lexicography. Jesenšek, Vida (Ed.). 2012. *5. Lexikographische Kolloquium.* Berlin: De Gruyter.

Hartmann, R.R.K. and Gregory James. 1998. *Dictionary of Lexicography.* London/New York: Routledge.

Hausmann, Franz Josef. 1989. Die gesellschaftlichen Aufgaben der Lexikographie in Geschichte und Gegenwart. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand and

Ladislav Zgusta (Eds.). 1989. *Wörterbücher. Dictionaries. Dictionnaires. An International Encyclopedia of Lexicography* I: 1-19. Berlin/New York: Walter de Gruyter.

Hausmann, Franz Josef. 1990. Das Kinderwörterbuch. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta (Eds.). *Wörterbücher, Dictionaries, Dictionnaires. An International Encyclopedia of Lexicography* II: 1365-1368. Berlin/New York: Walter de Gruyter.

Malkiel, Y. 1967. A Typological Classification of Dictionaries on the Basis of Distinctive Features. Householder, Fred W. and Sol Saporta (Eds.). 1967. *Problems in Lexicography*: 3-24. Bloomington: Indiana University.

Martínez de Sousa, José. 1995. *Diccionario de lexicografía práctica.* Barcelona: Biblograf.

McArthur, Tom. 1986. *Worlds of Reference. Lexicography, Learning and Language from the Clay Tablets to the Computer*. Cambridge: Cambridge University Press.

Tarp, Sven. 2008a. The Third Leg of Two-legged Lexicography. *Hermes. Journal of Linguistics and Communication Studies* 40: 117-131.

Tarp, Sven. 2008b. *Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Max Niemeyer.

Tarp, Sven. 2011. Pedagogical Lexicography: Towards a New and Strict Typology Corresponding to the Present State-of-the-Art. *Lexikos* 21: 217-231.

Tarp, Sven. 2012: New Experiences in Pedagogical Lexicography: Two Cuban School Dictionaries. *Hermes, Journal of Linguistics* 48: 95-107.

Tarp, Sven and Rufus Gouws. 2010. Skoolwoordeboeke vir huistaalleerders van Afrikaans. *Lexikos* 20: 466-494.

Welker, Herbert Andreas. 2008. *Panorama geral da lexicografia pedagógica*. Brasilia: Thesaurus Editora.

Wiegand, Herbert Ernst. 1985. Eine neue Auffassung der sog. lexikographischen Definition. Hyldgaard-Jensen, Karl and Arne Zettersten (Eds.). 1985. *Symposium on Lexicography II. Proceedings of the Second International Symposium on Lexicography May 16–17 1984 at the University of Copenhagen*: 15-100. Tübingen: Max Niemeyer.

Wiegand, Herbert Ernst. 1989. Der gegenwärtige Status der Lexikographie und ihr Verhältnis zu anderen Disziplinen. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta (Eds.). 1989. *Wörterbücher. Dictionaries. Dictionnaires. An International Encyclopedia of Lexicography* I: 246-280. Berlin/New York: Walter de Gruyter.

Wiegand, Herbert Ernst. 1989a. Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta (Eds.). 1989. *Wörterbücher. Dictionaries. Dictionnaires. An International Encyclopedia of Lexicography* I: 409-462. Berlin/New York: Walter de Gruyter.

Zgusta, Ladislav. 1971. *Manual of Lexicography*. The Hague: Mouton.

A Study of the Use of the *HAT Afrikaanse Skoolwoordeboek* by Primary School Children

Michele F. van der Merwe, *Department Curriculum studies: Afrikaans, Faculty of Education, University of Stellenbosch, Stellenbosch (michelevdm@sun.ac.za)*

Abstract: A research project on the use of the *HAT Afrikaanse Skoolwoordeboek* by primary school learners is described. (In South Africa the term **learner** is used to describe a pupil at school). Studies on the way in which learners use dictionaries provide valuable information for both lexicographers and language teachers. The assessment of users' skills regarding the use of the *HAT Afrikaanse Skoolwoordeboek* was made in the learners' home language. In order to design a test to assess dictionary skills regarding the language learning outcomes of Grade 4 learners, the language learning assessment outcomes in the Grade 4 Afrikaans home language curriculum had to be unwrapped. The process was undertaken in cooperation with Grade 4 teachers to ensure that a realistic assessment of dictionary skills was made and an assessment of dictionary skills of 200 learners ensued. The assessments were marked, marks were calculated and results were analysed. Teachers were trained in teaching dictionary learning strategies and language skills. Learners then received training from their teachers to extract semantic information from a dictionary, as this proved to be the most difficult skill to master, according to the results of the assessment.

Keywords: PRIMARY SCHOOL LEARNERS, LANGUAGE TEACHERS, ASSESSMENT, DICTIONARY SKILLS, LANGUAGE LEARNING OUTCOMES, AFRIKAANS HOME LANGUAGE CURRICULUM, DICTIONARY FUNCTIONS, SCHOOL DICTIONARY, SPELLING, READING, DESIGN-BASED RESEARCH, LEARNING ECOLOGY, SEMANTIC INFORMATION

Opsomming: 'n Studie van die gebruik van die *HAT Afrikaanse Skoolwoordeboek* deur laerskoolleerders. 'n Navorsingsprojek oor die gebruik van die *HAT Afrikaanse Skoolwoordeboek* deur laerskoolleerders word beskryf. (Die term **leerder** word in Suid-Afrika gebruik om na skoolleerlinge te verwys). Navorsing oor die manier waarop leerders woordeboeke gebruik, verskaf waardevolle inligting vir leksikograwe sowel as vir taalonderwysers. 'n Assessering van gebruikersvaardighede t.o.v. die gebruik van die *HAT Afrikaanse Skoolwoordeboek* is in die leerders se moedertaal gedoen. Die leeruitkomste van die graad 4-kurrikulum vir Afrikaans as huistaal is ontleed om 'n toets op te stel om woordeboekvaardighede rakende taalonderriguitkomste van graad 4-leerders te assesseer. Die proses het in samewerking met graad 4-onderwysers geskied, om te verseker dat 'n realistiese assessering van woordeboekvaardighede plaasvind. Assessering van die woordeboekvaardighede van 200 leerders is uitgevoer. Die assesserings is gemerk, punte is bereken en resultate is ontleed. Onderwysers het opleiding ontvang in die onderrig van woordeboekvaardighede, en ook in die onderrig van taalvaardighede. Leerders het spesi-

fieke onderrig ontvang om semantiese inligting in 'n woordeboek te ontsluit, aangesien dit, volgens die uitslae van die assessering, die moeilikste vaardigheid was om te bemeester.

Sleutelwoorde: LAERSKOOLLEERDERS, TAALONDERWYSERS, ASSESSERING, WOORDEBOEKVAARDIGHEDE, TAALONDERRIGUITKOMSTE, AFRIKAANS HUISTAALKURRIKULUM, WOORDEBOEKFUNKSIES, SKOOLWOORDEBOEK, SPELLING, LEES, ONTWERPGEBASEERDE NAVORSING, LEEREKOLOGIE, SEMANTIESE INLIGTING

Introduction

The primary motivation for this research arose from the fact that studies on the way learners use dictionaries provide valuable information for both lexicographers and language teachers. Lexicographers constantly strive to improve dictionaries and language teachers are able to improve the dictionary skills of the users. The notion of the user perspective has become prominent in modern lexicography and several studies have been directed towards the needs of the users of dictionaries. These needs will be understood better if researchers and lexicographers are acquainted with the skills and levels of skill of such users. Steyn (2006: 138) recognises this when she remarks, "an extensive gathering and analysis of information about the intended users' needs and skills on language and reference level should precede any new dictionary". Beyer (2010: 54) states that, although research on pedagogical dictionaries, especially school dictionaries, has not received much attention from researchers in lexicography, there are indications that school dictionaries are better planned than in the past.

Valid assessment tools

The primary research question of the study was formulated as follows: What is the level of dictionary skills of learners using a school dictionary in their home language? In order to assess such dictionary skills, a valid assessment tool needed to be developed.

One of the biggest challenges for research in lexicography, however, is to establish a sound approach to determining the dictionary skills of users by making use of an assessment tool. Tono (2001: 84) remarks with scepticism that "not a single reliable measurement of dictionary using skills has become available so far". Results that record subjects' behaviour are often difficult to interpret and seem artificial in a non-natural environment. Tono suggests having results verified "in a more natural classroom setting" or in real-life situations, both formal and informal. Welker (2010: 265) describes tests conducted by Benson (1989), Bogaards (1994) and Sobrinho (1998) and Nesi and Haill (2002) point out that most studies on dictionary use have relied on the retrospection of users via questionnaires or interviews (e.g. Tomaszczyk 1979; Béjoint 1981; Bogaards 1988; Atkins and Varantola 1988). Researchers such as Summers

(1988), Tono (2001), Knight (1994), McCreary and Dolezal (1999) (Nesi and Haill 2002: 277) have created artificial occasions for more readily observable dictionary consultation under controllable conditions. Questionnaire-based research has been criticised for relying strongly on the respondents' desire to please and conform.

According to Welker (2010: 18), tests in a scientific or school context are conducted to measure or evaluate. One may employ the term **test** whenever subjects in an investigation have to execute some task, the results of which are to be evaluated. Tarp and Gouws (2010: 288) describe a test in the lexicographical sense of the word as a method to evaluate to what degree the consultation of a dictionary or other lexicographical tool can help users to satisfy their needs. The problem with such a description of a test is that it does not make provision for evaluation or assessment of the skills of the dictionary user; it seems that the success of dictionary consultation is shifted to the dictionary as a user-friendly tool to accommodate the users' needs. A test, or an assessment tool then, needs to be objective, reliable and valid to evaluate whether a user (in this case a learner) has met the standard that has been set. According to Wiegand (in Smit 2001: 301), three types of tests are possible; namely, (i) the user test, (ii) a test to determine metalexicographical knowledge, and (iii) a dictionary test. A user test can test user skills, language competence, user experience, knowledge of dictionaries, knowledge of patterns of user actions, and the skills to perform user actions of a specific type, thereby demonstrating the mastering of a user in practice. In this study, a monothematic dictionary test was used to access the dictionary skills of Grade 4 learners. Learners were assessed while consulting the *HAT Skoolwoordeboek*, making it a monothematic dictionary test, because it concerned the use of one dictionary only. Such a test could inform the didactical component of dictionary use, as was the case in the assessment conducted. Wiegand (in Smit 2001: 301) divides such tests into classes of (i) tests for users-in-action (i.e. while they are consulting a dictionary); (ii) tests for users-not-in-action (i.e. usually comprising tests concerning users' knowledge of dictionaries); and (iii) a mixture of (i) and (ii). In the current study, the user-in-action test was chosen, as learners were assessed whilst using the *HAT Skoolwoordeboek*.

The assessment outcomes mentioned in the national home language curriculum were used to compile an assessment tool to ensure the validity of the research method. In developing this assessment tool, the researcher worked together with qualified teachers to avoid compromising the validity of the tool. The project described in the study focused on unwrapping the learning assessment outcomes in the Grade 4 Afrikaans Home Language Curriculum, designing a test to assess dictionary skills regarding language learning outcomes of Grade 4 learners, and analysing the results of the assessment of dictionary skills. The study reported on in this paper attempted to describe and assess the dictionary skills of Grade 4 learners using the *HAT Afrikaanse Skoolwoordeboek* (referred to as *HAT Skoolwoordeboek* for the purpose of this article) in their home

language. According to the introduction, the *HAT Skoolwoordeboek* was compiled for Afrikaans Home Language and Afrikaans First Additional Language learners. The compiler claims that the dictionary meets all the requirements of the National Curriculum and that the choice of lemmas, the formulation of the definitions, as well as the selection of examples, were determined by the school environment.

Unwrapping the language curriculum

The South African language curriculum is an outcomes-based curriculum focused on different outcomes to be achieved by learners. These outcomes are assessment outcomes and provide a valid assessment tool that can be described as objective, reliable and valid. Wiegand (Welker 2010: 18) views objectivity as a very important component of the evaluation of tests. He distinguishes between the objectivity of the execution of the test, the objectivity of the measurement of results and the objectivity of interpretation. The last two criteria may be joined under the heading 'evaluation'; the second refers to the way in which the execution of the task is measured, while the third one refers to the fact that the evaluation should be made by more than one person. By making use of the outcomes described in the language curriculum in the execution of the assessment, the researcher ensured that the evaluation would be objective. That a research team was involved and the evaluation was not done by one person only also contributed to the objectivity of the assessment.

In order to make effective use of dictionaries as language learning tools, one needs to be aware of the learning assessment outcomes in the Home Language Curriculum. Learning assessment outcomes could be identified by 'unwrapping' the outcomes of the curriculum. Unwrapping refers to the skill associated with the outcome, the content of the outcome and the content for teaching. According to the National Curriculum Statement (2004) under Learning Outcome 3: reading and viewing, as well as Learning Outcome 6: language structure and use, learners should be able to show knowledge of the complexity of alphabetical order; record words in a personal dictionary; check spelling in a dictionary; and select relevant texts for own information needs (e.g. dictionaries, children's encyclopaedias and reference books).

Language outcomes were identified when teachers in a workshop decided which skills were important in language training and which skills needed to be assessed in an assessment of dictionary use. Dictionaries lend themselves excellently to be used as a productive language learning tool for language teaching in the classroom. It made sense to let teachers decide on the language skills, rather than relying on a researcher with a theoretical background. Humblé (2001: 21) notes that the point of view of experienced teachers is more unbiased and reliable than those of some lexicographers and academic researchers whose knowledge of dictionary use and language teaching may be limited.

After the curriculum for Afrikaans Home language was unwrapped, the

following language outcomes that could be used in teaching dictionary skills were identified from Learning Outcome 6: language structure and use:

— Spell frequently-used words correctly
— Use orthography correctly in commonly-used words
— Use capital letters correctly
— Develop vocabulary by recognising word families and words from similar fields
— Explore the origin of words
— Use abbreviations
— Understand the use of adjectives: inflection, degrees of comparison and intensive forms
— Understand the use of nouns: number, gender and diminutives
— Use everyday synonyms, antonyms, homonyms and homophones
— Use the appropriate formality register
— Use simple idiomatic expressions correctly.

To answer the research questions posed at the beginning of the paper, an assessment tool was developed to assess the dictionary skills of learners. After the analysis of the curriculum had been done, Grade 4 teachers drew up an assessment tool with the purpose of making use of the *HAT Skoolwoordeboek* to achieve the identified language outcomes. The assessment was focused on the identified language outcomes for home language that could be applied to dictionary usage. The assessment tool comprised a test to be taken by the learners without any prior instruction in dictionary skills. Individual teachers assessed their own learners, at their own schools and in their own time during lessons. This was done to avoid creating the feeling that an outside evaluator was monitoring and assessing the learners' work and to ensure that the assessment situation remained as natural as possible. A total number of 200 Grade 4 learners was assessed.

Dictionary functions

According to Abecassis (2007: 252), it is necessary to be aware of the different motivations that impel users to consult a dictionary when the best possible strategies for successful looking up of words are evaluated. Linguistic activities such as speaking, reading and translating are behind triggering the use of a dictionary. Gouws (2006: 75) distinguishes two kinds of functions within the theory of lexicographical functions, namely cognitive and communicative functions. The purpose of cognitive functions is to assist the user with general and encyclopaedic data; data on specific subjects; and data on language. Communicative functions support the user in the solution of problems regarding text production in the mother language and in a foreign language, as well as with the translation of texts from the mother language to the foreign language and

vice versa. Tarp (2008) adds another function, namely an operational one which describes an additional, underlying function of dictionaries, namely to help learners to make use of dictionaries and to develop dictionary skills. Tarp and Gouws (2010: 470) distinguish seven dictionary functions that need to be present in school dictionaries for home language learners. These include the following: Communicative functions relevant for linguistics to help learners with text reception (written or oral) and text production (written or oral); cognitive functions relevant for linguistics to help learners with grammar, vocabulary and learning more about the world; cognitive functions not relevant for linguistics to assist learners to learn more about the etymology of a word; and operational functions not relevant for linguistics to help learners develop dictionary skills.

Dictionary functions can play an important role in describing the transfer of information from dictionaries to users. The notion of dictionary functions was used in this study to enhance the assessment process, by linking dictionary tasks performed by learners to appropriate dictionary functions. When learners were assessed on the basis of the assessment criteria that were developed, dictionary functions were attached to certain tasks in order to make connections with dictionaries and pedagogical lexicography.

Assessment of dictionary reference and language skills

The assessment is supplied in Addendum A. An analysis of the assessment of the questions posed, as well as the dictionary reference and language skills featured in the assessment, is supplied. (A repetition of skills occurs, as several questions were used to test the same skills). Where applicable, the dictionary functions associated with a specific task are also indicated.

1. Arrange words in alphabetical order — alphabetisation as a dictionary reference skill. (operational function)

2. Arrange words in alphabetical order within the same letter, up to three letters after the initial letter.

3. (a) Look up the guide word at the top of the page — understand the access structure of the dictionary. (operational function)
   (b) Look up the guide word at the top of the page — understand the access structure of the dictionary. (operational function)

4. (a) Look up a word in its alphabetical order. (operational function)
   (b) Look up a word in its alphabetical order. (operational function)

5. Look up the spelling of a word — alphabetisation as a dictionary reference skill and checking spelling to find the correct orthographic form of the word. (cognitive function)

6. Find a word in its alphabetical order.
   (a) Divide a word into syllables — understand the search structure of the dictionary. (operational function and cognitive function)
   (b) Divide a word into syllables — understand the search structure of the dictionary. (operational function and cognitive function)
7. Find a word in its alphabetical order.
   (a) Find etymological information on the word — understand the microstructure of the dictionary. (cognitive function)
   (b) Find the part of speech of the word — understand the microstructure of the dictionary. (cognitive function)
   (c) Find the plural of the word — understand the microstructure of the dictionary. (cognitive function)
8. Find a word in its alphabetical order. (operational function)

   Find the inserted text associated with the word — understand the search structure of the dictionary. (operational function)

   Find an applicable synonym in the inserted text — understand the search structure of the dictionary. (cognitive function)
9. Find a word in its alphabetical order. (operational function)

   Find a synonym for the word — understand the microstructure of the dictionary. (cognitive function)
10. Find a picture in the dictionary (a range of page numbers were supplied) — search reading/read to find information.
    (a) Find the word associated with the picture — search reading.
    (b) Find the definition of the word, read it, extract meaning from the definition and apply the found information — understand the internal search structure of the dictionary, use critical thinking skills about the semantic information supplied. (cognitive and communicative function)
11. Find a word in its alphabetical order.
    (a) Find where the stress is in the word when pronounced — understand the internal search structure of the dictionary. (operational and cognitive function)
    (b) Find the diminutive of the word — understand the internal search structure of the dictionary. (cognitive function)

The test focused on assessing reference skills, i.e. knowledge of alphabetical ordering and the ability to understand the access structure of the dictionary, as well as the ability to use a dictionary for definitions, spelling, pronunciation and obtaining grammatical and etymological information. Inference was as-

sessed as well, i.e. the ability to establish and interpret parts of speech, finding meanings, as well as the ability to correctly interpret meanings. Thus, the learners' abilities to use mostly operational and cognitive functions of the dictionary were assessed.

Table of results of assessment of Grade 4 learners using the *HAT Skoolwoordeboek*

In Table 1 results of the assessment are shown. The skills assessed, as well as the percentage of correct answers are indicated.

Question number	Skill assessed	Number of correct answers (from 200 learners)	Percentage of learners supplying correct answers
1	Alphabetisation	186	93%
2	Alphabetisation	113	56%
3a	Look up guide word	185	92%
3b	Look up guide word	155	77%
4a	Look up a word	118	59%
4b	Look up a word	116	58%
5	Look up spelling	150	75%
6a	Divide word into syllables	119	60%
6b	Divide word into syllables	88	44%
7a	Find etymological information	156	78%
7b	Find the part of speech	102	51%
7c	Find the plural of the word	124	62%
8	Find the inserted text and applicable synonym	115	58%
9	Find synonym	71	36%
10a	Find word associated with the picture	159	80%
10b	Extract meaning from the definition and apply the found information	71	36%
11a	Find stress in the word	87	44%
11b	Find diminutive	117	59%

Table 1: Results of assessment

In figure 1 the number of learners is shown per mark scored out of a total of 20 marks.

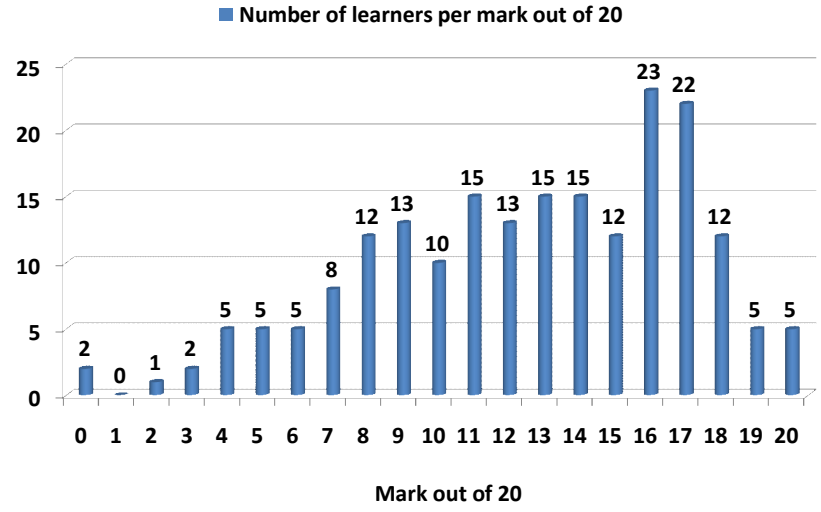


Figure 1: Results of assessment

A mark indicating that less than 50% of the class gave the correct answer was viewed as problematic. According to the results of the assessment, the following questions and related skills posed challenges to Grade 4 learners:

(i) Question 10(b): Find the definition of the word, read it, extract meaning from the definition and apply the found information — understand the internal search structure of the dictionary, think about the semantic information supplied. A total of **64% of the learners** assessed gave a wrong answer. For the question, learners had to look up a picture of **anys** (aniseed) in the dictionary, read the definition and explain what its use is. The article is presented as follows in the *HAT Skoolwoordeboek*:

> **anys** *s.nw.* [geen mv.] 'n plant waarvan die saad as geursel (bv. in beskuit) of as medisyne gebruik word.

Learners had difficulty in answering the question "For what can it be used?", although a clear answer to the question is supplied in the definition. The dictionary's cognitive and communicative functions had to be used to answer this question. The cognitive function refers to the fact that the learner's vocabulary could have been increased by the use of the picture as well as the semantic information supplied. The communicative function refers to text production, by which the learner should have been able to explain the use of the plant. This point indicates that learners were not able to apply an adequate reading strategy to extract semantic information from a definition. This also is an indication of a

low literacy level among children.

(ii) Question 9: Find a synonym for the word — understand the internal search structure of the dictionary. A total of **64% of the learners** could not supply the correct answer to the question posed. Learners had to find a synonym in the article **biefstuk**. The article is presented as follows in the *HAT Skoolwoordeboek:*

bief.stuk *s.nw.* [**-ke**] 'n dik stuk (bees)vleis, gesny om gerooster/gebraai te word➔SINONIEM: **steak**

For this question, learners had to read the whole article to find the synonym at the end of the article. The synonym is clearly marked with a synonym marker; the definition of the word is separated from the synonym with an arrow and the word 'sinoniem' appears in capital letters, but learners could not successfully apply a reading strategy to find the answer. Operational and cognitive functions of the dictionary were assessed in this question. The operational function refers to the dictionary skill of finding a synonym of a word in the dictionary, whereas the cognitive function refers to the extension of the learner's vocabulary through realising the semantic relations between different words.

(iii) Question 11(a): Find where the stress is in the word when pronounced — understand the internal search structure of the dictionary. Learners were directed to look up the word **emoe**, with the page number given, referring to the operational function of the dictionary. They then had to supply an answer concerning which part of the word is stressed, as indicated in the article. This refers to the cognitive function of the dictionary, as learners had to learn more about the grammar of the language. A total of **56% of the learners** assessed gave a wrong answer**.** Learners could not successfully copy the answer out of the dictionary, or did not understand the dictionary convention involved with how word stress is indicated. The article is presented as follows in the *HAT Skoolwoordeboek:*

e·moe *s.nw.* [~**s; ~tjie**] 'n groot Australiese voël wat vinnig kan hardloop, maar nie kan vlieg nie: Die groot loopvoël, die emoe lyk baie soos 'n volstruis.

(iv) Question 6(b): Divide the word **dubbelmediumonderrig** into syllables — understand the search structure of the dictionary. A total of **56% of the learners** could not supply the correct answer to the question posed. The word is divided into syllables in the dictionary article, making use of the cognitive function of the dictionary whereby learners could learn more about the grammar of the word. The operational function of the dictionary refers to the fact that learners had to familiarise themselves with the convention of dividing words into syllables by making use of full stops to distinguish between different syllables. Learners either did not understand the dictionary convention of indicating division of sylla-

bles, or the word was too long for them to understand and write down correctly. Interestingly enough, a total of 60% of the learners could provide the correct answer to Question 6(a), for which they had to divide a shorter word, namely **blokkiesraaisel,** into syllables. The article is presented as follows in the *HAT Skoolwoordeboek:*

dub.bel.me.di.um.on.der.rig *s.nw. [geen mv.]* onderrig in twee tale in dieselfde skool.

(v) Question 7(b): Find the part of speech of the word **canyon —** understand the internal search structure of the dictionary. A total of **49% of the learners** assessed answered the question wrongly**.** Learners did not understand the dictionary convention regarding the supply of a part of speech, which refers to the cognitive function of the dictionary. The article is presented as follows in the *HAT Skoolwoordeboek:*

can.yon *s.nw.* **[~s]** *(Spaans)* 'n rivierbedding met regaf kante wat diep in die landskap gevreet het: *die Visrivier-canyon in Namibië.*

Notable is the fact that learners had difficulty with questions related to semantic information. The two most problematic questions were about understanding the definition of a word and finding a synonym for a word. Both of these questions involved operational as well as cognitive functions of the dictionary.

Questions with which learners coped well are the following**:**

(i) Arrange words in alphabetical order — alphabetisation as a dictionary reference skill referring to the operational function of the dictionary. A total of **93% of** learners assessed answered the first question correctly. This serves as an indication that learners have mastered the alphabet excellently.

(ii) Question 3(a): Look up the guide word at the top of the page — understand the access structure of the dictionary referring to the operational function of the dictionary. A total of **92%** of the learners assessed answered the question correctly; this shows that learners have grasped the concept of a guide word at the top of a page.

(iii) Question 10(a): Find the word associated with the picture — do search reading and understand that a word can be associated with a picture. A total of **80%** of the learners assessed gave the correct answer and this shows that learners showed understanding of pictures and illustrations in a dictionary.

(iv) Question 7(a): Find etymological information on the word **canyon —** understand the internal search structure of the dictionary. The article is presented as follows in the *HAT Skoolwoordeboek:*

can.yon *s.nw.* **[~s]** *(Spaans)* 'n rivierbedding met regaf kante wat diep in die landskap gevreet het: *die Visrivier-canyon in Namibië.*

A total of **78%** of the learners assessed answered the question correctly and this serves as indication that learners had mastered the dictionary search skill of finding etymological information, as well as a reading skill. Etymological information is indicated in italics and in brackets in the *HAT Skoolwoordeboek* and the information type is thus clearly distinguished from semantic information. This refers to the cognitive dictionary function not applicable to the learning of linguistics.

(v) Question 3(b): Look up the guide word at the top of the page — understand the access structure of the dictionary, referring to the operational function of the dictionary. A total of **77%** of the learners assessed had the correct answer, which shows that learners have grasped the concept of a guide word at the top of a page. In the question they had to apply search reading skills, as they had to decide whether a certain word occurred on the dictionary page.

With regard to dictionary functions, it is evident from the results that learners excelled in questions regarding operational functions. Learners experienced the most difficulties with questions where more than one dictionary function (either cognitive and communicative, or cognitive or operational) had to be applied.

Dictionary skills activities

After the assessment was conducted, learners were given training in dictionary skills. Language and dictionary-orientated activities were compiled, aimed especially at extracting semantic information from the *HAT Skoolwoordeboek*, as this proved to be the major problem that learners experienced in the assessment. Activities were developed, for example to find semantic information in definitions and to successfully apply reading strategies. Learners were taught semantic relations between words, such as synonyms and antonyms, as well as to extract information from a dictionary article. After learners were taught semantic skills using a dictionary, teachers reported an increase in the ability of learners to do a successful dictionary consultation. Learners completed activities focused on different language outcomes as prescribed by the curriculum that could be applied to dictionaries, such as spelling, identifying the etymology of words, the use of adjectives and the use of nouns. Teachers reported that learners were also more aware of information provided by dictionaries and dictionary use was more frequent in class.

Conclusion

The assessment tool that was designed is a utility tool to be used for assessment with different dictionaries and in different dictionary user situations. It was

designed in cooperation with experienced educators and can be adapted to different dictionary environments.

It is clear from the analysed assessment results that learners experienced problems with the extraction of semantic information from the dictionary. Most of the learners could not convey the correct meaning of a word from the dictionary definition. They had difficulty in understanding the definition of the word and applying the information that was presented. This process required the use of cognitive as well as critical thinking skills. Learners also had difficulty finding a synonym in the lemma article, as well as in the inserted text. The relation of synonymy forms an important part of the semantic information supplied in dictionaries and most lexicographers would agree that semantic information is the most important category of information supplied in a school dictionary. The fact that learners had difficulty with this has far-reaching implications for the lexicographer, the user and the language teacher. Lexicographers have to ensure that (school) dictionaries are user-friendly, so that learners can find the information they are looking for. They have to ensure that definitions of words are clear, easy to understand and child-friendly. Learners need to have dictionaries in the classroom; they need to be trained in the use of dictionaries, especially where more than one dictionary function is concerned. One of the earliest information and reference tools with which children need to be familiarised is the dictionary. Sufficient dictionary use can improve a learner's reference skills, vocabulary, conceptual frame of reference, and facilitate the correct spelling of words.

Language teachers need to be trained in order to be aware of dictionaries as problem-solving tools in the classroom and they need to take cognisance of the dictionary skills of learners.

Bibliography

Abecassis, M. 2007. Is Lexicography Making Progress? On Dictionary Use and Language Learners' Needs. *Lexikos* 17: 247-258.

Beyer, Herman L. 2010. Kontekstuele bepalers by die beplanning van skoolwoordeboeke. *Lexikos* 20: 52-72.

Gouws, R.H. 2006. Leerdergedrewe leksikografie. *Tydskrif vir Taalonderrig* 40(1): 72-84.

Humblé, P. 2001. *Dictionaries and Language Learners.* Frankfurt am Main: Haag und Herchen.

Luther, Jana. (Ed.). 2009. *HAT Afrikaanse Skoolwoordeboek.* Cape Town: Pearson Education South Africa.

Nesi, H. and R. Haill. 2002. A Study of Dictionary Use by International Students at a British University. *International Journal of Lexicography* 15(4): 277-305.

Smit, M. 2001. Herbert Wiegand with a "Metalexicographical Panga" in the Jungle: An Unlocking of Wörterbuchforschung. *Lexikos* 11: 296-310.

Steyn, M. 2006. Aspects of the Micro-Architecture in Learner's Dictionaries. *Lexicographica* 22: 138-151.

Tarp, S. 2008. The Third Leg of Two-legged Lexicography. *Hermes, Journal of Language and Communication Studies* 40: 117-131.

Tarp, S. and R.H. Gouws. 2010. Skoolwoordeboeke vir huistaalleerders van Afrikaans. *Lexikos* 20: 466-494.

Tono, Y. 2001. *Research on Dictionary Use in the Context of Foreign Language Learning. Focus on Reading Comprehension*. Lexicographica. Series Maior 106. Tübingen: Max Niemeyer.

Welker, H.A. 2010. *Dictionary Use. A General Survey of Empirical Studies.* Brasilia: Author's Edition.

Addendum A

Weet jy hoe om 'n woordeboek te gebruik? Gr. 4

Gebruik die *HAT Skoolwoordeboek* om die volgende vrae te beantwoord:

1. Rangskik al die diere volgens die alfabet:
   bok, esel, wurm, donkie, arend, olifant, renoster, mier (1)

2. Rangskik die name alfabeties:
   Megan, Melanie, Markus, Monja, Miriam, Michael, Marina, Michelle (1)

3. Soek die gidswoorde: "begrip" en "beheptheid" op bl. 50.
   a. Is die woord "beheer" op die bladsy?
   b. Is "bejaard" op die bladsy? (2)

4. a. Watter woord kom na "bestryk" in die woordeboek?
      SKRYF DIE BLADSYNOMMER NEER. (1)
   b. Watter woord kom voor "billik" in die woordeboek?
      SKRYF DIE BLADSYNOMMER NEER. (1)

5. Onderstreep die korrekte spelling nadat jy die woord in die woordeboek opgesoek het:
   deliekaat, dellikaat, delikaat, deeliekaat (1)

6. Soek die woorde "blokkiesraaisel" en "dubbelmediumonderrig" in die woordeboek op.
   a. Verdeel "blokkiesraaisel" in lettergrepe.
      SKRYF DIE BLADSYNOMMER NEER. (1)
   b. Verdeel "dubbelmediumonderrig" in lettergrepe.
      SKRYF DIE BLADSYNOMMER NEER. (1)

7. Soek die woord "canyon" in die woordeboek op.
   a. Uit watter taal kom die woord?
   b. Watter woordsoort is die woord?
   c. Wat is die meervoud van "canyon"? (3)

8. Soek die woord "dadelik" in die woordeboek op. Soeklees in die blokkie onder "dadelik". Gee drie woorde wat dieselfde beteken (sinonieme) as "dadelik". (2)

9. Gee 'n sinoniem vir "biefstuk" (1)

10. Soek tussen bladsy 34-36 vir dié prentjie!
    a. Wat is dit?
    b. Waarvoor kan dit gebruik word? (3)

11. Soek die woord "emoe" op bl. 34.
    a. Waar lê die klem van die woord?
    b. Wat is die verkleinwoord van "emoe"? (2)

[20]

A Critique of the Controlled Defining Vocabulary in *Longman Dictionary of Contemporary English*

Hai Xu, *Centre for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou, China (xuhai1101@yahoo.com.cn)*

Abstract: This article critically analyses the Longman Defining Vocabulary (LDV) in relation to its size, range and frequency, senses, parts of speech, affixes, and multiword expressions. The recent versions of the LDV contain a relatively fixed number of items. Over 85% of those items were found to be highly frequent, and for the defining purpose, genus terms, grammatical terms, etc. have also been included. The number of affixes in the LDV has been greatly reduced, and some common derivatives have been listed separately. On the other hand, the actual size of the LDV is much larger than was reported, for LDOCE did not distinguish between the LDV items with different senses or forms. It was found that the claim of using the 'most common meanings' of the LDV items is not always held true. The parts of speech of the LDV items have not been systematically indicated. Many multiword expressions, which have been used in the definitions in LDOCE, are not part of the LDV. This study sheds some light on the improvement in the practice of using a controlled defining vocabulary in an English learner's dictionary.

Keywords: CONTROLLED DEFINING VOCABULARY, LONGMAN DEFINING VOCABULARY, SIZE, RANGE, FREQUENCY, SENSE INDICATION, POS INDICATION, INCLUSION OF AFFIXES, INCLUSION OF MULTIWORD EXPRESSIONS

Opsomming: 'n Beoordeling van die beperkte definisiewoordeskat in die *Longman Dictionary of Contemporary English*. In hierdie artikel word 'n kritiese ontleding gedoen van die Longman Definisiewoordeskat (LDW) met betrekking tot die grootte, omvang en frekwensie, betekenisse, woordsoorte, affikse en meerwoordige uitdrukkings daarvan. Die onlangse weergawes van die LDW bevat 'n relatief vaste aantal items. Daar is bevind dat meer as 85% van hierdie items 'n hoë frekwensie het, en dat genusterme, grammatiese terme, ens., ook ingesluit is vir die doeleindes van definiëring. Die aantal affikse in die LDW is grootliks verminder, en sommige algemene afleidings is apart gelys. Aan die ander kant was die werklike grootte van die LDW baie groter as wat aangedui is, aangesien LDOCE nie onderskei het tussen die LDW-items met verskillende betekenisse of vorms nie. Daar is bevind dat die bewering dat die 'algemeenste betekenisse' van die LDW-items gebruik is, nie altyd gegeld het nie. Die woordsoorte van die LDW-items is nie sistematies aangedui nie. Baie meerwoordige uitdrukkings wat in die definisies van LDOCE gebruik is, vorm nie deel van die LDW nie. Hierdie studie werp lig op die verbetering in die gebruik van 'n beperkte definisiewoordeskat in 'n Engelse aanleerderswoordeboek.

Sleutelwoorde: BEPERKTE DEFINISIEWOORDESKAT, LONGMAN DEFINISIEWOORDESKAT, GROOTTE, OMVANG, FREKWENSIE, BETEKENISAANDUIDING, WOORDSOORT-

AANDUIDING, INSLUITING VAN AFFIKSE, INSLUITING VAN MEERWOORDIGE UITDRUKKINGS

1. Introduction

In the English learner's dictionary market, the five major competitors except for COBUILD have all acknowledged the role of a controlled defining vocabulary (CDV) in the decoding task, and compiled definitions on the basis of around 2,000 to 3,500 CDV items. The practice of using a CDV in definitions originates from the *New Method English Dictionary* (NMED), a learner's dictionary which had 'an entirely original feature — definitions based on a "minimum adequate vocabulary"' (Cowie 1999: 24). As a response to the vocabulary control movement during the late 1920s and early to mid-1930s, NMED used 1,490 words to define 23,898 vocabulary items. The practice of using a CDV had not been continued until the birth of LDOCE in 1978. LDOCE used a CDV of approximately 2,000 items. While for the first four editions of OALD, no attempt was made to set up a restricted defining vocabulary (Cowie 1999), the editorial policy of OALD5 changed radically. OALD5 used some 3,500 CDV items. The number of CDV items in OALD6 was claimed to have fallen to under 3,000 words, and they were called the 'Oxford 3000' in OALD7 and OALD8. CIDE and its successive editions — CALD — restricted the use of a defining vocabulary to less than 2,000 words. MED, a relatively new addition to the English learner's dictionary family, used a CDV of under 2,500 words to write definitions.

The strength of a CDV lies in its comprehensibility and simplicity. The study of MacFarquhar and Richards (1983) shows that over half of their subjects judged the definitions in LDOCE1 more helpful and easier to understand than those in either OALD3 or WNWD2. Herbst's study (1986) also demonstrates that LDOCE1 was far more successful in providing easily comprehensible and sufficiently accurate definitions than OALD3 and CULD. A CDV is believed to lighten the learning burden of foreign students and to facilitate their decoding task. Even for native speakers, some definitions in a learner's dictionary are considered to be more comprehensible than those in a collegiate dictionary. McCreary and Amacker (2006) reported that in a comprehension task of hard words, the groups of American college students who used an advanced learner's dictionary (i.e. MED1) performed slightly better than the groups using a collegiate desk dictionary (i.e. MW11). One of the reasons that some MW11 entries led students to induced errors lies in that they used low frequency words in the defining language.

On the other hand, there are some criticisms of CDV. Some definitions using CDV items are deemed to be oversimplified, vague, unnatural or convoluted (Stein 1979: 6; MacFarquhar and Richards 1983: 115; Béjoint 2000: 70; Fontenelle 2009: 419-420). Some CDV items may achieve simplicity at the expense of accuracy. 'The lack of preciseness is particularly conspicuous for

names of animals, plants, substances, and games.' (Stein 1979: 6) In order to avoid non-'core' vocabulary, the lexicographers of LDOCE1 sometimes had to resort to 'syntactically more complex, convoluted (or less natural) constructions' (Fontenelle 2009: 419). For instance, instead of defining **tabasco** as 'a very hot sauce […]', LDOCE1 compiled an unnatural definition like 'a very hot-tasting liquid […]', for **sauce** was not part of the controlled vocabulary.

In the literature, the CDVs, particularly the Longman Defining Vocabulary (LDV), have been critically examined in relation to their size, range and frequency, senses, parts of speech, affixes, multiword expressions, etc. (Stein 1979; Michels and Noël 1984; Neubauer 1984; Herbst 1986, 1996; Jansen, Mergeai and Vanadroye 1987; Whitcut 1988; Fox 1989; Bogaards 1996, 2003, 2008; Cowie 1999; Clark 2003; Bullon and Leech 2007; Fontenelle 2009). As a follow-up, this study will focus on those aspects in the recent LDV, and consider its developments over its past five versions and in comparison with other CDVs.

2. Size of the LDV

Each version of the LDV included approximately 2,000 vocabulary items: the LDV1 contained 2,215 word forms[1]; LDV2, 2,244; LDV3, 2,091; LDV4, 2,109; and LDV5, 2,107. Compared with the LDV1, the LDV2 removed 203 items, and added 232 new items. On the basis of the second version, the third deleted 359 items, and entered 206 new items. There were slight changes between the third and fourth versions: 25 items were excluded, and 43 were included. The changes between the fourth and fifth were the smallest: only one word (i.e. *thousandth*) and one phrasal verb (i.e. *look sth up*) were removed.

Of the 230 defining vocabulary items that occurred only once in the five versions of the LDV, 169 items (73.5%) were found in the first version, 60 (26.1%) in the second, and 1 (0.4%) in the third. The changes of defining vocabulary items over the last three versions were negligible.

The above results indicate that the core of the LDV items has been established. Actually, there are 1,769 defining vocabulary items (including 30 affixes) in common in the five versions of LDV.

However, it should be pointed out that the figure of around 2,000 may not have such a magic power. The actual size of a CDV is often underestimated. 'Partly in order to suggest that the learning burden represented by a CV is lighter than in fact it is, there is a tendency on the part of their designers to conceal their actual size.' (Cowie 1999: 110) The count of CDV items in a learner's dictionary is often fraught with some problems: not distinguishing between words with different senses or forms, not indicating parts of speech, and not including multiword expressions. That is why Stein (1979: 6) argued that, 'it is more important that this vocabulary be semantically self-sufficient than that it be restricted to an arbitrary fixed number of items'.

3. Range and frequency of the LDV items

To determine the range and frequency of the LDV items, the present study used the computer programs RANGE and FREQUENCY, designed by Heatley, Nation and Coxhead (2002), to compare the LDV lists with the BNC Word Family Lists. The programs 'can be used to find the coverage of a text by certain word lists' (Heatley, Nation and Coxhead 2002), and the BNC Word Family Lists reflect the nature of language use of typical English native speakers.

Table 1 shows the range and frequency of the LDV items[2] in comparison with the BNC Word Family Lists. In the five versions of the LDV, over 45% of the vocabulary items belong to the most frequent 1,000 words of English (i.e. the Base Word List 1 of the BNC Word Family Lists). More than 26% of the LDV items fall into the second range of the 1,000 most frequent words of English (i.e. the Base Word List 2). The LDV lists contain about 12% of the items which are in the range of the third 1,000 most frequent words of English (i.e. the Base Word List 3). To put it simply, over 85% of the items in the LDV lists are extremely frequent. Therefore, most of the items in the LDV lists should be familiar to dictionary users.

Table 1: Range and frequency of the LDV items

Word Lists	LDV1 (1978)	LDV2 (1987)	LDV3 (1995)	LDV4 (2003)	LDV5 (2009)
Basewrd 1	994 (45.3%)	1067 (47.8%)	1080 (52.0%)	1079 (51.5%)	1078 (51.5%)
Basewrd 2	583 (26.6%)	606 (27.2%)	605 (29.1%)	607 (29.0%)	607 (29.0%)
Basewrd 3	304 (13.9%)	296 (13.3%	250 (12.0%)	255 (12.2%)	255 (12.2%)
Subtotal	85.8%	88.3%	93.1%	92.7%	92.7%
Not in the list	311 (14.2%)	262 (11.7%)	144 (6.9%)	155 (7.3%)	154 (7.3%)
Total	2192	2231	2079	2096	2094

We further conducted a study to find out which items in the LDV5 are outside the first three base word lists of the BNC. Running the *Mark texts* option in the RANGE program, we can easily identify those words. Altogether 154 items were found (See Appendix 1). Those items include 25 affixes (e.g. *-al*, *-ance*, *-ation*, *dis-*, *-ence*), 12 grammatical or metalanguage terms (e.g. *adjective*, *noun*, *particle*, *abbreviation*, *singular*), some genus terms (e.g. *cattle*, *creature*, *military*, *mineral*, *profession*), some words on science and technology (e.g. *atom*, *biology*, *network*, *software*, *spacecraft*), some emotional words (e.g. *anxiety*, *deceive*, *loyal*, *romantic*, *ugly*), and some common words on people (e.g. *girlfriend*, *opponent*, *poet*, *president*, *priest*), on animals (e.g. *beak*, *goat*, *lion*, *monkey*, *snake*) and on food (e.g. *corn*, *flour*, *onion*, *rice*, *sauce*), etc.

The inclusion of those words in the LDV demonstrates that high-frequency words alone are not adequate for the defining purpose (cf. Michels and Noël 1984; Neubauer 1984; Whitcut 1988). To explain word meanings, lexicographers need to use some genus or high-level words, high-generality words and a set of grammatical terms.

4. Senses of the LDV items

One of the main problems with the current CDV lists is that they do not indicate which senses a CDV item are included, and which are excluded. The LDV items have not been differentiated according to their senses. Although LDOCE claimed to use 'only the most common meanings', one is not always certain about the so-called 'most common meanings' of an LDV item.

Consider how the LDV item *note* was used in the definitions in LDOCE5. The item was used 180 times, which respectively correspond to 7 senses as indicated in LDOCE5. Table 2 lists the senses of *note* and the frequency of each usage.[3]

Table 2: Use of the LDV item *note* in the definitions

Meaning of the word *note*	Instance of usage in the definitions	Frequency of usage
1 TO REMIND YOU	**jottings** *n* [plural] *informal* short notes, usually written to remind yourself about something	14
2 FOR STUDYING	**review**[2] *v* [T] **3** *AmE* to look again at something you have studied, such as notes, reports etc	5
3 SHORT LETTER	**leave**[1] **8** [T] to deliver a message, note, package etc for someone or put it somewhere so that they will get it later	3
4 OFFICIAL LETTER	**chit** *n* [C] *BrE* **1** an official note that shows that you are allowed to have something	8
5 ADDITIONAL INFORMATION	**gloss**[2] *v* [T] to provide a note in a piece of writing, explaining a difficult word, phrase, or idea	7
6 MUSIC	**pizzicato** *n* [U] musical notes played by pulling on the strings of an instrument	134
7 MONEY	**legal tender** *n* [U] coins or bank notes that people can officially use as money in a particular country	9

Table 2 indicates that the meaning of 'the sound/sign in music' predominates in the usage of *note* in the definitions, accounting for 74.4% of all the occurrences. In other words, the sixth sense of *note* rather than the first one has the highest frequency of usage in the definitions. Since an English learner's dictionary generally arranges word senses according to their frequency of usage, the sixth sense of *note* in LDOCE5 should not be regarded as the most common meaning.

The above phenomenon also exists in some items printed in small capitals[4]. Consider the use of the word *command* in the following definitions:

execute *v* **4** COMPUTER *technical* if a computer executes a program or COMMAND (= instruction), it makes the program or command happen or work

execution *n* **6** [C, U] *technical* when you make a computer program work, or a COMMAND (= instruction) happen

toolbox *n* [C] **2** a set of COMMANDs or FUNCTIONs which do various things in a computer program

user interface *n* [C] how a computer program looks on screen and how the user enters COMMANDs and information into the program

The above entries show all the usage of the small capital COMMAND in the definitions in LDOCE5.[5] Clearly, COMMAND refers to 'an instruction to a computer to do something'. It corresponds to Sense 3 in LDOCE5. As for the first sense 'the control of a group of people or a situation', there is not an instance of usage. Hence, the claim of using the most common meanings of a defining vocabulary item in LDOCE is not always held true.

Some researchers thus suggested that CDV items should be cross-referenced to their appropriate senses in the dictionary (Neubauer 1984: 120; Herbst 1986: 114; Jansen, Mergeai and Vanadroye 1987: 84). In that case, however, the size of a CDV would rise substantially. A vocabulary item is more often than not polysemous, and only 20.19% of the LDV items are monosemous (Jansen, Mergeai and Vanadroye 1987: 84). As Petöfi (1977) pointed out, 'West's 1480 [*sic*, 1490] word forms of the defining vocabulary correspond to 4607 senses if one only counts the senses indicated by West himself in his dictionary' (cited from Neubauer 1984: 118). Herbst (1986: 105) also noted that 'the number of words used in LDOCE definitions could be estimated to lie between 5,000 and 10,000'. Hence, most learner's dictionaries are loath to measure the size of a CDV in terms of the number of senses of CDV items. That partly explains why the senses of CDV items are not differentiated and indicated in an English learner's dictionary.

Nevertheless, CIDE made some positive changes in the sense differentiation of polysemous CDV items. For instance, while all the CDV lists included the polysemous word *set*, only CIDE indicated that it was used in the meanings of [get ready] and [group], not in the senses of [position], [condition], [establish], [fix], etc. The sense coding system used in CIDE is called 'guide words'. Although OALD, LDOCE and MED have a similar sense coding system, respectively called 'short cuts', 'signposts' and 'menu', to guide the meanings of a polysemous headword, those dictionaries did not use the technique to differentiate the meanings of a CDV item.

5. POS indication of the LDV items

LDOCE introduced the policy that a POS label was given when there was a

restriction of some kind on a CDV item. For example, the POS label *n* was shown for the word *value*, indicating that in the definitions, the word was used only as a noun and not as a verb.

Yet, 'the parts of speech (POS) are not systematically mentioned' in the LDV list (Jansen, Mergeai and Vanadroye 1987: 83). Although the word *left* was included in the LDV list, some confusion over its POS would arise: it can be used in the definitions as an adjective, noun, adverb or past participle.

While the POS indication policy in the recent LDV lists did not show any change, it was noted that the Oxford 3000 labelled all its defining vocabulary items with the POS. Dictionary users will benefit from the clear indication of the POS of CDV items.

6. Affixes in the LDV lists

A derived word, such as *laziness*, though not on the LDV list, was still used in the definitions in LDOCE, for it can be formed by adding an affix to a CDV item. LDOCE included 54 affixes in the LDV1, 49 in the LDV2, and 30 respectively in the LDV3, LDV4 and LDV5.

The reason for the decrease of affixes in the LDV lists is that some affixes like *-al*, *-ly* and *-er* are too productive (Herbst 1986: 104). 'As a result of excessive reliance on affixation, some words like *free-dom* do not appear on the list. Others are: *busi-ness*, *for-th* and *un-less*.' (Jansen, Mergeai and Vanadroye 1987: 83) Although *independence* was not part of the LDV, it was still used in LDOCE3, for 'it can be constructed from its elements (**in-** + **depend** + **-ence**)' (Bogaards 1996: 289). That is why Herbst (1986: 114) argued that 'word formations whose meanings are not easily derivable ought to be listed as separate elements of the defining vocabulary'.

It was noted that the affix inclusion policy changed in the LDV3: 'the forms which are common, or which change their meaning when a prefix or suffix is added, (such as **acceptable** and **agreement**) are included in the full list.' The other English learner's dictionaries also limited their number of affixes in their CDVs. There are only 7 affixes in the Oxford 3000, namely 4 prefixes (*anti-*, *ex-*, *non-*, and *re-*), 1 suffix (*-ish*), and 2 combining forms (*mid-* and *self-*). Thus, some derivatives, such as *carefully*, *certainly/uncertain* and *encouragement*, were listed separately in the Oxford 3000. In the list of the Macmillan Defining Vocabulary (MDV), not a single affix was found. MED adhered to the principle that only inflected forms of the CDV items were used in the definitions. So some derivatives like *formally*, *improvement* and *leadership* became part of the MDV.

7. Multiword expression in the LDV lists

The role of multiword expressions in a CDV should not be underplayed, for a large proportion of language used in discourse is made up of formulaic se-

quences (Biber et al. 1999; Erman and Warren 2000; Wray 2002, 2008). Formulaic language such as phrasal verbs and collocations contributes to the naturalness of the defining language (Cowie 1999: 158). Phrasal verbs are, in particular, quite often used in a defining role (Cowie 1999: 111), but surprisingly there was only one phrasal verb in the LDV1 list, namely *wrap (up)* (Jansen, Mergeai and Vanadroye 1987: 83-84). LDOCE1 and LDOCE2 did not accord 'to each phrasal verb its full status as a separate lexical item in the defining vocabulary list' (Whitcut 1988: 52-53).

The recent editions of LDOCE paid more attention to the multiword expressions in the LDV. In terms of phrasal verbs, whereas there was only one in the LDV1, the number increased to 8 or 9 respectively in the LDV3, LDV4 and LDV5. Table 3 shows all the multiword expressions appearing in the LDV lists.

Table 3: Multiword expressions in the LDV lists

Types	LDV1	LDV2	LDV3	LDV4	LDV5
Phrasal verbs	*wrap (up)*		*deal with* *let go of* *lie down* *look after* *look for* *look sth up* *make into* *make up* *pick up*	*deal with* *let go of* *lie down* *look after* *look for* *look sth up* *make into* *make up* *pick up*	*deal with* *let go of* *lie down* *look after* *look for* *make into* *make up* *pick up*
Compounds	*postage stamp* *so-called* *upside-down*	*old-fashioned* *upside down*	*old-fashioned* *only just* *upside down*	*old-fashioned* *only just* *upside down*	*old-fashioned* *only just* *upside down*
Other multiword expressions	*all right* *according (to)* *no one* *owing to* *worthy (of)*	*according (to)* *no one* *owing to* *worthy (of)*	*according (to)* *as opposed to* *in spite of* *no one*	*according (to)* *as opposed to* *in spite of* *no one* *relating to*	*according (to)* *as opposed to* *in spite of* *no one* *relating to*

It is likely that the use of the multiword expressions in the definitions in LDOCE was not limited to those as shown in Table 3. To investigate the use of multiword expressions in the definitions, we compared those expressions in the LDV5 with those in the Oxford 3000. Among all the CDVs, the Oxford 3000 included the largest number of multiword expressions — altogether 126 ones (see Appendix 2). An overlap of 8 multiword expressions was found between the LDV5 and the Oxford 3000, namely *deal with*, *in spite of, look after*, *look for*, *make (sth) up*, *pick (sth) up*, *old-fashioned*, and *(as) opposed to*. Of the remaining 118 multiword expressions, it was found that 105 ones[6] were actually used in LDOCE5. Only the following 13 were not used in the definitions in LDOCE5: *associated with*, *insist on*, *look forward to*, *rely on*, *go bad*, *on board*, *bound to*, *capable*

(of), *a couple (of)*, *for instance*, *on purpose*, *Yours sincerely*, and *Yours Truly*. In other words, LDOCE5 used quite a large number of multiword expressions in the definitions, but did not acknowledge them in the LDV list.

It is not clear why LDOCE5 listed the multiword expressions as shown in Table 3, rather than the other ones that had been used. The LDV5 did include *consist*, *instead*, *rather*, *refer*, *relate*, *such*, etc., but when these items occurred in the definitions, they were, more often than not, used in the form of a set phrase, viz. *consist of*, *instead of*, *rather than*, *refer to*, *relate to*, *such as*, etc. It would be more helpful to learners if the fixed expressions were instead listed in the LDV.

In addition, the meanings of some multiword expressions are not transparent. Consider the meanings of *as well as* and *by accident* in the following definitions:

> **boarder** *n* [C] **1** a student who stays at a school during the night, as well as during the day
>
> **bang**[2] *v* **4** [T] to hit a part of your body, or something you are carrying, against something, by accident

The meanings of *as well as* and *by accident* can hardly be deduced from the meanings of their components *well* and *accident*. It is beyond the capabilities of some foreign learners to decode the meanings of such multiword expressions. Thus, it is a good idea to list them separately in the LDV and call learners' attention to them.

8. Conclusion

The recent LDV lists show some trends. Compared with the earlier versions of the LDV, the changes of CDV items in the recent lists are small. There is a relatively fixed number of LDV items. Over 85% of the LDV items fall into the ranges of the first three thousand most frequent words, and they should be familiar to ESL learners. For the defining purpose, some genus terms, grammatical terms as well as a few multiword expressions are also included in the LDV. The number of affixes in the LDV lists has been greatly reduced, and some common derivatives have been included in the full list.

There is still room for improvement in the LDV. It is insignificant to claim the use of only around 2,000 items in the definitions, for the actual number of the LDV items used in the definitions is much larger than was reported. The dictionary could indicate the senses, POS and multiword expressions of the LDV items more clearly. LDOCE can follow CIDE by using 'signposts' to show the senses of polysemous LDV items as used in definitions. The indication of POS of the LDV could be more consistent. Multiword expressions are indispensable to the definitions, and should be clearly indicated in the LDV list.

This article concerns CDVs in paper dictionaries. As we are entering a new era of electronic dictionaries, will CDVs still play a role? It is true that technical innovations offer the potential to extend a repertoire of modes of meaning representation from verbal explanation to multimodality, employing such devices as pictorial and animated illustrations, audio recordings and video clips (De Schryver 2003; Lew 2010). However, even in electronic dictionaries, definition still plays 'a prominent role' (Lew 2010: 292). 'Nor should the lexicographic definition contain words more difficult to understand than the explained word itself' (Zgusta 1971: 257). The basic principle of using simple words (i.e. CDV) in definitions is equally applicable to electronic dictionaries. Compared with printed dictionaries, one noticeable change in electronic dictionaries is that there might be more liberal use of words outside a CDV or a larger set of CDV items (cf. Lew 2010: 293). With mouse clicking or even hovering, hypertext links or pop-up windows in the electronic medium will provide instant access to the meaning of some words as used in definitions. Nevertheless, a caveat is that 'too many infrequent words in a definition may create comprehension problems' (Lew 2010: 293). Research issues, such as the pros and cons of a CDV in electronic dictionaries and the appropriate number of CDV items, deserve further investigation.

Acknowledgements

This study was supported by the MOE Project of Key Research Institute of Humanities and Social Sciences at Universities in P.R. China (Project No. 12JJD740010), by the Programme for New Century Excellent Talents in University (NCET-11-0912), and by the Project of Humanities and Social Sciences at Universities in Guangdong Province (Project No. 11WYXM025). I am grateful to two anonymous reviewers for their constructive comments and to Ms Janice McAlpine, Prof. Chuming Wang, Prof. Jianhua Huang and Prof. Alison Wray for their encouragement.

Endnotes

1. The figures reported in this section are the count of word forms (including affixes) rather than lemmas. Therefore, *actor* and *actress* were treated as two word forms, though they were listed together in the LDV lists. Similarly, *according (to)*, *alcohol(ic)*, *arrangement(s)*, etc. were respectively regarded as two word forms. On the other hand, some phrases (e.g. *deal with*, *let go of*, *no one*), though being separated in space, were given full status as one word form. The homographs in the list, such as *bear* (*n.*) and *bear* (*v.*), were respectively calculated.
2. The figures listed in Table 1 are the frequency of word types, not the frequency of tokens or word families (for the differences among the terms 'token', 'type' and 'word family', see Nation 2001: 7-8).

3. In Table 2, the senses of *note* were coded by the signposts as used in LDOCE5, and the sense ordering is identical with that of the dictionary.
4. In the interests of brevity and precision of a definition, almost all learner's dictionaries occasionally resort to some words outside a CDV list, and those words are printed in small capitals.
5. Another instance of usage of *command* in the definitions can be found, but it was not printed in small capitals:

   processor *n* [C] **1** the central part of a computer that deals with the commands and information it is given.

   As for non-LDV words in the definitions, there are some more typographical inconsistencies: *compound* is printed in small capitals in the definitions of **chloride**, **dioxide**, **halogen**, **halon**, **hydroxide**, **polymer**, **present participle**, **reagent**, **silica**, and **valence**, whereas in defining **hydrocarbon**, *compound* is in normal print; while *grave* in small capitals is used to define **archaeology**, **barrow**, **burial**, **bury**, **disinter**, **epitaph**, **ghoul**, **graveside**, **gravestone**, **headstone**, **tablet**, **tombstone**, **tumulus**, and **violate**, it is in normal print for the definition of **gravesite**. On the other hand, *physics*, part of the LDV5, has been carelessly set in small capitals in the following entries: **biophysics**, **Nobel Prize**, **physical**[1], **physical science**, **physicist**, **quantum**, **relativity**, **science**, and **solid-state**.

   In addition, it is not clear why LDOCE4 and LDOCE5 removed from their CDV lists the items *compound* and *grave*, which respectively occurred 11 and 15 times in the definitions, while adding such an item as *physics*, which was used 13 times.
6. This includes some multiword expressions in small capitals (e.g. *credit card* and *make-up*).

References

A. Dictionaries

CALD3 = Walter, E. (Ed.). 2008. *Cambridge Advanced Learner's Dictionary* (Third Edition). Cambridge: Cambridge University Press.

CIDE = Procter, P. (Ed.). 1995. *Cambridge International Dictionary of English* (First Edition). Cambridge: Cambridge University Press.

COBUILD1 = Sinclair, J. (Ed.). 1987. *Collins COBUILD English Language Dictionary* (First Edition). London/Glasgow: Collins.

COBUILD2 = Sinclair, J. (Ed.). 1995. *Collins COBUILD English Dictionary* (Second Edition). London/Glasgow: HarperCollins.

COBUILD6 = Sinclair, J. (Ed.). 2008. *Collins COBUILD Advanced Learner's English Dictionary* (Sixth Edition). Glasgow: HarperCollins.

CULD = Kirkpatrick, E.M. (Ed.). 1980. *Chambers Universal Learners' Dictionary*. Edinburgh: Chambers.

LDOCE1 = Procter, P. (Ed.). 1978. *Longman Dictionary of Contemporary English* (First Edition). Harlow: Longman.

LDOCE2 = Summers, D. (Ed.). 1987. *Longman Dictionary of Contemporary English* (Second Edition). Harlow: Longman.

LDOCE3 = Summers, D. (Ed.). 1995. *Longman Dictionary of Contemporary English* (Third Edition). Harlow: Longman.

LDOCE4 = Summers, D. (Ed.). 2003. *Longman Dictionary of Contemporary English* (Fourth Edition). Harlow: Pearson Education.

LDOCE5 = Mayor, M. (Ed.). 2009. *Longman Dictionary of Contemporary English* (Fifth Edition). Harlow: Pearson Education.

MED1 = Rundell, M. (Ed.). 2002. *Macmillan English Dictionary for Advanced Learners* (First edition). Oxford: Macmillan Education.

MED2 = Rundell, M. (Ed.). 2007. *Macmillan English Dictionary for Advanced Learners* (Second Edition). Oxford: Macmillan Education.

MW11 = Mish, F. (Ed.). 2003. *Merriam-Webster's Collegiate Dictionary* (Eleventh Edition). Springfield, Massachusetts: Merriam-Webster.

NMED = West, M.P. and J.G. Endicott. (Eds.). 1935. *The New Method English Dictionary*. London: Longmans, Green.

OALD3 = Hornby, A.S., A.P. Cowie and L.J. Windsor. (Eds.). 1974. *Oxford Advanced Learner's Dictionary of Current English* (Third Edition). Oxford: Oxford University Press.

OALD4 = Cowie, A.P. (Ed.) 1989. O*xford Advanced Learner's Dictionary of Current English* (Fourth Edition). Oxford: Oxford University Press.

OALD5 = Crowther, J. (Ed.). 1995. O*xford Advanced Learner's Dictionary of Current English* (Fifth Edition). Oxford: Oxford University Press.

OALD6 = Wehmeier, S. (Ed.). 2000. O*xford Advanced Learner's Dictionary of Current English* (Sixth Edition). Oxford: Oxford University Press.

OALD7 = Wehmeier, S. (Ed.). 2005. O*xford Advanced Learner's Dictionary of Current English* (Seventh Edition). Oxford: Oxford University Press.

OALD8 = Turnbull, J. (Ed.). 2010. O*xford Advanced Learner's Dictionary of Current English* (Eighth Edition). Oxford: Oxford University Press.

WNWD2 = Guralnik, D.B. (Ed.). 1980. *Webster's New World Dictionary of the American Language* (Second College Edition). New York: Simon and Schuster.

B. Other literature

Béjoint, H. 2000. *Modern Lexicography: An Introduction*. Oxford/New York: Oxford University Press.

Biber, D. et al. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.

Bogaards, P. 1996. Dictionaries for Learners of English. *International Journal of Lexicography* 9(4): 277-320.

Bogaards, P. 2003. MEDAL: A Fifth Dictionary for Learners of English. *International Journal of Lexicography* 16(1): 43-55.

Bogaards, P. 2008. Frequency in Learners' Dictionaries. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15-19 July 2008*: 1231-1236. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.

Bullon, S. and G. Leech. 2007. Longman Communication 3000 and the Longman Defining Vocabulary. Bullon, S. and G. Leech (Eds.). 2007. *Longman Communication 3000*: 1-7. Harlow: Pearson Education.

Clark, G. 2003. Analyzing Word Meaning and Representation by Navigating Dictionary Definitions. http: //www.crsc.uqam.ca/en/SiteGraham/index.html.

Cowie, A.P. 1999. *English Dictionaries for Foreign Learners: A History*. Oxford: Oxford University Press/Clarendon Press.

De Schryver, G-M. 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.

Erman, B. and B. Warren. 2000. The Idiom Principle and the Open-choice Principle. *Text: An Interdisciplinary Journal for the Study of Discourse* 20(1): 29-62.

Fontenelle, T. 2009. Linguistic Research and Learner's Dictionaries: the *Longman Dictionary of Contemporary English*. Cowie, A.P. (Ed.). 2009. *The Oxford History of English Lexicography. Volume II: Specialized Dictionaries*: 412-435. Oxford: Oxford University Press.

Fox, G. 1989. A Vocabulary for Writing Dictionaries. Tickoo, M.L. (Ed.). 1989. *Learners' Dictionaries: State of the Art*: 153-171. Singapore: SEAMEO Regional Language Centre.

Heatley, A., I.S.P. Nation and A. Coxhead. 2002. RANGE and FREQUENCY Programs. http: //www.victoria.ac.nz/lals/staff/paul-nation.aspx.

Herbst, T. 1986. Defining with a Controlled Defining Vocabulary in Foreign Learners' Dictionaries. *Lexicographica: International Annual for Lexicography* 2: 101-119.

Herbst, T. 1996. On the Way to the Perfect Learners' Dictionary: A First Comparison of OALD5, LDOCE3, COBUILD2 and CIDE. *International Journal of Lexicography* 9(4): 321-357.

Jansen, J., J.P. Mergeai and J. Vanadroye. 1987. Controlling LDOCE's Controlled Vocabulary. Cowie, A.P. (Ed.). 1987. *The Dictionary and the Language Learner. Papers from the EURALEX Seminar at the University of Leeds, 1–3 April 1985*. Lexicographica. Series Maior 17: 78-94. Tübingen: Max Niemeyer.

Lew, R. 2010. Multimodal Lexicography: The Representation of Meaning in Electronic Dictionaries. *Lexikos* 20: 290-306.

MacFarquhar, P.D. and J.C. Richards. 1983. On Dictionaries and Definitions. *RELC Journal* 14(1): 111-124.

McCreary, D.R. and E. Amacker. 2006. Experimental Research on College Students' Usage of Two Dictionaries: A Comparison of the *Merriam-Webster Collegiate Dictionary* and the *Macmillan English Dictionary for Advanced Learners*. Corino, E., C. Marello and C. Onesti (Eds.). 2006. *Proceedings XII EURALEX International Congress, Turin, Italy, September 6th–9th, 2006*: 871-885. Alessandria: Edizioni dell'Orso.

Michels, A. and J. Noël. 1984. The Pro's and Con's of a Controlled Defining Vocabulary in a Learner's Dictionary. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983.* Lexicographica. Series Maior 1: 385-394. Tübingen: Max Niemeyer.

Nation, I.S.P. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Neubauer, F. 1984. The Language of Explanation in Monolingual Dictionaries. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83. Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983.* Lexicographica. Series Maior 1: 117-123. Tübingen: Max Niemeyer.

Stein, G. 1979. The Best of British and American Lexicography. *Dictionaries: Journal of the Dictionary Society of North America* 1: 1-23.

Whitcut, J. 1988. Lexicography in Simple Language. *International Journal of Lexicography* 1(1): 49-55.
Wray, A. 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
Wray, A. 2008. *Formulaic Language: Pushing the Boundaries*. Oxford: Oxford University Press.
Zgusta, L. 1971. *Manual of Lexicography*. The Hague: Mouton.

Appendix 1: The LDV items outside the 3,000 most frequent words

abbreviation, absent, adjective, adverb, -al, alphabet, -ance, anxiety, anxious, -ation, atom, beak, beneath, biology, broadcast, bubble, bullet, carriage, cattle, ceremony, chemistry, chin, clay, comb, consist, corn, creature, criticize, curve, decay, deceive, decrease, defeat, dis-, dismiss, eager, email, -ence, enclose, -er, explosion, explosive, fasten, fever, flesh, flour, former, -ful, girlfriend, goat, herb, hollow, horizontal, horn, -ic, -ical, im-, importance, in-, -ing, ink, insect, inwards, -ion, ir-, -ish, -ity, -ive, -ize, kilogram, kilometre, kiss, kneel, laughter, leather, lion, loyal, loyalty, -ly, -ment, military, mineral, mist, monkey, navy, -ness, network, non-, noun, obey, ocean, old-fashioned, onion, opponent, oxygen, parallel, participle, pepper, phrase, physics, plural, poet, poetry, preserve, president, priest, profession, pronunciation, protest, quarrel, re-, rice, romantic, sauce, scatter, scissors, self-, sideways, singular, skirt, slippery, slope, snake, soap, software, sour, spacecraft, spice, spicy, stem, sting, swell, sword, thirsty, thread, tobacco, tomato, transparent, tribe, tropical, ugly, un-, underwear, universe, upright, valley, verb, vertical, vowel, waist, wealth, weave, website, wheat (154)

Appendix 2: Multiword expressions in the Oxford 3000

Phrasal verbs: *approve (of), associated with, based on, give birth (to), take care (of), care for, deal with, depend (on), disapprove (of), fall over, find out (sth), make fun of, get on, get off, give sth away, give sth out, give sth up, go down, go up, involved in, insist (on), leave out, look after, look at, look for, look forward to, make sth up, take notice of, pick sth up, put sth on, put sth out, refer to, relate (to), rely on, get rid of, set fire to sth, sit down, stand up, stick out, switch off, switch on, take sth off, take (sth) over, take part (in), throw sth away, tie sth up, wake (up), wind sth up* (48)

Compounds: *credit card, ice cream, well known, make-up, old-fashioned, post office, prime minister, swimming pool* (8)

Other multiword expressions: *one another, by accident, take action, in addition (to), in advance, all right, apart from, fall asleep, pay attention, go bad, on behalf of /on sb's behalf, a bit, on board, be born, bound to, be called, capable (of), in case (of), in charge of, in common, consist of, in control of, under control, a couple, of course, in detail, in the end, in exchange (for), fall asleep, in favour/favor (of), a few, at first, in front of, in general, be going to, good at, good for, have to, in honour/honor of, in a hurry, for instance, instead of, at least, a little, a lot (of)/lots (of), by means of, in memory of, next to, opposed to, in order to, out (of), in public, on purpose, rather than, the rest, be sick, feel sick, Yours sincerely, so that, as soon as, in spite of, such as, make sure, take place, thank you, Yours Truly, used to sth/doing sth, used to, as well (as), go wrong* (70)

Optimising Data Utilisation in Lexicography: The Case of the *Khoekhoegowab Dictionary*

Wilfrid Heinrich Gerhard Haacke, *Department of Language and Literature Studies, University of Namibia, Windhoek, Namibia (whaacke@gmail.com)*

Abstract: Despite of rapid progress in Southern Africa in the direction of multifunctionality of lexical databases through the advent of generic lexicographic software, a considerable number of lexicographic projects — especially in Khoe and Saan languages — still use or have recently used a word processor with the sole objective of compiling a printed dictionary. Hence the present paper expounds on the case of the *Khoekhoegowab Dictionary Project,* how in the early 1990s some off-the-shelf DOS-based database software was configured as part of a "home-grown" custom-made dictionary writing system. It is demonstrated in a non-technical way that the use of a structured database with fully-fledged retrieval facilities allows for the far-reaching elimination of human error in a dictionary, for the automatisation of processes like language reversal and sorting, and, finally, for the significantly enhanced usability of the data for purposes other than fixed media dictionary compilation. Compiling a dictionary without extensive query facilities as offered by tabular databases, is argued to be a lost opportunity, as it should be possible to utilise lexicographic data for more than just lexicography. By 2010 the data was accommodated in open source software to ensure its optimal survival in digital form for future use.

Keywords: AUTOMATISATION, COMPILATION SOFTWARE, DATA RETRIEVAL, DATABASE CONFIGURATION, DATABASE REPORT, FLAT-FILE DATABASE, FORM, INFORMATION GENERATION, KHOEKHOE, KHOESAAN DICTIONARIES, LEXICOGRAPHY, LOOKUP FACILITIES, MULTIFUNCTIONALITY, QUERY FACILITIES, RETRIEVAL FACILITIES, SOFTWARE, TONES

Opsomming: Die optimalisering van datagebruik in die leksikografie: Die geval van die *Khoekhoegowab Woordeboek*. Verskeie leksikografiese projekte — veral in die Khoe- en Saantale — gebruik 'n woordverwerker met die uitsluitlike doel om 'n gedrukte woordeboek saam te stel, of het dit tot onlangs nog so gebruik. Dit gebeur ondanks die opkoms van generiese leksikografiese sagteware, wat gelei het tot die vinnige vooruitgang van die multifunksionaliteit van leksikale databasisse in Suider-Afrika. Daarom gaan hierdie artikel in op die geval van die *Khoekhoegowab Woordeboekprojek,* hoe gekoopte DOS-gebaseerde databasissagteware vroeg in die 1990's as deel van 'n "tuisgemaakte" doelgemaakte woordeboekskryfstelsel opgestel is. Daar word op 'n nietegniese manier getoon dat die gebruik van 'n gestruktureerde databasis met volwaardige herwinningshulpmiddels verreikende uitskakeling van menslike foute in 'n woordeboek bewerkstellig, die outomatisering van prosesse soos taalomkering en sortering toelaat, en uiteindelik die bruikbaarheid van die data aansienlik verhoog vir doeleindes buiten woordeboeksaamstelling in vaste mediums. Daar word geargumenteer dat die samestelling van 'n woordeboek

sonder uitgebreide navraaghulpmiddels soos voorsien deur tabellariese databasisse, 'n verspeelde geleentheid is, aangesien dit moontlik behoort te wees om leksikografiese data te benut vir meer as net leksikografie. Teen 2010 is die data in oopbronsagteware geberg om te verseker dat dit optimaal voortleef in digitale vorm vir toekomstige gebruik.

Sleutelwoorde: OUTOMATISERING, SAGTEWARE VIR WOORDEBOEKSAMESTELLING, DATAHERWINNING, DATABASISKONFIGURASIE, DATABASISVERSLAG, PLATLÊERDATABASIS, VORM, INLIGTINGSGENERERING, KHOE, KHOESAANWOORDEBOEKE, LEKSIKOGRAFIE, NASLAANHULPMIDDELS, MULTIFUNKSIONALITEIT, NAVRAAGHULPMIDDELS, HERWINNINGSHULPMIDDELS, SAGTEWARE, TONE

1. Introduction

In a time of rapidly developing computational practices in lexicography it should amount to carrying coal to Newcastle to argue for the need of multifunctionality of lexical databases. However, despite of the very recent advances in software design there are still a number of recent if not current lexicographic projects in Southern Africa that operate with the sole objective of producing a printed dictionary, and that have pursued or still pursue this aim by means of a common word processor without resorting to an underlying database format or any dedicated lexicographic software. Most of these dictionaries deal with either Khoe or Saan languages, or Bantu languages outside South Africa, as the Bantu languages of South Africa are now catered for by the National Lexicography Units of the *Pan South African Language Board*, which by now all use *Tshwanelex*. The following titles are instances of Khoe or Saan dictionaries published in the last twenty years by means of common word processors (or, at best, *Toolbox*), or which are still in preparation:

— 1994. Dickens, P. *English–Ju|'hoan Ju|'hoan–English Dictionary*. (371 p.)

— 1994/2009. Traill, A. *A !Xóõ Dictionary*. (292 p.)

— 2001 (fourth prelim. edition). Visser, H. *Naro Dictionary* (240 p.)

— 2003. Kilian-Hatz, C. *Khwe Dictionary* (395 p.) based *i.a.* on the fieldwork data of Oswin Köhler collected since the 1960s)

— 2004. Weich, F. *San Dictionary/San-woordeboek. San–Afrikaans–English/ English–San–Afrikaans/Afrikaans–San–English*. (377 p.)

— 2008. König, C. and Heine, B. *A Concise Dictionary of Northwestern !Xun* (186 p.)

— Traill, A., A. Chebanne and H. Nakagawa (in prep.) *A Trilingual Dictionary in !Xóõ, English and Setswana*.

— Nakagawa, H., K. Sugawara and J. Tanaka (in prep.) *Ghui Dictionary*.

— Andy Chebanne uses Toolbox to collect lexical data from Shua and Tciretcire of the Eastern Kalahari;

— so do Christfried Naumann and Tom Gueldemann in their ongoing fieldwork on Western !Xoon.

The main reason for not using dedicated software is that projects like these usually are hamstrung by a historic legacy in that the compilation of the (lexical) corpus was started single-handedly in index card mode when — in several cases — personal computers did not exist yet, let alone generic software for dictionary compilation.[1] This use of obsolete methods may warrant a closer look at the case of the *Khoekhoegowab Dictionary Project* (originally registered as *Nama Dictionary Project* in 1981). This paper will expound on the way how in the early 1990s some off-the-shelf DOS-based database software was configured as part of a "home-grown" custom-made dictionary writing system, and how by 2010 the data was accommodated in open source software to ensure its optimal survival in digital form. The paper will then provide instance how such lexical data can be utilised significantly beyond its primary purpose of serving as lemmas in a fixed media dictionary or glossary. It is suggested that the time, cost and effort spent on converting lexical data that have been compiled in a word processor document into a structured database with fully-fledged retrieval facilities, is — in the long run — amply compensated for

— by the far-reaching elimination of human error in the dictionary;

— by the automatisation of processes like language reversal and sorting; and, finally,

— by the significantly enhanced usability of the data for purposes other than dictionary compilation.

To today start compiling a dictionary without extensive retrieval facilities as offered by databases, is a waste of opportunities that can no longer be justified. It should be possible to utilise lexicographic data for more than just lexicography.

2. The Beginnings of the *Nama Dictionary Project*

The "*Nama Dictionary Project*" (henceforth ***NDP***) commenced with data collection in 1981; that is, at least in Namibia, in the era of typewriters and handwritten index cards.

Pastor Eliphas Eiseb as Khoekhoe speaker and co-author, worked for the project full-time for eleven years (until 1992), collecting data by perusing the existing literature in Khoekhoegowab, when the team did not work in plenum (in the afternoons). Electronic text corpora for Khoekhoegowab did not exist in

those days of off-set printing, nor were corpus-query tools available for the analysis of text corpora and the compilation of concordances (cf. Van Sterkenburg 2003: 195 ff. and De Schryver and De Pauw 2007 on the development of digital resources). Even if electronic text corpora had been produced by scanning, they would have been of only limited value as the revised and standardised orthography of Khoekhoegowab had not yet established itself in the literature and lemmas would hence not have been recognised automatically. So the selection of lemmas depended almost solely on the discretion and memory of Eiseb while perusing the literature and recordings. The electronic database comprises over 24 500 Khoekhoe records, of which some 2 700 are illustrative examples. In this paper the term database will be used not in its widest sense of any accumulation of data but in the more technical sense of a tabular, structured database.

2.1 The Lexicographic Concept of the *Khoekhoegowab Dictionary*

The aim of the *NDP* was (and is) of a dual nature:

— firstly *practical*, to provide a comprehensive bilingual dictionary for general use by Khoekhoe as well as non-Khoekhoe speakers;

— secondly *academic*, to document the atrophying lexicon of the last surviving language of the KHOEKHOE branch of the KHOE (Central Khoesaan) family for comparative and other linguistic purposes.[2]

These two aims require rather diverging lexicographic approaches and procedures, and a combination of them unavoidably requires compromises. To do them full justice such aims can hardly be pursued in one and the same dictionary. Circumstantial constraints and urgency, however, demanded that several needs be addressed in the same project, as is so often the case in the early stages of lexicographic documentation of a language. The fact that the target users were to be of a maximally diversified nature required even more systematic planning of the lexicographic procedures, in an attempt to strike an acceptable balance. For an elaboration on these considerations see Haacke (1998).

The *Khoekhoegowab Dictionary* (***KhD***) was planned to cover the lexicon as widely as possible, including even archaic and obsolete words, albeit labelled accordingly. One reason is that it is hoped that valuable data for comparative reconstruction of proto-Khoe might be preserved. The other reason is that it is hoped that some of the culture-specific words are not replaced by modernisms but may be revived and thus saved from oblivion — like the glossonym *Khoekhoegowab* itself.

From its inception in 1981 to 1989 the *KhD* was compiled on handwritten index cards (with a back-up set), to be eventually typed after completion of all index cards. In 1988, when I was engaged in doctoral studies in London, I met

the phonologist and programmer Professor Jonathan Kaye at the School for Oriental and African Studies. When I approached him about my needs for customised lexicographic software, he immediately consented to develop such software.

The main reasons why I needed *customised* software was that

— lemmas were to be sorted by phonemes, rather than by letters; i.e. phonemes represented by polygraphs, such as clicks and their releases, should have their stipulated position in the alphabet;

— the Lepsius click symbols (/, ǁ, !, ≠) were to be integrated into the officially stipulated alphabetising sequence;

— lemmas were to be sorted according to four tones, after they had been sorted by segmental phonemes.

Kaye's concept provided for two main stages, strictly distinct from each other: firstly, a **pre-dictionary stage** in which the data would be entered into a database by means of an off-the-shelf generic database management program. Secondly, a **dictionary compilation stage** in which these data are converted into a print-ready presentation layout format by means of compilation software using a conventional word processor and software that he would write. Separate compilation programmes serve the compilation of the *Khoekhoegowab–English Dictionary*, the conversion into a *Khoekhoegowab–English Glossary*, and the reversal to an *English–Khoekhoegowab Index* or *Glossary*. The strength of this concept rests in the provision that the data is to be stored in one common database file, rather than entering it into one or more text files from the start. The fact that this strategy was adopted in 1989 is evidence of farsighted programming. It made a pivotal difference to the project.

The database management program to be used was *File Express*; the word processor *Word Perfect 6.0a* (mainly because of its superior macro facilities in those early years of Windows platforms), and the programming language was *Icon*.

3. The Configuration of the *FE* Database Format for *NDP5*

File Express (***FE***) is described by its producers *Expressware* as flat-file database management software with basic relational lookup facilities, first published in 1984; it uses a DOS platform and was written for the IBM Personal Computer. In a flat-file database the data are organised in a single two-dimensional matrix consisting of rows and columns, as opposed to a relational database, where further tables may be incorporated in the main table in a hierarchical structure. A typical flat-file application would be an address-book in tabular format, with the data for each person filling one row, while

information *categories* like street name, town, postal code, etc. would be assigned to respective columns.

FE was, at its time, considered to be the "friendliest, most intuitive database program available anywhere, at any price", according to the prolific freelance author on computer matters, Alfred Glossbrenner. This user friendliness manifests itself right away in the arrangement that the primary user interface is not the tabular database layout with rows and columns. Rather, an input screen or Form is displayed which transparently presents one record at a time, for easy entering or editing of data. The layout of the form is designed by the user. It is one of the credits of *FE* that even novices without programming skills can configure the structure of the database by simply designing a Form. It was not common in those days that software would allow novices to create a database without the assistance of a programmer.

The configuration of this Form is of utmost importance as it is this design of the database format that determines what categories of information can be retrieved and hence, what determines the extent of multifunctionality of the database. The header file of the *Nama Dictionary Project* uses the acronym "NDP**5**". This indicates that it was only the fifth version of the Form layout that was found to be satisfactory. It was progressively improved by trial and error.

In the record format a record is made up of "Fields" that contain specific types of information (in the case of the *NDP* displayed in horizontal lines; cf. also Figure 6 below). These fields of the Form display the data of the respective *column* of the table format.

Figure 1 displays the configuration of a record as defined on the database definition screen for the *NDP*. This set-up amounts to a Form in current database software like *OO BASE*. 18 fields were specified, each identified by a number and an arbitrary "field name" that serves as descriptive mnemonic for the user.

Only the fields displayed in bold print (F2, F4, F5 and F7-F13) eventually appear in the compiled dictionary. The other fields have housekeeping functions. The sequence of data types within a record was already determined by the ordering of the fields in the present Form. Of interest here is only the column for –length–. In "length" the field length has to be specified (in bytes), as *FE* uses fixed-length fields. Hence it is important in the interest of saving data volume, to configure a field as short as possible. One record in the above layout has a length of 563 bytes, irrespective of whether the space has been maximally populated or not. The complete data file of the *NDP* as used for the *Dictionary* had a size of just under 13.5MB. More will be said later about the saving of field length through the use of codes. All fields of a record need manual completion, but need not be obligatorily populated, as long as at least one field has data.

-field name-	-type-	-length-	-format-
1 KEY ENTRY	C	15	Character
2 ENTRY	**C**	**80**	**Character**
3 ENTRY ALLOC	C	30	Character
4 EXAMPLE	**C**	**50**	**Character**
5 PRONUNC	**C**	**50**	**Character**
6 CATCHWORD	C	22	Character
7 PT SPEECH	**C**	**2**	**Character**
8 LABEL	**C**	**2**	**Character**
9 AFR RENDERING	**C**	**130**	**Character**
10 SCIENTIFIC	**C**	**40**	**Character**
11 ADD COMMENT	**C**	**3**	**Character**
12 X-REFERENCE	**C**	**33**	**Character**
13 LOAN	**C**	**32**	**Character**
14 SOURCE	C	4	Character
15 TONE CONSTIT	C	16	Character
16 ENGL ENTRY	C	45	Character
17 NOTES?	C	2	Character
18 STAGE/DATE	C	6	Character

Record Length: 563 bytes

Figure 1: Database Specifications for a Record in *NDP5*

It should be kept in mind that this database structure was conceived in 1989 primarily, if not solely, with the intention to compile a bilingual Khoekhoe–English dictionary with tone marking. Initially the ideal of multifunctionality did not feature prominently in the design, other than using the lexicographic data for a tonological analysis. After all, the concept of multifunctionality was, at that stage, not very prevalent in lexicography in general. Once the compilation software for the *NDP*, i.e. *NDP5* had been finalised by 1992 it was no longer possible to add further fields to the *FE* database, as a change of field number would have lead to mismatches in the compilation software (unless the field was non-printable and was added after the last field that was addressed by the compilation software). The only possibility to extend options of information generation that does not involve major surgery is to add further metalinguistic codes that would be accommodated in existing fields. Examples will be presented in the next section.

The fields are now briefly introduced.

F1, *KEY ENTRY*, merely serves for sorting to subsume all records into one article under the main lemma as specified in F1. The contents of F1 do not appear in print. Any record that has an empty F1 serves as main lemma. The Khoekhoe–English dictionary contains 4 273 main lemmas, i.e. 4 273 articles.

F2, *ENTRY*, contains the Khoekhoe lemma to be printed, unless the record contains an example in F4, in which case F2 is empty. F2 and F4 thus are mutually exclusive. F2 is specified for citation tone. The mode of tone codification will be explained further on. Dialect variants are marked in the running text with one of the following labels:

[Bz]	Bondelzwarts	**[V]**	Vaalgras Nama
[D]	Damara	**[T]**	Topnaar, ǂAonîn
[Hm]	Hai‖om	**[ǂA]**	ǂĀkhoe
[S]	Sesfontein Damara	**[ǂD]**	ǂAodama

Figure 2: Labels for Dialect Variants

Data for a specific dialect can thus be extracted by searching for these labels.

F3, *ENTRY ALLOCATION*, is perhaps a misnomer, as more aptly it should have been called Example Allocation. It specifies the lemma (F2) that an example as provided in F4 has to illustrate, i.e. that F4 has to follow in the sort during compilation. F3 thus is not printed. As the lemma of F2 in turn is assigned to a head lemma by means of F1, all records from main lemma to sub-lemmas with their examples will be assigned to the same article in ordered fashion.

F4, *EXAMPLE*, contains example sentences to illustrate mainly usage (esp. verb valency) or collocations. 2 703 records (11%) of the total of 24 520 Khoekhoe records are examples.

F5, *PRONUNCiation*, provides pronunciation of a lemma if the standardised spelling (in F2) does not readily reflect it; cf. Figure 3.

Pre-dictionary database input:

F2 ENTRY	aama
F5 PRONUNC	**(pronunc.: a2a2mâ3a2)**
F7 PT SPEECH	t
F9 RENDERING	(go on) pub-crawl, ...

Dictionary output:

aama {āma} (pronunc.: ààmáˆà) *v.t* (go on) pub-crawl,

Figure 3: Information on Pronunciation

F6, *CATCHWORD*, is a very important retrieval tool. It represents lemmas of F2 without tone marking, so that the record can be searched for in a simple query without the tone pattern having to be known — which is the normal situation when one wants to consult a Khoekhoe entry. F6 is not printed.

F7, *PT SPEECH*, specifies the word category of lemmas. As certain abbreviations contain up to eight characters, e.g. *v.i.stat* (stative intransitive verb), all abbreviations are assigned a pre-ordained code consisting of maximally two characters, so as to save space in the fixed-length records. These codes of the database are automatically replaced with the full abbreviation by the dictionary compilation software. While codes were used to save space, they — as a significant spin-off — bring the advantage that the abbreviations that the codes are automatically replaced with will be absolutely consistent with no scope for human error.

F8, *LABEL*, specifies usage labels. Again, preordained codes of maximally two characters are used, so as to save space (while ensuring consistency). The following labels (in italics) appear in print after the conversion:

arch.	archaic	*lit.*	literal
bot.	botanical	*math.*	mathematics
colloq.	colloquial	*med.*	medical
derog.	derogatory	*meteor.*	meteorology
did.	didactic	*mil.*	military
dign.	dignified	*mod.*	modern
euphem.	euphemism	*mus.*	music(al)
exagg.	exaggeration	*myth.*	mythol./myth.
fig.	figurative	*obs.*	obsolete
gen.	general	*obsc.*	obscene
geom.	geometry	*ornith.*	ornithology
hort.	horticulture	*phys.*	physiology
id.	idiomatic	*poet.*	poetry/poetical
inf.	informal	*pr.n.*	praise name
joc.	jocular	*vulg.*	vulgar
jur.	juridicial	*zool.*	zoology
ling.	linguistics		

Figure 4: Usage labels

Obviously any of these labels can be summoned to create a report for some kind of semantic or stylistic investigation, e.g. to create a list of zoological

terms.

F9, *RENDERING*, contains the English ready equivalents. Different senses of a lemma are not structured into different fields of sub-fields of a record. Rather, each sense is numbered and accommodated in a separate record. The sorting procedure will, during compilation, sequence these senses in succession in the article. The following extract from an article is, for instance, compiled out of three records:

> **!gãũ**ǀ {**!gâu**} *v.t/i* 1 cross (*e.g.* river, street); go across (*e.g.* field); s.a. ǂHĺĺ; 2 *fig.* bec. drunk/intoxicated/inebriated, s.a. ǀHÒRŐ5; 3 ferment.

Should one, for whatever purposes, wish to extract a report on Khoekhoe words with multiple senses, one can simply submit a query searching for the figure "1" in F9.

F10, *SCIENTIFIC*, contains the scientific names for zoological, ornithological or botanical words.

F11, *ADDitional COMMENT*, provides a slot for encyclopedic information, for instance about cultural practices.

F12, *X-REFERENCE*, accommodates cross-references. More will be said about this later.

F13, *LOAN*, provides sources of loan words, occasionally also recipients of Khoekhoe words, or it draws attention to calques; e.g.

> **pűrúkhòè**b … (< Dutch pl. broeken),[3] or
> **ànídànà** … (cf. Afr. hoenderkop).

Needless to say, as a spin-off this field provides a rich source of material for a discussion of language contact or loaning. F13 is the last field that provides text for the printing template.

F14, *SOURCE*, merely serves for "household" purposes, as it refers to the source of a lemma for purposes of verification. This field, like all subsequent fields, is not printed.

F15, *TONE CONSTITuents*, provides information in numerical form on the underlying tone of lemmas, as well as on tone rules that apply to a compound lemma. This field is of utmost importance to the data base and has provided the tool for analysing the tonological system of Khoekhoegowab, as will be further discussed below.

F16, *ENGLish ENTRY*, provides a hand-picked selection of key words from the English renderings provided in F9. Each of these entries in F16 will serve for the reversal of Khoekhoe–English to English–Khoekhoe in the compilation program, that is, each respective English word in F16 will be paired with the Khoekhoe contents of F2, *ENTRY* to form a lemma in the English–Khoekhoe Index. The contents of F16 will thus not appear in print in the Khoekhoe–English *Dictionary*, only in the English–Khoekhoe *Index* or *Glossary* as separate entries in alphabetically ordered sequence.

F17, *NOTES?*, is a general purpose household field in which the editor can *ad libitum* provide for the identification of various types of data, but all represented by a preordained code of maximally two characters. The metalinguistic notes that were made mostly convey linguistic or cultural information in order to allow the extraction of relevant data. This field also contains notes concerning the stage of editing, e.g., if a term needed further investigation or discussion with certain key consultants.

The following list of codes (Figure 5) gives an idea what kind of information can be retrieved at the present stage (the choice of mnemonic label is trivial). As said, further codes can be added *ad libitum* in this field in future, as they do not affect the basic configuration of the database:

Phonotactic information:

s	segmental peculiarities/variants
v	variants (regional or dialectal), e.g. *!a**a**b/!a**e**b* (river)
r	historical reduction (by elision of segments, esp. CV**C**V > CVV)
i	insertion (epenthesis, vowel anticipation, inversion), e.g. *horaga/ ho**a**raga* (whole)
3	trisyllabic root

Tonological information:

f	derivation of deverbal noun by means of flip-flop

Lexical information:

p	incorporation (of object noun into verb)
o	incorporation (of postposition/postpositional phrase into verb)
d	ideophone
m	idiomatic usage
g	"ergative" pair

Editing tasks:

?	problematic, needs further attention
z	scientific identification required
=	re-checking completed
x, e	manual editing required in Word Perfect file
T	enquire with Topnaars

Figure 5: Labels for household purposes

F18, *STAGE/DATE* provides the month and year when a record was discussed by the team. The date is inserted manually, as this field also provides for some codes indicating the stage of completion. These codes, as also the relevant codes of F17, allowed final rounds of editing to deal with remaining problems.

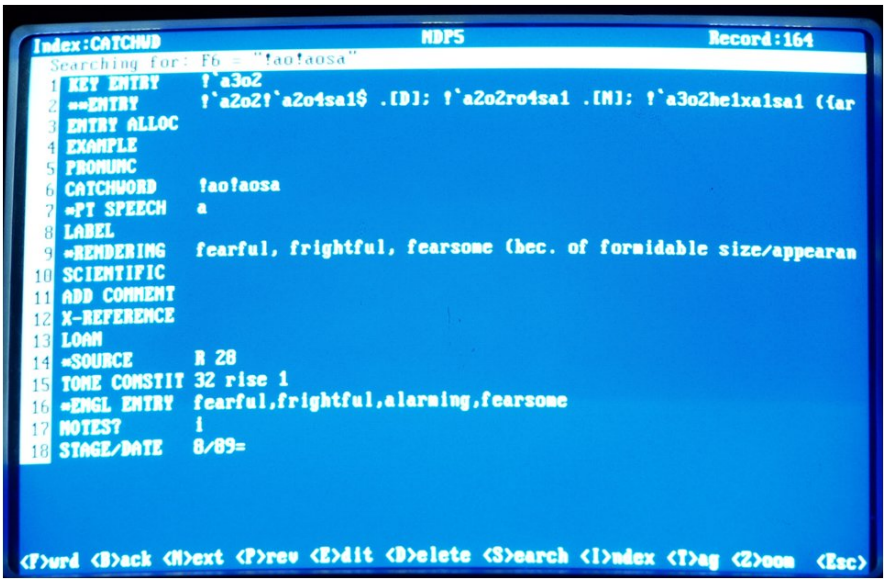


Figure 6: Photo of a Record as displayed on the *File Express* Database Screen

Figure 6 presents an example of a record as displayed in the *NDP5* configuration of *File Express* followed by Figure 7 displaying an extract of the finally compiled dictionary with the same record (underlined) in context:

intimidation; **!àò!àò-àò**b/s, **~rò-àò**b/s {**!ao!ao-ao**b/s, **~ro-ao**b/s} *n.* intimidator; s.o. who frightens/*etc.*; **!àò!àò(hè)sà**, **~rò(hè)sà** {**!ao!ao(he)-sa**, **~ro(he)sa**} *a.* (wh. h.b.) frightened/scared; **!àò!áósà** [D]; **!àòrósà** [N]; **!áòhèxàsà** (arch.) {**!ao!aosa**; **!aorosa**; **!aohexasa**} *a.* fearful, frightful, fearsome (bec. of formidable size/appearance); alarming; **!àò!áó-sàsì**b [D]; **!àòrósàsì**b [N]; **!áò(hè)sàsì**b {**!ao!aosasi**b; **!aorosasi**b; **!ao(he)sasi**b} *n.* fearfulness, frightfulness, fearsomeness; ...

Figure 7: The same Record (Figure 6) converted by the Compilation Software into a Lemma within an Article

4. The Benefits of a Database

In this section some of the major benefits of the Khoekhoe database are introduced.

4.1 The Query System

FE uses an extremely user-friendly and intuitive syntax for queries. There is no need to use standard SQL. It is a conversational program as natural language can be used to a large extent, e.g. "F1 is not empty" (which would yield all records that are not sub-lemmas, thus main lemmas). The beginning of the query is provided on screen: "*Find all records where:*" (for finding individual records) or "*Print all records where:*" (for producing reports) and the user must complete it with a minimal adherence to formulas, e.g. "F8 = z" or "F8 is z" would yield all records that are labelled as zoological terms.[4]

The power of database queries, as opposed to linear word processor searches, lies in the possibility to combine searches in *multiple* fields; e.g. *Print all records where:* "F6 = "..si" and F8 = "t" and F8 is not "o".[5] This query would generate a report listing all Khoekhoe transitive verbs (in the database) ending with the suffix "*-si*" that are not marked as "obsolete". The example should convey an impression what a powerful resource for especially but not only morphological investigation the database is. It is possible to combine more than two fields in a complex query.

4.2 Instances of further data utilisation of the *NDP*

Data can be selectively utilised for all sorts of publications on specific domains, ranging from smaller articles to books and different versions of the *KhD*. Below (4.2.2) it will be expounded on how the present Khoekhoegowab-oriented database has already been used for three dictionaries/glossaries, each with reversal of source and target languages. Currently the database is being extended to some endangered dialects for a second, enlarged edition (4.2.3).

An instance of a smaller publication was a list of some 350 botanical names: *A preliminary list of Khoekhoe (Nama/Damara) plant names* (Eiseb, Giess and Haacke 1991). Such intermediary publications may furthermore be useful for satisfying funding agencies that a project produces tangible results, even if the final dictionary is still years away.

Lexicographic databases are attractive to builders of computer applications of various kinds. The *NDP5* database is currently used by a postgraduate student in South Africa to extract morphological data for a Master's thesis on the development of a morphological parser for Khoekhoegowab.[6]

A Khoekhoe spell checker is an obvious utility that should emanate from this electronic database. At this stage, however, there are no substantive agreements with experts yet.

Plans are under way to use the *NDP5* database extensively for writing a reference grammar of Khoekhoegowab. The *NDP5* database provides the most comprehensive Khoekhoe corpus on morphological information that exists. Data have been extracted for university study guides on Khoekhoegowab grammar, and have been supplied to other researchers for their purposes.

4.2.1 Tonological Analysis of Khoekhoegowab

The *KhD* was to be marked for tone, as Khoekhoegowab is a fully-fledged tone language. Impressionistic annotation of words with diacritical tone marks, was out of the question, however, as a tonological system is as systematic and rule governed as the segmental phonology of a language. Hence a systematic investigation of the tonology of Khoekhoe was a prerequisite for marking tone in the dictionary. On the other hand, the lexicon of the database was to provide the data for such a study. It is for this reason that the field F15, *TONE CONSTITuents*, was introduced. This field, which is not printed, lists the underlying lexical tone melodies that serve as input to compound lemmas in F2, ENTRY; e.g. Figure 8:

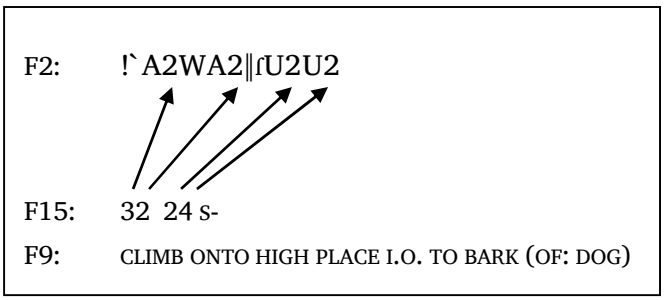


Figure 8: Tonal Input (F15) and Output (F2) for *!àwà ǁhùù*

The four tonemes that were postulated for Khoekhoegowab are marked in the *FE* database by Arabic numbers from 1 to 4 for the lowest to the highest toneme respectively, and the tone number follows on the t.b.u. (as ASCII codes do not provide for diacritics). F15 in the above example (Figure 8) indicates that the two disyllabic roots *!áwà* (climb) and *ǁhùúí* (bark) underlyingly have the bimoraic tone melodies 32 and 24 respectively. The codes "s" and "-" in F15 respectively indicate that in the compound as provided in F2 the first root undergoes a "**s**witching" rule (better known as "flip-flop", by analogy to Chinese), and that the second root changes from the citation melody (+) to a sandhi melody (-).[7] The correlation of the tonal input and output information in F15 and F2 was instrumental for the analysis of the lexical tonology by way of a doctoral thesis (Haacke 1999). No entry in F2 was marked for tone unless the relevant rule application was understood. In this way the reliability of the tonal marking was significantly enhanced. The metalinguistic codes for tone rules in F15 furthermore permit queries that provide all instances of a particular rule application (or combinations thereof) contained in the database. Such a retrieval is not possible with a word processor.

The device to enter tones as numeric characters in the database moreover

allows for sorting of homographs in consistent tonal sequences. In the compilation program a specific routine moves all numeric characters to the end of the word, before character-by-character sorting is done, e.g. *o2ra2* > *ora**22*** and *o1ra3* > *ora**13***. This causes *o1ra3* (***òrá*** "eat raw") to be placed before *o2ra2* (***òrà*** "raw"). After sorting, the conversion to diacritics is done on the original entry.

Again, the possibility to combine different fields in one query was of utmost importance in the investigation of tonology, so as to establish in optimal detail the interrelation between the segmental and suprasegmental phonology. A query like

"F2 = "m.." and F15 = "4..""

would, for instance, generate a report listing all lemmas that have "m" as initial consonant/segment and a "double high" toneme "4" as first tone, i.e. on the t.b.u. following the specified first consonant; cf. Figure 9:

Find all records where:
F2 = "m.." and F15 = "4.."
i.e.
Find all records where:
the first segment is "m" and the first tone is "4".

Figure 9: Correlating Segmental with Tonal Data to establish Depressor Consonants

The following Table from Haacke (1999: 56), Figure 10, presents a synoptic overview of the quantitative distribution of C_1 of roots with the six major tonal melodies (listed in the top line) of Khoekhoegowab in the 2 150 roots that were in the database at that stage. This list conveys crucial data of the phenomenon that Khoekhoegowab has depressor consonants that lower existing tonal melodies: The depressor consonant ***h***, ***m***, ***kh***, ***ts*** and the click (X) releases X***h***, and X***n*** depress the tonal melodies /22/ and /32/ to /13/ and /12/ respectively (instances underlined). This phenomenon of pitch-evolution through phonetically conditioned tone splits is known as tonogenesis — a phenomenon prevalent in *i.a.* Sino-Tibetan languages. The present database was instrumental in investigating to what extent tonogenesis occurs in Khoekhoegowab (and, hence, is likely to be found in related Khoe languages); it allows retrieval of every supporting example but also every counter example, and the record counting facility allows for statistical assessment of the frequency of instances.

	/12/	/13/	/22/	/32/	/24/	/43/	Total
b/p	0	7	1	0	2	1	*11*
d/t	12	16	2	8	12	5	*55*
g/k	12	15	6	9	12	10	*64*
ʔ(V)	3	13	11	15	25	13	*80*
h	**5**	**9**	**0**	**0**	6	7	*27*
m	**3**	**3**	**0**	**0**	4	3	*13*
n	5	5	5	2	12	4	*33*
s	**1**	**12**	**12**	**14**	21	13	*73*
x	**0**	**5**	**3**	**14**	19	14	*55*
kh	**14**	**17**	**0**	**0**	3	7	*41*
ts	**19**	**10**	**0**	**0**	14	8	*51*
Total:	*74*	*112*	*40*	*62*	*130*	*85*	***503***
ǀ	**0**	**6**	**20**	**21**	36	17	*100*
ǀg	22	16	3	12	15	12	*80*
ǀh	**31**	**24**	**1**	**2**	26	16	*100*
ǀkh	**0**	**4**	**8**	**15**	26	23	*76*
ǀn	**8**	**14**	**0**	**1**	15	8	*46*
Total:	*61*	*64*	*32*	*51*	*118*	*76*	***402***
ǁ	**0**	**6**	**16**	**21**	29	15	*87*
ǁg	14	18	10	15	24	12	*93*
ǁh	**19**	**23**	**0**	**0**	17	7	*66*
ǁkh	**0**	**1**	**16**	**17**	24	15	*73*
ǁn	***14***	***14***	**1**	**5**	21	13	*68*
Total:	*47*	*62*	*43*	*58*	*115*	*62*	***387***
!	**1**	**10**	**21**	**25**	34	9	*100*
!g	25	28	9	18	23	14	*117*
!h	**28**	**26**	**1**	**3**	16	19	*93*
!kh	**0**	**4**	**12**	**22**	30	15	*83*
!n	**25**	**24**	**1**	**11**	30	17	*108*
Total:	*79*	*92*	*44*	*79*	*133*	*74*	***501***
ǂ	**0**	**8**	**13**	**16**	28	11	*76*
ǂg	13	18	12	9	21	13	*86*
ǂh	**25**	**18**	**0**	**2**	14	18	*77*
ǂkh	**0**	**0**	**21**	**14**	15	6	*56*
ǂn	**17**	**15**	**1**	**3**	15	11	*62*
Total:	*55*	*59*	*47*	*44*	*93*	*59*	***357***
Grand Total:	***317***	***389***	***206***	***294***	***588***	***356***	***2 150***
Percentage:	**14,7**	**18,1**	**9,6**	**13,7**	**27,3**	**16,5**	**100%**
		27,7					
		28,4			**		

Figure 10: Distribution of C_1 with the Six Major Tonal Melodies in Disyllabic Skeleta

4.2.2 Tagging of Records for producing different versions of the dictionary

Databases are re-usable for other kinds of dictionaries or reference works. This possibility is enhanced by *FE*'s facility to tag records, i.e. to earmark them in an *ad hoc* way for a particular purpose. This facility was used for extracting the 10 150 records that have served as input to the Khoekhoegowab–English part of the *Khoekhoegowab–English/English–Khoekhoegowab Glossary/Mîdi Saogub* (Haacke and Eiseb 1999), and subsequently to the *Khoekhoegowab–Afrikaans Afrikaans–Khoekhoegowab Glossarium/Mîdi Saogub* (Haacke, Eiseb and Gericke 2010).

These glossaries are important by-products that had not been planned for initially, but could be created thanks to the software design. When the first trial page of the *KhD* in book format was produced by means of the new software, co-author Eliphas Eiseb was of the opinion that the Khoekhoe speakers would not use the dictionary, as it was too complicated. The tone marking with diacritics, and also the arrangement of lemmas in articles would be disconcerting for users with limited referencing skills. The project leader took heed of his opinion and requested the programmer, Jonathan Kaye to write another compilation program that would produce a Khoekhoe–English glossary without tone marking, using just the standard orthography. This book is now used in schools by MT-speakers. This solution of providing for two books derived from one database satisfactorily addresses the rather ambitious if not unrealistic aim to cater for a wide-fanned spectrum of users ranging from moderately literate MT-speakers to academics.

A further by-product that resulted from the tagged records via the extracted database of the *Glossary* is the *Khoekhoegowab–Afrikaans Afrikaans–Khoekhoegowab Glossarium/Mîdi Saogub*, which was commissioned by PANSALB (*Pan South African Language Board*) for use by the Khoekhoe descendants in South Africa. The commission was to replace English with Afrikaans. As by then contact with Jonathan Kaye had been lost, who had moved on from London to China, the compilers were obliged to use the existing database configuration, since the compilation software is linked to specific fields of it. Hence improvisation was called for:

The database with its 18 fields was cloned and

— the English data of F9 RENDERING transferred to the non-printable F11, which in the original database has the ADDitional COMMENT;

— the data of the original F16 ENGL ENTRY transferred to the non-printable F15, which originally has the TONE CONSTITuents.

This freed F9 for entering the AFR RENDERING and F16 for the AFR ENTRY to accommodate the key words for the Afrikaans–Khoekhoe reversal. The contents of the original F11 and F16 could be dispensed with in the compilation of the *Glossarium*, as this information was not needed. As the *Glossarium* was intended for a target group that wants to re-acquire the Khoekhoe language, it

was imperative, however, that tonal information was supplied somewhere. This required that the tonal decodification was re-activated in the macro that produces the Khoekhoe–Afrikaans part of the *Glossarium*, while the Afrikaans–Khoekhoegowab part was kept in the standard orthography without tone marking, as in the original *Khoekhoegowab–English Glossary.*

-field name-	-type-	-length-	-format-
1 KEY ENTRY	C	15	Character
2 ENTRY	**C**	**80**	**Character**
3 ENTRY ALLOC	C	30	Character
4 EXAMPLE	**C**	**50**	**Character**
5 PRONUNC	**C**	**50**	**Character**
6 CATCHWORD	C	22	Character
7 PT SPEECH	**C**	**2**	**Character**
8 LABEL	**C**	**2**	**Character**
9 AFR RENDERING	**C**	**130**	**Character**
10 SCIENTIFIC	**C**	**40**	**Character**
11 RENDERING	C	130	Character
12 X-REFERENCE	**C**	**33**	**Character**
13 LOAN	**C**	**32**	**Character**
14 SOURCE	C	4	Character
15 ENGL ENTRY	C	16	Character
16 AFR ENTRY	C	45	Character
17 NOTES?	C	2	Character
18 STAGE/DATE	C	6	Character

Figure 11: Database Modification for the *Glossarium*

4.2.3 New Uses: Updating and Expansion of Dialect Data

Haacke has for several years now been working on eliciting dialect equivalents for over 6 785 English lemmas of the *Glossary.* The dialects concerned are endangered dialects on the northern and northwestern periphery of Khoekhoegowab, viz. ǂĀkhoe and Haiǁom, as well as Sesfontein Damara in Kaokoland. These dialect variants, which are to be integrated into the main database, are to be added to a second, enlarged edition of the *KhD*. Obviously such data might also find their way into further publications like a comparative dictionary.

4.3 Editorial Advantages of the Database

The editorial advantages of a database are immeasurable. It is hard to say whether the main *KhD* with over 24 000 entries would ever have been published if no dedicated software had become available. The reversal of the Khoekhoe–English dictionary to English–Khoekhoe — albeit in *Index* form — would, because of time constraints, not have taken place. For certain, the late co-author Eliphas Eiseb would never have seen the fruit of his life's vocation, and certainly neither of the two glossaries would ever have appeared.

The editorial advantages of a database can only be identified very briefly here:

4.3.1 Reversal of Source and Target Language

The automatic reversal of source and target language, i.e. in our case Khoekhoegowab–English to English–Khoekhoegowab, obviously is one of the most compelling reasons for using a Dictionary Writing System. The reversal lists the English (or Afrikaans) lemmas that were identified in F16, in alphabetical order with the entire contents of F2 ENTRY or F4 EXAMPLE attached, whichever the case may be. As there will be numerous repetitions in F16 of specific English lemmas, but with different associations of F2 or F4, careful editing is required to conflate such multiple entries wherever the meaning warrants it. As this automatic reversal is bound to be riddled with lacunae of English concepts that do not occur in the English rendering (F9, F16) of Khoekhoe concepts — for no Khoekhoe terms for English concepts were coined by the authors, it is called an *Index*. Ideally users should follow an onomasiological approach by also consulting the Khoekhoe–English *Dictionary* for each English lemma, to get a more contextualised understanding of the Khoekhoe word.

4.3.2 Automatic Conversion from Tonal to Standard Orthography

As it is an essential feature of the database that all lemmas are marked for tone, the standard orthography had to be systematically adapted so as to provide the required tone-bearing units by replacing single vowels marked with a macron for length, with two identical vowels (as actually is historically correct, moreover). Each vowel, being a syllable peak, then bears a tone mark. The compilation software for the *KhD* automatically provides the version in standard spelling without tone marks after the entry with tone marks, indicated by curly brackets; e.g.

áàb/-i {**ā**b/-i} *n.* drink, beverage.

This was considered an essential aid for the less versatile user. In the *Glossary* the compilation software replaces the tone-marked entry and provides solely

the versions in standard orthography, both, in the Khoekhoegowab–English part and the English–Khoekhoegowab part, e.g.

āb/-i *n.* drink, beverage.

These automatisation processes are a very significant editorial aid in that they not only save immense amounts of time but also preclude human error in transcribing.

4.3.3 Speeding up editing and proofing

Using a database is a pivotal device of drastically reducing human error and inconsistency in editing.

A typographical error in spelling or tone assignment in the raw corpus data will more often than not be revealed by a resultant missort of the record in the database, which by itself usually is conspicuous. Proofing of spelling thus happens *en passant* over the years of data compilation and should have been largely concluded by the time the dictionary compilation stage is initiated.

Checking of editorial matters like bolding, italicisation, metalinguistic labels and abbreviations, numbering of sub-senses or of homographs, can be done systematically by filtering out the form fields (or spreadsheet columns) concerned. The fact that the preordained codes in the database are automatically replaced by the metalinguistic labels and abbreviations in the dictionary, already ensures consistency, unless a faulty code was typed for a start and hence will not convert. As *FE* does not allow selective use of font styles or diacritics and is limited to ASCII type characters, such editorial choices have to be configured by preordained coding in the database, so that they can be automatically converted by the writing program during the dictionary compilation phase. Italics, for instance, were indicated in the database by curly brackets.

Pre-dictionary database input:	
F10 SCIENTIFIC:	{Olea europea} subsp. {africana}
Dictionary output:	
	Olea europea subsp. *africana*

Figure 12: Italics in Scientific Names

The use of a database form will ensure adherence to the style manual and guide the analysis of a record; in short, it will significantly enhance data integrity. Detecting all such inconsistencies and omissions in dictionary articles that

were drafted with a word processor is a Sisyphean task indeed. When I once replied to the question how often I had proofread the print-version of the *KhD*, that I had proofread it only once because of time constraints, I was looked at silently in disbelief. I had to explain that the use of a database that had been fine-tuned over years had largely obviated proofreading of the compiled dictionary file, other than for checking that the compilation software had not slipped up on editorial matters.

4.4 Limitations of the *NDP5* Dictionary Writing System

It should be emphasised at the outset that the DWS as custom-designed by Jonathan Kaye in the early 90s was never intended as a commercial application. It was designed exclusively to meet my requirements as specified earlier, at a time when no commercially available software would have met my particular requirements. So there was never an intention to make it commercially viable. However, for the sake of objectivity a superficial comparison to current state-of-the-art expectations of DWSs as set out by, for instance, Joffe and De Schryver (2004) and by De Schryver and De Pauw (2007) may be opportune.

It was stated before that the software does not provide for a Corpus Query Package as at the time of its creation electronic text corpora hardly existed in Khoekhoegowab.

Most outstanding is the absolute separation of the pre-dictionary database stage and the dictionary compilation phase. While working in the database there is no possibility to obtain a WYSIWIG preview of a lemma as it will appear in print.

The software does not provide for automatic tracking and updating of cross-references. Checking that all cross-reference links are correct — let alone exist, consumes a considerable amount of time and invites human error. The *KhD* contains 3 706 cross-references. They needed manual checking of the paired records — a task best done on two PCs standing next to each other.

The software thus provides no automated controls for editing and consistency checks. Yet — as pointed out above — the actual editing is significantly facilitated by taking place in the database instead of in the final dictionary text, as would be the case when using a word processor.

Having pointed out the above limitations I wish to come to the cardinal advantage of the ***NDP*** software: that the ageing legacy software of the early 1990s has been replaced by open source software that should ensure the long-term survival of the primary database, while essentially still following the same basic concepts of the original database configuration and compilation software.

5. Conclusion: Transition to Open Source Software

In 2001 a contractual agreement was set up between PANSALB, the Publisher

and Haacke that the *Khoekhoegowab–English/English–Khoekhoegowab Glossary/Mîdi Saogub* should be converted to a *Khoekhoegowab–Afrikaans Afrikaans–Khoekhoegowab Glossarium/Mîdi Saogub* (see above). When in 2007 the project was ready to embark on the Dictionary Compilation Stage so as to convert the modified database into print-ready format, disaster struck: The custom-made software failed to perform because of its dependence on outmoded operating systems. Unless we could get it to work, the project was doomed to failure and the converted database of no use. I was fortunate to re-establish contact with Jonathan Kaye, who by then had abandoned Windows for Linux (Debian). In an almost daily shuttle of files per email over three months he admirably converted the DOS database into the camera-ready Word Perfect dictionary format in a multitude of ad hoc measures employing LINUX, thanks to his advanced programming skills.

The all too obvious lesson of the crisis was that the data had to be rescued into non-proprietary software before it would become inaccessible through the advance of Microsoft operating systems. Again Jonathan Kaye came to the rescue. He transferred the *File Express* database to an *OpenOffice CALC* spreadsheet and replaced the dictionary compilation procedures that were written with the aid of Icon and Word Perfect facilities originally, by formulas and macros that operate within CALC. ASCII characters were replaced by Unicode characters, which required that the Khoekhoe data are encoded differently now. In order to add and/or edit data for a next publication, as well as to retrieve data (the processes formerly done in *File Express*) the data have to be transferred from the *OpenOffice CALC* spreadsheet into an *OpenOffice BASE* database. *BASE* is a graphical front-end for accessing databases. It allows the execution of queries and reports, and the editing and adding of data by using a customised Form as user-interface. Forms are data input and output masks and are the actual editing tool. Once the pre-dictionary data compilation phase has eventually been completed, the data has to be transferred back into the *CALC* spreadsheet in order to compile the *Khoekhoegowab–English Dictionary* and the *English–Khoekhoegowab Index* in print-ready format.

The conversion of the database and dictionary compilation software from legacy formats to robust non-proprietary software subject to a GNU General Public License, thus has made possible if not ensured the survival of the only lexical corpus in database format of a Khoesaan language after the demise of the compilers. Apart from ensuring that more publications can be generated by the present proprietor of the database while it is not public domain as yet, the way is open to eventually subscribe to "best practice" aspirations. "Best practices", according to the E-MELD School of Best Practises, are "practices which are intended to make digital language documentation optimally long-lasting, accessible, and re-usable by other linguists and speakers".[8] This, to the best of my knowledge, is a challenge that state-of-the-art lexicographic software currently in vogue in South Africa and lexicographic projects in South Africa still have to face.

To conclude: This case history was presented in the hope to show that, firstly, lexicographers in this millennium need to avail themselves to dedicated lexicographic software with complex retrieval facilities, software that allows them to get an optimal return for their investment by ensuring that these usually formidable repositories of knowledge can be tapped not only for producing a dictionary but are reusable for other purposes.

Secondly, lexicographers should take heed of the current trend of following "best practices" in Language Documentation by resorting to open source software so as to ensure optimal survival of the multipurpose database, for coming generations to build on.[9]

Endnotes

1. As recently as 2004 Joffe and De Schryver (2004: 99) stated that "dedicated tools to assist lexicographers are not readily available".
2. *Khoekhoegowab* (literally: "*Khoekhoe*-language") is the revived original name of the language that in the late 19th century became better known as "Nama", and subsequently as "Nama/Damara" (cf. Haacke 2011). This glossonym was officially reinstated through the initiative of Eliphas Eiseb. For the sake of brevity the language is mostly referred to here as *Khoekhoe*.
3. The bolded vowels *oe* of the plural form of *purukh**oe**n*, (pairs of trousers), just as in the loan *rokh**oe**n* (skirts, < Du. *rokk**en***), indicate that these words are not loans from the more recent Afrikaans, but date back to first contacts of Khoekhoe with Europeans. These words in turn were borrowed from Khoekhoegowab by several Bantu languages in Southern Africa.
4. "z" is the metalinguistic code used in the database for the printed abbreviation "*zool.*".
5. As F6 lists lemmas without tone specifications, all relevant lemmas are listed irrespective of their tonal profile; two dots ".." represent a wildcard meaning "anything"; "t" is the metalinguistic code used for "*v.t.*", and "o" is the metalinguistic code for "*obs.*".
6. Hendrina du Plessis *Natural Language Processing of Khoekhoegowab (Previously Nama/Damara).* UNISA.
7. Sandhi: tonal modification by syntactic/morphological context.
8. **E_MELD:** ***E****lectronic* ***M****etastructure for* ***E****ndangered* ***L****anguages* ***D****ata*; http://emeld.org/school/what.html.
9. For an introduction to the relatively new discipline of Language Documentation cf. *i.a.* Himmelmann (2006).

References

De Schryver, G.-M. and G. de Pauw. 2007. Dictionary Writing System (DWS) + Corpus Query Package (CQP): The Case of *TshwaneLex*. *Lexikos* 17: 226-246.

Dickens, P. 1994. *English–Ju|'hoan Ju|'hoan–English Dictionary.* Quellen zur Khoisan-Forschung 8. Cologne: Köppe Verlag.

Eiseb, E., W. Giess and W. Haacke. 1991. A Preliminary List of Khoekhoe (Nama/Damara) Plant Names. *Dinteria* 21: 17-30.

Haacke, W.H.G. 1998. A Khoekhoe Dictionary in the Making: Some Lexicographic Considerations. Schladt, M. (Ed.). 1998. *Language, Identity, and Conceptualization among the Khoisan*: 35-64. Quellen zur Khoisan-Forschung 15. Cologne: Köppe Verlag.

Haacke, W.H.G. 1999. *The Tonology of Khoekhoe (Nama/Damara)*. Quellen zur Khoisan-Forschung 16. Cologne: Köppe Verlag.

Haacke, W.H.G. 2011. *Nama* als Sprachbenennung in der Koloniallinguistik Deutsch-Südwestafrikas: zwischen Endonym und Exonym. Stolz, Th., Chr. Vossmann and B. Dewein (Eds.). 2011. *Kolonialzeitliche Sprachforschung. Die Beschreibung afrikanischer und ozeanischer Sprachen zur Zeit der deutschen Kolonialherrschaft*: 139-160. Berlin: Akademie Verlag.

Haacke, W. and E. Eiseb. 1999. *Khoekhoegowab–English/English–Khoekhoegowab Glossary/Mîdi Saogub.* Windhoek: Gamsberg-Macmillan.

Haacke, W. and E. Eiseb. 2002. *A KHOEKHOEGOWAB DICTIONARY with an English–Khoekhoegowab Index.* Windhoek: Gamsberg-Macmillan.

Haacke, W., E. Eiseb and C. Gericke. 2010. *Khoekhoegowab–Afrikaans Afrikaans–Khoekhoegowab Glossarium/Mîdi Saogub.* Windhoek: Macmillan.

Himmelmann, N.P. 2006. Language documentation: What is it and what is it good for? Gippert, J., N.P. Himmelmann and U. Mosel. 2006. *Essentials of Language Documentation*: 1-30. Berlin: Mouton de Gruyter.

Joffe, D. and G.-M. de Schryver. 2004. Tshwanelex — A State-of-the-Art Dictionary Compilation Program. Williams, Geoffrey and Sandra Vessier (Eds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6–10, 2004*: 99-104. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.

Kilian-Hatz, C. 2003. *Khwe Dictionary. With a Supplement on Khwe Place Names of West Caprivi by Matthias Brenzinger.* Namibian African Studies 7. Cologne: Köppe Verlag.

König, C. and B. Heine. 2008. *A Concise Dictionary of Northwestern !Xun.* Quellen zur Khoisan-Forschung 21. Cologne: Köppe Verlag.

Nakagawa, H., K. Sugawara and J. Tanaka. In preparation. *Ghui Dictionary.*

Traill, A. 1994, 2009. *A !Xóõ Dictionary*. Quellen zur Khoisan-Forschung 9. Cologne: Köppe Verlag.

Van Sterkenburg, P. (Ed.). 2003. *A Practical Guide to Lexicography*. Terminology and Lexicography Research and Practice 6. Amsterdam/Philadelpia: John Benjamins.

Visser, H. 2001. *Naro Dictionary: Naro–English, English–Naro*. Fourth edition. D'Kar: Kuru Development Trust.

Weich, F. 2004. *San Dictionary/San-woordeboek. San–Afrikaans–English/English–San–Afrikaans/Afrikaans–San–English*. Johannesburg: Macmillan South Africa.

ANNA: A Dictionary with a Name (and what Lies Behind it)*

W. Martin, *Department of Language and Communication, VU University Amsterdam, The Netherlands (w.j.r.martin@vu.nl)*

Abstract: In 2011 the *Groot Woordenboek Afrikaans en Nederlands (Large Dictionary Afrikaans and Dutch)*, commonly known as ANNA, appeared. Contrary to so-called *difference dictionaries*, bilingual dictionaries of narrowly related languages which describe *only differences* between the two languages, ANNA describes both differences and similarities between Afrikaans and Dutch, not only on the semantic level but on the combinatorial and pragmatic level as well. In this sense ANNA is a unique project, based on an original *amalgamation* model. In this article first some background information will be given about the ANNA project and its results, followed by a presentation of the underlying model and an evaluation of it.

Keywords: AMALGAMATED DICTIONARY, BILINGUAL DICTIONARY, AMALGAMATION MODEL, AFRIKAANS, DUTCH, ANNA, COGNATES, FALSE FRIENDS, CONTRASTIVE DICTIONARY, READING DICTIONARY, TRANSLATION DICTIONARY

Samenvatting: ANNA: een Woordenboek met een Naam (en wat er achter steekt). In 2011 verscheen het *Groot Woordenboek Afrikaans en Nederlands,* met als roepnaam *ANNA.* In tegenstelling tot een traditioneel tweetalig woordenboek beschrijft ANNA de twee talen (Afrikaans en Nederlands) als één taal en behandelt zij niet alleen de onderlinge verschillen maar ook de (vaak vermeende) gelijkenissen op het vlak van betekenis, combinatoriek en pragmatiek. Dit maakt ANNA tot het eerste ge-amalgameerde tweetalige woordenboek gebaseerd op een origineel *amalgamatiemodel.* In dit artikel wordt allereerst achtergrondinformatie gegeven over het ANNA-project en zijn resultaten. Daarna wordt het onderliggende *amalgamatiemodel* voorgesteld en geëvalueerd.

Sleutelwoorden: AMALGAMATIEMODEL, GEAMALGAMEERD WOORDENBOEK, TWEETALIG WOORDENBOEK, AFRIKAANS, NEDERLANDS, ANNA, COGNATES, VALSE VRIENDEN, CONTRASTIEF WOORDENBOEK, LEESWOORDENBOEK, VERTAALWOORDENBOEK

0. Introduction

In March 2011, the *Groot Woordenboek Afrikaans en Nederlands* (Large Dictionary Afrikaans and Dutch, 2228 pgs.), known as ANNA, the *first amalgamated (bilingual) dictionary* ever, appeared. Since then I have given many lectures, seminars

* Two other articles where ANNA has been presented extensively are: Martin 2012 (in Genis et al. 2012) and Martin (to appear in *Tydskrif vir Nederlands en Afrikaans* in 2012).

and presentations on ANNA. In this contribution then, I will proceed as I have done in most of those speeches up till now: starting with some background and inside information about ANNA which, otherwise, is difficult for readers to get (section 1, *factual* part), to continue with a more *reflective*, *evaluative part* in which most attention will be paid to the model underlying the dictionary (sections 2 and 3).

1. ANNA: Background Information

1.1 What's in a name?

ANNA is a translation or bilingual dictionary Afrikaans–Dutch v.v. which, contrary to what is generally the case, does not treat the two languages as different ones but as the *same.* From this point of view the acronym AN-NA is rather misleading as it suggests that the dictionary consists of two parts, one part Afrikaans–Dutch (AN), — N standing for Nederlands (=Dutch) — and one part Dutch–Afrikaans (NA). ANNA, however, contains only one part, implying that both Afrikaans and Dutch can function in one and the same volume as a source language.

In a traditional bilingual dictionary Afrikaans–Dutch v.v. an entry such as *robot*, for instance, would look as follows:

In the Dutch–Afrikaans part:

robot
(automaat die arbeid verricht) [automaton carrying out work] *robot, blikman*

In the Afrikaans–Dutch part:

robot
(paal met lig wat verkeer reël) [pole with light regulating traffic] *stoplicht, verkeerslicht*
(outomaat wat werk verrig) [automaton carrying out work] *robot*

In ANNA the information from the two parts is brought together, *amalgamated,* resulting in an entry like this:

robot/*robot*
A/N (automaat die arbeid verricht) *robot, blikman*
A (paal met lig wat verkeer reël) stoplicht, verkeerslicht
[in ANNA roman font is used for Dutch; italics for Afrikaans]

As I will deal more extensively with amalgamation in the next part (sections 2 and following), it may suffice for the moment to draw the attention to the fact that *amalgamation* has the following effects:

— First of all, it leads to a (more) direct comparison of the two languages:

one does not have to look up the entry *robot*, for instance, in both parts to get a complete picture of both the differences and the similarities between this word in both Afrikaans and Dutch.

— Secondly, amalgamation also leads to a decrease of redundancy in the description of the two languages: in the case of *robot*, for instance, the common meaning 'automaton' need not be repeated in both parts, one mention now suffices.

— Finally, since amalgamation is a novel approach in lexicography[1], ANNA, as the first amalgamated bilingual dictionary ever, can serve as a kind of litmus test or touchstone for the evaluation of the model itself.

1.2 Facts and Figures

1.2.1 Temporal and financial aspects

The (editing) work on ANNA started in January 2000 and was completed in March 2011. Before that, a pilot-study was undertaken (see Martin, Gouws and Renders 1999) in order to define the project and to investigate its feasibility. Time was underestimated in the pilot as only a period of six years was foreseen. On the other hand, the financial prognosis proved to be rather realistic: the budget was overrun by only 25%. All in all, ANNA was a rather 'cheap' project as the project costs did not exceed 400,000 euros[2]. This financial asset was mainly due to the fact that use could be made of two invaluable pieces of *lexicographical infrastructure,* namely an existing Dutch Database (the *RBN* = Referentie Bestand Nederlands = Reference Database of Dutch) and an editor with reversal function at semantic level, namely *OMBI* (= editor for OMkeerbare BIlinguale Bestanden, editor for Reversible Bilingual Databases)[3]. Indeed, the fact that the RBN could be used provided us already with the macro-and the microstructure of the Dutch side of the N-A part, and, while semantically linking Dutch to Afrikaans by means of OMBI, the A-N part was being constructed for the greater part (see below).

However, the fact that this project left the down-trodden lexicographical paths to follow a model of its own (the amalgamation model) lead to the usual teething troubles and entailed some delay. In fact the project has been developed in four steps/phases:

— step one: the elaboration of the Dutch–Afrikaans part
This part of the project took most of the time (4,5 years: from beginning 2000 to mid-2004), among others, because of the teething problems just mentioned. So, for instance, it took some time before the editors got acquainted with the new model and with working with OMBI. The work was carried out at the University of Stellenbosch in close collaboration with the *VU* University of Amsterdam.

— step two: the elaboration of the Afrikaans–Dutch part
The work for this part was done at Port Elizabeth in close collaboration with Amsterdam. It took 3,5 years (from mid-2004 to end 2007).

— step three: the amalgamation and its editing
This step involved the amalgamation of the two previous parts and its editing; it was mainly carried out in Amsterdam in close collaboration with Port Elizabeth. It took 2 years (2008–2009) to finalise this phase.

— step four: final correction and production phase
The overall correction took one full year (2010) and was mainly carried out in Amsterdam and in Houten near Utrecht (publisher's place).

A couple of words of comments on these aspects may be in order here:

1. Work on this project has been carried out at locations often at a distance of several thousands of miles from each other. Thanks to modern ICT this has not been a major problem.

2. This project has been developed *stepwise.* In doing so, it was important to provide in step 1 for the information needed for steps 2 and 3. In the case of 'robot', for instance, it was necessary to indicate (by means of the (non)-appearance of the marker c2) that items were cognates or not:

   N. robot (= automaat) = A. robot (= outomaat) = c2
   N. stoplicht (= verkeerslicht) = A. robot (= verkeerslig)

Mainly on the basis of this information, OMBI and the amalgamation program, the data needed for step 2 (Afrikaans–Dutch) and step 3 (amalgamation) could be *automatically derived*, so to come to:

(step 2)
robot
1. (verkeerslig) = stoplicht (verkeerslicht)
2. (outomaat) = robot (automaat) = c2

(step 3)
robot/*robot*
A/N 1.(automaat) *robot*
A 2.(*verkeerslig)* stoplicht

See section 2 for more details on the amalgamation.

1.2.2 ANNA: a multinational/multilingual project

ANNA, as is the case with most bilingual dictionary projects, is a *multinational*

project, involving a small team of editors spread over three countries: South Africa, the Netherlands and Flanders. The fact that the editorial staff was relatively small was not a matter of principle but simply a pragmatic fact: it was not easy to find people with the necessary qualifications willing to engage in a long-running project of this nature. On the other hand, the small size of the team (ten members) offered the advantage of greater consistency. The project was lead by an editorial board consisting of five people[4] in which I functioned as the project leader.

ANNA is not only a multinational but also a *multilingual* project. At first sight, there is nothing remarkable in that, given the fact that all bilingual projects are, by definition, concerned with more than one language. However, ANNA, right from the start, has been set up as a project that should supersede the two languages any bilingual dictionary project involves, and in this sense ANNA is more than just *bi-lingual.* This has to do with the infrastructure developed for this project and which is meant to be (re-)usable for other languages than those involved in the language pair at stake (Afrikaans and Dutch). I refer to the section 1.3.4 (on the reasons to start up this project) for further details.

1.2.3 ANNA in numbers

ANNA counts 2228 pages and has a weight of 2,2 kg. Next to this 'heavy' print version there is also a 'light' version of ANNA in the form of a CD-ROM[5]. The latter, in fact, does not differ from the former as to contents, but offers, of course, the usual facilities and advantages electronic dictionaries have over their print companions.

The book contains 59051 dictionary entries among which there are:

— 29840 *cognates* or *combination words*
These are words which have the same form in Dutch and Afrikaans and, at least one common meaning (for 'sameness' of form see below, section 2.2).
Examples: tafel/*tafel* [E. table], ontsnappen/*ontsnap* [E. escape], gel/*jel* [E. gel].

— 15011 *unique Afrikaans items*
These are words that only occur in Afrikaans. As a rule they have in Dutch a non-cognate translation equivalent or a paraphrase.
Examples: *trapsoetjies* [N. kameleon, E. chameleon], *verkleurmannetjie* [N. kameleon, E. chameleon], *suurlemoen* [N. citroen, E. lemon].

— 14200 *unique Dutch items*
These are items that only occur in Dutch, the counterparts of the unique Afrikaans items.
Examples: citroen [A. *suurlemoen*], kameleon [A. *verkleurmannetjie, trapsoetjies*], giraf [A. *kameelperd,* E. giraffe].

As one will have noticed, on a total number of about 45000 items both in Afrikaans and in Dutch, nearly two thirds of them are *cognates,* which means that both languages have a large common core and two exclusive parts which balance each other. However, this common core of cognates does not exclude mutual differences. 'Robot', for instance, is both similar and different in Afrikaans and Dutch. Such cognates are called *partial cognates* or *partial false friends:* they share the same form, have at least one common meaning but also at least one different one.

Absolute false friends, which share form only but no meanings, such as 'amper' [A. = E. nearly ; N. = E. hardly] or neuken/*neuk* [A. = E. beat; N. = E. fuck] also occur. Although there are not many of them (about 500), they are, indeed rather striking because they lead to misunderstandings, therefore in ANNA they have a special marker (!!).

The number of meanings or senses in ANNA amounts to 73619 of which there are:

— 36311 cognate or common ones

— 18049 unique Afrikaans ones

— 19259 unique Dutch ones

The above numbers confirm the image of a large degree of formal and semantic cognateness (2/3) between Afrikaans and Dutch. The greater amount in forms in Afrikaans is balanced by the greater amount of unique meanings in Dutch.

Finally, ANNA contains 90008 examples/combinations from which there are:

— 64196 contrastively relevant

— 25812 not contrastively relevant
Combinations were considered contrastively relevant

— either when there was a difference in the combination words in both languages;

— or when the translation of the entry word differed from the translation equivalents given;

— or when both these criteria were met.

This number should be interpreted against a functional background: not contrastively relevant combinations are only taken up when they have an illustrative and/or discriminatory function (which is the case with polysemes for instance), contrastively combinations, on the other hand, are always taken up because of their role in the understanding or production process of a foreign language. For examples see the appendix where contrastive combinations are marked by ≠, whereas non contrastive ones are marked by =.

1.3 To start or not to start (a bilingual dictionary project)

An important question at the beginning of any dictionary project and, in particular, a long term and rather expensive one, is whether there are good reasons to start up the project at all.

Understandably, this question was raised at the beginning of the ANNA project and in the pilot study mentioned above. There we came to the conclusion that, as a rule, for a bilingual dictionary project four arguments could be taken into account to start or not to start the project. In the next sections these arguments will be dealt with.

1.3.1 The communicative argument

If (many) speakers of two different language communities for economical, cultural, political or any other reason, often come into contact with each other, then a bilingual dictionary can be called upon as an interlingual instrument[6] to facilitate the communication between the two groups. However, given the fact that speakers of Afrikaans and Dutch can communicate with each other, each of them using his/her mother tongue, be it with the inevitable misunderstandings, miscommunications and problems (see, among others, the false-friends-cases), the communicative argument in itself is *not a sufficient argument*, in the case of Afrikaans and Dutch, to start up such a project[7]. On the other hand, there are at least *three good reasons* that do apply in the case of Afrikaans and Dutch: the functional, the descriptive and the infrastructural argument.

1.3.2 The functional argument

If one accepts that language is a vehicle not only for basic communication, but also one to properly and fully express oneself in, be it in literature, in science or in everyday situations, then a bilingual dictionary is an important instrument to understand/express the subtleties and nuances of the other language, the other culture. Before the appearance of ANNA however, only small or mini-dictionaries existed between Afrikaans and Dutch such as: Dekker and Paardekooper (1990), Prisma Miniwoordenboek (2004) and Veltkamp-Visser (1998, 5th edition). So, in order to carry out important language functions such as speaking, understanding, translating and learning the foreign language at an advanced level, a large dictionary was needed and ANNA could fill that gap.

1.3.3 The descriptive argument

With most bilingual dictionary projects the 'gap-in-the-market'-argument suffices. In the case of Afrikaans and Dutch one very specific other argument can be added. Because of the fact that Afrikaans is strongly indebted to Dutch —

up till 1925 Dutch was still an official language in South Africa — an empirically based *confrontation/comparison* with Dutch could *free* Afrikaans and Afrikaans monolingual dictionaries, further from Dutch and Dutchisms. With this I mean words and meanings that still apply in Dutch but no longer do in Afrikaans, such as, for instance, the word 'aardig' that in HAT4 (*Verklarende Handwoordeboek van die Afrikaanse Taal*, Explanatory Desk Dictionary for Afrikaans) still gets as one of its meanings 'aantrekkelik, aangenaam' [E. nice], which has since long passed out of use in Afrikaans where it has been replaced by 'gaaf' or 'nice'.

The paradox of ANNA is that she can make speakers of Afrikaans, by confronting them with Dutch, much more conscious of the existence of 'Dutchisms' (also see Houwelingen and Carstens 1998).

1.3.4 The infrastructural argument

From the beginning ANNA has been set up as a project that should supersede the two languages in question. In other words, the aim was not only to produce a contrastive dictionary Afrikaans–Dutch, but also to lay the foundation for an *exportable model,* one that could be used for other closely related languages, such as the 'black' languages in South-Africa: Xhosa and Zulu, and North-Sotho, South-Sotho and Tswana etc.

The infrastructural argument, i.e. the wish to provide for *dictionary technology* that could lead to a 'new lexicography in South-Africa' (see Martin 2005), thus played an important role in setting up the ANNA project, and within this setup the amalgamation model in its turn took a central position. In the next section we will focus on this aspect by trying to give an answer to the following questions:

What is the amalgamation model?

When can it be used?

What does it look like in actual practice?

2. The amalgamation model

2.1 What is the amalgamation model?

The amalgamation model is not an explanatory model but a descriptive one. It aims to describe the lexemes of two languages in a bilingual dictionary in a *directly contrastive way*, contrary to the *indirect way* found in a traditional bilingual dictionary. In this sense the model is innovative, a real novelty in lexicography. As a rule however, there is no innovation at will: I had to see a couple of hundreds of Dutch–Afrikaans items made in a traditional way first, to fully understand that this was not the way to proceed and only thereafter I was able

to design the model. Moreover, when I designed it, it was my aim not to draft a model that could apply for Afrikaans and Dutch only, but one that would be *generic*, at least generic for any pair of closely related languages (also see Martin and Gouws 2000 where the model is mentioned for the first time). What this means is explained in the following section.

2.2 Scope of the amalgamation model

As stated above, although the amalgamation model is not bound to the language pair Afrikaans and Dutch, it can only be applied to closely related languages. In order to successfully qualify as a pair of closely related languages the following 'operational' qualitative and quantitative criteria need to be met:

— Qualitative aspects: both the 'form' of the words (spelling) needs to be the 'same' and at least one of the meanings.

— Quantitative aspects: not all words from the two languages need to show the above characteristics; however, in order for the model to be applied successfully, there has to be a sufficient critical mass.

In what follows these characteristics are dealt with in more detail.

A. 'Sameness' of form

'Same' items in this context cover three groups:

— Items with a fully identical spelling form both in Afrikaans and Dutch such as 'tafel' [E. table].

— Items with a small, systematic spelling or morphological difference in the two languages such as N. 'zalm'/A. 'salm' [E. salmon] or N. 'ontsnappen'/A. 'ontsnap' [E. escape].

— Items with a bigger, non-systematic difference between Afrikaans and Dutch, but which are still recognizably similar in form, such as N. 'gel'/A. 'jel' [E. gel] or N. 'pompoen'/A. 'pampoen' [E. pumpkin] or N. 'penguïn'/A. 'pikkewyn' [E. penguin].

Although this last group may give rise to different interpretations, the pragmatic approach that was followed in ANNA did not lead to dramatic difficulties (see further under References 3.2.1).

B. 'Cognates' versus 'non-cognates'

Words that are considered the same are amalgamated, i.e. treated together in

one entry. They are called *cognates. Non-cognates* on the other hand, such as N. 'kameleon'/A. 'verkleurmannetjie' [E. chameleon] are treated as in any traditional bilingual dictionary, which means that they are treated separately, in two entries.

C. Sharing (at least one) meaning

In order to qualify as a cognate, and so to be treated in one entry, the items in question must also meet one semantic criterion: they need to share at least one meaning.

In other words, sameness/similarity of form is a necessary, but not a sufficient feature to qualify for amalgamation. Consequently, absolute false friends, which only share form, no meaning, such as 'amper', meaning 'hardly' in Dutch and 'almost' in Afrikaans, will be two entries. 'Robot' on the other hand, with one common meaning and one which applies to Afrikaans only (see above), will be treated as a cognate and will, as such, be dealt with in one entry.

D. Critical mass

To apply the amalgamation model 'successfully' there needs to be a sufficient critical mass of cognates. Of course, 'success' is a relative concept. Yet one could argue that the degree of (relative) success correlates directly with the number of cognates there is to be found between the two languages. If this number is (much) smaller than half of the total number of words to be described, the degree of amalgamation will be smaller than 0,5 too and the smaller this coefficient the less the dictionary will differ from a 'normal', traditional, dictionary.

In ANNA the cognates/non-cognates ratio is about 50% as there are:

59051 dictionary entries, of which there are
29840 cognates
15011 unique Afrikaans items
14200 unique Dutch items

The above numbers make clear that there is an overlap between Afrikaans and Dutch of about 2/3. This seems to satisfy sufficiently the 'critical mass' criterion.

2.3 Illustration of the model: macro and micro

Instead of the traditional double macrostructure distributed over two volumes (A-B, B-A), the *macrostructures* of the two languages in an amalgamated bilingual dictionary (A *and* B) become *unified,* combined as one whole. So, for instance, in ANNA the section 'lekker-lekkerte' (E. tasty-tastiness) looks as fol-

lows (Dutch items are in roman, Afrikaans ones in italics, followed by English translations between brackets):

lekker, *lekker*	(nice, tasty)
lekkerbek, *lekkerbek*	(gourmet)
lekkerbekje, *lekkerbek*	(fried fillet of haddock)
lekkerbekkig	(finicky)
lekkergoed	(sweets)
lekkerig	(who likes to lick)
lekkerkry	(pleasure)
lekkerlyf	(squiffy)
lekkernij, *lekkerny*	(delicacy)
lekkerruik	(scented)
lekkers	(sweets)
lekkerte	(nice thing)

This unification allows not only for a direct illustration of morphological similarities and differences but also for the generation of hypotheses at this level.

The micro-amalgamation reveals itself in the cognates or amalgamated items. To clearly indicate both differences and similarities at the semantic as well as at the combinatorial level, use is made of the following markers:

A/N	indicates a common meaning
N	indicates a meaning which occurs in Dutch only
A	indicates a meaning which occurs in Afrikaans only
≠	marks a combination or example which shows contrast
=	marks a combination or example which shows no contrast

The following somewhat simplified ANNA-entry can illustrate how this system works.

WERF

A/N (werkplaats voor schepen) *skeepswerf,* [m.g.] *werf*
= op de werf *op die skeepswerf* ≠ een schip van de werf laten lopen *'n skip te water laat*
N ([BN] bouwterrein) *bouterrein*
A *(oop stuk grond rondom die huis)* erf
≠ *'n pragtig geleë plaashuis met werf en tuin* een fraai gelegen woonboerderij met erf en tuin; *'n motor het die werf opgery* een auto reed het erf op; (fig.) *elke mens moet sy eie werf skoonhou* ieder moet zijn eigen tuintje wieden

[The three meanings are: shipyard; building site; yard. The translation of the examples reads: on the shipyard; launch a ship; a nice farmstead with yard and garden; a car drove into the premises; everyone has to clean up one's own backyard.]

For more elaborated examples of entry words I refer to the *appendix*. From

the data presented there it becomes clear that, actually, there are three kinds of cognates: *absolute cognates, absolute cognates with form difference* and *partial cognates.*

Absolute cognates are words which are completely identical in Afrikaans and Dutch, both in form and in meaning, such as is the case with, for instance, opwinding/*opwinding.* However, as one can observe, this does not mean that these items are used in completely the same way, neither collocationally nor pragmatically.

Absolute cognates with form difference are, in fact, a subset of the absolute cognate-class. They have the same meaning(s) in both languages and the form difference they show is small enough to be (easily) recognizable. See, for example, hartinfarkt/*hartinfark,* stikken/*stik.*

Partial cognates are words which share the 'same' form having at least one meaning in common and one meaning that differs. See, for example, taai/*taai* and geil/*geil.*

In a way these *partial cognates* are also *partial false friends* (also see above, section 1.2.3). Contrary to *absolute false friends* such as stoep1 and *stoep2* or *lemoen* and limoen, which only share forms, no meanings, they share forms and *part of their meanings.*

A last type of items is the *non-cognates:* they have the same meaning but a form which is clearly different. *Non-cognates*, as in any traditional bilingual dictionary, are treated as separate entries. See, for instance, appelsien, sinaasappel, and *lemoen.*

3. Amalgamation: pros, cons and pitfalls

3.1 Advantages

Two of the advantages of the model have already been mentioned in section 1, I briefly repeat them here.

a. The amalgamation facilitates the direct comparison between the two languages.

Differences and similarities at the level of orthography, morphology, semantics, pragmatics and combinatorics become, so to speak, clear immediately and not through the combination of two volumes.

In the above case of 'werf', for instance, all meanings are brought together, yielding a global overview, whereas in a traditional treatment they would be distributed over the two volumes as follows:

— meanings 1 and 2 would appear in the Dutch–Afrikaans part
— meanings 1 and 3 would appear in the Afrikaans–Dutch part

In this respect it is interesting to refer to Reneé Marais, who, in a review of ANNA, points at the advantages this approach offers in the context of lan-

guage learning. Marais writes the following:

> The fact that the two languages in this dictionary are mixed is very convenient. The contrastive approach yields many useful data. One learns a non-related language from scratch, from *tabula rasa,* from point zero. Although learning a closely related language also implies that one simply has to learn certain grammatical aspects and unknown lexical items, one, especially, has to be conscious of similarities, small differences and false friends. (Marais 2011: 191; my English)

b. The amalgamation *reduces redundancy*.

In the case of 'werf', for instance, the first meaning, shared by both languages, needs only to be mentioned once.

In addition to these 'known' advantages, I will in what follows, elaborate on two other advantages:

c. In an amalgamated bilingual dictionary it is easier to detect *contrastive patterns* (at morphological, semantic or combinatorial level) than in a traditional bilingual dictionary. These patterns then can subsequently serve as *hypothesis generators.*

One can, for instance, wonder whether there are different systematic patterns for meaning extensions for certain classes of words. Compare, for instance, 'N. woest' to 'A. woest' where the 'desolate' meaning applicable to landscapes in both languages, 'extends' in Dutch to an attribute applicable to 'people', whereas in Afrikaans it applies to 'situations'. Or think of differences in figurative extensions. In this respect one can test more general hypotheses concerning stereotypes or opinions about the two languages. Dutch speakers, for instance, consider Afrikaans as a cute, charming language, richer in imagery than Dutch. However, speakers of Afrikaans, when asked about the 'richness' of Dutch, also consider Dutch very rich in metaphors and images. The fact that figurative expressions are marked for both languages and brought together can help in 'objectifying' these opinions. In the entry 'muur' (E. wall), for instance, figurative expressions in Dutch appear (such as: 'geld uit de muur halen', literally: get money out of the wall, meaning: 'get money from a vending machine') which have no figurative counterpart in Afrikaans and vice versa (for instance: '*oor die muur wees'*, literally*:* to be beyond the wall, meaning: to be worn-out) (on figurative expressions in Afrikaans and Dutch, also see Swanepoel 1997).

In other words, often, by leaving one's own frame of reference, one can better see oneself. ANNA, in confronting the two languages from both angles, can certainly offer data to test certain claims and beliefs in more detail.

d. As always, however, the proof of the pudding is in the eating. The fact that there is now an example of amalgamation makes it possible to *evaluate the model*. This not only allows us to stress the advantages of the

model but also to detect certain side effects and/or pitfalls, two phenomena I will deal with in the next section.

3.2 System side effects

Just like any system, the amalgamation model shows certain side effects due to the system itself. These, however, need not be harmful to the user.

3.2.1 References

The fact that cognates are treated together raises the question where to deal with those cognates which show a form difference. In other words: in which entry will the user have access to items such as gel/*jel* or *pikkewyn*/pinguïn? Will he/she find this information under 'gel' or under 'jel', under 'pikkewyn' or under 'pinguïn'?

As these are combined entries or cognates with difference in form, there must be a *preferred* first item to guarantee access. This problem is typical for paper dictionaries as in an electronic version (e.g. the CD-ROM version of ANNA) the user can access the data by means of both language forms.

In the paper version we have chosen a representation in which the Dutch form is followed by the Afrikaans one, for instance:

drijven, *dryf*	[float]
drijvend, *drywend*	[floating]
drijver, *drywer*	[driver, drover]
drillen, *dril*	[drill]
dringen, *dring*	[push] etc.

To help the user who wants to consult these cognate entries via the Afrikaans form, a pragmatically functioning reference system had to be worked out. To put it briefly: if, in a purely alphabetical ordering, the Afrikaans item would precede or follow the Dutch item immediately or occur in the immediate neighbourhood, then no reference is included. If, however, the Afrikaans and Dutch item would be separated by more than seven items, then a reference is included, leading the user from the Afrikaans to the Dutch form. As a consequence, no reference is made from '*dring*' to 'dringen', nor from '*dril*' to 'drillen' (those items would follow each other). Reference is, however, made from '*dryf*' to 'drijven', from '*drywend*' to 'drijvend' and from '*drywer*' to 'drijver', as these items are too far removed from each other.

ANNA contains about 5000 such references, always going from Afrikaans to Dutch. In this way users who want to look up a cognate word using the Afrikaans form will find what they are looking for, not being hindered by too many references.

3.2.2 Meaning order

With polysemous cognates the meaning order is not defined by, for instance, frequency of occurrence of meanings, but by the 'system' itself. Indeed it is the amalgamation system that in the case of polysemous cognates 'dictates' the following order:

> The meanings Afrikaans and Dutch have in common precede
> those which are exclusively Dutch, if any, which in their turn precede
> those which are exclusively Afrikaans, if any.

See, for instance, the 'werf'-example in section 2.3 and the 'taai'-example in the appendix.

Such a 'predefined' path is needed if one wants to guide the user through the semantic wood of an amalgamated dictionary. Of course, this path will not necessarily coincide with the meaning order in an (Afrikaans) monolingual dictionary based on frequency. For instance, in ANNA, one will find that in the entry 'pad' [E. path] the 'small road'-reading will precede the 'road'-reading (which is by far the most frequent in Afrikaans) because of the fact that the former is the reading Dutch and Afrikaans share. Other examples are: 'robot', where the 'automaton'-reading precedes the 'traffic light'-reading, 'kar', where the 'vehicle on two wheels'- precedes the 'car'-reading, 'brander', where the 'apparatus'- precedes the 'wave'-reading etc., etc.

Consequently, the ANNA-user has to be conscious of the fact that he/she is consulting a bilingual contrastive dictionary with a logic of its own regarding the order in polysemous cognates.

3.2.3 The place of cognate equivalents in macro and micro

If an Afrikaans item has a Dutch cognate equivalent, then, even if there is another more usual equivalent in Afrikaans, it is the cognate equivalent that will be put in the macro. For instance, one will find as a macro-entry the 'head' taxateur/*taksateur* although in Afrikaans the more common word for 'taxateur' [E. appraiser, valuer] is *waardeerder.* However, in the microstructure the most frequent equivalent will be put first or, alternatively, the less frequently used (but cognate) equivalent will be given a restrictive label, such as, for instance, [formal], in the case of *taksateur.*

3.2.4 Pitfalls

More serious perhaps than the system side effects just mentioned, which, after all, can be 'overcome' within the system, are, what I will call, pitfalls. What I am referring to are cases where there is a meaning overlap or minor meaning/ usage differences between Afrikaans and Dutch. As a rule, in ANNA, we have

chosen to abstract away from these differences, in other words, to 'lump', not to 'split', quite in line with the amalgamation approach itself. This does not mean, however, that these differences have been ignored: although these 'differences' are treated under 'common' meanings, use is made of examples, pragmatic labels and/or comments to make these often more subtle differences clear.

In ANNA, for instance, the Dutch item 'tergen' with the meaning 'sarren' [E. provoke] and the Afrikaans item 'terg' with the meaning 'pla' [E. tease] are considered cognates rather than false friends, and, consequently, they are treated under one common *A*/N-meaning and not in separate entries. However, the following comments are added (in Afrikaans, I translate them here into English):

> Although 'terg' in Afrikaans can also mean 'provoke', compared to Dutch, its meaning is less strong mostly having a jocular rather than a nasty connotation.

The difference between the two items is also made clear by means of examples. For instance: N. iemand op alle mogelijke manieren tergen [E. provoke s.o. in all possible ways] A. *iemand op alle moontlike maniere tart;* A. *graag terg* [E. like to tease] N. graag plagen.

Some other examples:

— rukken/*ruk*
In Afrikaans as well as in Dutch this verb can be used both transitively and intransitively. In the latter case, Dutch needs to use a prepositional object as a complement. For instance: 'aan iets rukken' [E. to pull at something]. In Afrikaans this needs not be the case. The verb *ruk* can also be used without a prepositional object, meaning 'to make a jerk, moving with jolts, as if pulled'. This specific Afrikaans usage is not treated in a different entry nor as a different sense, but by means of the example *die vliegtuig ruk* (N. het vliegtuig schokt, E. the air plane is shaking/jolting).

— aanbieder/*aanbieder*
Afrikaans and Dutch share the common general meaning 'someone who offers something, a provider'. However, depending on the context, this word can get a more specific semantic load in Afrikaans. For instance: the *aanbieder* of a TV-programme is called a 'presentator' in Dutch [E. presenter, host], the *aanbieder* of a course is a 'cursusleider' in Dutch [E. course instructor]. As one can observe, here too, the more specific usages in Afrikaans are accommodated under the more general, common meaning and the differences between Afrikaans and Dutch are made clear by means of examples.

— ontkennen/*ontken*
In this case, depending on the language, different semantic actants are used with the same item, without changing the basic meaning:

	Affected Object
N. ontkennen	iets
A. *ontken*	*iets, iemand*

If the Affected Object is [-human], then the semantic load is 'deny'.
If the Affected Object is [+human], then the semantic load is 'deny/ refuse to recognise'.

Onthouden/*onthou* is a similar case: here again the general meaning of 'keep in memory' if applied to a human object is specified to 'remember someone'. Such cases are 'lumped' in ANNA, the difference made clear by means of examples.

	Affected Object
N. onthouden	iets
A. *onthou*	*iets, iemand*

To conclude: it is important for users to be aware of the fact that the amalgamation approach entails the lumping of meanings rather than the splitting of them when small, more subtle differences are at stake. In ANNA we have tried to deal with these cases in a sensible way, although it is quite well possible that in this approach some differences have been overlooked or could be dealt with in a more explicit way.

4. Conclusion

In this article I have presented ANNA and the underlying amalgamation model. In addition, I have provided a first evaluation of the model. As for its applicability to other languages, within the ANNA-project itself the focus was on the application possibilities for so-called 'black' languages in South Africa, where it seems to be promising.

Whether the model is also applicable to other 'black' languages such as, for instance, Kirundi and Kinyarwanda spoken in, respectively Burundi and Ruanda, or Scandinavian and Slavic ones, like Danish and Norwegian or Russian and Byelorussian, is up to the specialists in the field to find out, although it seems at first sight, that his could be the case[8].

If this proves true then between these languages a new type of *contrastive dictionary,* comparable to ANNA, could emerge.

However, there is more. As I pointed out in the *Introduction* to the ANNA dictionary (p. 11, my translation into English):

> Last but not least, an amalgamated dictionary is also a *reading dictionary:* a dictionary in which one cannot only look for translations or contrasts, but one in which one can also read and browse. Anyone looking in ANNA for N. 'bloot' [E. naked] for instance, will not only find that the most common equivalent in Afrikaans is *kaal*, but ANNA will take him further from 'bloot' to N. 'blootshoofds' [A. *kaalkop,* E. bald] and from N.' blootshoofds' to A. *bloots ry* [N. het (iem.) lastig

maken, E. give (s.o.) a dressing-down], and from A. *bloots ry* to N. 'blootsvoets' [A. *kaalvoet*, E. bare-footed] etc. In other words, before you know it, ANNA takes you with her from word to word to word and that is exactly where she has quite a lot to offer.

It therefore seems justified to conclude that if ANNA could inspire others, to new, more adequate, types of bilingual dictionaries, be it amalgamated, contrastive, reading or some other type, I would feel more than rewarded for the effort made by the ANNA-team in realizing this project.

Notes

1. In a sense monolingual diachronic dictionaries also show a kind of amalgamation as they bring 'same items' from different time periods together. However, here all similarity with the amalgamation model stops as the latter focuses on synchronic similarities and differences of two different languages and not on the evolution of lexical items through time.
2. ANNA has been financed mainly by private sponsors, the main sponsor being the ZASM foundation (ZASM = Zuid-Afrikaanse Spoorwegmaatschappij (South African Railway Company)). Other sponsors were the PUK Vice-Chancellor's Trust of the North West University (Potchefstroom), the University of Stellenbosch, The Dutch Language Union, The department of Foreign Affairs of Flanders, the Van den Bergh van Heemstede Foundation, the Joan Louw Trust, The LW Hiemstra Trust and the Prince Bernhard Cultural Foundation.
3. For more information on the RBN, see Van der Vliet 2007; for OMBI, see Maks 2007.
4. The ANNA-team consisted of nine members. The editorial team itself consisted of five members: E. Boekkooi (Port Elizabeth), R. Gouws (Stellenbosch), I. Maks (Amsterdam), L. Renders (Hasselt) and myself acting as an editor-in-chief and project leader.
5. The CD-ROM version has been produced by Pharos Publishers, who also distribute the South African version of ANNA.
6. Of course also a 'foreign' language can be used as an 'interlingua'.
7. It goes without saying that any project can be started up if there are no financial constraints. In most cases however one has to prioritise because of financial limitations.
8. Notice that next to a certain degree of 'sameness', the two languages should also show a certain degree of 'difference'. Cognates should, for instance, show sufficient differences at combinatorial, pragmatic and/or semantic level.

References

Dekker, L. en P.C. Paardekooper. 1990. *Nederlands–Afrikaanse Woordeboek.* Pretoria: J.L. van Schaik.

Groot Woordenboek Afrikaans en Nederlands. 2011. Martin, W. (Editor-in-chief). 2011. *Groot Woordenboek Afrikaans en Nederlands*. Houten/Antwerpen: Het Spectrum B.V.

HAT. 2000. Odendal, F.F. and R.H. Gouws. 2000. *HAT: Verklarende Handwoordeboek van die Afrikaanse Taal*. Cape Town: Perskor.

Maks, I. 2007. OMBI: The Practice of Reversing Dictionaries. *International Journal of Lexicography* 20(3): 259-274.

Marais, R. 2011. Een mooie dikke dame. "ANNA", het Groot Woordenboek Afrikaans en Nederlands. *Ons Erfdeel*, nr. 1: 190-192.

Martin, W. 2005. ANNA and a 'New' Lexicography in South Africa. Schutte, G.J. and H. Wels (Eds.). 2005. *The Vrije Universiteit and South Africa. From 1880 to the Present and Towards the Future: Images, Practice and Policies*: 49-57. Amsterdam: Rozenberg Publishers.

Martin, W. 2012. Amalgamated Bilingual Dictionaries. Genis, René et al. (Eds.). 2012. *Between East and West. Festschrift for Wim Honselaar on the Occasion of his 65th Birthday. Pegasus Oost-Europese Studies* 20*:* 437-449. Amsterdam: Uitgeverij Pegasus.

Martin, W. To appear. ANNA:een woordenboek met een naam en wat er achter steekt. *Tydskrif vir Nederlands en Afrikaans*. 2012.

Martin, W., R. Gouws and L. Renders. 1999. *Haalbaarheids- en Definitiestudie voor een Woordenboek Afrikaans–Nederlands/Nederlands–Afrikaans.* Amsterdam: Vrije Universiteit.

Martin, W. and R. Gouws. 2000. A New Dictionary Model for Closely Related Languages: The Dutch–Afrikaans Dictionary Project as a Case-in-point. Heid, Ulrich et al. (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress*. *EURALEX 2000. Stuttgart, Germany, August 8th–12th, 2000*: 783-792. Stuttgart: Stuttgart University.

Prisma Miniwoordenboek Afrikaans–Nederlands/Nederlands–Afrikaans. 2004. Utrecht: Het Spectrum.

Swanepoel, P. 1997. Idiomatiese taalgebruik en die afstand tussen Afrikaans en Nederlands. *Tydskrif vir Nederlands en Afrikaans* 4(1): 128-161.

Van der Vliet, H. 2007. The Referentiebestand Nederlands as a Multi-purpose Lexical Database. *International Journal of Lexicography* 20(3): 239-257.

Van Houwelingen, F. and A. Carstens. 1998. "Nederlandismes" in HAT. *Literator* 19(2): 1-12.

Veltkamp-Visser, S. 1998. *Afrikaans op reis:* taalgids voor de Nederlandssprekende toerist in Zuid-Afrika. Fifth Edition. Amsterdam: NZAV.

Appendix: ANNA

appelsien [nw.] [de; mv: -en]

N (<BN> sinaasappel) *lemoen*

geil [bnw.] , ***geil*** [b.nw.] {*Opm.: In Afr. het 'geil' 'n ruimer betekenis as in Ned. en beteken nie alleen paarlustig nie.*}

A/N (<inf.> met een zeer sterke geslachtsdrift) *jags*, *bronstig*, *katools*, *geil* = geile gedachten *jagse gedagtes* ≠ een geile bok *'n bronstige bokram*, *'n geil bokram*; geil zijn op iemand *katools wees oor iemand*; een geile griet *'n katoolse meisiekind*

A (*welig, vrugbaar*) welig, vruchtbaar ≠ *geil grond* vruchtbare/vette grond; *geil groei* welig tieren; *'n geil jaar* een jaar met een goede oogst

hartinfarct [nw.] [het; mv: -en] , ***hartinfark*** [nw.] [mv: *-e*] {*Opm.: In Afr. kom 'hartinfark' uitsluitlik in die mediese register voor.*}

A/N (het afsterven v.d. hartspier) *hartversaking*, <form.> *hartinfark* ≠ roken verhoogt ook de kans op een hartinfarct *rook verhoog ook die kans op hartversaking*; een hartinfarct krijgen *hartversaking kry*

lemoen [nw.] [mv: *-e*] {*!! Opm.: Sien ook Ned. 'limoen'.*}

A (*oranje/goudgeel sitrusvrug*) sinaasappel, <BN> appelsien ≠ *'n lemoen skil/eet* een sinaasappel pellen/eten ≠ <fig.> *weggegooi word soos 'n uitgesuigde lemoen* weggeworpen worden als een uitgeknepen citroen; <fig.> *iemand uitsuig soos 'n lemoen* iemand als een citroen uitknijpen

opwinding [nw.] [de; mv: -] , ***opwinding*** [nw.] [mv: -]

A/N (spanning) *opwinding*, *opgewondenheid* = er heerst (grote) opwinding *daar heers (groot) opwinding*; zorgen voor (de nodige) opwinding *sorg vir (die nodige) opwinding*; grote/enorme opwinding *groot/enorme opwinding* ≠ van opwinding niet meer kunnen slapen *van opwinding nie meer kan slaap nie*; er was veel opwinding om de nieuwe auto *daar was groot opwinding oor die nuwe motor*

sinaasappel [nw.] [de; mv: -s, sinaasappelen]

N (sappige zuidvrucht) *lemoen* ≠ een sinaasappel pellen/eten *'n lemoen skil/eet*

stik zie stikken

stikken$_{1}$ [ww.intr.] , ***stik*** [ww.intr.]

A/N (door ademnood sterven) *stik* = <fig., inf.> stikken van de warmte/het lachen/woede/jaloezie *stik van die warmte/die lag/woede/jaloesie* ≠ <inf.> stik! *gaan bars!*, *bokker jou!*; stikken door de rook *stik van die rook*; <inf.> iemand laten stikken *iemand kwaai in die steek laat* ≠ <fig., inf.> stikken in het geld *stink van die geld*; <fig., inf.> het stikt hier van de soldaten/muggen/kroegen *dit wemel hier van die soldate/muskiete/kroeë*

stikken$_2$ [ww.tr.] , ***stik*** [ww.tr.]

AN (naaien met eenvoudige steek) *stik* = een zoom in een broekspijp stikken *'n soom in 'n broekspyp stik*

stoep$_1$ [nw.] [de; mv: -en] {*!!*}

N (strook langs weg voor voetgangers) *sypaadjie* ≠ denk erom, op de stoep blijven met je step! *onthou, bly op die sypaadjie met jou skopfiets!* ≠ <fig.> bij iemand op de stoep staan *voor iemand se deur staan*; <fig.> stoepetje/stoepje spelen *'n tipe balspeletjie op straat speel*

stoep$_2$ [nw.] [mv: *-e*] {*!!*}

A (*verhoogde vloerarea buite huis, veranda*) veranda ≠ *aan die voorkant was 'n stoep wat met wingerd omrank was waar 'n mens heerlik kon sit* aan de voorzijde was een met wingerd omrankte veranda waar het heerlijk zitten was ≠ <fig.> *(eers) voor jou eie stoep vee* (eerst) je eigen straatje schoonvegen

taai [bnw.] , ***taai*** [b.nw.]

AN 1 (stug) *taai* = dat vlees is bijzonder taai *daardie vleis is besonder taai* ≠ <fig.> zo taai als een schoenzool zijn *so taai soos (skoen)leer wees* ≠ <fig.> *'n (taai) turksvy* een heet hangijzer; <fig.> *taai in die bek wees* hardleers zijn, weerspannig zijn; <fig.> *taai gesprekke/onderhandelinge* moeizame gesprekken/onderhandelingen 2 (volhardend, sterk) *taai* = je taai houden *jou taai hou* ≠ een taaie oude man *'n taai ou man*, *'n seningtaai ou man*; houd je taai! *uithou en aanhou!*, *vasbyt!* ≠ *jou taai hou* je kras houden ≠ <fig.> *'n taai klap* een flinke optater; <fig.> *so taai soos 'n ratel wees* uitzonderlijk taai zijn

N (vervelend) *moeilik* = taaie leerstof *moeilike leerstof*; een taaie roman *'n moeilike roman* ≠ een taaie beklimming *'n moeilike klim*

A (*klewerig*) kleverig, plakkerig, klef = *taai deeg* plakkerig deeg; *taai vingers* plakkerige vingers; *taai hande* kleverige handen, kleffe handen; *taai brood/gebak* klef brood/gebak ≠ *'n taai spul(letjie)* een kleverig goedje

Publikasieaankondigings / Publication Announcements

English Dictionary for South Africa. 2011, 1108 pp. ISBN 978-1-86890-119-7 (Hardback with CD-ROM). Cape Town: Pharos Dictionaries. Price: R375.

Arnold Fischer, et al. *Oxford English–Xhosa Dictionary.* 2012, 752 pp. ISBN 978 0 19 904342 2 (Paperback). Goodwood: Oxford University Press Southern Africa. Price: R199.95.

Pedro A. Fuertes-Olivera and Henning Bergenholtz (Editors). *e-Lexicography. The Internet, Digital Initiatives and Lexicography.* 2011, xiv + 341 pp. ISBN 978-1-4411-2806-5 (Hardback). London/New York: Continuum International Publishing. Price: £85.00.

Lorna Hiles (Editor). *Oxford South African Illustrated Dictionary.* 2011, 384 pp. ISBN 978 0 19 59957 9 (Hardback). Goodwood: Oxford University Press Southern Africa. Price: R104.95.

Phillip Louw (Hoofredakteur). *Oxford Afrikaanse Skoolwoordeboek.* 2012, 684 pp. ISBN 978 0 19 599795 8 (Paperback). Goodwood: Oxford University Press Southern Africa. Price: R109.95.

Jana Luther (Redakteur). 2011. *Longman-HAT Basic Dictionary/Basiswoordeboek English–Afrikaans/Afrikaans–Engels*. 2011, xiv + 224 pp. ISBN 978-1-77578-138-7 (Sagteband). Pinelands: Pearson Education SA. Prys: R50.

Jana Luther (Redakteur). 2011. *Longman-HAT Pocket Dictionary/Sakwoordeboek English–Afrikaans/Afrikaans–Engels*. 2011, xiv + 558 pp. ISBN 978-1-77578-126-4 (Sagteband). Pinelands: Pearson Education SA. Prys: R90.

Jana Luther (Redakteur). 2011. *Longman-HAT School Dictionary/Skoolwoordeboek English–Afrikaans/Afrikaans–Engels*. 2011, xiv + 558 pp. ISBN 978-1-77025-780-1 (Sagteband). Pinelands: Pearson Education SA. Prys: R120.

Willy Martin (Hoofredakteur). *Groot Woordeboek Afrikaans en Nederlands.* 2011, 2228 pp. ISBN 978-1-86890-120-3 (Hardeband). Kaapstad: Pharos Woordeboeke, Prys: R999.95. (Artikel in hierdie uitgawe.).

Luvuyo S. Martins and Pumlani Sibula. *Cricket Terminology/iSigama seQakamba/ Kriekketterminologie.* 2011, iv + 71 pp. ISBN 978-0-620-500289 (Paperback). Stellenbosch: Stellenbosch University Language Centre/IZiko leeLwimi le Univesithi yaseStellenbosch/Universiteit Stellenbosch Taalsentrum.

Luvuyo Martins and Mkhonto van Zyl. *Masincokole ngeMpilo ngesiXhosa./Let's chat about health in isiXhosa./Kom ons gesels oor gesondheid in isiXhosa.* 2011, ix + 137 pp. ISBN 978-0-620-51030-1 (Paperback). Stellenbosch: IZiko leeLwimi le Univesithi yaseStellenbosch/Stellenbosch University Language Centre/Universiteit Stellenbosch Taalsentrum.

Grant Ngidi and Pumlani Sibula. *Rugby Terminology/iSigama soMbhoxo/Rugbyterminologie.* 2011, iv + 84 pp. ISBN 978-0-620-50029-6 (Paperback). Stellenbosch: Stellenbosch University Language Centre/IZiko leeLwimi le Univesithi yaseStellenbosch/Universiteit Stellenbosch Taalsentrum.

Helga Schaberg and Phillip Louw. *Oxford South African School Dictionary*. Third edition. CD-ROM. 2011. ISBN 978 0 19 904064 3 (Paperback with CD-ROM). Goodwood: Oxford University Press Southern Africa. Price: R189.95.

South African Trilingual Wine Industry Dictionary/Suid-Afrikaanse Drietalige Wynbedryfwoordeboek/Isichazi-magama saseMzantsi Afrika seshishini lewayini ngeelwimi ezintathu. http://www.sawis.co.za/dictionary

Dennis Spohr. *Towards a Multifunctional Lexical Resource. Design and Implementation of a Graph-based Lexicon Model.* 2012, XIV + 206 pp. ISBN 978-3-11-027115-7 (Hardback). Lexicographica. Series Maior 141. Berlin/New York: Walter de Gruyter. Price: €89.95.

Martin Stark. *Bilingual Thematic Dictionaries.* 2011, IX + 486 pp. ISBN 978-3-11-025889-9 (Hardback). Lexicographica. Series Maior 140. Berlin/New York: Walter de Gruyter. Price: €99.95.

Tweetalige Skoolwoordeboek/Bilingual School Dictionary. 2012, 672 pp. ISBN 978-1-86890-128-9 (Sagteband). Kaapstad: Pharos Woordeboeke. Prys: R129.95.

VOORSKRIFTE AAN SKRYWERS

(*Tree asseblief met ons in verbinding (lexikos@sun.ac.za) vir 'n uitvoeriger weergawe van hierdie instruksies of besoek ons webblad:* http://www.wat.co.za)

A. REDAKSIONELE BELEID

1. Aard en inhoud van artikels

Artikels kan handel oor die suiwer leksikografie of oor implikasies wat aanverwante terreine, bv. linguistiek, algemene taalwetenskap, rekenaarwetenskap en bestuurskunde vir die leksikografie het.

Bydraes kan onder enigeen van die volgende rubrieke geklassifiseer word:

(1) **Artikels:** Grondige oorspronklike wetenskaplike navorsing wat gedoen en die resultate wat verkry is, of bestaande navorsingsresultate en ander feite wat op 'n oorspronklike wyse oorsigtelik, interpreterend, vergelykend of krities evaluerend aangebied word.

(2) **Resensieartikels:** Navorsingsartikels wat in die vorm van 'n kritiese resensie van een of meer gepubliseerde wetenskaplike bronne aangebied word.

Bydraes in kategorieë (1) en (2) word aan streng anonieme keuring deur onafhanklike akademiese vakgenote onderwerp ten einde die internasionale navorsingsgehalte daarvan te verseker.

(3) **Resensies:** 'n Ontleding en kritiese evaluering van gepubliseerde wetenskaplike bronne en produkte, soos boeke en rekenaarprogramme.

(4) **Projekte:** Besprekings van leksikografiese projekte.

(5) **Leksikonotas:** Enige artikel wat praktykgerigte inligting, voorstelle, probleme, vrae, kommentaar en oplossings betreffende die leksikografie bevat.

(6) **Leksikovaria:** Enigeen van 'n groot verskeidenheid artikels, aankondigings en nuusvrystellings van leksikografiese verenigings wat veral vir die praktiserende leksikograaf van waarde sal wees.

(7) **Verslae:** Verslae van konferensies en werksessies.

Bydraes in kategorieë (3)-(7) moet almal aan die eise van akademiese geskrifte voldoen en word met die oog hierop deur die redaksie gekeur.

2. Wetenskaplike standaard en keuringsprosedure

Lexikos is deur die Departement van Onderwys van die Suid-Afrikaanse Regering as 'n gesubsidieerde d.w.s. inkomstegenererende navorsingstydskrif goedgekeur.

Artikels sal op grond van die volgende aspekte beoordeel word: taal en styl; saaklikheid en verstaanbaarheid; probleemstelling, beredenering en gevolgtrekking; verwysing na die belangrikste en jongste literatuur; wesenlike bydrae tot die spesifieke vakgebied.

3. Taal van bydraes

Afrikaans, Duits, Engels, Frans of Nederlands.

4. Kopiereg

Nóg die Buro van die WAT nóg die African Association for Lexicography (AFRILEX) aanvaar enige aanspreeklikheid vir eise wat uit meewerkende skrywers se gebruik van materiaal uit ander bronne mag spruit.

Outeursreg op alle materiaal wat in *Lexikos* gepubliseer is, berus by die Direksie van die Woordeboek van die Afrikaanse Taal. Dit staan skrywers egter vry om hulle materiaal elders te gebruik mits *Lexikos* (AFRILEX-reeks) erken word as die oorspronklike publikasiebron.

5. Oorspronklikheid

Slegs oorspronklike werk sal vir opname oorweeg word. Skrywers dra die volle verantwoordelikheid vir die oorspronklikheid en feitelike inhoud van hulle publikasies.

6. Gratis oordrukke en eksemplare

Skrywers ontvang vyf gratis oordrukke van elke artikel of resensieartikel van hulle wat gepubliseer is asook een gratis eksemplaar van die uitgawe waarin sodanige artikel(s) verskyn het. Skrywers van suiwer evaluerende resensies en van bydraes tot die rubrieke Leksikonotas, Leksikovaria, Projekte en Verslae ontvang vyf gratis oordrukke van hulle bydraes. In laasgenoemde vier kategorieë kan die redaksie egter, afhangend van die aard en omvang van die bydraes, besluit om ook 'n eksemplaar van die betrokke uitgawe aan 'n skrywer toe te ken.

7. Uitnodiging en redaksionele adres

Alle belangstellende skrywers is welkom om bydraes vir opname in *Lexikos* te lewer en aan die volgende adres te stuur:

Die Redakteur: LEXIKOS
Buro van die WAT
Posbus 245
7599 STELLENBOSCH
Republiek van Suid-Afrika

B. VOORBEREIDING VAN MANUSKRIP

Die manuskrip van artikels moet aan die volgende redaksionele vereistes voldoen:

1. Lengte en formaat van artikels

Bydraes moet verkieslik nie 20 getikte A4-bladsye met teks in dubbelspasiëring en ruim kantlyne (ongeveer 2,5 cm) oorskry nie. Manuskrip moet verkieslik in elektroniese formaat as ASCII-teks, as volledig geformateerde Microsoft Word (DOS of Windows) lêers of as WordPerfect (DOS of Windows) lêers op rekenaarskyf (360 KB tot 1.44 MB) voorgelê word. 'n Rekenaardrukstuk van die artikel moet die skyf vergesel. Elke artikel moet voorsien wees van 'n Engelse opsomming van tussen 150 en 250 woorde, sowel as tussen 10 en 30 Engelse sleutelwoorde.

2. Grafika

Een stel duidelike oorspronklike illustrasies, tabelle, grafieke, diagramme, of kwaliteitsafdrukke daarvan, moet voorgelê word. Die plasing van grafika binne die teks moet duidelik aangedui word.

3. Bibliografiese gegewens en verwysings binne die teks

Kyk na onlangse nommers van *Lexikos* vir meer inligting.

INSTRUCTIONS TO AUTHORS

(*For a more detailed version of these instructions, please contact us (lexikos@sun.ac.za) or refer to our website:* http://www.wat.co.za)

A. EDITORIAL POLICY

1. Type and content of articles
Articles may deal with pure lexicography or with the implications that related fields such as linguistics, general linguistics, computer science and management have for lexicography.

Contributions may be classified in any one of the following categories:

(1) **Articles:** Fundamentally original scientific research that has been done and the results that have been obtained, or reflecting existing research results and other facts in an original, synoptic, interpretative, comparative or critically evaluative manner.
(2) **Review articles:** Research articles presented in the form of a critical review of one or more published scientific sources.

Contributions in categories (1) and (2) are subjected to strict anonymous evaluation by independent academic peers in order to ensure the international research quality thereof.

(3) **Reviews:** An analysis and critical evaluation of published scientific sources and products, such as books and computer software.
(4) **Projects:** Discussions of lexicographical projects.
(5) **Lexiconotes:** Any article containing practice-oriented information, suggestions, problems, questions, commentary and solutions regarding lexicography.
(6) **Lexicovaria:** Any of a large variety of articles, announcements and press releases by lexicographic societies which are of particular value to the practising lexicographer.
(7) **Reports:** Reports on conferences and workshops.

Contributions in categories (3)-(7) must all meet the requirements of academic writing and are evaluated by the editors with this in mind.

2. Academic standard and evaluation procedure
The Department of Education of the South African Government has approved *Lexikos* as a subsidized, i.e. income-generating research journal.

Articles will be evaluated on the following aspects: language and style; conciseness and comprehensibility; problem formulation, reasoning and conclusion; references to the most important and most recent literature; substantial contribution to the specific discipline.

3. Language of contributions
Afrikaans, Dutch, English, French or German.

4. Copyright
Neither the Bureau of the WAT nor the African Association for Lexicography (AFRILEX) accepts any responsibility for claims which may arise from contributing authors' use of material from other sources.

Copyright of all material published in *Lexikos* will be vested in the Board of Directors of the Woordeboek van die Afrikaanse Taal. Authors are free however to use their material elsewhere provided that *Lexikos* (AFRILEX Series) is acknowledged as the original publication source.

5. Originality
Only original contributions will be considered for publication. Authors bear full responsibility for the originality and factual content of their contributions.

6. Free offprints and copies
Authors will receive five free offprints of each of their articles or review articles published, as well as one complimentary copy of the issue containing such article(s). Authors of purely evaluative reviews and of contributions to the categories Lexiconotes, Lexicovaria, Projects, and Reports receive five free offprints of their contributions. In the case of the latter four categories, the editors may, however, depending on the nature and scope of the contributions, decide to grant the author a copy of the issue concerned.

7. Invitation and editorial address
All interested authors are invited to submit contributions for publication in *Lexikos* to:

The Editor: LEXIKOS
Bureau of the WAT
P.O. Box 245
7599 STELLENBOSCH
Republic of South Africa

B. PREPARATION OF MANUSCRIPTS

Manuscripts of articles must meet the following editorial requirements:

1. Length and format
Contributions should not exceed more than 20 typewritten A4 pages with double spacing and ample margins (about 2,5 cms). Manuscript should preferably be in electronic form on a (360 KB to 1.44 MB) floppy disk as either ASCII text, fully-formatted Microsoft Word (DOS or Windows) or WordPerfect (DOS or Windows) files. A computer printout of the article should accompany the disk. Each article must be accompanied by an English abstract of 150 to 250 words, and between 10 and 30 English keywords.

2. Graphics
One set of clear original drawings, tables, graphs, diagrams or quality prints thereof must be submitted. The locations of graphics must be clearly indicated in the text.

3. Bibliographical details and references in the text
Examine recent issues of *Lexikos* for details.

HINWEISE UND RICHTLINIEN FÜR AUTOREN

(Nehmen Sie bitte uns Kontakt auf (lexikos@sun.ac.za) für eine ausführlichere Wiedergabe dieser Hinweise oder besuchen Sie unsere Webseite: http://www.wat.co.za)

A. REDAKTIONELLE ZIELSETZUNGEN

1. Art und Inhalt der Artikel
Es können Artikel aufgenommen werden, die sich mit Themen der Lexikographie befassen oder mit Zusammenhängen, die zwischen der Lexikographie und benachbarten Fachgebieten wie z.B. Linguistik, allgemeiner Sprachwissenschaft, Lexikologie, Computerwissenschaft und Management bestehen.

Die Beiträge sollten einer der folgenden Kategorien entsprechen:

(1) **Artikel**, die grundlegend über neue Forschungsansätze und deren Ergebnisse berichten, oder die bestehende Forschungsergebnisse und andere Informationen selbständig, interpretativ, vergleichend oder kritisch bewertend wiedergeben.
(2) **Rezensionsartikel**, die in der Form eines Forschungsartikels eine oder mehrere veröffentlichten wissenschaftlichen Quellen kritisch rezenzieren.

Beiträge in Kategorien (1) und (2) werden streng anonym von unabhängigen wissenschaftlichen Experten begutachtet, um ein internationales fachliches Niveau in *Lexikos* zu gewährleisten.

(3) **Rezensionen**, die veröffentlichte wissenschaftliche Quellen und Produkte, wie z.B. Bücher und Software, analysieren und kritisch bewerten.
(4) Lexikographische **Projekte**, die vorgestellt werden.
(5) **Notizen zum Lexikon**, die praxisbezogene Informationen, Vorschläge, Probleme, Fragen, Kommentare und Lösungen hinsichtlich der Lexikographie enthalten.
(6) **Lexikovaria**, die unterschiedliche Beiträge, Ankündigungen und Pressemitteilungen lexikographischer Vereinigungen, die dem praktischen Lexikographen wichtig sein können, einschließen.
(7) **Berichte** über Konferenzen und Workshops.

Beiträge in Kategorien (3)-(7) müssen im akademischen Stil abgefaßt werden. Sie werden von der Redaktion unter diesem Gesichtspunkt beurteilt.

2. Wissenschaftliche Standards und das Beurteilungsverfahren
Das Erziehungsministerium der südafrikanischen Regierung hat *Lexikos* als eine subventionierte, d.h. einkommenerzeugende Forschungszeitschrift anerkannt.

Artikel werden auf Grund der folgenden Gesichtspunkte bewertet: Sprache und Stil; Sachlichkeit und Verständlichkeit; Problembeschreibung, Argumentation und Schlußfolgerung; Hinweise auf die neueste und wichtigste Literatur; wesentlicher Beitrag zum besonderen Fachgebiet.

3. Sprache der Beiträge
Afrikaans, Deutsch, Englisch, Französisch oder Niederländisch.

4. Das Urheberrecht
Weder das Büro des WAT noch die African Association for Lexicography (AFRILEX) übernehmen Verantwortung für Ansprüche, die daraus entstehen könnten, daß Autoren Material aus anderen Quellen benutzt haben.

Das Urheberrecht aller in *Lexikos* publizierten Artikel wird dem Direktorium unseres Büros übertragen. Es steht Autoren jedoch frei, ihren Beitrag anderweitig zu verwenden, vorausgesetzt, *Lexikos* (AFRILEX-Serie) wird als Originalquelle genannt.

5. Originalität
Nur Originalbeiträge werden begutachtet. Autoren tragen die volle Verantwortung für die Originalität und den sachlichen Inhalt ihrer Beiträge.

6. Sonderdrucke und Freiexemplare
Autoren erhalten fünf Sonderdrucke ihrer veröffentlichten Artikel oder Rezensionsartikel gratis sowie ein Freiexemplar der betreffenden Ausgabe. Rezensenten und Autoren von Beiträgen zu den Kategorien Notizen zum Lexikon, Lexikovaria, Projekte und Berichte erhalten fünf Sonderdrucke ihrer Beiträge gratis. Die Redaktion kann sich jedoch, abhängig von der Art und dem Umfang der Beiträge der letztgenannten vier Kategorien, vorbehalten, dem Autor ein Freiexemplar der Ausgabe zu überlassen.

7. Einladung und redaktionelle Adresse
Alle Autoren, die interessiert sind, Beiträge für *Lexikos* zu liefern, sind herzlich willkommen. Sie werden gebeten, ihre Artikel an die folgende Adresse zu schicken:

Der Redakteur: LEXIKOS
Buro van die WAT
Postfach 245
7599 STELLENBOSCH
Republik Südafrika

B. VORBEREITUNG DES MANUSKRIPTS

Ein Artikelmanuskript muß den folgenden redaktionellen Anforderungen entsprechen:

1. Umfang und Format
Beiträge sollen nicht länger als 20 getippte A4-Seiten in zweizeiligem Abstand und mit Randabständen von ca. 2,5 cm sein. Das Manuskript sollte möglichst als elektronischer Text auf einer (360 KB bis 1.44 MB) Diskette vorgelegt werden, entweder im ASCII-Format, oder in formatiertem Microsoft Word (DOS oder Windows) bzw. WordPerfect (DOS oder Windows). Ein Ausdruck des vollständig formatierten Artikels soll mit der Diskette eingereicht werden. Jedem Artikel ist eine Zusammenfassung im Umfang von 150-250 Wörtern beizufügen. Ferner sollen etwa 10-30 inhaltskennzeichnende Stichwörter zu jedem Artikel angegeben werden.

2. Abbildungen
Ein reproduktionsfähiger Satz der originalen Abbildungen, Illustrationen, Tabellen, Graphiken und Diagramme oder Qualitätsabdrucke muß vorgelegt werden. Der Text selber sollte klare Hinweise auf die Position der Abbildungen enthalten.

3. Bibliographische Einzelheiten und Hinweise im Text
Zu Einzelheiten des bibliographischen Systems sind neuere Ausgaben von *Lexikos* einzusehen.

INSTRUCTIONS AUX AUTEURS

(*Pour une version plus détaillée de ces instructions, contacter le Bureau du WAT (lexikos@sun.ac.za) ou consulter notre website:* http://www.wat.co.za)

A. POLITIQUE ÉDITORIALE

1. Caractéristiques et contenu des articles
Les articles seront consacrés à la lexicographie pure, ou aux rapports entre la lexicographie et les disciplines voisines telles que la linguistique, la linguistique générale, l'informatique et le management.

Les contributions pourront appartenir à l'une des catégories suivantes:

(1) **Articles:** Recherches scientifiques originales, avec leurs résultats; ou présentations originales, synoptiques, interprétatives, comparatives, évaluatives et critiques des résultats de recherches en cours;

(2) **Articles bilans:** Articles de recherche présentés sous forme de bilan critique de travaux scientifiques déjà publiés.

Les contributions appartenant aux catégories (1) et (2) seront soumises de manière anonyme à des experts spécialistes indépendants afin d'en assurer la qualité scientifique au niveau international.

(3) **Recensions:** Analyses et évaluations critiques de travaux de recherche et de productions scientifiques, telles que livres ou logiciels;
(4) **Projets:** Présentations de projets lexicographiques;
(5) **'Lexiconotes':** Textes contenant des informations pratiques, ou des suggestions, des problèmes, des questions, des commentaires et des solutions concernant des activités lexicographiques;
(6) **'Lexicovaria':** Articles, annonces, communiqués de presse émanant de centres de lexicographie et qui revêtent un intérêt particulier pour les lexicographes;
(7) **Rapports:** Rapports sur des colloques et ateliers.

Les contributions dans les catégories (3) à (7) devront répondre aux exigences de qualité des publications scientifiques et seront évaluées dans cette optique.

2. Critères et procédures d'évaluation
La revue *Lexikos* est reconnue et subventionnée par le Ministère de l'Éducation du gouvernement Sud-Africain comme revue devant générer des revenus.

Les articles seront évalués selon les critères suivants: langue et style, concision et clarté, formulation de la problématique, raisonnement et conclusion, référence aux travaux les plus importants et les plus récents, contribution substantielle à la discipline.

3. Langue des contributions
Afrikaans, allemand, anglais, français ou néerlandais.

4. Copyright
Le Bureau du WAT ou l'*African Association for Lexicography* (AFRILEX) décline toute responsabilité en cas de réclamations motivées par l'utilisation d'autres sources par les auteurs.

Les droits d'auteurs des documents publiés dans *Lexikos* appartiennent au Conseil d'administration du *Woordeboek van die Afrikaanse Taal* (*WAT*). Cependant, les auteurs sont libres d'utiliser leurs textes dans d'autres publications, à condition d'indiquer *Lexikos* (collection AFRILEX) comme source de la publication originale.

5. Originalité
Seules les contributions originales seront acceptées pour la publication. Les auteurs conservent l'entière responsabilité de l'originalité et du contenu factuel de leur texte.

6. Tirés-à-part et exemplaires gratuits
Les auteurs recevront gratuitement cinq (5) tirés-à-part de chaque article ou article de bilan, ainsi qu'un exemplaire gratuit de la publication contenant cet article.

Les auteurs des recensions et des publications dans les catégories 'Lexiconotes', 'Lexicovaria', Projets et Rapports recevront cinq (5) tirés-à-part de leur texte. Dans ces quatre dernières catégories, le responsable de la revue pourra néanmoins décider, en fonction de la nature et des dimensions des publications, d'accorder un exemplaire gratuit de la revue à leurs auteurs.

7. Adresse de la revue
Les auteurs intéressés sont invités à soumettre leurs propositions à:

L'éditeur: LEXIKOS
Bureau du WAT
Boîte postale 245
7599 STELLENBOSCH
République d'Afrique du Sud

B. PRÉSENTATION DES MANUSCRITS

Les manuscrits se conformeront aux exigences suivantes:

1. Longueur et format
Les contributions ne devraient pas excéder 20 pages dactylographiées, de format A4, avec double espacement et marges suffisantes (environ 2,5 cm), si possible en format électronique (disquette 360KB à 1,44MB), sous forme de fichier ASCII, complètement formaté sous Microsoft Word ou sous WordPerfect (DOS ou Windows). La disquette sera accompagnée d'un tirage papier. Chaque article sera pourvu d'un résumé en anglais de 150 à 200 mots, et de 10 à 30 mots-clés.

2. Tableaux et graphiques
Les dessins, tableaux, graphiques et diagrammes seront envoyés, soit sous leur forme originale soit sous forme d'une copie de bonne qualité. Leur place dans le texte devra être clairement indiquée.

3. Bibliographie et références dans le texte
Voir les exemplaires récents de *Lexikos*.