AFRILEX -reeks series 7:1997

LEXIKOS 7

BURO VAN DIE WAT

AFRILEX

Lexikos 7

Lexikos 7

Redakteur

Editor

J.C.M.D. du Plessis

Resensieredakteur

Review Editor

E. Botha

AFRILEX African Association for Lexicography

BURO VAN DIE WAT

STELLENBOSCH

Menings wat in artikels en resensies uitgespreek word, is nie noodwendig dié van AFRILEX of die Buro van die WAT nie

Opinions expressed in the articles and reviews are not necessarily those of AFRILEX or of the Bureau of the WAT

Adviesraad / Advisory Board

Redaksiekomitee / Editorial Committee

Inhoud / Contents

Beskouende artikels / Contemplative Articles

Projekte / Projects

Leksikovaria / Lexicovaria

Verslae / Reports

Resensieartikels / Review Articles

Resensies / Reviews

Publikasieaankondigings / Publication Announcements

Voorwoord

Sedert die verskyning van die eerste nommer van *Lexikos* het die tydskrif in 'n toenemende mate daarin geslaag om kommunikasie tussen leksikografiese vakgenote, nie alleen in Afrika nie, maar ook in die res van die wêreld te laat plaasvind. Met die totstandkoming van AFRILEX waarvan *Lexikos* nou die offisiële mondstuk geword het, kon dié kommunikasie tot 'n nog groter mate voortgesit word. Bekendheid met en kennis van mekaar se projekte en werksaamhede bring die moontlikheid mee om kundigheid uit te ruil en samewerking te bewerkstellig. Hieraan het die leksikograwe in Afrikalande met hul ryke taalverskeidenheid 'n groot behoefte.

In die verlede is daar beskrywings van leksikografiese en elektroniese projekte en verslae oor konferensies en werkseminare in *Lexikos* gepubliseer. Nou wil *Lexikos* 'n nog groter geleentheid vir sodanige kommunikasie bied. Om dié rede is twee bykomende rubrieke in hierdie nommer van *Lexikos* ingevoer: **Projekte / Projects** en **Verslae / Reports**. Lesers word uitgenooi om beskrywings van hul projekte en verslae oor konferensies en werkseminare voor te lê. Bydraes tot hierdie rubrieke moet aan die eise van akademiese geskrifte voldoen, maar sal nie noodwendig deur keurders geëvalueer word nie, aangesien hulle nie as navorsingsbydraes bedoel is nie.

Proff. A. Delbridge, R.H. Gouws en M.H. Heliel word as nuwe lede van die Adviesraad verwelkom. 'n Woord van verwelkoming gaan ook aan die nuwe lede wat tot die Redaksiekomitee toegetree het. Die uittredende lede word bedank vir die bydrae wat hulle gelewer het om artikels vir *Lexikos* te keur. Die Redaksiekomitee vervul 'n baie belangrike rol om die akademiese gehalte van die tydskrif te verseker. Dit word weerspieël in die bydraes in hierdie nommer van *Lexikos*. Die lede van die Redaksiekomitee moet gelukgewens word met die deeglike en gesaghebbende wyse waarop hulle artikels gekeur het.

Die nuwe resensieredakteur van *Lexikos*, mnr. Etienne Botha, word bedank vir sy bydrae. Hy het gesorg dat 'n interessante verskeidenheid publikasies bespreek en vermeld word in die rubrieke waarvoor hy verantwoordelik is: **Resensieartikels / Review Articles, Resensies / Reviews** en **Publikasieaankondigings / Publication Announcements**.

Die lede van die afdeling Redaksionele Steundienste van die Buro van die WAT, mev. Hermien van der Westhuizen, mnr. Etienne Botha en mej. Tanja Harteveld, verdien 'n spesiale woord van dank vir hul aandeel in hierdie nommer van *Lexikos*. Nie alleen het hulle al die tydrowende administrasie en korre-

spondensie behartig nie, maar mev. Van der Westhuizen en mej. Harteveld het ook met professionalisme, entoesiasme en baie geduld die dikwels moeilike elektroniese setwerk gedoen, terwyl mnr. Botha met groot noukeurigheid gehelp het met die taalversorging van 'n aantal van die artikels.

J.C.M.D. du Plessis

Buro van die Woordeboek van die Afrikaanse Taal

Foreword

Since the publication of the first issue of *Lexikos* the journal has increasingly succeeded in effecting communication among lexicographic colleagues not only in Africa, but also in the rest of the world. With the establishment of AFRILEX, of which *Lexikos* has now become the official mouthpiece, this communication can be pursued to an even greater degree. Awareness and knowledge of each other's projects and activities open up the possibility of exchanging know-how and bringing about collaboration. There is a great need for this among lexicographers in African countries with their rich diversity of languages.

In the past, *Lexikos* published discussions of lexicographical and electronic projects and reports on conferences and workshops. Now, however, *Lexikos* seeks to offer an even greater opportunity for such communication. For this reason two additional categories have been introduced in *Lexikos*: **Projekte / Projects** and **Verslae / Reports**. Readers are invited to submit discussions of projects and reports on conferences and workshops. Contributions to these categories must meet the requirements of academic writing, but will not necessarily be evaluated by the judges, because they are not meant to be research articles.

Proff. A. Delbridge, R.H. Gouws and M.H. Heliel are welcomed as new members of the Advisory Board. A word of welcome also goes out to new members who have joined the Editorial Committee. The Editorial Committee fulfils a very important function in ensuring the academic standard of the journal. This standard is reflected in the contributions in this issue of *Lexikos*. The members of the Editorial Committee must be congratulated on the thorough and expert way in which they judged the articles.

The new review editor of *Lexikos*, Mr Etienne Botha, is thanked for his contribution. He managed to have an interesting range of publications discussed and listed in the three categories for which he is responsible: **Resensieartikels / Review Articles, Resensies / Reviews** and **Publikasieaankondigings / Publication Announcements**.

The members of the division Editorial Support Services in the Bureau of the WAT, Mrs Hermien van der Westhuizen, Mr Etienne Botha and Miss Tanja Harteveld, deserve a special word of thanks for their part in this issue of *Lexikos*: not only did they handle all the time-consuming administration and correspondence, but Mrs Van der Westhuizen and Miss Harteveld also did the often

difficult electronic typesetting with professionalism, enthusiasm and much patience, while Mr Botha helped to edit the language of some of the articles with great accuracy.

J.C.M.D. du Plessis

Bureau of the Woordeboek van die Afrikaanse Taal

'n Woord van AFRILEX

As 'n vakvereniging vir leksikograwe van Afrika is dit vir AFRILEX belangrik om aan sy lede 'n gespreksforum te bied waar leksikografiese aangeleenthede wat vir Afrika relevant is, aan die orde kan kom. AFRILEX streef ook samewerking met ander leksikografieverenigings na en beywer hom vir 'n ledetal wat verder as die grense van Afrika strek. Gevolglik is dit belangrik dat die Afrikagerigtheid van die vereniging ondersteun word deur 'n inklusiewe benadering wat AFRILEX ook toeganklik en aantreklik maak vir belangstellendes van buite Afrika. By AFRILEX se tweede internasionale kongres wat in Julie 1997 in Durban gehou is, is dit dan ook beklemtoon dat die vereniging se bedrywighede op sowel die Afrika as die algemene leksikografie gerig moet wees. 'n Gespreksgeleentheid vir taalspesifieke en algemeen teoretiese aspekte moet deur AFRILEX geskep en bevorder word.

Dit is belangrik dat die inklusiewe benadering ook in *Lexikos*, as offisiële mondstuk van AFRILEX, op die voorgrond moet tree. Met *Lexikos* 7 gebeur juis dit. Selfs 'n vlugtige kyk na die inhoudsopgawe gee blyke van 'n verskeidenheid onderwerpe wat behandel word. Die Afrika en die algemene, die teoretiese en die praktiese leksikografie kom duidelik in hierdie aflewering van *Lexikos* na vore.

Lexikos is AFRILEX se tydskrif en lede word versoek om die tydskrif aktief te ondersteun deur navorsingsartikels en resensies vir moontlike publikasie aan die redaksie voor te lê.

Vir sy voortgesette betrokkenheid by AFRILEX en sy entoesiastiese handhawing van *Lexikos* se posisie as 'n toonaangewende vaktydskrif, spreek AFRILEX sy hartlike dank uit teenoor die Buro van die WAT, en veral dr. J.C.M.D. du Plessis wat as redakteur van hierdie sewende uitgawe opgetree het.

Rufus Gouws
Voorsitter: AFRILEX

A Few Words from AFRILEX

As an association for lexicographers of Africa, it is important that AFRILEX should offer its members a discussion forum where the emphasis is on matters relevant to lexicography in Africa. However, AFRILEX also seeks cooperation with other lexicographic associations and endeavours to accommodate members from beyond the borders of this continent. It is therefore essential that the focus on Africa be complemented by an inclusive approach which makes AFRILEX accessible and attractive to interested lexicographers not working in Africa. At the second AFRILEX international conference held during July 1997 in Durban, it was emphasised that the activities of the association should be aimed at both African and general lexicography. AFRILEX has to create and promote opportunities to discuss language-specific as well as general theoretical issues.

It is important that this inclusive approach should also come to the fore in *Lexikos*, the official mouthpiece of AFRILEX. This is exactly what happens in *Lexikos* 7. Even a cursory glance over the table of contents gives evidence of the variety of issues treated in this volume. African and general, theoretical and practical lexicography are well represented in this issue of *Lexikos*.

Lexikos is the journal of AFRILEX and members are invited to support it actively by submitting research articles and reviews for possible publication.

The Board of AFRILEX expresses its sincere thanks to the Bureau of the WAT and especially to Dr J.C.M.D. du Plessis, editor of this seventh issue, for continued commitment to AFRILEX and for enthusiastically maintaining the status of *Lexikos* as a leading professional journal.

Rufus Gouws
Chairperson: AFRILEX

Redaksionele doelstellings

Lexikos is 'n tydskrif vir die leksikografiese vakspesialis en word in die AFRI-LEX-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 leksiko-grafiese 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, rekenaarweten-skap, 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* (AFRI-LEX) 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 afsonder-like 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 han-del, 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 spora-dies 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 ist 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 dient *Lexikos* als amtliches Mundstück des *African Association for Lexicography* (AFRILEX), u.a. weil das Büro des WAT gerade das ausgesprochene Ziel mit der Ausgabe der AFRILEX-Serie hatte, die Gründung solches 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 zum lexikographischen Gespräch 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.

Issues in the Planning of a Multilingual Explanatory Dictionary of Chemistry[1] for South African Students

Adelia Carstens, *University of Pretoria, Pretoria, South Africa*

Abstract:. Developing human potential and actively promoting science and technology are among the priorities of the present South African government. Significant progress in these areas can only be made if relevant education and training are provided timeously. Surveys conducted by overseas as well as local researchers indicate that mother-tongue education is one of the measures to improve learning. While mother-tongue education at secondary and tertiary level is at present not provided to speakers of African languages, compromise solutions have to be offered, such as multilingual explanatory special-field dictionaries. By providing linguistic and encyclopedic information in English, Afrikaans and two or more of the African languages the concepts of the subject-field are made accessible via the language(s) of wider communication as well as via the mother tongue of the student. Although there are many difficulties and potential pitfalls awaiting terminographers and special-field experts who attempt a project of this nature, substantial preliminary work that has already been done with regard to a quadrilingual explanatory dictionary of chemistry indicated that this ideal is not out of reach. The compilers believe that successful completion of the project, as well as favourable results, will verify the hypotheses that served as points of departure for the project.

Keywords: CHEMISTRY, EXPLANATORY DICTIONARY, MULTILINGUAL DICTIONARY, TERMINOGRAPHY, PEDAGOGICAL DICTIONARY, SPECIAL-FIELD DICTIONARY, SCIENCE AND TECHNOLOGY, TRANSLATORY DICTIONARY

Opsomming: Aspekte van die beplanning van 'n meertalige verklarende Chemiewoordeboek vir Suid-Afrikaanse studente. Die ontwikkeling van menslike potensiaal en die aktiewe bevordering van wetenskap en tegnologie ressorteer onder die belangrikste prioriteite van die huidige Suid-Afrikaanse regering. Betekenisvolle vooruitgang op hierdie terreine kan egter slegs gemaak word indien relevante onderrig en opleiding tydig voorsien word. Ondersoeke deur sowel oorsese as plaaslike navorsers het getoon dat moedertaalonderrig een van die maniere is waarop groter sukses met leeraktiwiteite behaal kan word. Aangesien moedertaalonderrig op sekondêre en tersiêre vlak tans nie vir sprekers van die Afrikatale beskikbaar is nie, moet kompromie-oplossings aangebied word, byvoorbeeld die voorsiening van meertalige, verklarende vakwoordeboeke. Deur linguistiese en vakkundige inligting in Engels, Afrikaans en sommige van die Afrikatale aan te bied, word toegang tot die konsepte van die betrokke vakkundige terrein verleen deur middel van sowel die breër kommunikasietale as die student se moedertaal.

2   Adelia Carstens

Hoewel daar vele probleme en slaggate op die leksikograwe en vakkundiges wag wat so 'n projek aanpak, het die werk wat reeds in verband met 'n meertalige verklarende Chemiewoordeboek gedoen is, getoon dat hierdie ideaal tog haalbaar is. Die samestellers is daarvan oortuig dat die suksesvolle afhandeling van die projek, asook gunstige resultate, die hipoteses sal verifieer wat as vertrekpunte vir die studie gedien het.

Sleutelwoorde: CHEMIE, MEERTALIGE WOORDEBOEK, PEDAGOGIESE WOORDE-BOEK, TERMINOGRAFIE, VAKWOORDEBOEK, VERKLARENDE WOORDEBOEK, VERTA-LENDE WOORDEBOEK, WETENSKAP EN TEGNOLOGIE

1.   Introduction

Recently published Government policy documents on transformation (e.g. the *White Paper on Reconstruction and Development* (1994), *South Africa's Green Paper on Science and Technology* (s.a.) and the *White Paper on Science and Technology* (1996)) strongly emphasise education and training in the fields of science, engineering and technology. This process of developing human potential through access to information is not merely a measure to redress imbalances created by the past political system, but also a way to empower the people of the country and to stimulate the economy.

This article argues that multilingual, explanatory special-field dictionaries can be implemented to realise these objectives by providing easy access to new or incompletely learnt concepts[2]. For those who have expressed scepticism about terminographical work for the African languages (cf. Louwrens 1997; Mutasa 1996[3]) there is the consolation that the dictionary type in question does not exclude participation in scientific and economical domains by means of a language of wider communication such as English (cf. Cluver 1996: 1, 7). By selecting English as the source language of the main lemma list and by also providing accessible definitions in English, such a dictionary could indeed contribute towards improving the user's proficiency in the English special language for that specific subject-field.

In the subsections below the relationship between the objectives of terminology and terminography (the domains concerned with the documentation of special languages), and the needs which gave rise to the dictionary concept in question, will be expounded.

2.   Objectives of terminography

According to Cluver (1989: 8) "the primary objective of terminographical work is (thus) to ensure, firstly, that each identifiable technical concept is clearly named by a technical term, and that the concept is adequately described in a definition". "Secondly," he adds, "terminography attempts to ensure that the term and its definition are accepted by all or the majority of the practitioners of

a specific subject field." A third objective of terminographical work is, in his opinion, deduced from the second, namely to ensure maximum clarity and exactness in technical or scientific communication.

Cluver, however, criticises this view of terminography as a "traditional" or "structuralist" view which only focuses on the documentation, systematising, defining, standardisation and recording of technical terms in dictionaries, thesauri or term banks — thereby restricting the scope of the discipline to fully-standardised languages with well-developed terminologies (cf. Cluver 1989: 10). He advocates a sociolinguistic approach (cf. Cluver 1989: 9) according to which terminography is not only concerned with standardisation, but also with the elaboration of the technical vocabularies of developing standard languages. The development of technical vocabularies should, however, never be an end in itself, but rather (according to Cluver 1996: 1) an instrument for the speakers of that language "to gain access to modern information and technology that is needed for them to become independent members of the modern information society". Cluver (1996: 1) further asserts that "this objective includes enabling people to gain access to learning institutions via their own language".

A third contribution that the creation of terms in the vernacular of the people can make, according to Cluver (1989: 9; 1996: 1), is the promotion of nationalism by enhancing the status of the language with its own speakers. This could again lead to increased unity within the speech community, provided that the speakers are ready to put their vernacular to use in "prestigious" spheres where only a majority language such as English has heretofore been considered appropriate (cf. Cluver 1987: 27). Rey (1995: 52) echoes this view by contending that "the specific task of terminological intervention is to improve the quality of linguistic communication and the relations between speakers and their language".

These extended views of the objectives of language elaboration and functional expansion by means of terminology takes into account systematic and normative considerations as well as sociolinguistic and ideological concerns. All of these views, however, focus primarily on the naming aspect of terminology, and not on cognitive functions. Terminographical work with regard to developing languages also include the formulation of definitions in the mother tongues of the people, which may serve as vehicles of conceptualisation and possibly also as starting points for technical communication in the vernaculars of the people of South Africa. This dimension is implicitly reflected in Rey's (1995: 105) multi-faceted motivation for comprehensive terminological work in developing languages:

> From this viewpoint the analysis of the terminological needs arising from cultural, didactic, scientific and technological change depends on specific sociolinguistic factors. In some cases there is not even an admission of needs and the languages concerned are confined by historical circumstances to other functions, their speakers being obliged to learn one of

the dominant languages of the respective subject-field. Each language is capable of naming everything; it is a political decision whether people are allowed to develop terminologies in their mother tongue, or in a 'national' language, or whether they have to resign themselves to borrowing a vehicular language for a particular subject field. The impression that certain languages cannot supply the needs of conceptual structures is purely ideological.

3.   Terminographical needs of African language-speakers

3.1   Pedagogical needs

According to the *White Paper on Science and Technology* issued by the Department of Arts, Culture, Science and Technology in 1996, democracy (participation of all the people of the country in all public domains — AC) is dependent upon the availability and accessibility of information:

> Democracy implies being aware of choices and making decisions. The extent to which this is possible depends largely on how much information is available to the people and how accessible it is (1996: 50).

Considering the low grades in science and mathematics achieved by matriculants of underprivileged backgrounds during the almost three years since the official demise of apartheid, one realises that the current system of education is not yet geared to meet the requirements of a science and technology-oriented South Africa. It seems that mother-tongue speakers of African languages continue to experience immense difficulties in mastering the empirical and the natural sciences, thereby remaining outside the mainstream of economic activity.

The hypothesis that language is one of the most important stumbling blocks in gaining access to the natural sciences is verified by a survey that Kwesi K. Prah undertook in 1992 among university students of Botswana, Namibia, Swaziland-Kwaluseni, the Western Cape, the Transkei and Lesotho (Prah 1995). Regarding knowledge of different languages, the answers to his questionnaire indicated that all the respondents knew English, 79% knew other languages apart from English and/or Afrikaans, and 58% were familiar with three or more languages. For the overwhelming majority of the students from the different universities, the language they knew best was their mother tongue. About 14% of the students indicated that they had problems studying in English (which is of course a subjective perception, and the percentage may not be an accurate reflection of the realities). The following categories of responses were given to describe the types of problems faced by students studying in English:

- Poor teaching of English at primary and secondary school levels
- Primary and secondary education were received in Afrikaans-medium schools
- Writing pace in English too slow for lectures
- Problems with English spelling
- Problems with English grammar and syntax
- Textbook-English complicated
- Problems with pronunciation
- Language problems undermine concentration in class
- Limited vocabulary
- Frequent need to translate from mother tongue to express ideas

To the question "If science and technological studies were taught and conducted in your mother tongue, do you think you could find your studies easier?" more than 59% indicated that studies would be easier if the programme were conducted in the mother tongue. Questioned regarding the wider societal benefits they could foresee if modern scientific and technological ideas were rendered into African languages, the statistically leading views of the respondents were the following:

- The rural masses will be rapidly educated
- Africans will gain more confidence in their cultures and history
- Africans will become inventive
- African society as a whole will come to understand scientific and technological ideas
- The power and influence of the present elite will be diminished
- Africans will do better in their studies
- There will be wider and broader job markets
- African languages will be greatly enriched
- There will be a great increase in African scientists and technicians

These statistics compare well with the findings of a preliminary survey on the medium of instruction used in science classes at a number of high schools in Pretoria where Northern Sotho is the mother tongue of most of the pupils. It became evident that teachers quite often switch to the vernacular during science lessons and practical sessions, primarily for two reasons:

- the teacher "feels more comfortable when presenting classes in the mother-tongue", which may probably indicate that he/she is not proficient enough in English to give a precise and clear explanation of a particular concept, procedure, etc.;
- many pupils are not proficient enough in the second language (English) to follow explanations, instructions and arguments, and to express themselves in English.

It is therefore not surprising that more than 80% of Prah's respondents chose "very much so" on a five point scale in answer to the question "Do you think the task of rendering and translating modern scientific and technological ideas into African languages should be a key national issue?"

On the basis of Prah's (1995) survey among university students, as well as the preliminary survey conducted among secondary school teachers of science in Pretoria, it is assumed that mother-tongue education could play a major role in providing access to science and technology. It is further believed that multilingual explanatory special-field dictionaries could serve as important learning aids during the secondary and tertiary phases of education, especially if the emphasis is placed on clear and simple conceptual definitions in the mother tongue. Such explanations could assist students in the process of conceptualisation which has, due to the apartheid education system, not developed naturally in many speakers of African languages.

There are, however, those who argue that term creation (translatory activities) is a prerequisite for explanatory activities such as teaching and defining. Matšela (1987: 80) is but one of the scholars who seems to hold such an opinion:

> One of the major problems of introducing into African languages the teaching of such subjects as Agriculture, Biology, Chemistry, Physics, Mathematics, Engineering, Linguistics, Philosophy, Psychology and the like, is the lack or dearth of relevant specialized technical terminologies in those languages.

This view seems logical if it is assumed that a person needs at least a partially developed special language to undertake actions such as teaching or defining. On the other hand none of the terms used in dictionary definitions for pedagogical purposes need to claim terminological status. Furthermore, if non-standardised terms or terms of different styles and registers are used in definitions there ought not to be any objection to entering them as surrogate equivalents in the macrostructure of a multilingual dictionary in order to avoid lemmatic gaps. Such forms could be clearly marked (by means of a specific graphic symbol) to indicate their temporary status.

Another danger that could be associated with focusing solely on the translatory aspect of terminological work for the African languages is that a proliferation of "artificial" terms could be coined by terminographers and special-field experts, which may be rejected entirely by the speakers of the language in question. Rey (1995: 18-22) warns that a mere desire to describe (for instance by large-scale coining of translation equivalents — AC) "would condemn terminology to impotence, or to the modest state of a translation or documentation aid, when in reality it is an indispensable body of knowledge for satisfying a fundamental need which precedes all social planning".

In a domain such as chemistry there are also the requirements of international standardisation organisations, such as IUPAC (International Union of

Pure and Applied Chemistry), that have to be considered when assigning terminological status to linguistic forms. According to Alain Rey (1995: 14) the differentiating descriptive systems and the organisation of relevant characteristics in modern chemical nomenclature are based on "a knowledge of processes and functions and no longer only on characteristics which can be described in terms of space and appearance according to their place in a table". These universal guidelines are not to be ignored.

Regardless of whether a terminographer places the main emphasis of his / her work on defining or term-creation, all morphological and syntactic coinings, borrowings, meaning transfers and indigenised terms (transliterations), etc., should be tested widely for their acceptability with mother-tongue speakers, as well as for their true reflection of the internal structure of the subject-field. The editorial team of any multilingual dictionary in South Africa should therefore include mother-tongue speakers of the languages in question who also have profound knowledge of the subject-field in question.

3.2   Economic needs

As mentioned above, two of the primary aims of the Reconstruction and Development Programme of the South African Government are redressing the past unequal distribution of resources, and developing science and technology for economic purposes (see chapter 3, paragraph 3.13.2, of the Government's *White Paper on Reconstruction and Development* of September 1994).

This view is broadened in the *Green Paper on Science and Technology* which characterises science, engineering and technology (SET) as the "absolutely vital components of economic and social progress", the development of which is largely dependent on the educational sector. The *Green Paper* assigns great importance to interaction and co-operation between industry, scientific councils, universities and state departments in order to create a coherent system that will contribute towards achieving the national, social and economic goals of the country (p. 7). Prominence is given to the creation and maintenance of a "scientific infrastructure" which will provide a trained work force as well as the necessary hardware and software. It is stressed (p. 15) that apartheid education has denied black people access to science-based careers, resulting in the current situation "that there are few black engineers and technologists, and not enough black students studying for postgraduate qualifications in these fields". Quick and effective action has to be taken to diminish the imbalances regarding the way SET is practised in schools, colleges, technikons and universities. "We must therefore stop considering terminology as being limited to conceptual and linguistic analyses recorded on cards or in databases," stresses Rey (1995: 53). "The social, political and economic connections are essential because they create … financial support."

In the following subsection the relationship between user-needs and ter-minographical principles, as pertaining to the quadrilingual[4] dictionary of chemistry discussed in this article, will be expounded briefly.

4.   Special user-needs and the gradedness of terminographical principles

Two of the most important steps in the planning of a special-field dictionary are considering the users and functions of the intended dictionary (cf. Bergenholtz and Tarp 1995: 19). These considerations must include:

(a)   the text types for which the dictionary is intended;
(b)   the user-groups the dictionary is aimed at; and
(c)   the communicative function of the dictionary, namely reception, produc-tion and/or translation.

4.1   Text type

According to Bergenholtz and Tarp (1995: 18) the text corpus of the dictionary must be composed in such a way that it corresponds with the text types to be encoded or decoded by the users. First of all the fact should be recognised that scientific and technical language exhibit "different degrees of expertise". To structure the principle of technical gradability the following text types (applica-ble to specific user-groups) are distinguished by Bergenholtz and Tarp (1995: 19):

(a)   from expert to expert (expert language, such as used in scientific journals — AC)
(b)   from expert to semi-expert (the language for special purposes (LSP) of textbooks, etc.)
(c)   from semi-expert to semi-expert (jargon and the language used in texts by experts from related areas)
(d)   from semi-expert to layman (the language used in popularised texts with scientific or technical topics)

Textbooks for senior secondary pupils and undergraduate university or tech-nikon students could be categorised under (b) above, if it is argued that "semi-experts" also include "semi-laymen". This text type should then play a central role in the collection of data for a dictionary which is to fulfill the terminologi-cal needs of the mentioned user-group. If the study matter comprises a specific textbook or textbooks the terminographer and the expert(s) representing the subject-field should systematically check the alphabetical index of such a publi-cation. The metalanguage of a dictionary compiled for the tertiary education

sector (semi-experts) and the secondary education sector (ranging from semi-experts to almost laypersons) should moreover not be more complex than that of the prescribed textbook(s).

4.2   User-group

Besides text types and functions which have to be considered, every technical dictionary must be compiled with a specific user-group in mind. When a dictionary aimed at a foreign target language is compiled, two important factors have to be taken into account, namely encyclopedic knowledge (knowledge of the subject-field) and foreign-language competence. Using binary values with regard to these parameters, yet recognising that the transition between the different types is fluid, Bergenholtz and Tarp (1995: 21) distinguish four main user-types:

encyclopedic competence

experts	high level of encyclopedic and low level of foreign-language competence	high level of encyclopedic and high level of foreign-language competence
laypeople	low level of encyclopedic and low level of foreign-language competence	low level of encyclopedic and high level of foreign-language competence
	noncompetent	competent

foreign-language competence

Although communication in technical and scientific fields ideally requires that "the user already possesses the configuration of knowledge which determines the role of the term in a structured system ... the limiting case of this restriction is, of course, the requirement that a new term be learnt at the same time as new knowledge, which may include the addition of one or two additional dimensions to an imperfectly learnt concept" (cf. Sager, Dungworth and McDonald 1980: 75). In the case of undergraduate South African students with fragmented knowledge of the conceptual system of the subject-field, even more than one or two additional aspects of the concept might have to be learnt together with a term.

Presenting information in an accessible format for the student with incomplete encyclopedic and second-language competence seems to be one of the problem-solving agents in the current situation, but may require certain concessions from the lexicographer. Two of the basic terminographic requirements, namely precision and economy, might have to be compromised in order

to accommodate the third, which is appropriateness (Sager, Dungworth and McDonald 1980: 206 ff.).

Economy may have to be compromised by presenting more encyclopedic information, and even duplicating conceptual information by using illustrations which partially overlap with the conceptual information supplied verbally. Furthermore, the lexicographer(s) may deem it necessary to label certain information categories explicitly by means of symbols or abbreviations. In a subject-field such as chemistry *empirical formulae, symbols, synonyms, opposites* and *abbreviations* might need labelling in order to facilitate unambiguous transfer of information. Such explicitness is, however, uneconomical in terms of dictionary-space and will necessarily have an influence on the price.

The principle of precision may have to be relinquished partially in order to attain understandability, transparency and lucidity (i.e. accessibility). Precision is a concept that should be given content with reference to the situation, which include *inter alia* the level of scientific and technological communication, the intention of the encoder, the competence of the decoder, etc. (cf. Swanepoel 1990: 27). However, given the fact that scientific and technical dictionaries are aimed at providing an accurate picture of a highly structured conceptual field, precision is only negotiable in as far as standardisation is maintained; in other words as long as the definition remains "universally understood" within that specific subject-field (cf. Sager, Dungworth and McDonald 1980: 207). This aspect will be discussed again later with regard to specific examples in the dictionary project described here.

4.3   Communicative function of the dictionary

The third consideration to be taken into account is the communicative function(s) which the dictionary is intended to fulfil, namely that of production (native language or foreign language), reception (native language or foreign language) or translation (into or from a foreign language), and the information categories that will support the intended function(s).

Bergenholtz and Tarp (1995: 25) warn that where a specialised dictionary is designed for users with different mother tongues, an integration of the above functions presents problems. The problems are magnified when the user-group is heterogeneous with regard to encyclopedic and second-language proficiency. Such is e.g. the case with South African undergraduate students of the natural sciences who come from diverse educational backgrounds and language groups, having to conduct their studies entirely through English or Afrikaans.

The following section will deal with problems faced by the terminographer who wishes to compile a scientific dictionary for pedagogical purposes aimed at a heterogeneous user-group such as outlined above.

5. Problems regarding the overall planning of a multilingual explanatory dictionary of chemistry for pedagogical purposes

5.1 Data-collection

According to Bergenholtz and Tarp (1995: 90) a dictionary may be empirically based on introspection, existing literature and/or texts.

5.1.1 Introspection

Introspection indicates that the lexicographer relies solely on his/her own competence. Special-field dictionaries do, however, require a more differentiated competence as a basis of introspection, since general-language competence cannot automatically be translated into special-language competence. Introspection should, therefore, also involve active participation by experts in the subject-field (cf. Arntz and Picht 1989: 224). According to Wiegand (1990: 2207) one special-field expert should not alone decide which terms to include as lemmas in a general dictionary. This principle is just as valid for special-field dictionaries: in the case of a subject-domain such as chemistry it is important to involve at least one expert on each of the main fields, namely organic, inorganic, physical and analytic chemistry. The contributions of the various experts must be mentioned explicitly in the outside matter of the dictionary, as they are the persons who are ultimately responsible for the scientific correctness of subject-field (encyclopedic) information included in the dictionary. Lexicographers and terminographers can only intervene in an auxiliary capacity, e.g. advising on linguistic appropriateness (cf. Rey 1995: 58). If a multilingual dictionary is planned the introspective approach requires the participation of a considerable number of lexicographers, preferably mother-tongue speakers of the languages in question.

The "plural" form of introspection is multispection, which entails that questions have been asked of a large number of informants. In a multilingual country where different major languages enjoy different statuses and have different functions multispection could be used for more than one purpose: firstly to legitimise a dictionary among the speakers of the various language groups involved, secondly to keep in touch with linguistic realities (e.g. by not recklessly coining equivalents where terms already exist and already enjoy a measure of standardisation in the relevant linguistic communities), and thirdly continuously to monitor the relevance and usefulness of the dictionary with regard to the communicative functions it has to perform and the level at which these functions have to be performed. In the case of special-field dictionaries intended for the senior secondary and tertiary levels, lecturers and teachers should be included in multispection so as to ensure that the dictionary stays in tune with the encyclopedic and the linguistic proficiency of the user-group.

5.1.2   Existing literature

According to Bergenholtz and Tarp (1995: 93) part of the empirical basis of an LSP dictionary may not only be the utilisation of existing literature, primarily dictionaries and encyclopedias, but also other reference works, such as handbooks, textbooks, scientific articles and monographs. The use of existing dictionaries has the advantage of being fast, but the method cannot be considered reliable if the lemma list of the dictionary consulted is not checked against the present stage of development of the subject-field in question as well as the scope intended for the dictionary in preparation. This does, however, become less of a problem if consultation of existing literature is combined with multi-spection which includes careful scrutiny by experts and other role-players in the subject-field.

Although it may not be necessary to state titles that have been used only a few times, the dictionary front matter should contain a list of the most frequently consulted sources. The preliminary research for the dictionary project described in this article revealed that in some existing dictionaries of chemistry there are only very vague references to sources used. The editor of the *Dictionary of Chemical Terminology in Five Languages* for instance only states in the preface that "the dictionary was compiled on the basis of the contemporary literature in chemistry and related disciplines in the relevant languages (handbooks, monographs, encyclopedias, scientific journals, various IUPAC and ISO publications, etc.)". In other dictionaries the requirement of "proving the source" (one of the principles of dictionary-making on which H.E. Wiegand places a very high premium) is completely ignored, such as *Longman Illustrated Dictionary of Chemistry* which mentions Arthur Godman as the sole author/editor/compiler without reference to any co-author or written sources that were consulted. It is hard to believe that one single lexicographer could have had the lexicographical, linguistic and encyclopedic proficiency to compile an entire special-field dictionary.

5.1.3   Texts

Building up a text corpus which is statistically representative of a particular subject-field could be regarded as the most reliable empirical basis. This is, however, a time-consuming, costly and labour-intensive operation which few multilingual countries in Africa can afford. Furthermore many of the terms of the exact and the natural sciences have fixed scientific definitions determined by international standardisation organisations, such as IUPAC. Such terms need not be represented in a citation corpus in order to elicit their meanings, but corpora could be helpful in determining the collocability of such terms if illustrative examples are given in a dictionary. If one takes into consideration that there is no (or almost no) existing literature on the natural sciences in any

one of the African languages, text corpora are not the answer to all the questions.

In the current South African situation, with special reference to the making of a multilingual explanatory dictionary of chemistry, a combination of the above-mentioned methods seems to be advisable, namely scanning textbooks and other authentic study material, verifying this material by comparison with existing dictionaries of chemistry, and making use of multispection. The latter could *inter alia* include:

1. delimitation of the macrostructure by requesting departments of chemistry from a wide spectrum of universities and technikons to comment on a preliminary term list;
2. identification of the criterial conceptual features for each terminological category;
3. determining information categories by making use of the metalexicographic knowledge of lexicographers combined with knowledge of the subject-field and its requirements;
4. deciding whether to include existing terms in the African languages that are not completely consistent in terms of IUPAC requirements, or rather to form transliterations of English terms, but risking rejection by speakers of the languages; and
5. testing coinages in the African languages among students as well as professional chemists with an African language as their mother tongue.

5.2   Problems regarding general organisation

In most traditional dictionary typologies, distinctions such as monolingual roughly correlated with explanatory, and bi-/multilingual correlated with translatory. Special-field/technical correlated with encyclopedic and general correlated with linguistic. Apart from these accepted distinctions and correlations, technical/scientific unambiguously entailed standard and excluded learners'/pedagogical.

The increased emphasis on user-orientation has, however, caused the borders of the traditional typological categories to become fuzzy, permitting the combination of dictionary functions. Certain subject-fields are seemingly better suited to typological hybridisation than others. Bergenholtz and Tarp (1995: 72) say in this regard:

> On the whole, it is much easier to combine mono- and bilingual encyclopedic and linguistic functions in scientific and technical dictionaries than in dictionaries of law and economics, the reason being, of course, that the former are not culture bound and therefore do not have to provide com-

parative encyclopedic information, as do bilingual dictionaries for cul-
ture-dependent LSPs.

Smit (1996: 65) is strongly in favour of hybrid dictionaries for special purposes,
but following Wiegand (1988a: 751) she suggests that the purposes of the dic-
tionary should be made clear, preferably in the title and the subtitle of the dic-
tionary and/or in the dictionary introduction.

Apart from the advantages of multilingual dictionaries — namely the com-
bination of several bilingual dictionaries into one, the possibility of comparing
different languages, and saving space — multilingual dictionaries are liable to
reach excessive proportions and the danger does exist that the microstructure
may become extremely complex. The introduction of semantic/pragmatic
information such as definitions and usage examples should therefore be care-
fully considered before they are included in the dictionary plan of such a dic-
tionary.

Another organisational aspect to be considered carefully is the choice of a
language of explication, i.e. the language of continuity used for headings and
texts belonging to the outside matter of the dictionary. On the one hand, one
could argue that English is the obvious choice in South Africa, being widely
regarded as the national language of the country, the second language of most
dictionary-users, the primary medium of instruction, and the international lan-
guage of science and technology. On the other hand, using English as the
pivotal language could be interpreted as a violation of the democratic language
rights of the speakers of the other languages represented in the dictionary. The
dictionary could resultantly fail as a facilitator of language equality.

A related problem concerns the decision on whether the dictionary should
be bi- or unidirectional, or primarily unidirectional with some bidirectional
characteristics (cf. Bergenholtz and Tarp 1995: 55-56). One variation could for
instance be a monolingual English defining dictionary, to which has been
added a multilingual dimension in the form of equivalents as well as reverse
word lists, e.g. *Dictionary of Chemical Terminology in Five Languages*. Reverse
word lists here serve only the purpose of directing the user to the appropriate
English lemma, in other words they are only reference lists. A second type
could be a multilingual explanatory dictionary with English serving as the lan-
guage of lemmatisation in the primary word list and in which all the linguistic
and encyclopedic information supplied for English lemmas are also supplied
for the translation equivalents. Reverse word lists for the other three languages
could then be supplied after the primary word list. This is the model chosen for
the dictionary discussed in this article. A third variation would be a multidi-
rectional, multilingual explanatory dictionary, providing the same information
as the second type above, but allowing access to all microstructural information
via the terms of each language at the appropriate alphabetic position in a spe-
cific section. A dictionary of this type obviously contains much redundancy,

and is also not advisable where some of the languages in question do not yet have fully standardised terminologies.

5.3   Issues regarding data-fields (information categories)

5.3.1   Scope of the discussion

It is obviously not possible to give a full account of all the problems relating to data-fields that were encountered during the first phases of the *Quadrilingual Explanatory Dictionary of Chemistry* (henceforth abbreviated as *QDC*). I shall therefore restrict the discussion to a few specific problems. Furthermore, since the empirical work (data-collection and -processing) on Sepedi[5] and IsiZulu have only just begun, I shall concentrate on issues related to the treatment of the English terms and their definitions. The focus will primarily be on the needs of the intended user-group coupled with an evaluation of the extent to which metalexicography and existing dictionaries of chemistry address these needs. The dictionaries concerned are *McGraw-Hill Dictionary of Chemical Terms* (henceforth M-HDCT), *Glossary of Chemical Terms* (henceforth *GCT*), *Dictionary of Chemical Terminology in Five Languages* (henceforth *DCT*), and *Longman Illustrated Dictionary of Chemistry* (henceforth *LIDC*).

5.3.2   Issues relating to form

Traditional monolingual science dictionaries provide very little linguistic information. This may be ascribed either to the fact that many dictionaries of science and technology have been designed by experts in the subject-field without any linguistic background or interest in grammar, or to the fact that experts, especially in culture-independent subject-fields such as the natural sciences, may consider grammar as less important and not crucial for communication in a scientific or technical domain. One could argue that successful communication does, after all, not depend on correct grammar, but on economy, conceptual precision and accessibility.

Although it may be true that the transfer of information about a specific factual domain is the most important aspect of terminography, a dictionary of the type in question is not adequate without linguistic information such as part of speech, morphological formation, contextual information and information on linguistic and stylistic usage (i.e. labels indicating frequency, temporal markedness, preference, etc.). Brief attention will be given to issues regarding (linguistic) form.

(a)   Syntactic and morphological information

For a learner with low second-language proficiency the listing of derivations can be useful, also from a morphological point of view. Of the above-listed

dictionaries only *LIDC* supply part of speech, e.g. **acid** (*n*). As a bonus to the user with low second-language proficiency *LIDC* supplies, at the end of the article, derived forms of the lemma belonging to other syntactic categories, e.g. **acidify** (*v*) and **acidic** (*adj*). It will, however, still have to be established how this issue should be handled regarding the African languages.

Morphological information in a dictionary of chemistry for semi-experts could also include affixes such as *di-* and *-lysis;* and combining forms such as *allo-, amphi-, morph-, -therm, ortho-* and *-philic.* Sublexical lemmas could either be listed alphabetically in the back matter of a special-field dictionary or as part of the main lemma list. Firstly, their inclusion could explain opaque terms of Greek or Latin origin for the user, and consequently make them easier to remember. Secondly, they could serve as keys to the decompositional interpretation of complex terms which have not been entered as lemmas. Although *LIDC*'s categorisation of sublexical lemmas into "prefixes" and "suffixes" (affixes), and "word parts" (combining forms) is quite arbitrary, the mere inclusion of a comprehensive, fully defined list of sublexical lemmas is commendable.

(b)   Linguistic labelling

Linguistic labelling is the explicit marking of deviations from the major part of the terms described in the dictionary. The deviation may be of different types, such as frequency (e.g. *rare/rarely used*), systematicity (e.g. *trivial name*) and the temporal deictic centre (e.g. *old/obsolete*). For a learner such information is extremely valuable as it explicitly encourages or discourages use of a specific term.

Unfortunately none of the above dictionaries seem to have a clearly defined and consistent policy on labelling. None of them for instance supplies any list of labels in the outside matter of the dictionary. *LIDC* does, however, provide usage information in some cases. **Alcohol** is for instance defined as "trivial name for ethanol". Unfortunately *ethanol* has not been entered as a lemma and the user is left without a conceptual definition. He/she is also not informed that *ethanol* is the preferred IUPAC term. Although *M-HDCT* implicitly refers to *ethanol* in the definition of **alcohol,** no indication is given of the usage restriction on the latter.

In the case of **metalloid** it is only *GCT* that explicates the usage restriction by opening the definition with "An obsolete term formerly used to ...". *DCT* furnishes the synonym *semimetals* in brackets after the lemma **metalloids,** and then includes the following semi-implicit usage information as a "*Remark*" in small print after the definition: "Term was used to denote non-metals."

A consistent way of treating usage-restricted terms in a pedagogical special-field dictionary may be to define a small set of labels in the front matter and then to use them systematically for marking lemmas as well as synonyms

(supplied in the synonym field). The lemma **olefin** could for instance be followed by its part of speech (*n*), then by a usage label (*old term*)/(*trivial name*), after which a cross-reference "*See* **alkene**" could be given. On its part the lemma **alkene** could be followed by the part of speech, the appropriate conceptual definition, the empirical formula C_nH_{2n}, and the synonym *olefin*, labelled as (*old term*) or (*trivial name*).

(c)　Nonlinguistic forms

According to Sager, Dungworth and McDonald (1980: 277) "special languages are able to compress information both syntactically and lexically". In the vocabulary of chemistry this process is *inter alia* facilitated by the creation of abbreviations such as E (energy), STP (standard temperature and pressure) and *amu* (atomic mass unit), letter symbols such as Ω (ohm) and Fe (iron), and by the combination of letters and numbers into short designations (chemical formulae) such as H_2SO_4 (sulphuric acid) and NaCl (sodium chloride). Although these compressed forms may at face value be adding to the text density of a special-field dictionary, they are actually more semantically transparent than their linguistic counterparts which often (especially the trivial names and semi-trivial names) give no indication of the number of atoms present in a specific compound. Compare for instance *sulphuric acid* with its formulaic equivalent H_2SO_4. Formulae have, moreover, become internationalised "so that they can no longer be said to belong to any particular language" (Sager, Dungworth and McDonald 1980: 79). Such forms actually respond to the need for eliminating the references created by one language (cf. Rey 1995: 55). Furthermore the use of acronyms, symbols and other abbreviations facilitate "the most obvious economy" (Sager, Dungworth and McDonald 1980: 316), and also reflect the internal structure of the discipline more precisely (cf. Sager, Dungworth and McDonald and 1980: 320). However, other than their treatment in general bi- and multilingual dictionaries of chemistry, these compressed forms need to be entered consistently into separate data fields. As proposed above, the field name or an abbreviation thereof, e.g. *symbol* or *symb.*, could be printed in the dictionary to facilitate clarity and unambiguous information transfer.

5.3.3　Problems relating to meaning

(a)　The definition

According to Rey (1995: 42) the terminological definition is a compromise between the lexicographical definition and an encyclopedic description. Definitions in the natural sciences are especially problematic in the sense that "the objects are always remodelled constructions which have however to satisfy

conditions which can only be defined by experience and observation." This fact entails that new dictionaries for the natural sciences can never completely rely on the definitions of existing dictionaries, and that thorough introspection by special-field experts is an absolute requirement.

(i)   Types of definitions

The literature on conceptual definitions in terminography reflect different foci, such as the degree of correspondence between the content of the definition and the content of the concept, and the nature of the conceptual features used in the definition. Resultantly various definition typologies have seen the light, as *inter alia* reflected in Arntz and Picht (1989), Bierwisch and Kiefer (1969), Cluver (1989), Dahlberg (1976, 1978) Felber (1985), ISO (1969), Mönke (1978) and Swanepoel (1990).

Swanepoel (1990: 165-166) regards the following definition types as important for terminography:

- the intensional definition (defining a term by mentioning the genus concept as well as those features distinguishing it uniquely from other concepts at the same level of abstraction)
- the extensional definition (the enumeration of species which are at the same level of abstraction, or of all objects belonging to the concept defined)
- the contextual definition (definition by way of an example from actual usage, often a full-sentence containing the term and equating it implicitly with a descriptive definition (cf. Picht and Arntz 1989: 68))
- the operational definition (describing a process or operation by which the referent of the term is realised)
- the ostensive definition (the use of illustrations)
- the synonym definition (the description of the concept by using a synonym)

While the point of view expounded in this article is that conceptual definitions should — in terms of contents, scope and style (cf. Bergenholtz and Tarp 1995: 145) — be prepared with the intended user in mind, the appropriateness of the first four definition types given above will be discussed briefly with regard to QDC.

From a stylistic point of view, taking into consideration the intended users' low encyclopedic proficiency as well as their low proficiency in English, contextual definitions seem to be the most appropriate for QDC. In a multilingual dictionary already carrying a load of translatory information, a contextual definition may facilitate economy by fulfilling the roles of both definition and illustrative example. By including contextual information the dictionary could then also be useful for encoding activities. The following example from QDC shows the advantage of using this type of definition:

> **amphiprotic** *adj.* An amphiprotic solvent has both acidic and basic proper-
> ties.

Apart from its conceptual content the user may also learn from the definition
that the adjective *amphiprotic* often co-occurs with the noun *solvent* (although it
sometimes also co-occurs with *molecule, substance* and *material*).

The contextual definition does not stand in opposition to either the inten-
sional, extensional or operational definitions. In terms of the features included,
a contextual definition may belong to any of these types.

In a subject-field such as chemistry, nouns denoting objects and substances
are usually defined intensionally, e.g.

> **anion** *n.* An anion is an ion with negative charge.
> **barometer** *n.* A barometer is a device used for measuring atmospheric
> pressure.
> **colloid** *n.* Colloids are particles that are larger than normal molecules, but
> are nevertheless small enough to remain suspended in a dispersing
> medium for an indefinite period.
> **enzyme** *n.* Enzymes are proteins that act as catalysts in biochemical reac-
> tions.

Simple but concise definitions like the above, however, often provide too little
conceptual information to enable the student with low encyclopedic proficiency
to conceptualise adequately. In such cases it may be necessary to supplement
the intensional definition with either an extensional definition[6], or with an illus-
tration, or both. The compilers of *QDC* found that a quasi-extensional defini-
tion (naming a few prototypical examples of the concept in question) could in
most cases be integrated elegantly with the intensional definition by using *e.g.*
as a linking-device:

> **base metal** *n.* A base metal is an ordinary metal that is oxidised by air and
> reacts with mineral acids, e.g. iron, lead, tin, zinc.
> **enzyme** *n.* Enzymes are proteins that act as catalysts in biochemical reac-
> tions, e.g. rennin, diastase, amylase, pepsin, zymase.
> **gas** *n.* Gas is one of the states of matter with a definite mass but no definite
> volume and no definite shape. A gas expands to fill the volume of
> its containing vessel, e.g. hydrogen (H_2), carbon dioxide (CO_2), me-
> thane (CH_4).

In this way the abstractness of the definition can be decreased to some extent.
In a number of cases, however, it was found that extensional definitions could
not be linked to intensional definitions by means of a conjunctive device, in
which case it was decided simply to juxtapose the two definition types:

> **ferromagnetism** *n.* Ferromagnetism is a property of certain solid sub-
> stances which causes them to be strongly attracted by magnetic
> fields. Well-known ferromagnetic materials are the elements iron,
> cobalt and nickel and many of their alloys.

It may, however, be argued that the first sentence of the above definition is not
actually of the intensional type, but of the operational type. Operational defini-
tions are especially useful for defining the concepts of variables, functions,
states and processes, as illustrated by the following examples from *QDC*:

> **potential energy** *n.* Potential energy is the energy that an object possesses
> as a result of its position with respect to another object and is equal
> to the work done in reaching that position.
> **reaction rate** *n.* Reaction rate is defined in terms of the decrease in concen-
> tration of reactant molecules or the increase in concentration of
> product molecules with time.

According to Arntz and Picht (1989: 67) the operational definition shows strong
parallels with the intensional definition while the different "operations" may be
viewed as differentiating features (differentia). The compilers of *QDC* decided
to separate intensional and operational definitions from extensional definitions
by keeping them in different data-fields; not only because of the mentioned
similarity between intensional and operational definitions, but also to separate
conceptual information pertaining to the distinctive features of a concept from
information pertaining to subordinate or instantiating concepts (extensional
information). It was argued that this measure could also make computer
searches more exact.

(ii)   The inclusion of technical terms in definitions

A question concerning the foreign-language competence of target-users (if it is
assumed that they would prefer to use the English definitions as well) is to
what extent the metalinguistic vocabulary of definitions could or should be
simplified for pedagogical purposes. Cluver (1989: 105), following Hutchinson
and Waters (1987), seems to be of the opinion that scientific language may be
simplified in cases when nontechnical explanations of technical concepts are
required, such as in science journalism, and that a core vocabulary for defini-
tions should be identified. However, in the majority of cases where the lexi-
cographers involved in the *QDC* project felt that a technical term could be sub-
stituted by a more general term, some special-field experts were of the opinion
that simplification led to incorrect or vague information transfer. Compare the
following definition of **gel**:

> **gel** *n.* A gel is the dispersion of a hydrophilic colloid in water. The result
> appears to be solid but it is easily deformed.

The lexicographers suggested that *hydrophilic* be replaced by *soluble*, as *hydrophilic* is treated as follows in the *Collins English Dictionary*:

> **hydrophilic** *adj. Chem.* tending to dissolve in, mix with or be wetted by water: *A hydrophilic colloid.*

The chemists were, however, adamant that the particles of a colloid are per definition **not soluble; but they attract water and are wetted by it.** Furthermore, some of the experts were of the opinion that the term *particle(s)* is too vague to be used as a substitute for *colloid* and that relinquishing precision in this case would result in a loss of necessary conceptual information.

Also in the case of **strong acid** chemistry experts felt that simplifying the definition would result in incomplete information transfer. According to them a formulation such as "A strong acid is an acid of which the molecules easily undergo chemical changes in solution" is not at all synonymous with "A strong acid is an acid with a high degree of dissociation in solution" (although *dissociation* means "a reversible chemical change of the molecules of a single compound into two or more other molecules, atoms, ions or radicals").

Possible ways of getting around the problem of a metalanguage containing technical terms are to provide paraphrases of difficult terms in brackets after these terms, or to use a type-face/symbol to indicate that such a term is defined elsewhere in the dictionary. *LIDC* uses a system of indicating cross-references by means of up or down arrows. If any term is immediately followed by (↑) or (↓) the user knows that it is entered as a lemma elsewhere in dictionary and can be looked up.

Frequent use of bracketed paraphrases within a definition could however result in large-scale duplication of information, and a proliferation of symbol markings could result in the text becoming too busy. The latter could also be frustrating as the user has to look up multiple lemmas in the dictionary before he/she is able to interpret the definition of the original term.

There does not seem to be a quick and easy solution to this problem, especially in the case of higher-niveau terminology. Each and every problematic case requires careful introspection by lexicographers and chemists.

(b)   Usage examples

A text-field seldom found in traditional special-field dictionaries, especially dictionaries of science and technology, is that of the usage example. None of the dictionaries of chemistry mentioned in this article provides explicit contextual examples such as collocations, competence examples or citations. However, the fairly elaborate and encyclopedic definitions in *GCT* and *LIDC* often implicitly provide usage information, especially collocations. Compare the definition of **initial** in *LICD*:

> **initial** (*adj*) the first part of a process, the first event in time, e.g. the first reading on a burette is the *initial reading*, taken before the contents (↑) are run into a flask; the mass of a substance before it undergoes (p. 213) a chemical change is its *initial mass*; the first temperature recorded in an experiment is the *initial temperature*. (author's italics)

Here *initial reading*, *initial mass* and *initial temperature* not only have an encyclopedic function, but also a linguistic function.

As demonstrated in paragraph 5.3.3(a)(i) above, the consistent use of a contextual definition style compensates for an absence of explicit usage examples by embedding the term in a syntactic structure. Although the inclusion of separate usage examples could be useful, one should (with regard to a multilingual dictionary for South African students) keep in mind that almost no written special-field texts exist in the African languages, and that mere translations of usage examples taken from authentic English texts could result in highly artificial constructions. Furthermore contextual information is seemingly not so important for culture-independent subject-fields such as chemistry as it is for culture-dependent fields such as law and music.

6.   Conclusion

This article intended to give an overview of the role multilingual explanatory special-field dictionaries could play in providing educationally disadvantaged South African students with access to information on science and technology. The linguistic and the encyclopedic proficiency of target-users, as well as the needs and preferences of the particular linguistic community were identified as key issues in determining the structure and content of such a dictionary.

Although the initial phases of compiling a *Quadrilingual, Explanatory Dictionary of Chemistry* have been completed satisfactorily, the acid test will seemingly be the measure of acceptability with which it is met, and its usefulness in attaining the goals and objectives of government, industry, education and the people of the country.

Notes

1. The dictionary project referred to in this article is a joint project of the Chemistry Division of the Suid-Afrikaanse Akademie vir Wetenskap en Kuns and the Departments of Afrikaans and African Languages of the University of Pretoria.

2. The term concept is used throughout the article to denote "A mental construct for classifying the individual objects of the outer or inner world by means of a more or less arbitrary level of abstraction" (ISO, Recommendation R704: 1968, revised 1986).

3. Louwrens is not against the development of terminologies in the African languages. He does however argue that "terminographic work in the African languages stands in danger of

becoming a wasteful endeavour if the language issue is politicised to the extent that linguistic diversity is subjected to the ideal of nation-building" (1997: 251).

Mutasa regards the African languages in their current usage as mere "vernaculars", restricted to traditional contexts such as occupational guidance, traditional medicine, ancestor worship and general conversation (1996: 26). He seems to regard the creation of terminologies as a prerequisite for using these "vernaculars" as media of instruction for subjects such as mathematics, medicine and chemistry (1996: 30).

4. The four languages to be included in the first edition of the dictionary are English, Afrikaans, Sepedi and IsiZulu.

5. The use of "Sepedi" instead of "Northern Sotho" in this article can be motivated by the fact that the standard language — such as represented by most textbooks, grammars and general dictionaries — is based on the Pedi dialect.

6. A true extensional definition explicates the entire conceptual scope of the term by naming all the immediately lower concepts in the hierarchy, e.g. "The planets of the solar system are Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, Neptune, Pluto" (cf. Arntz and Picht 1989: 66).

Bibliography

(a)   Dictionaries

Godman, A. 1982. *Longman Illustrated Dictionary of Chemistry (LIDC)*. Essex: Longman.

Grandison, A., D. McGinley, T. Shearer, L. Knight, E. Summers, C. Lyons and C. Forde (Eds.). 1991. *Collins English Dictionary (CED)*. Third edition. Glasgow: Harper Collins.

Hampel, C.A. (Consulting chemical engineer) and G.G. Hawley (Editor). 1982. *Glossary of Chemical Terms (GCT)*. Second edition. New York/Cincinnati/Toronto/London/Melbourne: Van Nostrand Reinhold.

Kryt, D. (Ed.). 1980. *Dictionary of Chemical Terminology in Five Languages (DCT)*. Amsterdam / New York/Warsaw: Elsevier Scientific Publishing Company.

Parker, S.P. (Editor in Chief). 1984. *McGraw-Hill Dictionary of Chemical Terms (M-HDCT)*. New York/St. Louis/San Francisco: McGraw-Hill.

(b)   Other literature

Arntz, R. and H. Picht. 1989. *Einführung in die Terminologiearbeit*. Band 2. Hildesheim/Zürich/New York: Georg Olms.

Bergenholtz, H. and S. Tarp. 1995. *Manual of Specialised Lexicography*. Amsterdam/Philadelphia: Benjamins.

Bierwisch, M. and F. Kiefer. 1969. Remarks on Definitions in Natural Language. Kiefer, F. (Ed.). 1969: 55-79.

Cluver, A.D. de V. 1987. A Sociolinguistic Approach to the Study of Technical Languages. *Logos* 7(2): 13-30.

Cluver, A.D. de V. 1989. *A Manual of Terminography*. Pretoria: Human Sciences Research Council.

Cluver, A.D. de V. 1996. Language Development. *Proceedings of a LANGTAG Workshop on Issues in Language Development: Focusing on the Future of the Apartheid Language Boards and The Harmonisation and Development of African Languages.* State Language Services. Department of Arts, Culture, Science and Technology.

Dahlberg, I. 1976. Über Gegenstande, Begriffe, Definitionen und Benennungen. *Muttersprache* 2: 81-117.

Dahlberg, I. 1978. A Referent-orientated, Analytical Concept Theory for INTERCONCEPT. *International Classification* 5(3): 142-151.

Department of Arts, Culture, Science and Technology. 1996. *White Paper on Science and Technology.* Pretoria: Department of Arts, Culture, Science and Technology.

Felber, H. 1985. *Terminology Manual.* Paris: Infoterm.

Felber, H. and G. Budin. 1989. *Terminologie in Theorie und Praxis.* Tübingen: Gunther Narr.

Hausmann, F.J. et al. (Eds.). 1989. *Wörterbücher: Ein internationales Handbuch zur Lexikographie / Dictionaries: An International Encyclopedia of Lexicography / Dictionnaires: Encyclopédie internationale de lexicographie.* Berlin: De Gruyter.

Hutchinson, T. and A. Waters. 1987. *English for Specific Purposes.* Cambridge: Cambridge University Press.

ISO. 1986. *Principles and Methods of Terminology.* DIS 704. Geneva: ISO.

ISO. 1969. *Vocabulary of Terminology.* R 1087. Geneva: ISO.

Kiefer, F. (Ed.). 1969. *Studies in Syntax and Semantics.* Dordrecht: Reidel.

Louwrens, L. 1997. On the Development of Scientific Terminology in African Languages: The Terminographer's Dilemma in a New Dispensation. *Lexikos* 7: 245-251.

Matšela, Z.A. 1987. The Problems of Modernizing the Development of Sesotho Scientific/Technical Terminologies. *Logos* 7(2): 79-88.

Munske, H.H. et al. (Eds.). 1988. *Deutscher Wortschatz: Lexikologische Studien.* Berlin: De Gruyter.

Mutasa, D.E. 1996. Constraints on the Promotion of African Languages to the Level of English, French and Portuguese. *S.A. Journal of Linguistics.* Supplement 32: 23-34.

Prah, Kwesi K. 1995. *Mother Tongue for Scientific and Technological Development in Africa.* Bonn: German Foundation for International Development Education, Science and Documentation Centre.

Rey, A. 1995. *Essays on Terminology.* Philadelphia: Benjamins.

Sager, J.C., D. Dungworth and P.F. McDonald. 1980. *English Special Languages.* Wiesbaden: Oscar Brandstetter.

Smit, M. 1996. Wiegand's Metalexicography as a Framework for a Multilingual, Multicultural, Explanatory Music Education Dictionary for South Africa. Unpublished doctoral dissertation. University of Stellenbosch.

South African Government. 1994. White Paper on Reconstruction and Development. Government's Strategy for Fundamental Transformation.

South African Government. s.a. South Africa's Green Paper on Science and Technology.

Swanepoel, P. 1990. *Definisies in vakkommunikasie.* Pretoria: Human Sciences Research Council.

Wiegand, H.E. 1988. Was eigentlich ist Fachlexikographie? Munske, H.H. et al. (Eds). 1988: 729-790.

Wiegand, H.E. 1989. Die deutsche Lexikographie der Gegenwart. Hausmann, F.J. et al. (Eds). 1989: 2100-2246.

The Microstructural Treatment of Sublexical Lemmas in Afrikaans Descriptive Dictionaries

Ilse Feinauer, *Department of Afrikaans and Dutch,*
University of Stellenbosch, Stellenbosch, South Africa

Abstract: Arguments in metalexicographic literature on the status of subword and multiword lexical items resulted in a more comprehensive lemmatic treatment of these lexical items in the latest editions of Afrikaans descriptive dictionaries, e.g. *Woordeboek van die Afrikaanse Taal* Volume IX (WAT), *Verklarende Handwoordeboek van die Afrikaanse Taal* (HAT), *Nasionale Woordeboek* (NW), *Verklarende Afrikaanse Woordeboek* (VA), and *Basiswoordeboek van Afrikaans* (BA).

This article seeks to investigate whether sublexical lemmas are microstructurally treated the same as lexical lemmas, or whether lexicographers still distinguish between these various types of lemmas in some way or other. If differences exist in the treatment of the various types of lemmas, Afrikaans lexicographers are still word-biased: although subword lexical items are sometimes microstructurally treated in the same way as lexical items in that they are given lemmatic status, they are nevertheless distinguished microstructurally. For example, in WAT IX and HAT, sublexical lemmas are treated in almost the same manner as word lemmas in that their pronunciation is consistently given. However, in most of the dictionaries mentioned, grammatical information on sublexical lemmas is found as part of the definiens. Except for BA, no Afrikaans dictionary illustrates the contextual use of sublexical lemmas by means of examples or quotations; only their formation products are provided. More labels, and even etymological information, should also be provided more frequently as part of the microstructure of sublexical lemmas.

On the basis of the preceding it can already be concluded that Afrikaans lexicographers have not nearly done enough in removing traces of word-bias in descriptive dictionaries. In addition to macrostucturally treating subword lexical items as lemmas, these items should, where possible, be dealt with on a microstructural level in exactly the same way as lexical lemmas.

Keywords: AFFIXES, AFRIKAANS DESCRIPTIVE DICTIONARY, DEFINIENS, ETYMOLOGY, EXAMPLES, HOMONYMY, LABELS, LEXICAL CATEGORIES, LEXICAL LEMMA, LEXICOGRAPHY, MACROSTRUCTURE, MICROSTRUCTURE, OPPOSITION, POLYSEMY, PRONUNCIATION, QUOTATIONS, STEMS, SUBLEXICAL LEMMA, SYNONYMY, TECHNO-STEMS, WORD BIAS

Opsomming: Die mikrostrukturele hantering van subleksikale lemmas in verklarende woordeboeke van Afrikaans. Argumente in die metaleksikografiese literatuur oor die status van subwoordelike en meerwoordelike leksikale items het aanleiding gegee tot 'n meer omvattende lemmatiese hantering van hierdie leksikale items in die jongste uitgawes van verklarende woordeboeke van Afrikaans, soos *Woordeboek van die Afrikaanse Taal*

Deel IX (WAT), *Verklarende Handwoordeboek van die Afrikaanse Taal* (HAT), *Nasionale Woordeboek* (NW), *Verklarende Afrikaanse Woordeboek* (VA), en *Basiswoordeboek van Afrikaans* (BA).

Hierdie artikel wil vasstel of subleksikale lemmas mikrostruktureel dieselfde as woordlemmas gehanteer word, en of leksikograwe steeds op die een of ander manier tussen hierdie verskillende tipes lemmas onderskei. Indien verskille bestaan in die hantering van die onderskeie tipes lemmas, bevoordeel leksikograwe van Afrikaans steeds die woord: alhoewel subwoordelike leksikale items makrostruktureel soms wel dieselfde as leksikale items gehanteer word deurdat hulle ook lemmastatus het, word hulle nogtans mikrostruktureel onderskei. In WAT IX en HAT byvoorbeeld, word subleksikale lemmas op byna dieselfde manier gehanteer as woordlemmas deurdat hul uitspraak konsekwent verskaf word. Daarenteen word grammatikale inligting vir subleksikale lemmas in die meeste van die genoemde woordeboeke as deel van die definiens aangetref. Afgesien van BA, illustreer geen woordeboek van Afrikaans die kontekstuele gebruik van subleksikale lemmas deur middel van voorbeelde of sitate nie; slegs hulle vormingsprodukte word verskaf. Meer etikette, en selfs etimologiese inligting, behoort ook meer dikwels as deel van die mikrostruktuur van subleksikale lemmas voorsien te word.

Op grond van die voorafgaande kan reeds gesê word dat die leksikograwe van Afrikaans nie naastenby genoeg gedoen het om tekens van woordvoorkeur in verklarende woordeboeke te verwyder nie. Benewens die makrostrukturele hantering van subwoordelike leksikale items op 'n soortgelyke manier as lemmas, behoort hierdie lemmas, waar moontlik, op 'n mikrostrukturele vlak op presies dieselfde wyse as leksikale lemmas behandel te word.

Sleutelwoorde: AFFIKSE, DEFINIENS, ETIKETTE, ETIMOLOGIE, HOMONIMIE, LEKSI-KALE KATEGORIEË, LEKSIKALE LEMMA, LEKSIKOGRAFIE, MAKROSTRUKTUUR, MIKRO-STRUKTUUR, POLISEMIE, SINONIMIE, SITATE, STAMME, SUBLEKSIKALE LEMMA, TEEN-STELLING, TEGNOSTAM, UITSPRAAK, VERKLARENDE WOORDEBOEK VAN AFRIKAANS, VOORBEELDMATERIAAL, WOORDVOORKEUR

Introduction

Arguments in metalexicographic literature (e.g. Gouws 1989, 1990, 1991, Rettig 1989, Stein 1985, Zgusta 1971) on the status of subword and multiword lexical items resulted in a more comprehensive lemmatic treatment of these lexical items, especially subword lexical items, in the latest editions of Afrikaans descriptive dictionaries, e.g. *Woordeboek van die Afrikaanse Taal* Volume IX (WAT), *Verklarende Handwoordeboek van die Afrikaanse Taal* (HAT), *Nasionale Woordeboek* (NW), *Verklarende Afrikaanse Woordeboek* (VA), and *Basiswoordeboek van Afrikaans* (BA). Gouws (1991: 75) states: "To ensure a sound treatment of all these lexical items the *traditional word-based lexicography* should be replaced by a broader lexicon-based approach that offers a more comprehensive reflection of the lexicon by listing and treating multilexical and sublexical lemmas" (own italics). This article seeks to examine whether this broader lexicon-based approach in more recent Afrikaans descriptive dictionaries has indeed resulted in a more sound treatment of sublexical lemmas, or whether the macrostruc-

tural changes merely ascribed to them the same lemmatic status as that of lexical items, while still distinguishing between them on a microstructural level.

The types of information to be discussed follow the system of Hausmann and Wiegand (1991: 343) for the specification of microstructural information in monolingual descriptive dictionaries. It includes **synchronic identification** dealing with the form of the lemma, e.g. pronunciation and lexical category, **diachronic identification** that provides the etymology, **labelling, descriptive information** such as the definiens, **syntagmatic information** that covers collocations and examples, and **paradigmatic information** on the semantic relationship between the lemma and other lexical items, e.g. synonymy, opposition, homonymy and polysemy.

Dictionaries generally treat subword lexical items more satisfactorily than multiword lexical lemmas, since in form subword lexical items resemble words more closely than multiword lexical items. For example, some stems are used both lexically and sublexically, the only difference being the hyphen attached to the sublexical stem. This inevitably leads to a freer inclusion of subword lexical items in the macrostructure, but does this necessarily mean the same microstructural treatment as that of lexical items?

In this paper the only distinction regarding sublexical lemmas will be the distinction between affixes and stems. Stems here also include techno-stems.

1.   Affixes

According to Müller (1989: 876), suffixes in German dictionaries generally show an inconsistent lemmatization. The same applies to Afrikaans dictionaries in that suffixes are lemmatized even more inconsistently than prefixes. The reason for the more frequent lemmatization of prefixes is that they are easier to alphabetize, and at first sight also appear more word-like than suffixes. In German dictionaries lexicographic treatment of suffixes is less intensive than that of prefixes. Müller adds that this is also true for French and English dictionaries. This is not the case with Afrikaans dictionaries. When affixes are lemmatized, the microstructural treatment of prefixes and suffixes is very similar. See -in[4] and in-[5] in HAT:

> -in[4] Onproduktiewe agtervoegsel — tans in die baie gevalle ongebruiklik — waarmee vroulike persoons- en diername gevorm is van manlike persoons- en diername, bv. in *Jodin, koningin, gemalin, vriendin, waardin, eselin, leeuin*.
>
> in-[5] Voorvoegsel voor 'n beperkte aantal b.nwe. en s.nwe. met versterkende waarde, bv. in *ingoed, ingierig; ingat, inskottel, inkoejawel, inkokke-wiet*.

Consequently, no distinction will be made between prefixes and suffixes in the following discussion.

2.   Stems

Sublexical stems, especially techno-stems, are more consistently lemmatized than affixes, and in particular suffixes (see Carstens 1995: 148-149). This could be ascribed to two factors: stems resemble words more closely than affixes, and since many of the techno-stems are of Latin or Greek origin, they require eluci-dation. As a result of the closer resemblance between words and stems, one would expect that if discrepancies in the microstructural treatment of words and suffixes occur, the inconsistencies between that of words and stems would be less.

3.   Synchronic Identification

3.1   Pronunciation

Although it is sometimes provided inconsistently, pronunciation is one type of information seldom neglected in sublexical lemmas. WAT IX is the only descriptive dictionary to provide phonetic transcriptions, and all sublexical items are consistently transcribed. However, they consistently lack an indica-tion of main stress. Compare the information for word lemmas and sublexical lemmas:

> **ligtekop** [ləxˈtəkɔp] vs **ligtekop-** [ləxtəkɔp]
> **-logie** [loxi / lugi] vs **logies** [loːˈxis]

Stress guidance is in fact required for sublexical **-logie**, since the main stress differs from that of **logies**.

In HAT, where main stress and syllabic divisions are given for lexical lem-mas, the same applies mostly to the sublexical items:

> **gal.va'no-, -i.seer' , i.o'ne-**

Carstens (1995: 151), however, identifies the following inconsistencies in HAT: for some techno-stems neither main stress nor syllabic division is given:

> **chiro-, filo-, fono-, hidro-, intro-, mega-, meta-, tele-,**

while for others no main stress is given, although the syllabic division still occurs:

e.lek.tro-, ga.mo-, he.te.ro-, ho.mo-, xe.no-

In VA, where only the main stress is given, inconsistencies are also found: for **in'tra-, io'ne-, he'tero-** and **me'de-** stress is given, but not for **iso-** and **kouewa-ter-, kouwater-**.

In NW too, stress patterns for sublexical lemmas occur inconsistently, but the main difference between NW and the other dictionaries is that stress patterns are shown for the products of sublexical lemmas. This information is more valuable since the sublexical lemma as such is seldom pronounced:

> **il-, im-, in-:** *on-, nie.* Il'logies, imper'feksie, inakku'raat, ...
> **bio-:** *met betrekking to lewende organismes, tot wat lewe.* Bioche'mie, -'fisika, -morfolo'gie.
> **kuns-:** *kunsmatige, nagemaakte.* 'Kunsaas, -arm, ...

Unfortunately, some products in articles of the same lemma sometimes show stress patterns, while others do not:

> **in-:** ... Inadekwaat, -akkuraat, -ak'tief, -effek'tief, -essensi'eel, ...
> **by-:** 1. Bybaantjie, -figuur, -gebou, ... 2. ... 'Bygooi, -kry, -las, ... 3. (krieket) ... 'Byglip. Diep-, half-, reg-, vlakby.

In other articles no stress pattern is given for any product, while the sublexical lemma itself shows the main stress:

> **'anti-:** 1. ... anti-Christelik, anti-republikeins, ... anti-militarisme, ... 2. ... anti-kritiek, ...
> **'hetero-:** ... Heteroseksueel, -sentries.

These inconsistencies do not point to an underestimation of sublexical lemmas; they merely show lexicographic carelessness. The complete omission of information such as the indication of the main stress in WAT IX seems to reveal a more fundamental problem: sublexical lemmas are treated differently than lexical lemmas, which could indicate a negation of their lemma status.

3.2   Lexical Category

One of the most noticeable anomalies in the treatment of sublexical lemmas is the statement of the lexical category both with regard to the slot in which it occurs and the terms used for this classification. In every Afrikaans dictionary except BA, the lexical classification for all types of sublexical lemmas is presented as part of the definiens, while the lexical categories for lexical lemmas occur in a specific slot, namely directly after the lemma (HAT and VA), or after

the phonetic transcription of the lemma (WAT IX). BA is the only descriptive Afrikaans dictionary to consistently categorize sublexical lemmas in the same slot as lexical lemmas, namely directly after the lemma (henceforth the appropriate part in the various articles used as examples will be underlined):

> **-toe** *agtervoegsel.* **1 -toe** is 'n rigtingaanduidende agtervoegsel. ...
> **aand-** *eerste deel van samestelling.* **1** Wanneer **aand-** saam met 'n naamwoord gebruik word, het dit betrekking ...

Lexicographically, the first three sublexical lemmas in HAT are treated correctly, with the lexical category in a separate slot and not part of the following definiens:

> **a-**[4] <u>voorv.</u> (*an-* voor vokale) Aanduiding van neutraliteit, houdingloosheid, ...
> **-a**[5] <u>agterv.</u> Meestal by verkorte manlike persoonsname, ...
> **-aan**[2], ook **-iaan** <u>agterv.</u> **1** Aanduiding van herkoms indien aan 'n pleknaam gevoeg: ...

Forty-six pages later, the fourth sublexical lemma is found. For **ante-**, and the rest of the sublexical lemmas in HAT, the unbiased treatment of these lemmas is, strangely enough, replaced by the lexicographically less correct method of including lexical categories in the definientia:

> **ante-** <u>Voorv.</u> met die bet. "voor" in posisie, ...
> **-ie**[2] <u>Substantiefvormende agtervoegsel,</u> afgelei van L. *-ium* en aanduidende 'n geval van ...
> **ko'le-** <u>Eerste lid van ss.</u> met die bet. "steenkool" ...

As for WAT IX and VA, there is no justification for devaluating the semantic status of the sublexical lemma by including grammatical and functional information in its definiens. The definientia of sublexical lemmas should and could be reserved mainly for semantic information. Compare:

> **-loos** ... <u>Agterv. waarmee byv. afl. gevorm word</u>, m.d. bet. "sonder" of "met 'n minimum van", ... (WAT)
> **-logie** ... **1** <u>Slotkomponent van selfst. komposita</u> m.d. bet. "leerstelling, teorie of wetenskap", ... (WAT)
> **laeweerstands-** ... Selde ook *laagweerstand-* en *laeweerstand-*. <u>Aanvangskomponent van selfst. komposita, wat aandui dat dit wat m.d. tweede komponent benoem word</u>, 'n lae weerstand besit of bied, ... (WAT)
> **-isme.** <u>Agtervoegsel om abstrakte begrippe te benoem</u> ... (VA)
> **kouewater-, kouwater-.** <u>Eerste deel van -ss,</u> wat slaan op koue water. (VA)

kwasi-. Asof, kamtig, kastig, <u>veral gebruik as voorvoegsel (woorddeel)</u> wat aan die grondwoord die betekeniswaarde gee van: ... (VA)

In considering the embedded lexical category of **kwasi-** (VA) in its definiens, Carstens (1994: 259) stresses that the user would be able to retrieve the correct grammatical information far more easily if only the lexicographer used the expected slot more consistently. However, VA treats its sublexical lemmas in more detail than it does its lexical lemmas, since no lexical categories are given for lexical lemmas, except when they are multifunctional.

In some instances the lexical categories of sublexical lemmas are completely omitted. No attempt is made to categorize sublexical lemmas in NW:

aan-: 1. vorentoe, verder. ...
tuin-: gekweekte. ...

In HAT, sometimes no lexical category is given:

bio- Met betrekking tot wat lewe, ...
tus'sen-² *Tussen* vorm talryke ss. waarvan slegs die vernaamste opgeneem word; die bet. kom gewoonlik ooreen met dié van die voorsetsel.

According to Stein (1985: 43), a well-argued theoretical distinction is needed between the different types of sublexical lemmas, e.g. prefixes, suffixes, technostems, and other bound stems (see also Gouws (1989: 85-96) for a classification of sublexical lemmas). Unfortunately, such a well-argued theoretical distinction does not exist in Afrikaans descriptive dictionaries. The most appropriately termed, with the least deviations, are the affixes. Interestingly enough, all the lexicographers, except those of BA, use the Germanic terminology **voorvoegsel** and **agtervoegsel** for the Latin-derived **prefix** and **suffix** respectively. This in itself creates problems in the distinction, since the Germanic terms are usually more vague than their counterparts of Latin origin. BA terms the lemma **-toe** (as in *agtertoe* and *huis toe*) an **agtervoegsel**, but it is not a **suffix**. Sometimes affixes are termed more descriptively, but this is usually the exception to the rule. See HAT for:

-ies², ook **-etjies, -jies, -kies, -tjies** <u>Affektiewe verkleiningsuitgang</u> ... wat meestal voorkom by (a) adjektiewe en bywoorde waarin dit die betekenis het van ...

Further inconsistencies are found in the designation of more foreign sublexical items. Van Niekerk (1991: 289) shows the variety of terms used for technostems in the 1979² HAT. Nothing much has changed in the 1994 edition of HAT, since, according to Carstens (1995: 150-151), 38 of the 55 techno-stems used as initial components are described as **woordelement(e)** (word ele-

ment(s)), 5 are termed **voorvoegsel(s)** (prefix(es)), and 6 are described as **eerste lid van samestellings / woorde** (first part of compounds / words). This does not, however, occur in HAT alone. BA also terms **hiper-** and **super-** respectively as **prefiks** (prefix) and **eerste deel van samestelling** (first part of compound). The compilers of VA, in their turn, give two possible terms, leaving it to the user to choose the more appropriate one:

> **iso-:** <u>Voorvoegsel (woorddeel)</u> ... (prefix (word part))
> **kwasi-:** ... <u>voorvoegsel (woorddeel)</u> ... (prefix (word part))
> **kuif-:** ... <u>voorvoegsel (woorddeel)</u> ... (prefix (word part))

VA is one of the few dictionaries to term a sublexical stem **kuif-** a prefix. If inconsistencies are ignored, terms for the designation of sublexical stems are generally more descriptive.

> **kuns-²** <u>Eerste lid van ss</u>. met die bet. "nagemaak ...", bv. ... (HAT)
> **lede-** ... 4 <u>Aanvangskomponent van komposita</u> m.d. bet. "van 'n lid (lede)", bv. ... (WAT)
> **Knysna-.** <u>Eerste deel (voorvoegsel) van samestellings</u> wat die gebied of streek Knysna in die Suidoos-Kaap aandui, ... (VA)
> **aand-** *eerste deel van samestelling*. 1 Wanneer **aand-** saam met 'n naamwoord gebruik word, het dit ... (BA)

BA is the only dictionary to record the particles of particle verbs as sublexical lemmas, and to term them categorically correctly as particles:

> **aan-** *deeltjie*. 1 Wanneer **aan-** saam met 'n werkwoord optree, ...
> **in-** *deeltjie*. In- tree as deeltjie in baie werkwoorde op, ...
> **uit-** *deeltjie*. Uit- word as deeltjie gebruik in baie werkwoorde, bv. ...

The difficulties which Afrikaans lexicographers experience in finding first of all appropriate slots in articles for the lexical categories of the various sublexical lemmas, and secondly the correct metalinguistic expressions for consistent terming of the same type of sublexical lemma, could all be related to the unfortunate grammatical terminology of Afrikaans for the designation of lexical categories. The cardinal term in this terminology is **woord** (*word*) as in: **woord**soort (*word class*), selfstandige naam**woord** (*noun*), voornaam**woord** (*pronoun*), werk**woord** (*verb*), by**woord** (*adverb*), byvoeglike naam**woord** (*adjective*), voeg**woord** (*conjunction*), and lid**woord** (*article*). According to Ponelis (1989: 53) there are, except for *adjektief* and *substantief*, no adequate alternative neutral terms derived from Latin in Afrikaans for the **woord**-constructions (see Ponelis (1989) for a full discussion on this). Since sublexical lemmas are not words in the true sense, their classification cannot fill the slot reserved for word-based

categorizations. The same applies to other lemmas which cannot immediately be categorized according to the pattern:

> is. Gesegdevorm van **wees** as selfstandige werkwoord, koppelwerkwoord en as hulpwerkwoord; ... (VA)

As was argued above, this might also be the reason for the difficulty in finding the correct terminology to describe sublexical lemmas. It varies from **first / second part / segment of compound, initial / final (word) element, initial / final component of compound**, to **word part / element**. Afrikaans grammarians, and subsequently lexicographers, find it difficult to create and establish new classificatory terms for lexical items smaller or larger than words unless **woord** forms part of the terms.

4.   Diachronic Identification

4.1   Etymology

Etymology is the type of microstructural information generally neglected in Afrikaans descriptive dictionaries. Obviously, it will be almost nonexistent in the articles of sublexical lemmas. No examples could be found in VA, BA and NW. In NW, the omission is more noticable, since the lexical lemma **pro** has an etymological specification, while the sublexical **pre-**, for instance, does not:

> **pro** L. *vir.* ...
> **pre-:** *voor-.* ...

In WAT IX, some sublexical lemmas, e.g. those derived from names, have quite comprehensive etymological information and it actually occurs in the correct slot, separate from the definiens:

> **lilliput-** ... *(na Lilliput, denkbeeldige land in J. Swift se Gulliver's Travels, waarin alles, ook die inwoners, baie klein is)* ...

HAT frequently provides etymological information in the articles of sublexical lemmas. Most of the foreign techno-stems are provided with etymological information such as:

> **super-** ... *(L.)* Eerste lid van ss. met die bet. ...
> **bio-** Met betrekking tot wat lewe ... [G. *bios* lewe]
> **meta-** Woordelement met die bet. ... [G. *meta*].

34   Ilse Feinauer

In the above examples, the treatment is identical to that of lexical lemmas, i.e. with the etymological information separate from the actual definiens. Unfortunately these examples are less common: for most of these sublexical lemmas etymological references are incorporated in the definientia. Stein (1985: 40) criticizes this deficiency in the articles of sublexical lemmas, where etymological or diachronic criteria apply instead of the required synchronic criteria such as functional and semantic aspects.

> **ideo-** Woordelement afgelei van <u>G. *idea*</u> idee.
> **hiper-** Voorvoegsel uit <u>G. *huper*</u>, met die betekenis ...
> **hidro-** ... Woordelement afgelei van <u>G. *hudor* water</u>, vloeistof, vog.

It can be added in HAT's favour that the etymology is also given for those sublexical lemmas which the average user would not consider foreign:

> **-isme**[1] Taamlik produktiewe agtervoegsel ontleen aan <u>F.< G. *-ismos*</u>, en gebruik om benamings te vorm ...
> **-ier**[2], ook **-enier, -nier, -r** Agtervoegsel ontleen aan <u>F. *-ier* < L. *-arius*</u> met die betekenis ...
> **-ies**[1] soms ook **-iek, -ities, -ties** Adjektief- en bywoordvormende agtervoegsel, afgelei van <u>F. *-ique* < L. *-icus*</u> met die betekenisse ...

HAT seems to be the only Afrikaans descriptive dictionary to recognize the important encoding function of etymological information, especially in sublexical lemmas (see also Van Niekerk 1991: 288). For example, the lexical **isme** does not have etymological information, but as can be seen above, the sublexical **-isme** has. The etymology enables the user to interpret unexplained lemmas or lexical items not included in the dictionary. The user can also create his or her own lexical items accurately and correctly if he or she knows the original meaning obtained from the etymology.

4.2   Labelling

According to Stein (1985: 39), restrictive usage labels are not applied as consistently in sublexical as in lexical lemmas. This is definitely the case in Afrikaans dictionaries. Where the user would expect a label, there is none, and where there are labels, their application is inconsistent and differs from that of lexical lemmas. Compare the following entries in HAT:

> **lui**[1] mv. *(w.g.; veroud.)* Mense: ...
> **-lui**[2] Agtervoegsel by 'n aantal s.nw. om die mv. aan te dui, ...

The sublexical **-lui** is nowadays used just as infrequently, and is therefore becoming equally archaic as the lexical **lui**, but only the latter is labelled. In WAT IX, however, **³-lui** is microstructurally correctly labelled:

> **³-lui** ... *(soms verhewe en verouderend, behalwe in 'n aantal gevestigde ss.)* Slotkomponent van selfst. komposita ...

Other WAT IX examples where the labelling of sublexical lemmas corresponds to that of lexical lemmas, are:

> **-liede** ... *(verhewe; verouderend, behalwe in 'n aantal gevestigde ss.; minder gebruiklik)* Slotkomponent van selfst. komposita ...
> **laeweerstand-** ... *(ongewoon)* Sien LAEWEERSTANDS- ...

In NW, the label for **by-** is also included in the slot reserved for labels in the articles of lexical lemmas:

> **by-:** ... 3. (krieket) een van verskillende posisies van veldwerkers aan ...
> **bybreekbal** (krieket) bal wat so geboul word dat dit ...

It has already been mentioned that etymological information is given for most of the foreign techno-stems in HAT. However, only a few are provided with labels to show in which scientific or technical field they are the most productive. Labelling would to an even greater degree enhance the encoding facility effected by the etymological information. When HAT does label sublexical lemmas, the method is as follows:

> **in-³ 1** *(w.g.)* Eerste lid van 'n aantal weinig gebruiklike samestellings ...
> **in-⁴** Onproduktiewe agtervoegsel — tans in die baie gevalle ongebruiklik — waarmee ...
> **-iet²** Agtervoegsel ontleen aan F. ... (c) in die chemie — 'n sout van 'n suur met minder as ...
> **by-³** ... 3 In krieket, posisie aan die rugkant van die kolwer; ...

Of these four examples, only the first is labelled in the same way as lexical lemmas. The others all have the restrictive usage indicated as part of the definiens. Even in the lexicographically correctly labelled example (in-³), the information of the label is repeated in the definiens.

5.   Descriptive Information

5.1   Definiens

According to Stein (1985: 39), "(a)ffix definitions ... are usually extremely brief and therefore often unsatisfactory". The definientia of affixes in Afrikaans dic-

tionaries are sometimes apparently quite erudite, but on closer inspection the semantic information is indeed extremely brief and in some instances even completely lacking. See **kwasi-** in VA:

> **kwasi-.** <u>Asof, kamtig, kastig,</u> veral gebruik as voorvoegsel (woorddeel) wat aan die grondwoord die betekeniswaarde gee van: <u>kamtig, soge-naamd, skynbaar,</u> ...

The semantically salient part in the definiens above, as well as in that of many affixes, consists only of a synonym or synonyms. Compare the definientia of **in-** in HAT, NW and VA:

> **in-²** ... met die betekenis (a) "<u>nie, on-, non</u>" ... (HAT)
> **in-:** gew. minder gewone wisselvorm van <u>on-</u> (kyk aldaar) in verskeie woorde. ... (NW)
> **in-⁵.** Voorvoegsel met die betekenis: *<u>nie, non, on-</u>* ... (VA)

on-, **il-** and **im-** are treated more or less in the same way regarding their semantic information. The only dictionary giving more explicit semantic information for affixes in this paradigm, is BA:

> **on-** ... Wanneer **on-** aan 'n adjektief gevoeg word, beteken dit "<u>nie soos die adjektief nie</u>", <u>byvoorbeeld iets wat onaangenaam is, is nie aangenaam nie.</u> ...

Even in the comprehensive WAT IX, the definientia for affixes consist mainly of synonyms:

> **-liks** ... **1** Agterv. van afgeleide bw., m.d. bet. "<u>elke</u>", waarby die stam gew. 'n temporele s.nw. is, ...
> **-logie** ... **1** Slotkomponent van selfst. komposita m.d. bet. "<u>leerstelling, teorie of wetenskap</u>", ...
> **-loos** ... Agterv. waarmee byv. afl. gevorm word, m.d. bet. "<u>sonder</u>" of "<u>met 'n minimum van</u>", ...

Since affixes are extremely difficult to define in isolation, compilers often exclusively use synonym definientia for this purpose. Affixes only acquire semantic substance when attached to a stem, therefore many affixes are defined only in relation to their stems, e.g.

> **-lik** ... Agterv. waarmee talle b.nw. gevorm is, maar wat tans byna heel-temal onproduktief in Afr. is en waarby die afl. dikw. <u>'n besondere graad of toepassing v.d. stambetekenis uitdruk of bygekry het,</u> sodat die bet. v.d. afl. <u>nie sonder meer in terme v.d. bet. v.d. stam omskryf kan word</u>

<u>nie</u>. ... **1a**   Agterv. waarmee b.nw. van oorganklike ww. gevorm is (word), m.d. bet. "<u>wat ge-(ww.) word of ge-(ww.) kan word, wat (ww.) word of (ww.) kan word, of wat (ww.) of kan (ww.)</u>", ...

With this in mind, HAT must be praised for its definientia of the following suffixes where more extensive semantic content of the suffix itself is found:

-aan² ook **-iaan** ... **1** <u>Aanduiding van herkoms</u> indien aan 'n pleknaam gevoeg: ... **2** <u>Aanduiding van 'n volgeling, aanhanger, e.d.</u> indien aan 'n persoonseienaam gevoeg: ...
-iet² Agtervoegsel ontleen aan F. ... en gebruik ter aanduiding van ... (c) ... <u>'n sout van 'n suur met minder as die normale suurstofatome,</u> ...

HAT should also be complimented on its definitions of techno-stems, where much more semantic information is given on the sublexical item:

bio- <u>Met betrekking tot wat lewe,</u> bv. biologie.
-skoop ... met die bet., "<u>middel, instrument om waar te neem of te toon</u>" ...

The techno-stem sublexical lemmas as such have more semantic content than affixes, and consequently more semantic value. Therefore, if the sublexical lemma resembles a lexical item more closely, its definiens will be semantically more extensive. The definientia of nonaffixal sublexical lemmas in the different dictionaries definitely show more variation. The dictionaries also show more inconsistencies in the definitions of the same type of sublexical lemma. For particles, NW sometimes provides synonyms in combination with more extensive definientia, whereas BA's definientia sometimes contain no semantic information at all:

in-: 1. *na binne.* ... 2. *deur die oppervlakte na binne.* ... 3. *bereik, tot by iem. of iets kom.* ... (NW)
uit-: 1. *na buite (weg).* ... 2. *(iets) weg na buite.* ... 3. *(iets) hard, luid, waarneembaar.* ... 4 *(iets) tot die end toe; tot die voltooiing; klaar.* ... 5. *weg, skoon.* (NW)
in- ... In- tree as deeltjie in baie werkwoorde op bv. **inbreek,** ... **inbring,** ... (BA)
uit- ... Uit- word as deeltjie gebruik in baie werkwoorde, bv. **uitbly, uitkom,** ... (BA)

Compared to the definientia of **in-** and **uit-** in BA, the definiens of **oor-** provides detailed semantic, in addition to functional, information:

oor- ... 3 ... Wanneer **oor-** vooraan 'n werkwoord gevoeg word, beteken dit die <u>werkwoord veroorsaak beweging van die een kant na die ander kant</u>

verby of bo-oor iets anders. ... 4 Wanneer **oor-** vooraan 'n werkwoord gevoeg word, beteken dit die werkwoord moet weer gedoen word. Om iets byvoorbeeld oor te doen, beteken om dit weer te doen. ...

The semantic information in the definientia of sublexical stems is better in most Afrikaans descriptive dictionaries:

hand-: *wat met die hand gebruik, gedra, gedryf of gehou word*. ... (NW)

Knysna-. Eerste deel (voorvoegsel) van samestellings wat die gebied of streek Knysna in die Suidoos-Kaap aandui, bv. ... (VA)

kuns-[2] Eerste lid van ss. met die bet. "nagemaak, nie deur die natuur gemaak nie", bv. ... (HAT)

laevolume- ... Aanvangskomponent van selfst. komposita, wat aandui dat dit wat m.d. tweede komponent benoem word, in klein, beperkte hoeveelhede gemaak word of voorkom; ... (WAT)

linker- ... **Linker-** verwys na die kant of rigting wat ooreenstem met daardie kant van 'n mens se liggaam waar die hart sit. ... (BA)

Unfortunately the definiens for **regter-** in BA is not semantically as explicit as the one for **linker-**. This is easily observed, since these two lemmas are cross-indexed.

regter- ... **Regter-** word gebruik voor naamwoorde wat aan die regterkant van iets is. ...

In WAT IX, for instance, circular definitions are used for some sublexical lemmas. The sublexical lemma is defined with reference to the lexical item:

lede- ... **1a** Aanvangskomponent van komposita waarin *lede-* in bet. parallel is aan *lede* ('LEDE), bv. ...

This is possible only if the user has access to the definiens of the lexical item. For **lede-** this is lexicographically sound, but for **laeweerstands-** this can be confusing, since **lae weerstand** is not referred to the section for specialized expressions of [2]**laag,** where it is defined:

laeweerstands- ... Aanvangskomponent van selfst. komposita, wat aandui dat dit wat m.d. tweede komponent benoem word, 'n lae weerstand besit of bied, of vir of van 'n lae weerstand is. ...

Stein (1985: 39) also misses information on collocation restrictions in the definientia of affixes. Collocation restrictions are found in two Afrikaans dictionaries, namely BA and WAT IX, but then only with sublexical stems:

> **aand-** ... **2 Aand-** beteken vir 'n besonder deftige geleentheid wat gewoon-
> lik in die aand plaasvind. <u>Dit word veral ten opsigte van kleredrag
> gebruik;</u> ... (BA)
>
> **laat-** ... **I 1 a** ... Aanvangskomponent van selfst. komposita wat 'n laat of
> latere stadium van 'n <u>geologiese tydperk, 'n tydperk i.d. kultuur- of
> kunsgeskiedenis,</u> e.d. aandui; ... (WAT)

Collocation specifications should be common lexicographic practice, not only
for lexical items, but particularly for all types of sublexical lemmas. This will
provide the user with specific encoding instructions regarding the type of stems
with which a specific sublexical lemma can be combined.

If lexicographers consistently and adequately define sublexical lemmas,
they would perhaps not feel the need to fill the definiens with nonsemantic
baggage such as grammatical and etymological information.

6.   Syntagmatic Information

6.1   Collocations and Examples

The absence of contextual guidance in the form of either collocations or exam-
ple sentences is one of the greatest lexicographic shortcomings in the articles of
sublexical lemmas, affixal and others. Syntagmatic information takes the form
of lists of one-word citations illustrating formation potentialities of the sublexi-
cal lemmas as initial or final components. According to Rettig (1989: 645), this
is "Information zur Wortbildungslehre" rather than actual syntagmatic informa-
tion. The word-generating processes of sublexical lemmas should be fully re-
corded, since the formation product as such is the environment within which
the full value of the sublexical lemma is realized. The formation product, how-
ever, is not used in isolation. It should be placed in a broader context in order
to demonstrate and complement all the other microstructural information
within the article of the sublexical lemma, e.g. labels, definientia, etc. This
would also lead to greater cohesion within the article.

BA is the only Afrikaans dictionary to not only record, but also illustrate
word formations by means of collocations or example sentences:

> **on-** *prefiks.* Wanneer **on-** aan 'n adjektief gevoeg word, beteken dit "nie
> soos die adjektief nie" ... **onaangenaam, onaktief, onbeleef(d), ondraag-
> lik,** ... <u>*Sy is 'n uiters onbeleefde persoon, want sy groet niemand nie, en sê vir
> niks dankie of asseblief nie. Laas somer was die hitte soms so ondraaglik dat 'n
> mens net in die swembad kon sit en niks doen nie.*</u> ...
>
> **aand-** ... **1** Wanneer **aand-** saam met 'n naamwoord gebruik word, het dit
> betrekking op iets wat in die aand gebeur. **aanddiens, aandete, aand-
> klas, aandkoerant, aandlug.** <u>*Aandete is van sesuur af beskikbaar. Die aand-
> klasse word nie juis getrou deur die studente bygewoon nie.*</u>

> **regter-** ... **Regter-** word gebruik voor naamwoorde wat aan die regterkant van iets is. **regterarm, regterduim, regterhand, regteroor, regterstuur,** ... *Die karre in Suid-Afrika het almal regterstuur teenoor dié in Europa wat linkerstuur het.* *Haar regterknie het kwaai opgeswel van die harde stamp.* ...

However, BA is inconsistent, and treats **linker-** in the same way the other dictionaries treat their sublexical lemmas:

> **linker-**   ... **Linker-** verwys na die kant of rigting wat ooreenstem met daardie kant van 'n mens se liggaam waar die hart sit. **linkerarm, linkerbaan, linkerhand, linkerkant, linkervleuel.** ... (BA)
> **ont-:** 1. *verwyder.* **Ontbas, -been, -blaar, -bos, -jaar, -horing,** ... 2. *losmaak.* **Ontkoppel, -bind.** ... (NW)
> **pro-.** Voorvoegsel met die betekenis: ten gunste van, partydig vir, ... bv. **~-Afrikaans, ~-Boer, ~-Duits, ~-Engels.** (VA)
> **lede-** ... 4   Aanvangskomponent van komposita m.d. bet. "van 'n lid (lede)", bv. *ledemosie, lederegister, ledevoorstel;* **ledebelange** s.nw., **ledefonds** snw., **ledekorps** s.nw. ... (WAT)

In HAT, only one example is given to illustrate the syntagmatic behaviour of the sublexical **bio-**, raising doubts (valid or not) whether **bio-** should ever have been recorded as sublexical lemma:

> **bio-** Met betrekking tot wat lewe, bv. in *biologie*. [G. *bios* lewe]

HAT gives no examples in some other cases, e.g. for **mega-, meta-, -gram**[1] and **-gram**[2]. This lack of examples is also present in VA for **hidro-, hipo-,** and **mega-.**

7.   Paradigmatic Information

7.1   Synonymy

According to Müller (1989: 880), the articles of sublexical lemmas could also include synonyms and semantic oppositions. Since all Afrikaans descriptive dictionaries supply these lexical relations for lexical lemmas, there is no reason why they should be omitted in the articles of sublexical lemmas. Naturally the synonym or antonym for sublexical lemmas need not be sublexical as well, since sublemmas could have lexical or multilexical items as synonyms or oppositions.

Synonyms are frequently used as definientia for sublexical lemmas (see paragraph 5 on descriptive information). In addition to other semantic information in the definientia of sublexical lemmas, synonyms are only found in BA

and WAT IX. In BA, a synonym occurs in the article of **aand-**. In WAT IX, synonyms are found in the articles of **linker-**, **links-**, and **laevet-** in the same slot as that for the lexical lemmas. Usage labels are sometimes given for the synonym in WAT IX, which points to good lexicographical practice.

> **aand-** ... **2** Aand- beteken vir 'n besondere deftige geleentheid ...; <u>formeel</u>. aandklere, aanddrag, aandpak, aandrok. (BA)
>
> **linker-** ... **1a** Aanvangskomponent van selfst. komposita wat aandui ...; <u>gew. sin. met *linkerkantse* en soms met *hot*-</u>; teenoor *regter-*; bv. *linkerbladsy, linkerbors,* ...
>
> **b i** Aanvangskomponent van selfst. komposita wat aandui ...; <u>soms sin. met *linkerhand-* (wsk. n.d. Eng.)</u>; teenoor *regter-*; bv. ... (WAT)
>
> **laevet-** ... Aanvangskomponent van selfst. komposita, wat aandui ...; <u>sin. *vetarm*</u>; bv. **laevetdieet** s.nw. ... (WAT)

Lexicographically, WAT IX handles the synonymy of this paradigm very well in that cross-references to e.g. **linker-** are made in the articles of **linkerkantse** and **linkerhand-**. Even **hot-** in WAT IV refers to **links-** in volume IX. Unfortunately there is no reference to **aand-** in the article of **formeel** in BA.

7.2 Opposition

The lexical relation of opposition is provided in most of the Afrikaans descriptive dictionaries, but rather infrequently for sublexical lemmas. One of the biggest shortcomings here is inconsistency, not only regarding the frequency or the slot used, but also, in most cases, regarding the incomplete cross-referencing of the two opposing terms.

Most of the dictionaries except HAT and VA provide the opposite terms for **linker-** / **links-** and **regter-** / **regs-**. NW, however, gives the opposite term for **regter-** very explicitly (the opposition relation to **linker-** is used as the definiens), but at **linker-** the opposite term is absent, making it impossible for the user who refers to **linker-** first, to find its opposition:

> **regter-:** <u>teenoorgestelde van linker</u>. Regterarm, -hand, -oog, -voet.
> **linker-:** aan die kant van die liggaam waar die hart is. Linkeragterbeen, -arm, -been, -hou, ...

The same incomplete lexicographical treatment is found in the article of the cricket term **by-**, where **weg-** is indicated as opposition, but for **weg-** no opposition (or label) is given:

> **by-:** ... **3.** (krieket) *een van verskillende posisies van veldwerkers* ...; <u>teenoor weg-</u>. Byglip. Diep-, half-, reg-, vlakby. (NW)
> **weg-:** **4.** *aan die regterkant van 'n regse kolwer.* Wegkant, -paaltjie. (NW)

BA certainly treats **linker-** and **regter-** well regarding opposition, but its treatment of particles shows inconsistencies: for two polysemous distinctions of **onder-**, oppositions are provided. However, for the first one **oor-**, there is no cross-reference in its article to its opposition **onder-**, and for the second one **bo-**, no sublexical lemma is recorded at all.

> **onder-** ... **1.** Wanneer **onder-** vooraan 'n woord staan, ... **ondergewig, onderproduksie; onderbelig,** ... *Die dorp is reeds onderbevolk,* ... <u>Teenoor: oor-.</u> ... **3.** Wanneer **onder-** vooraan 'n naamwoord voorkom, beteken dit die laer deel. ... **onderdeur, onderkant, onderlip, onderlyf, ondertande.** ... *Sy het 'n kort onderlyfie met 'n lang bolyf.* <u>Teenoor: **bo-.**</u>
> **oor-** ... **1.** Wanneer **oor-** vooraan 'n woord gevoeg word, ... **oorbekend, oorbelas,** ... **oorbevolk, oorbewei,** ... *Groot gedeeltes van ons plase is oorbewei,* ... **2.** Wanneer **oor-** vooraan ...

WAT IX provides quite a few sublexical lemmas with opposition terms, using the same slot as that used for lexical lemmas:

> **laat-** ... **I 1 a** Selde ook los geskryf v.d. daaropvolgende s.nw. Aanvangskomponent van selfst. komposita wat 'n laat ... ; <u>teenoor *vroeg-* of *vroeë*;</u> bv. **Laat-Barok** s.nw. ...
> **laevolume-** ... **1** Aanvangskomponent van selfst. komposita, wat aandui dat dit wat m.d. tweede ...; <u>teenoor *hoëvolume-*;</u> bv. laevolumekomponent s.nw., ...

Inconsistencies are found here as well, since sublexical lemmas on the same page as **laevolume-**, such as **laeweerstand-, laevesel-, laevet-** and **laevlak-** do not have their opposites indicated. The only inconsistency regarding the use of the slot of the opposing term in the articles of sublexical lemmas, compared to that of lexical lemmas, occurs in HAT. In the article for **anti-**, it is incorporated as part of the usage information. Incidentally, **anti-** is the only sublexical lemma found in HAT for which an opposition term is recorded.

> **anti-** Voorv. voor b.nw. en s.nw. met die betekenis "teen", bv. *antikerklik, anti-Russies:* .... *antibakteries,* ... Dikwels ook gebruik <u>(teenoor *pro-*)</u> by geleentheidsvorminge, bv. *antialles, antirook,* ...

7.3   Polysemy and Homonymy

Sublexical lemmas are not treated differently from lexical lemmas regarding the indication of polysemous or homonymous relations, either macrostructurally or microstructurally. In WAT IX, HAT and BA, they are treated in detail and comprehensively: polysemous varieties and homonomy are indicated. The

way in which these indications are done, is not always lexicographically equally sound. WAT IX records three separate homonyms for the sublexical -lui, as well as three homonyms for the lexical lui (see Feinauer (1996)). In HAT, however, sublexical -lui is indicated as part of the homonymic paradigm of lexical lui. HAT consistently records sublexical and lexical lemmas incorrectly as homonyms: they cannot be classified as homonyms since the hyphen of the sublexical lemma shows it to be different in form from the lexical lemma.

Stein (1985: 40) mentions that in the case of affixes, the etymological principle often overrides functional and semantic aspects. This leads to the indiscriminate grouping of affixes of the same form. What should synchronically be listed as homonyms are still treated as various polysemous senses on the grounds of diachronic semantic resemblances. WAT IX and HAT, e.g., both record -logie as a polysemous sublexical lemma, whereas their various meanings have drifted so far apart that they should have been recorded as homonyms:

> -logie ... 1 Slotkomponent van selfst. komposita m.d. bet. "leerstelling, teorie of wetenskap", bv. *antropologie, biologie, ekologie, ...*
> 2 Slotkomponent van selfst. komposita m.d. bet. "woord, rede, spraak of gesprek", bv. *doksologie, tetralogie, toutologie, trilogie.* (WAT)
> -logie ... Woordelement met die bet. (a) aard, handeling of afdeling van kennis, bv. in *sosiolgie, teologie;* (b) verhandeling, gesprek, bv. in *trilogie.* (HAT)

This treatment of homonymy in Afrikaans descriptive dictionaries, however, is not unique to sublexical lemmas, but has been criticised often with regard to lexical lemmas as well.

Conclusion

If all these inconsistencies are taken into consideration, it can safely be said that in all Afrikaans dictionaries, sublexical lemmas are still treated with less care than lexical lemmas. The word-bias against sublexical lemmas is unfortunately still alive in Afrikaans lexicography, admittedly less so in some dictionaries. This conclusion is based on the microstructural representation of these lemmas, particularly with regard to all the nonsemantic information included in their definientia and the microstructural information excluded from their articles. Sublexical lemmas are therefore not treated equally, since they are still being distinguished from lexical lemmas on a microstructural level. The wish of Müller (1989: 879) for the future treatment of sublexical lemmas in German dictionaries is equally applicable to Afrikaans dictionaries: "... für die Zukunft einen stärkeren, konsequenten und methodologisch-systematisierten Ausbau der bisherigen Ansätze und intensive Bemühungen um die Kodifizierung der

Wortbildung sowohl als Resultat wie auch als Prozeß ...". If this could be effected in Afrikaans dictionaries, no more articles of sublexical lemmas will hopefully be found that convey hardly any information such as **tussen-** in HAT:

> **tus'sen-**[2] *Tussen* vorm talryke ss. waarvan slegs die vernaamste opgeneem word: die bet. kom gewoonlik ooreen met dié van die voorsetsel.

References

Botha, T.J.R. (Ed.). 1989. *Leksikografie en leksikologie.* Pretoria: Serva.

Carstens, A. 1994. *Verklarende Afrikaanse Woordeboek*[4] as spieël van normverplasing. *Lexikos* 4: 249-281.

Carstens, A. 1995. 'n Kritiese beskouing van *HAT*[3]. *Lexikos* 5: 138-165.

De Villiers, M., J. Smuts, L.C. Eksteen and R.H. Gouws. 1987[7]. *Nasionale Woordeboek.* Cape Town: Nasou.

Feinauer, I. 1996. Die negende deel van die *Woordeboek van die Afrikaanse Taal. Lexikos* 6: 233-271.

Gouws, R., I. Feinauer and F. Ponelis. 1994. *Basiswoordeboek van Afrikaans.* Pretoria: Van Schaik.

Gouws, R.H. 1989. *Leksikografie.* Cape Town: Academica.

Gouws, R.H. 1990. Vaste uitdrukkings as multileksikale lemmas in verklarende Afrikaanse woordeboeke. *Tydskrif vir Geesteswetenskappe* 30(4): 265-283.

Gouws, R.H. 1991. Toward a Lexicon-based Lexicography. *Dictionaries* 13: 75-90.

Hausmann, F.J. and H.E. Wiegand. 1991. Component Parts and Structures of General Monolingual Dictionaries: A Survey. Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.). 1989-1991: 328-360.

Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.). 1989-1991. *Wörterbücher. Dictionaries. Dictionnaires. An International Encyclopedia of Lexicography.* Berlin: Walter de Gruyter.

Ilson, R. (Ed.). 1985. *Dictionaries, Lexicography and Language Learning.* Oxford: Pergamon Press.

Labuschagne, F.J. and L.C. Eksteen. 1993[8]. *Verklarende Afrikaanse Woordeboek.* Pretoria: Van Schaik.

Müller, W. 1989. Theorie der einsprachigen Lexikographie II: Beschreibungsprobleme. Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.). 1989-1991: 869-882.

Odendal, F.F., P.C. Schoonees, C.J. Swanepoel, S.J. du Toit and C.M. Booysen. 1994. *Verklarende Handwoordeboek van die Afrikaanse Taal.* Midrand: Perskor.

Ponelis, F.A. 1989. Leksikale elemente. Botha, T.J.R. (Ed.). 1989: 48-54.

Rettig, W. 1989. Die Wortbildungzusammenhänge im allgemeinen einsprachigen Wörterbuch. Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.). 1989-1991: 642-649.

Stein, G. 1985. Word-formation in Modern English Dictionaries. Ilson, R. (Ed.). 1985: 35-44.

Van Niekerk, A.E. 1991. Neo-klassieke en pseudo-sintaktiese komposita. *Lexikos* 1: 281-297.

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

Zgusta, L. 1971. *Manual of Lexicography.* The Hague: Mouton.

Lemmatisation of Adjectives in Sepedi

R.H. Gouws, *Department of Afrikaans and Dutch,*
University of Stellenbosch, Stellenbosch, South Africa and
D.J. Prinsloo, *Department of African Languages,*
University of Pretoria, Pretoria, South Africa

Abstract: One of the great challenges to compiling better dictionaries for the African languages is to develop sound strategies and procedures for planning the structure of the dictionaries. In this regard all the structural components of a dictionary, including the macrostructure, microstructure, mediostructure and access structure, come into play. Most dictionaries for African languages, including Sepedi dictionaries, fail even at this level. In this article the planning of especially the macrostructure in respect of one lexical category which has been unsatisfactorily treated in Sepedi dictionaries, namely the adjective, will be attempted. Secondly the lemmatisation of adjectives in six Sepedi dictionaries will be critically evaluated. This will be done with the emphasis on various metalexicographical aspects.

Keywords: LEMMATISATION, ADJECTIVES, SEPEDI, MACROSTRUCTURE, USER-PERSPECTIVE

Opsomming: Lemmatisering van adjektiewe in Sepedi. Een van die grootste uitdagings vir die samestelling van beter woordeboeke vir die Afrikatale is om gepaste strategieë en prosedures vir struktuurbeplanning te ontwikkel. Alle struktuurkomponente van 'n woorde-boek, insluitende die makro- en mikrostruktuur asook medio- en toegangstrukture is hier ter sprake. Die meeste woordeboeke vir Afrikatale, Sepedi ingesluit, faal in hierdie opsig. In hierdie artikel sal gepoog word om 'n uiteensetting te gee van die wyse waarop die makrostruktuur ten opsigte van die adjektief beplan moet word. Die adjektief is maar een van die kategorieë op hierdie vlak wat onbevredigend in bestaande woordeboeke hanteer is. Tweedens sal die lemmatisering van adjektiewe in ses Sepedi woordeboeke krities beskou word teen die agtergrond van verskeie metaleksikografiese aspekte.

Sleutelwoorde: LEMMATISERING, ADJEKTIEWE, SEPEDI, MAKROSTRUKTUUR, GEBRUIKERSPERSPEKTIEF

Introduction

According to Wiegand (1989: 251) lexicography is a practice aimed at the pro-duction of dictionaries in order to activate another practice, i.e. the cultural

practice of dictionary use. Any lexicographer compiling a dictionary has the obligation to present the contents of the dictionary in such a way that it will lead to the cultural practice of dictionary use. This can only be achieved if the construction of the specific dictionary adheres to the user-perspective by taking not only the linguistic needs but especially also the reference skills of the intended target user into account. User-friendliness in dictionaries implies that the contents of the dictionary is made as accessible to the user as possible. Attempts to enhance the retrievability of information are often impeded by a high degree of textual condensation. The utilisation of structural markers and other methods to assist the target user in his endeavour of reaching the desired data-presentation means that the internal search route has to be indicated quite clearly. Although this is an important facet of dictionaries, an improvement of the internal search route is not the only way to ensure a better retrievability of information. The macrostructure remains the main access structure of any dictionary with a strictly alphabetical ordering system. Lexicographers too often neglect the importance of a well-designed macrostructure as a functional component of the total linguistic contents of a dictionary by restricting their attempts to enhance user-friendliness to the microstructural level.

The first step towards the improvement of the lexicographic standard of dictionaries for African languages must be to do the groundwork right. Dictionaries are instruments of linguistic and communicative empowerment and therefore lexicographers have to make sure that their intended target users receive an optimal linguistic presentation. To achieve this goal every lexicographer has to rely on a sound theoretical knowledge, and the compilation of every dictionary has to be preceded by the formulation of a business plan, adhering to the aims of the typological criteria of that specific dictionary, and aimed at the specific needs and reference skills of a well-defined target user. This business plan has to be rooted in a general theory of lexicography. According to Wiegand (1984: 14-15) one of the components of a general theory of lexicography is the theory of organisation. This includes all the activities leading to the drawing up of a dictionary plan — that all-important activity that has to precede the compilation of each and every dictionary. The position of the target user may never be underestimated when compiling a dictionary or when drawing up the dictionary plan. Dictionaries are compiled to be used and therefore the target user should be placed in a position where he/she can utilise a dictionary for the successful retrieval of linguistic data.

The traditional, and often haphazard, approach according to which words were entered into a dictionary "as they cross the compiler's way" can no longer be justified. The user-perspective which determines the selection, presentation and treatment of lexical items compels the lexicographer to include those lexical items in the macrostructure that can contribute to the aims of the typological category to which the specific dictionary belongs. The way in which macrostructural elements are presented should also reflect their linguistic status. When dealing with a specific lexical category, the way in which these items

have to be lemmatised, has to be determined on linguistic grounds. The lexicographer has to do an exhaustive analysis of the phenomenon by firstly breaking it down into all its combinations and permutations. Once the compiler is satisfied that he has covered the full scope as viewed from the living language and not only the grammar book, he may start planning how to lexicographically treat the issue within crucial parameters such as the target user's needs, affordability of the dictionary, proper presentation and treatment of the lemma, decisions regarding the data categories to be given, etc. Apart from acquainting himself with sound basic lexicographic principles and practice, he has to study the problematic aspects that the African languages have in common as well as problematic aspects unique to a specific language.

It will be argued that in respect of the adjectives, most dictionaries fail to answer the questions most likely to be asked by their target users, who are usually defined as scholars and students who wish to learn the language. This is due to the lack of a proper needs assessment as part of the overall theory of organisation.

The presentation of adjectives

In the six Sepedi dictionaries used in this present survey, the extremes, with regard to the lemmatisation of adjectives, lie between the *Klein Noord-Sotho woordeboek* where only two forms of a specific adjective are entered into the dictionary without proper guidelines in the front matter on the one hand, and *Sediba* where all possibilities namely nine for each stem are included as lemmas in the central word list on the other.

In planning the macrostructure for a specific lexical category the first step will be to determine whether a limited or an unlimited number of lexical items, i.e. words or stems, are dealt with. The terms "limited" and "unlimited" will be used in a rather oversimplified way. Say, for example, that nouns, verbs, reflexive forms of verbs, etc. are unlimited in that an infinite number of such forms occur while subject concords are limited in that there is only a maximum of 15.

So, in respect of the adjective, the first step will be to determine whether the number of adjectives are limited or unlimited. Only about 30 adjectives of reasonable frequency, listed under (1), occur in Sepedi.

(1)

-bedi	"two"
-raro	"three"
-ne	"four"
-hlano	"five"

-so	"black"
-šweu	"white"
-tala	"green/blue"
-hubedu	"red"
-hwibidu	"red"
-sotho	"brown"
-sehla	"yellow/tawny/grey"
-koto	"thick"
-sese	"thin/narrow"
-thata	"hard/difficult"
-golo	"big"
-nyane	"small"
-telele	"long/tall"
-kopana	"short"
-be	"bad/evil"
-fsa	"new/young"
-swa	"new/young"
-ntši	"many"
-ngwe	"other/some"
-tona	"male"
-tala	"old"
-kaaka	"this big"
-botse	"beautiful/good"
-bose	"delicious/tasty/nice"
-bjalo	"such"
-kae?	"how many?"
-bjang?	"what kind/sort of?"

(Note in passing that some unusual words qualify as adjectives in Sepedi, for example the numbers 2, 3, 4, and 5 as well as the question words *kae?* and *bjang?*)

One of the issues on which the lexicographer has to make a decision is whether it will satisfy the needs of the target user if these adjectives under (1) were lemmatised in that form. What must be kept in mind when answering this question is the lack of typological diversity in Sepedi lexicography. Conse- quently, the target users of these dictionaries are defined as students and schol- ars with the inclusion of inexperienced learners. For these users the outer access structure has to provide a direct route to the item they are searching. The typi- cal item they will search, will be words encountered in written texts or oral conversations. None of the adjectives as they are listed under (1), will be found in Sepedi literature. This is due to the fact that these adjectives always have to

take the nominal prefixes of the different noun classes. Compare (2).

(2)   -golo "big/important"

Class:

1.	monna yo mogolo	monna "man"
2.	batho ba bagolo	batho "people"
3.	mohlare wo mogolo	mohlare "tree"
4.	mebotoro ye megolo	mebotoro "cars"
5.	lesogana le legolo	lesogana "young man"
6.	mahlo a magolo	mahlo "eyes"

The typical target user of the Sepedi dictionaries under discussion, who encounters any occurrence of the adjective -golo, will find this lexical item used as the stem of a complex form in which the item -golo is preceded by a prefix. This confronts the lexicographer with a dilemma. Pursuing a lexical-based approach to the compilation of the macrostructure (cf. Gouws 1991), the lexicographer will have to include lexical items like stems and affixes in the macrostructure if they have a productive occurrence in real language use. Hausmann and Wiegand (1989: 337) also argue in favour of the fact that all lexical units, including e.g. affixes and other elements of word-formation, may be lemmata. This would mean that lexical items like the stem -golo and the prefixes mo-, ba-, etc. should be included as lemmas in a Sepedi dictionary. The dilemma of the lexicographer is that the reference skills of the target user of the dictionaries under discussion may not equip the user with the expertise to apply the necessary word-formation rules in order to retrieve information about an adjective like mogolo from merely consulting the articles of the sublexical lemmas (cf. Gouws 1989) mo- and -golo. Theoretical soundness and practical realities oppose each other and the lexicographer has to make a difficult decision regarding the forms to be lemmatised.

When deciding on which form to include as macrostructural component, a lexicographer has to consider the theoretical status attributed to that form. According to Hausmann and Wiegand (1989: 329) lemmatisation refers to "the selection of one single morphological form whose function in the macrostructure is to represent the total set of grammatical and morphological forms of the linguistic sign treated in the microstructure". This implies that one lemma sign does not necessarily represent only one lexeme or only one morphological form. Dictionaries usually opt on a systematic basis for one type of item to be lemmatised, e.g. the first person singular form of a verb. Although the treatment is aimed at that lemma sign, it applies to other forms of the lemma as a member of the ordered set of items constituting the treatment units of the dic-

tionary as well. Hausmann and Wiegand (1989: 329) also point out that the inclusion of all irregular forms in the macrostructure is rare.

Adhering to the above-mentioned notion of the lemmatisation of one selected morphological form representing a whole set of forms, a lexicographer can be led to the point where the lemmatisation of adjectives in Sepedi dictionaries does not confront him with any problems. This will imply that only the stem form will be lemmatised and the dictionary user will have to rely on his own linguistic intuition to find the desired information and to apply it to complex words. As noted above, such a lemmatisation system will impede access to the presented data because the lemma sign will not represent a form that can be related to the words found in Sepedi literature. This will characterise the dictionaries as extremely user-unfriendly.

Contrary to the belief that only the stem should be lemmatised, it could also be argued that the complex adjectives consisting of a stem and a prefix are not irregular forms but rather the regular forms of the adjective with the stem as an item which is nonexistent as independent form. Such a word-based approach will not make provision for the lemmatisation of sublexical items like stems, but only for the inclusion of words as lemmas. This will lead to the lemmatisation of all the occurrences of the complex adjectives consisting of e.g. -*golo* plus a prefix. Such a complete list for the different classes will look like column 2 under (3):

(3)

Class	Column 1	Column 2
1.	mogolo	mogolo
2.	bagolo	bagolo
3.	mogolo	---
4.	megolo	megolo
5.	legolo	legolo
6.	magolo	magolo
7.	segolo	segolo
8.	kgolo	kgolo
9.	kgolo	---
10.	kgolo	---
14.	bogolo	bogolo
15.	gogolo	gogolo
16.	gogolo	---
17.	gogolo	---
18.	gogolo	---
	---------	---------
TOTAL:	15	9

If provision for each noun class is to be made, the cost in terms of macrostructural redundancy will be fairly severe. In principle, 15 times 30 = 450 articles, only to make provision for the adjectives in Sepedi.

This number can immediately be reduced to nine since classes 1 and 3, 8-10 and 15-18 respectively take similar forms. However, 9 times 30 still renders a large number of 270 possibilities. The crucial issue will be to maintain a delicate balance between user-friendliness and the possibility of redundancy getting out of hand which in turn directly effects economy and affordability of the dictionary. In simple terms it means that if all of the 270 possibilities are to be accommodated, the dictionary will be very user-friendly since no knowledge of the grammar will be presupposed and all adjectives could be found under the first letter, e.g. *mogolo* under *m-*, *segolo* under *s-*, etc. However, it could be very redundant. This problem could once again activate a tension between the dictionary and the dictionary-using public.

Economy efforts compel the lexicographer to employ space-saving mechanisms — like the lemmatisation of fewer forms. This leads to the professionalisation of lexicography and a high degree of textual condensation. It becomes increasingly difficult for the lay dictionary user to understand this professionalised instrument and to use it successfully. Hausmann (1989: 13) discusses this problem and refers to this conflict between dictionary and user as a conflict between *dictionary culture* and *user-friendliness*. Hausmann sees user-friendliness as the adaptation of lexicography to society whereas dictionary culture is the adaptation of society to lexicography. This means that user-friendliness demands that the contents and presentation of a dictionary should be determined by the needs and expertise — or lack thereof — of society. Dictionary culture means that society has to be educated to utilise more sophisticated dictionaries.

Possible solutions in Sepedi dictionaries

One extreme solution to the problem could be to reduce column 2 under (4) from nine possibilities to only TWO as in column 4.

The other extreme would be to enter the full range of 270 possibilities into the dictionary with exhaustive treatment in each case, which will of course be very user-friendly but extremely redundant. Lexicographers have to endeavour to make these extremes more viable. The major challenge will be to make the first extreme, namely to lemmatise only two forms as under (4) column 4, more user-friendly.

(4)

Class	Column 1	Column 2	Column 3	Column 4
1.	mogolo	mogolo	-golo	-golo
2.	bagolo	bagolo	-golo	---
3.	mogolo	---	-golo	---
4.	megolo	megolo	-golo	---
5.	legolo	legolo	-golo	---
6.	magolo	magolo	-golo	---
7.	segolo	segolo	-golo	---
8.	kgolo	kgolo	kgolo	kgolo
9.	kgolo	---	kgolo	---
10.	kgolo	---	kgolo	---
14.	bogolo	bogolo	-golo	---
15.	gogolo	gogolo	-golo	---
16.	gogolo	---	-golo	---
17.	gogolo	---	-golo	---
18.	gogolo	---	-golo	---
TOTAL:	15	9		2

Utilising the front matter

A possible way of coping with this problem is to utilise the front matter of the dictionary by including easy to read guidelines, e.g.:

"In this dictionary adjectives are entered on the stem, e.g. *mogolo* in an example such as *monna yo mogolo* 'a big/tall/important man' must be looked up under word minus stem, that is *mogolo – mo = -golo.*" Thus the complete table of guidelines would be as in (5):

(5)

1 and 3	mogolo	look up under	-golo
2.	bagolo	look up under	-golo
4.	megolo	look up under	-golo
5.	legolo	look up under	-golo
6.	magolo	look up under	-golo
7.	segolo	look up under	-golo
14.	bogolo	look up under	-golo
15-18	gogolo	look up under	-golo

The form for classes 8, 9 and 10, *kgolo*, will be lemmatised as *kgolo* and is no problem.

Within a target user community with a well-developed dictionary culture this approach could surely be defended. Dictionaries have to be regarded as carriers of texts (cf. Wiegand 1996). In a dictionary as a text carrier that displays a typical textual book-structure, the central word list is a compulsory text. All functional text parts preceding this central word list constitute the front matter and all the functional text parts following the central word list constitute the back matter of the dictionary (cf. Hausmann and Wiegand 1989: 330-331). Besides the central word list there is only one other obligatory text, i.e. the text in the front matter containing the user's guidelines. Because this is an obligatory text, the lexicographer may include information in this text which will assist the user to achieve an optimal retrieval of information from the central word list. When adjectives are treated in Sepedi dictionaries, there should, from a metalexicographic perspective, in principle be no objections to a limited lemmatisation of this word class if the front matter contains a text with user's guidelines in which a sound and systematic explanation of this word class is given.

Once again, however, the potential conflict between user-friendliness and dictionary culture has to be taken into account. Hartmann (1989: 103) argues that an analysis of user's needs should precede dictionary design. The lexicographer of a Sepedi dictionary should allow the outcome of a needs and reference skills analysis to determine a variety of characteristics of the dictionary. One aspect to be considered by the lexicographer is whether the typical target user is in the habit of utilising the texts in the front matter to improve his dictionary using skills or his access to the presented information.

Unfortunately lexicographers may seldom rely on the willingness or habit of their target users to utilise a text that does not form part of the central word list. Therefore Busane (1990) is in the right when he says that dictionary users are not known for consulting the guidelines to the dictionary, they want to find what they need instantly without referring to grammatical rules and guidelines in the front matter or even guidelines within the dictionary itself.

Alternative possibilities

The main weakness of the other extreme, namely to lemmatise all the adjectives, could be combated by attempts to reduce redundancy by, among others, (a) reduction based on frequency-of-use, (b) shorter articles including less data categories and (c) cross-references.

(a)   Reduction based on frequency-of-use

The compiler could decide to omit the adjective due to the fact that the overall

count of eleven occurrences for classes 15-18 under (6) is very low in comparison to the rest. This is especially the case for other adjectives which are less frequently used than -*golo*.

(6)

1 Class	2 'big'	3 Freq.	4 Sediba	5 Popular	6 Klein	7 Pukuntšu	8 Shuters	9 New
1 and 3	mogolo	611	mogolo	---	---	mogolo	mogolo	mogolo
2	bagolo	328	bagolo	---	---	bagolo	bagolo	---
4	megolo	70	megolo	megolo	---	---	megolo	---
5	legolo	243	legolo	---	---	legolo	legolo	---
6	magolo	136	magolo	magolo	---	magolo	magolo	magolo
7	segolo	161	segolo	segolo	---	segolo	segolo	segolo
8-10	kgolo	597	kgolo	kgolo	kgolo	kgolo	kgolo	kgolo
14	bogolo	241	bogolo	bogolo	---	bogolo	bogolo	bogolo
15-18	gogolo	11	gogolo	---	---	gogolo	---	gogolo
	-golo	0	---	golo	-golo	-golo	---	golo

The ways in which adjectives have actually been lemmatised in six Sepedi dictionaries will be evaluated with reference to (6).

Column 1 gives the noun class or classes related to the specific form of the adjective, column 2 the adjectives for classes 1-18.

In column 3 the overall frequency count on a one million corpus, compiled from fifty different books and magazines, is shown, followed in columns 4-9 by an indication of the inclusion or omission of the adjectives in the Sepedi dictionaries in question. It is clear from column 3 that this adjective is in principle highly used in Sepedi. (A total count for all the classes is 2398 which means that it is used more than 40 times on average in every single Sepedi book or magazine.) Furthermore it is clear that the forms *mogolo* (classes 1 and 3), *kgolo* (classes 8-10) and *bagolo* (class 2) are the most highly used.

As indicated in column 4, all the relevant forms are entered in *Sediba*, which represents one of the extremes. This is more or less as good as it can be in respect of user-friendliness.

According to column 5, all the relevant forms, except the forms for classes 1 and 3, 2, 5 and 15-18, are given in the *Popular Northern Sotho Dictionary*. In addition, the stem *golo* is given, but as a word, that is, without the hyphen indicating its status as a sublexical lemma. This is unacceptable, especially in view of its high frequency of use in classes such as 1-3.

As shown in column 6, the compilers of the *Klein Noord-Sotho woordeboek* opted for the other extreme, namely to enter only the form *kgolo* for classes 8-10 and the stem -*golo* for the rest. (Compare column 4 under (4) once again.) In fact, only -*golo* was entered and treated, while *kgolo* was entered with a cross-

reference to -*golo*.

According to column 7, *Pukuntšu* gives all the relevant forms, with the exception, for no apparent reason, of class 4. Also entered is -*golo*, properly marked as a stem.

As indicated in column 8, all the relevant forms are entered in the *Shuters New Sepedi dictionary*, with the exception of class 15-18 which was omitted on the basis of low frequency. As shown in the case of *Sediba* in column 4, it is unnecessary to enter the stem form -*golo* as well, since all the derivations have been covered.

Finally, in column 9 the entries for the *New English-Northern Sotho Dictionary* are given with the forms for classes 2, 4 and 5 missing and *golo* entered as a word instead of a stem.

(b) Shorter articles including less information categories

In addition to attempting reduction based on frequency-of-use, shorter articles could be employed. Articles could be shortened in various ways. A decrease of the data types would also decrease the density of information. If this is done on the basis of a needs analysis which results in the omission of redundant or less functional data categories, this option could lead to an increase in the users' comprehension. However, the articles can also be shortened by a process of textual condensation that does not omit data categories but retains them although in a more condensed presentation. Textual condensation, accompanied by a high degree of information density, results in a more complex microstructural presentation which impedes the retrieval of information and the successful interpretation of the articles. According to Kühn (1989: 112) the use of a dictionary has to be understood as a communicative act. The lexicographer has to endeavour to improve the quality of this communicative act. In a dictionary aimed at scholars, students and learners, textual condensation will definitely be detrimental for the user when employing the dictionary in a communicative act. The inclusion of all adjectives as lemma signs is a user-friendly option. However, if this is accompanied by a treatment that omits certain data categories or that condenses the presented data, the question arises whether it would not have been better to utilise the available space for a more extensive treatment of fewer lemmas.

(c) Cross-references

A lexicographic procedure that has not yet had an optimal employment in South African dictionaries, is the dictionary-internal mediostructure. According to Wiegand (: 196: 11) the dictionary-internal mediostructure interconnects the knowledge elements represented in different sectors of the dictionary on several levels of lexicographic description. Wiegand (1996: 11) continues:

A lexicographer *refers* the potential user from a *reference position* giving the *reference item* or other *reference transmitting items* to the *reference address*, which possibly provides access to the lexicographic data relevant for obtaining the user's objective. Thus, a *reference relation* is established either between the reference item or other reference transmitting items to one or more reference address(es).

One of the biggest advantages of the effective utilisation of a dictionary-internal mediostructure is that precious space can be saved by, for example, giving an exhaustive treatment of one entry with cross-references from the other skeleton entries. This could be regarded as user-unfriendly in a different way, as is the case in (5) where the user has to consult and rely on guidelines given in the nonalphabetical section. However, if the reference address is a lemma in the central word list, the system of cross-referencing can enhance the text-internal cohesion. This can also lead the user to experience the lexicon as network of relations.

In Sepedi dictionaries the employment of a procedure of dictionary-internal mediostructural relations will compel the lexicographer to give an explicit explanation of the system in the front matter of the dictionary. The application of a system of cross-referencing should be done in such a simple and explicit way that even the user who does not consult the front matter has to be able to follow the reference route and to retrieve the necessary information from the treatment of the reference address. This would mean that all the various occurrences of an adjective can be lemmatised but these lemmas will receive a limited lexicographic treatment and will primarily be used as reference items filling the reference position. Besides grammatical information, e.g. an indication of the nominal prefix and the specific noun class, the treatment will consist of an indication of the reference address. This reference address could be the stem which is the salient component of each adjective.

Conclusion

The lemmatisation of adjectives may no longer be done in an arbitrary way. A detailed analysis of the problems and possible solutions is a prerequisite for the compilation of a proper macrostructure. Each and every aspect should be subjected to a similar analysis before one could think of tackling the microstructure. In this regard the lexicographer has to rely on the results of metalexicographical research.

References

1.   General

Busane, M. 1990. Lexicography in Central Africa: The User Perspective with Special Reference to Egypt. Hartmann, R.R.K. (Ed.). 1990: 19-35.

Gouws, R.H. 1989. *Leksikografie.* Cape Town: Academica.

Gouws, R.H. 1991. Toward a Lexicon-based Lexicography. *Dictionaries* 13: 75-90.

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

Hartmann, R.R.K. 1989. Sociology of the Dictionary User: Hypotheses and Empirical Studies. Hausmann, F.J. *et al.* (Eds.). 1989: 102-111.

Hartmann, R.R.K. (Ed.). 1990. *Lexicography in Africa.* Exeter Linguistic Studies 15. Exeter: University of Exeter Press.

Hausmann, F.J. and H.E. Wiegand. 1989. Component Parts and Structures of General Monolingual Dictionaries: A Survey. Hausmann, F.J. *et al.* (Eds.). 1989: 328-360.

Hausmann, F.J. 1989. Die gesellschaftlichen Aufgaben der Lexikographie in Geschichte und Gegenwart. Hausmann, F.J. *et al.* (Eds.). 1989: 1-19.

Hausmann, F.J. *et al.* (Eds.). 1989. *Wörterbücher. Ein internationales Handbuch zur Lexikographie / Dictionaries. An International Encyclopedia of Lexicography. / Dictionnaires. Encyclopédie internationale de lexicographie.* Berlin: Walter de Gruyter.

Kühn, P. 1989. Typologie der Wörterbücher nach Benutzungsmöglichkeiten. Hausmann, F.J. *et al.* (Eds.). 1989: 111-127.

Wiegand, H.E. 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.R.K. (Ed.). 1984: 13-30.

Wiegand, H.E. 1989. Der gegenwärtige Status der Lexikographie und ihr Verhältnis zu anderen Disziplinen. Hausmann, F.J. *et al.* (Eds.). 1989: 246-280.

Wiegand, H.E. 1996. Über die Mediostrukturen bei gedruckten Wörterbüchern. Zettersten, A. and H. Pedersen (Eds.). 1996: 11-43.

Zettersten, A. and H. Pedersen (Eds.). 1996. *Symposium on Lexicography* VII. Tübingen: Max Niemeyer.

2.   Dictionaries

Kriel, T.J. 1985. *New English-Northern Sotho Dictionary.* Johannesburg: Educum.

Kriel, T.J. 1988. *Popular Northern Sotho Dictionary.* Pretoria: J.L. van Schaik.

Kriel, T.J. and E.B. van Wyk. 1989 *Pukuntšu.* Pretoria: J.L. van Schaik.

Lombard, D.P. *et al.* 1991. *Sediba.* Pretoria: Via Afrika.

Prinsloo, D.J. and B.P. Sathekge. 1996. *Shuters New Sepedi Dictionary.* Pietermaritzburg: Shuter and Shooter.

Ziervogel, D. and P.C. Mokgokong. 1979. *Klein Noord-Sotho woordeboek.* Pretoria: J.L. van Schaik.

A Study in the Lexicographical Treatment of Arabic Synonyms

Mohamed Helmy Heliel, Department of English,
Kuwait University, Kuwait

Abstract: Recently three dictionaries of Arabic synonyms were published with the aim of help-
ing Arabic learners, writers and translators. Though Classical Arabic lexicography distinguishes
itself in the field of synonymy, Modern Standard Arabic lacks reliable dictionaries in the field and
hence the importance of analysing these three dictionaries, identifying their deficiencies and sug-
gesting remedies to help establish a sound basis for a user-friendly dictionary of Arabic synonyms.

This paper deals with certain lexicographical features closely related to synonymy, i.e. con-
cerning corpus: selection of headwords and derivatives, arrangements of headwords, contextuali-
zation, collocations, figurative usage, equivalence, polysemy, register and style.

Keywords: HEADWORD SELECTION, DERIVATIVES, HEADWORD ARRANGEMENT,
CONTEXTUALIZATION, COLLOCATIONS, FIGURATIVE USAGE, EQUIVALENCE, POLYSE-
MY, REGISTER, STYLE

Opsomming: **'n Studie van die leksikografiese hantering van Arabiese**
sinonieme. Drie woordeboeke van Arabiese sinonieme is onlangs gepubliseer met die doel om
Arabiese aanleerders, skrywers en vertalers te help. Hoewel Klassieke Arabies homself op die
gebied van sinonieme onderskei, is daar geen betroubare woordeboeke op hierdie gebied in
Moderne Standaard-Arabies nie. Daarom is dit belangrik om hierdie drie woordeboeke te anali-
seer, hulle leemtes te identifiseer en oplossings aan die hand te doen om sodoende 'n stewige basis
te lê vir 'n gebruikersvriendelike woordeboek van Arabiese sinonieme.

Hierdie artikel handel oor sekere leksikografiese kenmerke wat nou verwant is aan sinonimi-
teit, d.w.s. betreffende korpus: seleksie van soekwoorde en afleidings, rangskikking van soek-
woorde, kontekstualisering, kollokasies, figuurlike gebruik, ekwivalensie, polisemie, register en
styl.

Sleutelwoorde: SOEKWOORDSELEKSIE, AFLEIDINGS, SOEKWOORDRANGSKIKKING,
KONTEKSTUALISERING, KOLLOKASIES, FIGUURLIKE GEBRUIK, EKWIVALENSIE, POLISE-
MIE, REGISTER, STYL

1. Introduction

Recently three dictionaries of Arabic synonyms have been published to fill a
niche in modern Arabic lexicography. In 1993 two dictionaries came out: *Mod-*

ern Arabic Thesaurus (MAT) and *A Pocket Dictionary of Synonyms and Antonyms (PDSA)*, to be followed by *A Mini Dictionary of Arabic Synonyms (MDAS)* in 1996. As a matter of fact, synonymy is a grossly neglected area, whether in Arabic linguistic research or in lexicographic investigation. The publication of these three dictionaries undoubtedly marks the beginning of attaching importance to the substitutional or paradigmatic relations in Arabic dictionaries.

A dictionary of Arabic synonyms has to address itself to students, writers and translators, and native and nonnative speakers. Moreover, it has to prove that it deals with current standard Arabic, i.e., the synonyms should be words that actually are in use. It also has to adopt a lexicographical methodology which would help the user find the sought-for Arabic synonyms *easily* and *quickly*. It should provide him/her with the means of differentiating between them and choosing the appropriate one for a particular context.

In this paper we aim to study certain lexicographical features of these dictionaries and suggest ways of treating them. Some of these features are concerned with corpus: selection of headwords and derivatives, the others with lexicographical treatment: arrangement of headwords, contextualization, collocations, figurative usage, equivalence, polysemy, register and style.

2.   Selection of Headwords

The policy adopted in this respect is indicated in both the *PDSA* and the *MDAS*. According to the *PDSA*, the words chosen are "the most frequently used ones". The *MDAS* indicates that it covers "8800 synonyms arranged under main entries of the most frequently used words".

When all the words starting with the letter ب [B] were counted, it was found, as indicated below, that the number of headwords in the *MDAS* far exceeds that of the other two dictionaries:

PDSA	MAT	MDAS
40	52	174

However, there is no agreement as regards

(a)   the derivatives indicated, and
(b)   the number of senses treated in polysemous words,

but what the three have in common is the host of *archaic* and *obsolete* words incorporated, sometimes in the headwords, but very often in the list of synonyms.

Examples

1.   *PDSA*
      Headwords

علندى ['alandii "crude, uncivil"]

أبيل [abiil "hermit"]

ثجر [thajr "breadth"]

ثجة [thujjat "orchard"]

حسبانة [ḥusbaanat "pillow"]

      Synonyms

بواء [buwaa'] under headword كفؤ [kuf'un "efficient"]

جمة [jummat] under headword هامة [haamat "forefront"]

مصدعة [maṣda'at] under headword مخدة [mikhadda "pillow"]

لاحب ، مخرف [laaḥib, makhraf] under headword طريق [ṭariiq "road"]

حلس [ḥils] under headword بساط [bisaaṭ "rug"]

2.   *MAT*
      Synonyms

مضرحى [maḍraḥiyy], هيثم [haytham], قشعم [qash'am], غداف [ghudaaf] under headword نسر [nisr "vulture"]

نبر [nabr] under headword وقح [waqiḥ "rude"]

فدفد [fadfad], نجوة [najwat], يفاع [yafaa'], نشز [nashaz] under headword نجد [najd "highland, plateau"]

معجم [mu'jam], مضرس [muḍḍarras], صيرفى [ṣayrafiyy] under headword مجرب [mujarrab "experienced"]

شنشنة [shanshana] and سجيحة [sajiiḥa] under headword فطرة [fiṭra "innate character, natural disposition"]

3.   *MDAS*
       Synonyms

      أبلم [ablama] under headword ورم [warima "became swollen"]

      ابن بقيع [ibn baqii'] under headword كلب [kalb "dog"]

      شلطاء [shalṭaa'] under headword سكين [sikkiin "knife"]

      سعراء [sa'raa'], أدماء ['admaa'] under headword سمراء [samraa' "a brunette"]

The criterion for selecting the headwords especially in a small-sized dictionary of synonyms like *PDSA* (100 pages), *MAT* (150 pages) or *MDAS* (323 pages) should be their frequency of use in the language. Instead of having these small-sized dictionaries with useless hosts of archaic words, it would be better to start with an experimental edition, choose only the ten thousand most common words in Modern Standard Arabic, omit synonyms related to animals, birds, swords etc. (cf. *MAT* and *MDAS*) which originally are either descriptive or dia-lectal words and by no means synonyms in Arabic and give the common terms the fullest semantic treatment.

2.1   Derivatives

Arabic is a highly derivative language and the translator may have special diffi-culties in finding a synonym for a particular derivative whose meaning is not closely related to the root. Thus derivatives in a dictionary of Arabic synonyms should be carefully selected. The three dictionaries vary a great deal in the way they deal with derivatives. The *PDSA* includes different derivatives from the root kh.w.f (خ.و.ف) such as خاف [khaafa "was afraid"] and the noun خوف [khawf "fear"]. However, this is not systematically practised in the dictionary, e.g. خان [khaana "to betray"] is given as a verb with no other derivatives. This is also typical of the *MAT* which sometimes gives only the verb, e.g. خاف [khaafa "was afraid"], with no other derivatives, and at other times both the verb and the noun, e.g. خاصم [khaaṣama "to quarrel with"] and خصم [khaṣm "adversary"]. The *MDAS* is to some extent more systematic. As derivatives it gives the verb خاف, the noun خوف, the verb خان, the active participle خائن [khaa'in "traitor"] and the *masdar* خيانة [khiyaana "betraying, be-trayal"]. As far as derivatives are concerned, it is the most comprehensive of the three. However, when one item, the verb برد [barada "be cold") in the

MDAS, is compared with derivatives given in an Arabic college-level monolingual dictionary *Majaanii al-Ṭullaab*, it proves to be quite poor (4:14).

<div align="center">

MDAS   *MJ*

بَرَدَ / بُرِد / بَرَّد / تبرَّد / اَبْتَرَد / بارِد /   بَرَد / بَرُد / بارِد / بَرّادة

بَرّاد / بَرْد / برودة / تبريد / بَرَد /

اسْتَبرد / تبرَّدَ (الماء) / تبارَد

4   14

</div>

It has to be noted that the *MAT* indicates that it caters for verbs, adjectives and nouns and uses symbols for them: ف (for verb), ص (for adjective) and س (for noun). Arabic derivatives are by no means confined to these categories. It seems that the authors are imposing an English categorization on Arabic. Arabic derivatives, among others, include اسم الفاعل [ism al-faa'il, the active participle], اسم الآلة [ism al-'aala, the instrumental noun], اسم المـرة [ism al-marra, nomen vicis], اسم المفعول [ism al-maf'uul, the passive participle] and اسم الهيئة [ism al-hay'a, the noun of manner].

3.   Arrangement of Headwords

The order followed in the three dictionaries is alphabetical, but each adopts a specific system to help the user find the Arabic synonyms.

The *PDSA* follows the alphabetical arrangement without any additional index or symbols except a few ones: : for synonyms, * for antonyms, ج for plurality and // for shift of meaning.

In the *MAT* the headword is followed by the root and its grammatical category noun, verb or adjective, e.g.

آب [aaba "to return"] أ . و . ب [a.w.b.] ف (verb): ['aada, raja'a, inqalaba ...]

An index of the roots of headwords is provided together with the words derived from them, e.g.

أ . و . ب [a.w.b.] مآب [ma'aab "return" (n)] آب [aaba "to return"]

Such an index is redundant and does not serve any purpose. The natural order in an Arabic alphabetical dictionary is that reference is made to the root, and not *vice versa*. There is no need for such an index, since the items in the body of the dictionary are arranged alphabetically. Moreover, many words are wrongly placed in the index, e.g.

(a)   بلادة (n)   [balaada "dullness, lethargy"]
(b)   بليد (adj)   [baliid "dull, lethargic"]
(c)   بلد (n)   [balad "country"]

where (c), unlike (a) and (b), is not related in meaning to the verb root ب . ل . د [b.l.d].

In both dictionaries retrieval of any of the synonyms of a particular headword is, in many cases, impossible, e.g.

1.   *PDSA*

اضطرام [idţiraam] and اتّقاد [ittiqaad] given as synonyms for the headword ضرام [ḍiraam "burning, fire"]

2.   *MAT*

إملاق [imlaaq], شظف [shaẓaf] and عوز ['awaz] given as synonyms for the headword فاقة [faaqa "poverty"]

The *MDAS* distinguishes itself by ease of retrieval. An index (188 pages) has been incorporated. It comprises, in alphabetical order, all the words in the dictionary (135 pages), headwords and synonyms. References to headwords pinpoint the particular sense(s) of the word being sought, enabling the user to go directly to the most appropriate main entry, e.g.

ابتز [ibtazza]: سلب [salaba "steal"]

where ابتز in the index refers the reader to the main entry سلب in the body of the dictionary and *ibtazza* is one of the words given in the synonym list of *salaba*.

By incorporating the index the *MDAS* managed to avoid the awkwardness of and waste of time spent on using the cross-references in the body of the dictionary, a defect also present in the otherwise excellent English dictionary of synonyms *Roget's II, the New Thesaurus* (1997) and another Arabic dictionary of synonyms, *Mu'jam al-Ma'aanii* (1971).

4.   Contextualization

The three dictionaries group undifferentiated synonyms together with unde-fined entry words. It is our belief that unless the user, both native and non-native speaker, can him-/herself distinguish correctly between the many semi-equivalents given, he/she can hardly make use of them (see Arnold's (1979) experiment). Though dictionaries of synonyms are restricted by the projected size of the dictionary, words cannot exist in a vacuum, completely divorced from context. Unfortunately, this is often the case in English college-level dic-tionaries of synonyms, even the most recent ones (see *Chambers English Thesau-rus*, 1995 and the *21st Century Synonym and Antonym Finder*, 1993).

   In order to write an adequate lexical entry in a dictionary of synonyms, the lexicographer must give a thorough treatment of four components:

(1)   A concise, carefully-worded denotation of the sense shared by the head-word and its synonyms should be provided (see *Roget's II* 1988 and 1997).

(2)   This should be supplemented by at least one illustrative example show-ing a typical usage context (see *The Oxford Concise Thesaurus* and *Random House Webster's College Thesaurus*).

(3)   In addition to denotation, a word may have a connotation, the suggestive or associative implications beyond its literal sense. To deal with the con-notation problem, status labels should be used, e.g. *derogatory, disap-proving, euphemistic, jocular,* etc.)

(4)   In the vast majority of cases, descriptive synonyms are given, i.e. the synonyms that can be substituted for one another in all declarative sen-tences without affecting their truth-conditions. However, synonyms dif-fer in their expressive or evocative meaning, e.g. قَتَلَ [qatala] which could be translated as *killed*, and ذبح [dhabaḥa] which indicates killing people rather violently or in large numbers and which could be trans-lated as *slaughtered* or *butchered*. Synonyms have different conditions of appropriateness and the lexicographer should shed light on this feature, otherwise the dictionary will be a lethal instrument in the hands of a translator and a foreign learner.

In the majority of cases, a lexical item is in some respects different from its cognitive synonyms or the items with which it shares certain properties. This difference is reflected in:

(a)   The semantic or propositional mode, i.e. the content of the message. Hence the importance of a definition and an illustrative example, e.g. جناية [Jinaaya "crime"], جنحة [junḥa "misdemeanour"] and مخالفة [mukhaalafa "fine"].

(b)   The expressive mode. Without it communication would be impossible. According to Cruse (1986: 274) "every communicative utterance must transmit as part of its meaning an indication of intended propositional attitude". Hence the need for indicating the expressive differences between the synonyms incorporated in the dictionary.

5.   Collocations

Apart from the interface between a speaker's intentions and language, there are interactions among linguistic items constituting discourse (Cruse 1986: 277) which are unpredictable and a source of difficulty for the translator, both native and nonnative speaker, of the language into which he/she is translating. These are collocations. Synonyms may be of the same semantic field, but may differ a great deal in their collocational range or the set of contexts in which they occur. Related to the field of "destruction", for example, we may have كسر [kasara], هدم [hadama], صدع [ṣada'a], حطم [ḥaṭṭama] and دك [dakka], for each of which a special collocate is used:

كسر [kasara] / door, leg, bread, plate
هدم [hadama] / building
صدع [ṣada'a] / wall, plate, glass
حطم [ḥaṭṭama] / (technical) the atom
دك [dakka] / wall, castle, enemy positions.

The fact that the three dictionaries do not indicate the collocates, detracts from their value (cf. *The Oxford Concise Thesaurus*). A collocate is an essential part of the meaning of the whole cohesive unit known as collocation and hence it should be included.

6.   Figurative Usage

When the three dictionaries were used for finding synonyms for the lexical item هد [hadda "destroy"] used figuratively in the sense of "ruin one's health", they

proved to be inadequate. Figurative usage is an area that has been neglected even in monolingual Arabic dictionaries. Items used to indicate "destruction" in Arabic could easily be used figuratively, in which case they need a special lexicographical treatment in a dictionary of synonyms. The following are illustrative Arabic verbs, first in their literal and then in their figurative sense, translated into English with the subject or object indicated:

Verb	Literal sense	Figurative sense
حطم [ḥaṭṭama]	destroy	wreck/ruin (one's hopes)
هد [hadda]	pull down, raze	(disease) ruined or weakened him
خرب [kharaba]	destroy	(a) ruin, bankrupt (economy)
		(b) sabotage (factory, machinery)

One of the dictionaries which attach importance to this aspect, is *Harrap's English Synonyms* (1989) where a clear distinction is made between literal and figurative meanings (see for example the verb *black out*).

7.   Equivalence

As a rule, the three dictionaries adopt the wrong principle of taking the word as a lexicographic unit. For instance, the headwords are, except in a few cases, confined to single verbs without prepositions, though the preposition in Arabic prepositional verbs (see Heliel 1992) determines and distinguishes the meaning of the verb, e.g.

أقبل (إلى) [aqbala (ilaa) "come to (a place or person)"]

أقبل (على) [aqbala ('alaa) "to take interest in"]

Although prepositional verbs are sparingly given as synonyms in the *PDSA* and *MAT*, they cannot unfortunately be retrieved. The proper prepositional particle normally accompanying the Arabic verb and determining its sense should be provided in dictionaries of synonyms, whether as headwords or in synonym lists. There should also be a means of retrieving them.

On the whole, the three dictionaries adopt the narrower classical sense which limits itself to the identification of meaning between words only. In this respect, a comparison should be made between Arabic and recent college-level English dictionaries of synonyms. The *Merriam-Webster Concise School and Office Thesaurus* (1991) and *The Oxford Concise Thesaurus* (1995) are sources that could

be utilized. The *MAT*, it has to be noted, sparingly gives construct phrases and idioms in the synonym lists, which is a step forward in Arabic lexicography, e.g. فريد [fariid "unique"], منقطع النظير [munqaṭi'u al-naẓiir] and نسيج وحده [nasiiju waḥdihi] which are construct phrases of noun plus noun.

Synonymity relations in natural languages, it is true, not only characterize single words, but also larger linguistic units. Examples of these units in Arabic are *construct phrases* (idaafa), *binomials, collocations, prepositional verbs* and *idioms*. These units have to be incorporated in an Arabic dictionary of synonyms.

8.   Polysemy

A worthwhile attempt at dealing with different senses of Arabic lexemes and their synonyms, especially the basic ones, has been made by the *MDSA*. Though it is restricted by the size limitations, it surpasses the other two dictionaries, e.g.

عدل ['adala "to act justly"]
عدل (عن) ['adala 'an "change one's mind"]
عدل [adala "put right"]
عدل ['adala "deviate"]

The *PDSA* and the *MAT* give only one sense of the verb "to act justly", though the prepositional structure عدل عن ['adala 'an] in the sense of "change one's mind" and the verb عدل ['adala] in the sense of "put right" are frequently used. Thus a semantically carefully designed plan for selecting, dividing senses and distinguishing between them should be drawn up before compiling an Arabic dictionary of synonyms.

9.   Register and Style

One of the basic differences between synonyms which has not been indicated in the three dictionaries, is register, i.e. a variety of language used by a single speaker which is considered appropriate to different occasions and situations of use (Cruse 1986: 283). Here we stress the importance of field which refers to the topic of discourse (*legal, scientific, medical, political,* etc.) and style (*derogatory, euphemistic, poetic, appreciative, formal, informal,* etc.). It is inadequate to list a

number of indiscriminate synonyms, one after the other without indication of register or style.

"The multiplication of synonyms," as Cruse (1986: 284) says, "is most marked in the case of referring to areas of experience which have a high emotive referring significance such as death, sex, excretory functions, money, religion, power relations and so on." From these areas of experience the following example of synonyms related to sex has been chosen. It could serve as an illustration of an entry with suggested labels of field and style, e.g. جماع [jimaa' "sexual intercourse"]: رفث [rafath], مباشرة [mubaashara], وطء [wat'], نكاح [nikaah], اتصال جنسي [iṭṭiṣaal jinsii], مضاجعة [muḍaaja'a], مباضعة [mubaaḍa'a], where جماع could be indicated as (neutral), رفث as (Islamic law, Koranic language), مباشرة as (euphemistic), وطء as (derogatory), derived from the verb وطأ [wata'a "to tread or trample underfoot, to step or walk over, to mount"], اتصال جنسي as (neutral), مضاجعة as (euphemistic "sleeping with"), مباضعة as (derogatory), derived from بضع [bud' "female sexual organ"].

10.   Conclusions

Through an analysis of the lexicographical treatment of synonymy in three recently published Arabic dictionaries, the deficiencies have been identified and remedies suggested. Most important among these deficiencies is the decontextualization of the lexical items which depreciates their lexicographical value. The *MDAS* which is the most recent of the three, breaks new ground in certain respects. However, much has to be done about linguistic multiword units, derivatives, item selection, denotation, connotation, equivalence, polysemy, collocations, figurative usage, register and style which would pave the way for a reliable dictionary of Arabic synonyms.

References

Arnold, Donna I. 1979. Synonyms and the College-Level Dictionary. *Dictionaries* 1: 103-112.

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

Heliel, Mohamed Helmy. 1992. Verb-Particle Combinations and the Arabic/English Translator. Robert de Beaugrande *et al.* (Eds.). *Language, Discourse, Translation in the West and Middle East*. Amsterdam: John Benjamins.

Dictionaries Consulted

Chambers English Thesaurus. 1995. Edinburgh: Chambers.

Harrap's English Synonyms. 1989. London/Paris: Harrap Books.

Majaanii al-Ṭullaab. 1995. Beirut: Daar al-Majaanii.

Merriam-Webster Concise School and Office Thesaurus. 1991. Springfield, Mass.: Merriam-Webster.

A Mini Dictionary of Arabic Synonyms (Arabic - Arabic). 1996. Beirut: Librairie du Liban.

Modern Arabic Thesaurus (Arabic - Arabic). 1993. Beirut: Librairie du Liban.

Mu'jam al-Ma'aanii. 1971. Baghdad: al-Zamaan.

The Oxford Concise Thesaurus. 1995. Oxford: Oxford University Press.

A Pocket Dictionary of Synonyms and Antonyms. 1993. Beirut: Librairie du Liban.

Random House Webster's College Thesaurus. 1997. New York: Random House.

Roget's II, the New Thesaurus. 1988. Expanded Edition. New York: Houghton Mifflin.

Roget's II, the New Thesaurus. 1997. Third Edition. New York: Berkeley.

21st Century Synonym and Antonym Finder. 1993. New York: Dell.

Collocability in Languages for Special Purposes (LSPs): Some Preliminaries

Francis E. Knowles, *Department of Languages and European Studies,*
Aston University, Birmingham, Great Britain

Abstract: This paper is concerned with the language professional discourse communities use for their internal communication. The characteristics of these languages for special purposes (LSPs) are many and varied, as well as being underresearched. The focus adopted here is to examine the phenomenon of multiword units, many of which are orthographic pluralities designating conceptual singularities. It is important to recognise at the outset that colocation is not the same as collocation. Analysis and systematisation of these textual "clusterings" is intended to separate them into two radically different types of entity: multiword segments possessing terminological status; and collocative material. The methods used to achieve the above objective are both qualitative, i.e. micro-environmental analysis, and quantitative, i.e. statistical patterning exhibiting a certain level of frequency and constancy. Collocational material quoted here also shows by its configuration that discourse communities use collocations to which the general public are not inured and with which they may not necessarily be familiar at all.

Keywords: CHUNKING, CO-OCCURRENCE, COGNITIVE ENTITY, COLLOCABILITY, CONCEPT, CORPORA, DISCOURSE COMMUNITY, DISTRIBUTION, DYAD, ENCYCLO-PAEDIC COMPETENCE, LANGUAGE FOR SPECIAL PURPOSES, LEXICOGRAPHY, MENTAL LEXICON, MULTIWORD UNIT, OCCURRENCE, SOCIOLECT, STATISTICS, TERMINOLOGY, TERM, TRIAD

Opsomming: Kollokeerbaarheid in tale vir spesiale doeleindes (TSD's): 'n Aantal inleidende aspekte. Hierdie artikel handel oor die taal wat professionele gespreks-gemeenskappe gebruik vir interne kommunikasie. Die kenmerke van hierdie tale vir spesiale doel-eindes (TSD's) is baie en veelsoortig, en word ook onvoldoende nagevors. Die fokus wat hier ge-kies is, is om die verskynsel van meerwoordige eenhede te ondersoek, baie waarvan ortografiese meervoudighede is wat konseptuele enkelvoudighede benoem. Dit is belangrik om van die begin af te besef dat kolokasie nie dieselfde as kollokasie is nie. Analise en sistematisering van hierdie tekstuele "klusterings" is bedoel om hulle in twee algeheel verskillende tipes entiteite te verdeel: meerwoordige segmente wat terminologiese status besit; en kollokatiewe materiaal. Die metodes gebruik om bogenoemde doel te bereik is sowel kwalitatief, d.w.s. analise van mikro-omgewings, en kwantitatief, d.w.s. statistiese patroonvorming wat 'n sekere vlak van frekwensie en konstant-heid vertoon. Kollokasionele materiaal wat hier aangehaal is, wys ook deur die konfigurasie daar-van dat gespreksgemeenskappe kollokasies gebruik waaraan die groot publiek nie gewoond is nie en waarmee hulle glad nie noodwendig bekend mag wees nie.

Sleutelwoorde: SEGMENTERING, MEDEAANWESIGHEID, KOGNITIEWE ENTITEIT, KOLLOKEERBAARHEID, KONSEP, KORPUSSE, GESPREKSGEMEENSKAP, VERSPREIDING, DIADE, ENSIKLOPEDIESE TAALVERMOË, TAAL VIR SPESIALE DOELEINDES, LEKSIKO-GRAFIE, MENTALE LEKSIKON, MEERWOORDIGE EENHEID, VOORKOMS, SOSIOLEK, STATISTIEKE, TERMINOLOGIE, TERM, TRIADE

> *D'autre part, les collocations ne sont pas des unités de lexique: tout ce qui est de l'ordre du mot — les mots composés, par exemple — devrait être écarté. Ceci est hélas plus facile à dire qu'à faire, particulièrement dans les langues de spécialité. Les associations nom + nom et nom + adjectif, par exemple, qui y sont nombreuses, sont souvent considérées, à cause de leur caractère de fixité et du fait qu'elles font référence à une notion qui occupe une place précise dans une taxinomie, comme des unités lexicales. Mais les critères de différenciation entre mot et "non-mot" (ou plutôt terme et "non-terme") ne sont pas toujours limpides. Dans un dictionnaire de collocations d'anglais général, il n'est pas trop difficile d'exclure des suites comme "generation gap" ou "battering ram", mais en langue de spécialité, les problèmes sont un peu plus complexes.*

> Henri Béjoint et Philippe Thoiron

Preamble

The above quotation — which, interestingly, chooses English examples to support its thesis — sets the scene admirably for what is to follow in due course on what is acknowledged to be a very significant aim, within text linguistics, discourse analysis and, of course, lexicography and terminology studies: firstly, the location within running text and the subsequent analysis — either by hand or by computer — of units of meaning which comprise more than one (ortho)-graphic word; and secondly and much more importantly, the overarching phenomenon of "chunking". However, before we can sensibly discuss these matters, the major focus of this paper, we must tread a path to the promontory from which we will best be able to view the scene.

Point de départ

Chunking is a primarily mental phenomenon, the symptoms of which are to be found in linguistic formulations. Put in other terms, chunking is a psychological cause, associativity, which has linguistic effects, juxtaposition or blending. It follows that any macro-investigation of chunking as a phenomenon is best pursued by the micro-investigation of "chunks", stretches of agglutinating linguistic material which are felt to represent segments of thought rather than just fragments. Chunks are also entities about which and about the use of which

there is some sort of social consensus. However, chunks hardly ever seem to be static in developmental terms; rather, they appear to be on a trajectory towards "explicitness" and involved in a systemic combat driven by language users and aimed at a prevaricating homeostatic trade-off between lexis and syntax. Once they have reached it, whatever that means in orthographic terms, they are no longer inchoate and they are no longer ambivalent: they have, in fact, merely become "normal", if somewhat lengthier, units of language — in common perspective, at the very least.

It is a well-known but regrettable fact that very, very few language communities possess satisfactory collocations dictionaries, the global mission of which — in addition to their strictly utilitarian function — is to demonstrate, reactively and proactively, that colocation is not collocation. The former, that is, mere juxtaposition, is either volatile or an example of a "fixed" multiword unit, whereas collocation manifests associative regularities — sometimes based on assonance, sometimes on prosodic effect, sometimes on less easily definable criteria — that can be shown to be statistically significant rather than "binary" in the sense of present versus absent. The normal unavailability of collocations dictionaries is a great pity because that is exactly what advanced learners need and, indeed, what many native speakers hanker after too. In fact, it is not stretching things too far to say that first-class collocational control is the hallmark of the true L2 expert; collocational control is, of course, normally the last linguistic subsystem to be mastered by L2 learners who proceed to an advanced level. Correct deployment of collocations is particularly important for anyone striving for authenticity of performance within a particular professional sociolect, such as the language of medicine or economics.

Lexicographic tools are urgently needed to help those who have not had the advantage, from their early years, of "statistical exposure" to lexical patterning which, although often analytically idiosyncratic, is habitual. It follows that investigation of the phenomenon of collocability relates to the hidden, subliminal patterns and rhythms of language which need to acquire a description having its basis rather in the statistics of occurrence and co-occurrence. Such lexicographical tools need to be primed — in the sense of acquiring the data they treat — by computational tools, the purpose of which is to identify, excerpt and prioritise relevant lexical material in a form suitable for subsequent lexicographical treatment. However constant this general aim may be, the actual means of fulfilling it will differ from one language to another. In all cases, of course, it is a question of locating stable and authentic multiword material — this is probably the only initial common denominator.

The "intrusion" of terminology — in a double sense!

In languages for special purposes (LSPs), a particular problem is posed by the terminological usages which proliferate in such modes of discourse. It so happens that technical terms are themselves often composed of more than one

orthographic word. Some criterion is therefore needed to discriminate between terms and collocations. The former are cognitive entities inserted as such into flowing text, the latter are, predominantly, inserted separately into text, yet their associativity helps that text to flow and to cohere. However, the position is given a further twist by the fact that genuine collocability also manifests itself in LSP discourse.

Let us commence by suggesting that Béjoint and Thoiron's formulation "lexical units" can be the cause of some confusion as this piece of nomenclature focuses on the messenger rather than the message, so to speak. Is it not preferable to refer to cognitive units or cognitive entities — which, naturally, require linguistic form, at least for the purposes of interpersonal communication? Quibbles apart, all of these terms attempt to capture the "truth" that it is not only words which retain their referents outside text; word-groups and phrases may also do so. This is entirely in accordance with the customs, habits and social compact established within societies for structuring individual and group experience. These lexical pointers along with whatever they point to constitute the mental lexicon of individuals and, by aggregation, the working consensus of society about how to interact linguistically with the world. Items in mental lexicons are mappings, complex rather than simple, between mental images and the names which point to them. In some ways the names are secondary and subservient: they may alter, leaving their referent unchanged, for instance. Yet they have power: names tame! The "downside" of this, via a different analogy, is that the natural dynamics of mental images are reduced to slow-motion film or even to still photographs. Yet names are also the major instrument of effability and they may acquire or discard referents as a result. Most of all, names, normally named "words", are at the very basis of social intercourse. Of course, much social intercourse takes place on a distinctly professional level, invoking/evoking — and instantiating by words — established concepts (Latin: ideas "taken together", i.e. holistically merged), phenomena or artefacts current within particular discourse communities (DCs) for the simple (!) reason that they have been professionally introduced, negotiated and subsequently validated by them. This process is necessary so that DC members can communicate with each other on the same terms, in a double sense, understanding without either external or internal mediation and not merely comprehending each other's thinking and reasoning. In this way a DC's sophistication increases, first of all encyclopaedically (at least for the pioneers) and then terminologically. The route from term to concept is immediate for DC members; for those outside the DC it is most likely to be mediate. It can, obviously, also be a cul-de-sac. The people of a DC and their purposes always have precedence over their technolect, so to speak.[1] Putting this point in the parlance of sociology, the DC — even represented by its neophytes! — is the independent variable, the linguistic resources that they call on to express themselves individually and corporately are the dependent variable(s). A less obvious corollary, even to those directly involved, of this system of socio-professional "information exchange" is the

continual growth — not necessarily linear! — in the given paradigm's cohesion. It should in fact be noted, in the margins, that this "growth" can sometimes lead to the sclerosis of orthodoxy, hopefully then followed by genuine renewal based on new relationships and a new dialectic, plus a new or at least revitalised discourse to go with them. In such extreme circumstances a DC's members may well be prisoners of their thought or "mind-set"; they are not normally prisoners of their terminology because they can realign and reseman-ticise it by agreement.

How and where to establish a "base" for investigative purposes?

Those who observe and analyse DCs because of an interest, even a utilitarian interest, in their technolects are almost exclusively never members of them. This, of course, puts them in the empirically best possible but still suboptimal position to carry out their work as dispassionate investigators. The position is still less than perfect because little or no regard is given to what is encyclopae-dically expressible (i.e. "encyclopaedic competence"?), only to what is actually expressed by linguistic means (i.e. "linguistic performance"?). However, that is an easy statement to make. After all, there are constraints. What is more diffi-cult is making a judicious and also successful choice of methods for the investi-gation. It seems that linguists are — very sensibly — increasingly resolving the problem of linguistically modelling discourse, not least LSP discourse, by pri-mary recourse to quantitative rather than qualitative methods. Qualitative ana-lysis of LSP discourse by "frontal attack" requires an exceedingly well-stocked arsenal of tools and techniques, many of them still evolving and maturing. The yield tends to be on a micro level, the analogy of intensive "case-study" ap-proaches which always seem to studiously eschew any mention of "compa-rators" when the real questions are always the following. First and foremost, just how representative of wider practice and habits are the "findings"? Second-ly, has the "gulf" between encyclopaedics and linguistics been successfully bridged? Quantitative methods offer a way round this dilemma — at a cost! These methods can summon from an arbitrarily large corpus exhaustive lists of linguistic "segments" along with their frequency of occurrence. Although the frequencies are numbers, this information is of high qualitative value because the individual items in the associated lexis can then be described in terms of their actual functional load in the running source text and also in terms of their putative "market share" in analogous "still-to-be-written" text. Such material is of prime value for pedagogical purposes, principally because it is direct and powerful evidence of authentic usage. This then is the nature and purpose of quantitative modelling in linguistics — specifically lexis in this case — and it is arguably the best route open to investigators, particularly those with strong pedagogical interests.

The crux of the matter

We are now at a point where we can begin to draw together the two main strands of this paper. The first strand is obvious enough: how do we isolate within running text, firstly, DC cognitive units expressed by single orthographic words and how do we isolate those cognitive entities which are by definition conceptual "singularities" even though they are at the same time orthographic "pluralities"? Obviously, such matters of procedure and the variety of contingent cruces will depend on the particular natural language involved in any such investigation.[2] In some languages, of which German may be taken as a suitable representative, the linguistic designation/configuration of many concepts, both within everyday life and within specialist DCs, often occurs via the compounding of separate lexical items into one orthographic unit. The linguistic process involved is one of holisticisation — who can say whether this is a true reflection and model of the Gestalt formed by the fusion of the latent (!) mental constructs? Just as the concept is "taken together" and becomes molecular, so the component [Latin: "putting together"] names are similarly batched and merged:

> *Farbfernsehempfänger, [Farb[{fern}{seh}[empfänger]]]:*
> *colour television receiver, [colour_ [{tele}{vision}_ [receiver]]];*

> *Preisfestsetzungsrichtlinien, [Preis[fest[setzungs[{richt}{linien}]]]]:*
> *price-fixing guidelines, [price-fixing_ [guidelines]].*

English achieves the same objective of terminologically nominating a concept by juxtaposition and/or hyphenation. Phrasal formulation is also often an option in English: *guidelines for price-fixing / guidelines for fixing prices.*[3] It should be apparent, even on this slender basis, that delineating technical terms in running text is not an easy business as far as English is concerned. It is, of course, not unimportant which of the three available methods a language uses, and in what "mix", for structuring syntactic meaning: element order, function words, or inflection. The characteristics of English, with its "residual" inflection and consequent reliance on "neighbourhood" are particularly impervious to analysis by traditional methods. The chunking phenomena[4] of interest yield only to subtle environmental analysis, often supported by statistical profiles. An instance of this is the English "phrase term"

> *stability augmentor pitch axis actuator housing support*

which brackets as

> *[[[[[stability augmentor] [pitch axis]] actuator] housing] support].*[5]

See also Appendix I for a brief excerpt of text on navigation[6] with terminological usages "coded up" — in an intuitively simple manner — as single ortho-

graphic words. The paragraph shown represents only about one tenth of the entire text but the artificial graphic system used cuts the token count by 70 items, while — because of the batching effect — reducing the number of types by only a mere half dozen!

A fundamental distinction and its implications

It is vitally important to understand that the formation/usage of such multiword units in English — and in many other languages — has nothing whatsoever to do with collocability, the chronologically second but thematically and substantively first strand in our discussion. It is a fact of life that, alongside everyday collocations of greater or lesser currency, professional collocations also exist. It is to this feature of language that we now turn our attention, encouraged and helped by the following dichotomising principle: If multiword units are terms they are not collocations; if they <u>are</u> collocations they cannot by definition be terms.

Operationally and computationally, this principle reduces to the need to distinguish in English between those spaces between orthographic words which separate and those which glue and bond. Fortunately, this problem normally solves itself in the course of computer-driven analyses by dint of the statistical force of grouping.

Collocation: a macro phenomenon with micro differentials

Collocation is a linguistic phenomenon which manifests itself via the habitual but often seemingly idiosyncratic or inexplicable[7] association by co-occurrence in a microcontext, in parole, of one word with another. The strength of such associations may vary, both within a community and over time. There is some basis for stating that collocation is a phenomenon which can partly be measured in terms of its surface systematicity. Collocation is a linguistic subsystem which individual L1 users acquire, as noted above, not by study but by sheer exposure to the phenomenon from their early years on. Each new stimulus adds incrementally to the speaker's ability to control his/her own linguistic behaviour and to make certain judgements about other people's linguistic behaviour. In many ways, collocational control — which also embraces the ability not to produce incorrect collocations! — is not merely a hallmark of linguistic maturity, it is also a kind of shibboleth. That is why it is so important — it is also why collocational control is difficult to master, seriously so if the person attempting to establish control is not a native speaker of the language concerned. Such a person has to adopt or devise a chronologically optimised regimen for acquiring collocational control: the problem is that emulation of native speakers is too slow and simulation is too risky. The simulation has to be con-

nectionist anyway, developing its own neural network, to use two terms in a literal rather than a metaphorical sense.

Given that collocations occur freely and naturally in all those sociolects which are "everyone's property", it would be somewhat strange if they did not also occur in special languages, the LSPs associated with and "owned" and regulated by DCs which group professional persons by gradations of various types. It is a very pertinent — and underresearched — question as to whether exposure to the collocational habits of technolect speakers instils collocational orthodoxy of linguistic behaviour in the same way as in general circumstances. One hypothesis must be that there are differences between the two situations alluded to. People joining obvious professional groupings do so at a time in their lives when their adult native language habits are more or less established. Nonnative speakers have a more obvious problem. Either way, there is a learning and adaptation process which has to be much more purposive and "supraliminal"; this, in its turn, requires deliberate reflection, learning and greater awareness about linguistic, sociolinguistic and metalinguistic matters. The urge to adopt "group-speak" is more keenly felt even if "group-thought" lags behind somewhat! Yet there is an ongoing tension between the demands of the logical and ontological framework of the professional DC and the requirements of illocutionarily successful linguistic formulations and terminological codifications. The "genre" problem within a DC's discourse is an added subtlety, not least in terms of the collocational practice of DC members vis-à-vis other DC members or "layfolk", to name only one such communication axis.

A LSP-based investigation of collocability

The (ongoing) research on which the above remarks are based is an investigation of LSP discourse in the areas of finance/banking and — partly — of medicine.[8] The texts used in the investigation come from two types of source. Firstly, the holdings of the Aston Scientific and Technical English Corpus (ASTEC) provided the medical text. Secondly, and quantitatively more importantly, the finance and banking material was read from a commercially available CD-ROM which contains the entire text of the 1993 editions of the *Financial Times*, classified into subcorpora designated by theme and genre.[9]

Once a *prima facie* satisfactory corpus of text has been identified, it can then be processed. For very large corpora (i.e. 2 million tokens) the machine processing is carried out in a UNIX environment, via a suite of programs expressly designed for the purposes described in some detail here. The method of initially selecting and then identifying multiword textual segments is simple enough. The objective is to find, inspect and count batches, no more than arbitrary fragments really, consisting of nine orthographic words. Once found, these fragments can be sorted and arrayed by their "middle", i.e. fifth, word. The scanning window then moves one word to the right, so to speak, and iterates the same basic process. Some further sophistication is available — at addi-

tional computational expense — by the following method which has been shown to be operationally viable with groups of two words (dyads) and three words (triads). The algorithm works as follows: in the case of dyads, the words comprising them must be within a sentence boundary, with no intervening punctuation marks. Neither member of the dyad may be a function word. For triads the strategy is similar, with the "concession" that the middle word in a triad may be the function word *of*. The yield, unlemmatised, from the algorithm is sorted by frequency and then inspected with a view to identifying segments rather than fragments. Many of the segments are proper names, either institutional or personal, of encyclopaedic significance to the DC concerned. Those that are not are either terms representing cognitive units or they are potential collocations. A list of 553 triads — representing occurrence frequencies descending from 308 to 3 inclusive — was derived, by the above "unintelligent" methods, from a sample of the World Stock Markets corpus. Of the items in this list 62% represented cognitive units (*interest rate cut*), proper names (*UBS Phillips Drew*), jargonistic turns of phrase (*dealers took profits*), or collocational expressions (*volume remained flat*). Similarly, a list of 1268 dyads — representing occurrence frequencies descending from 431 to 5 inclusive — was then derived from the World Stock Markets corpus. Of the items in this second list 44% represented the same set of lexical categories: cognitive units (*corporate earnings*), proper names (*Bill Clinton*), jargonistic turns of phrase (*depressed sentiment*), or collocational expressions (*Paris fell*). An analogous analysis of a medical corpus drawn from the *British Medical Journal* yielded 121 terms from a list of 883 triads (see Appendix II).

Computational experience has shown, at this stage, that the main operational requirement is copious output which can then be refined by classification even if the discard rate is high. A surfeit of information is better than a dearth thereof! The steps needed to achieve copious output are easy enough to understand and appreciate. A large corpus will yield a rather long list of types, to which is appended the number of tokens for the relevant type, that is, its occurrence frequency. No attempt is made at all to group the types together into their "lemma set" and hence attempt to look at the behaviour of the canonical form. This can be done separately for English text, but even for inflected languages (such as Polish) it is more profitable in fact to look at the types independently of their lemma "allegiance". The point is that the various forms of the lemma all lead different lifestyles in text, not just in terms of frequency but also in terms of the lexical company they keep. The grouping action of lemmatisation actually obscures important information about type frequency and functional load within text and — above all — about type behaviour in terms of semantic differentiation. Be that as it may, it is probably the type frequency list (plus token counts) which is the main priming material for further investigation. Clearly, the analyst has choices to make, invidious choices, possibly. Attention tends to gravitate either towards types with a fairly high frequency or towards items of almost curiosity value. Having once selected an item for

further scrutiny, the analyst goes back to the corpus and retrieves all the lines containing the item in question. This is the necessary preliminary to the study of the said item's distributional characteristics — that is, seeing how its meaning is "defined" by the constancy of its neighbours, its morphological variants and, if such applies, its — often Janus-like — polysemy (see Appendices III and IV). This is, in fact, the only logistically feasible way of studying words as "chameleons". It can be noted in passing that any search profile can be entered as a string rather than as a fully-fledged word if interest is actually focused on a lemma or hyperlemma (as in the set: *STRUCture, reSTRUCture, STRUCtural, STRUCturalism, deSTRUCtion, conSTRUCtive* etc.). The result of the actual search is a set of one-line concordance citations which may supply information that is adequate for the analyst's purposes. If this is not the case, the entire sentence from which the keyword has been somewhat artificially extracted can be retrieved for inspection. As a further specific aid to collocational studies, a so-called synoptic chart (see Appendix V) can be produced of the node's left and right neighbours to a depth of four items. The individual items arrayed are "decorated" with their own occurrence frequencies. This is the basic method of attack for the purposes of collocational analysis but it is worth mentioning in this precise context that an overview of the valency pattern of the node words selected often also comes across very forcefully. Valency patterning is, of course, held by some to be an integral pattern of collocation studies. The techniques enumerated here have also worked, incidentally, to good effect on material in languages other than English.

It is clear from the initial stages of the above ongoing programme of research into LSP collocability that, as regards text generation, a marked degree of formulaic writing exists in the World Stock Markets financial press. This is true with respect to sentence structure and to the choice of words used for the purposes of qualification. This is either a tendency to avoid any disorientation in readers' minds that might be caused by novelty of usage or — just possibly — the result of summoning up, by reason of time pressures, certain prefabricated utterance structures of an algebraic kind and leaving the few variables mostly unchanged (see Cowie 1991 and 1992).

The techniques described immediately above achieve their greatest effect in the way they present for inspection the left-hand and right-hand neighbour(s) of the nodes chosen. Scrutiny of these environments leads naturally and fairly rapidly to the identification of any collocations present. Let the professional collocations used to describe the quality of stock market trading in Appendix VI speak for themselves and invite reflection, but let not the merits of the material in Appendices II, III and IV be overshadowed in the process!

Implications and Renvoi

The implications of what precedes are numerous but can nonetheless be classified as either practical or theoretical. Let us deal with the practical implications

first. Computationally, the impediments impinging on both principle and practice are minimal. Largely, the name of the game continues to be the game of the name! In terms of particular investigations a great role is played by pedagogical impetus: in fact, the agenda can safely be driven by this as no gulf is likely to open up between those who seek and find material[10] and those who place it in the lexicographer's/terminographer's crucible.

On the level of theoretical considerations, there is much more to speculate about. Let us ask some questions which are not merely rhetorical. What is the precise nature of the mismatch between orthographies and the cognitive units they must find designations for? What is chunking — with respect to term formation and collocability — and how can synchronically- and diachronically-focused studies be pursued with both descriptive and analytical purposes? How does the primeval soup of language congeal in different ways and in what contrasting and competing ways within particular languages? What is the logical, semasiological, ontological and analogical nature of, say, the sets (or classes?) of adjectives collocating with certain nouns or groups of nouns? Is this all *sui generis* or is it describable in generic terms?

Acknowledgement

Notes

1.   For further discussion of the linguistic and sociological basis of DCs, see Knowles 1997.
2.   Let us, for the purposes of this discussion, set aside the cross-cultural differentials and the occasional mismatches between the encyclopaedic networks of "national" DCs. Such mismatches cause communication problems by cropping up in LSP discourse potentially addressed to all members of a supposedly international and "shared" specialist/intellectual culture.
3.   It may be that German's/Germans' tendency to generate complex/compound neologisms in the form of large numbers of one-word lexicalisations (their words are *Komposita* or *Zusammensetzungen*!) is explained not just by the linguistic resources available but also by holisticising perceptions.
4.   Languages such as Arabic possess a mechanism known as iDaafa (annexion) in which two (occasionally more) words, while remaining orthographically separate, unite to form a separate cognitive unit. An instance of this is Arabic *ra's maal*, literally *head of money*, i.e. *capital*. These units are fully lexicalised and are atomic in the sense that no other linguistic material may be interposed between the two orthographic words involved. Hebrew, Farsi and Turkish also have very similar structures. Note that these compounds are exactly that, they are not collocations, although collocation might well have been their origin.
5.   See Varantola 1984: 42.
6.   From the article "Navigation" in Amerongen 1977.
7.   Why do English speakers always *enter a caveat*? *Enter*, here, is a truly transitive verb, a usage very different from *"entering" a building*.

8.   I express a great debt of gratitude to Patricia Thomas and to my close colleague Dr. Peter Roe for insights gained from the many continuing conversations we regularly have on the matters discussed here. A particular vote of thanks is also due to Peter for the specification and implementation of the ATA (Aston Text Analyser) software package which is now under commercial development.

9.   We express our sincere gratitude here to the FT for permission to make extensive use of the CD-ROM materials.

10.   Terms or collocations — on this occasion we draw no contrast!

References and Bibliography

Aarts, J. and T. van Heuvel. 1985. Computational Tools for the Syntactic Analysis of Corpora. *Linguistics* 23: 303-335.

Amerongen, C. 1977. *How Things Work 1.* Translation by L. Zgusta of *Wie funktioniert das?* (Mannheim: Bibliographisches Institut, 1963). St.Albans: Granada.

Béjoint, H. and P. Thoiron. 1992. Macrostructure et microstructure dans un dictionnaire de collocations en langue de spécialité. *Terminologie et Traduction* 2/3: 513-522.

Burger, H. *et al.* 1982. *Handbuch der Phraseologie.* Berlin: De Gruyter.

Church, K. and P. Hanks. 1990. Word Association Norms, Mutual Information and Lexicography. *Computational Linguistics* 16(1): 22-29.

Cowie, A.P. 1991. Multiword Units in Newspaper Language. Granger, S. (Ed.). 1991. *Perspectives on the English Lexicon.* Louvain-la-Neuve: Cahiers de l'Institut de Linguistique de Louvain.

Cowie, A.P. 1992. Multi-word Lexical Units and Communicative Language Teaching. Arnaud, P. and H. Béjoint. 1992. *Vocabulary and Applied Linguistics*: 1-12. London: Macmillan.

Fillmore, C. and B. Atkins. 1994. Starting where the Dictionaries Stop: The Challenge of Corpus Lexicography. Atkins, B.T.S. and A. Zampolli (Eds.). 1994. *Computational Approaches to the Lexicon*: 349-393. Oxford: Oxford University Press.

Fontenelle, T. 1992. Collocation Acquisition from a Corpus or from a Dictionary: A Comparison. Tommola, H. *et al.* (Eds.). 1992. *EURALEX '92 Proceedings. Papers submitted to the 5th EURALEX International Congress, Tampere, Finland*: 221-228. Tampere: University of Tampere.

Knowles, F. 1997. Lexical Carthography. Somers, H. 1997. *Terminology, LSP and Translation*: 125-140. Amsterdam: Benjamins.

Morgenroth, K. 1993. *Le terme technique.* Tübingen: Niemeyer.

Pavel, S. 1993. La phraséologie en langue de spécialité. Méthodologie de consignation dans des vocabulaires terminologiques. *Réseau International de Néologie et de Terminologie (RINT)* 10: 67-82.

Pilz, K. 1974. *Phraseologie.* Stuttgart: Metzler.

Rousseau, L. 1993. Terminologie et phraséologie, deux composantes indissociables des langues de spécialité. *Réseau International de Néologie et de Terminologie (RINT)* 10: 9-13.

Sager, J.C. 1990. *A Practical Course in Terminology Processing.* Amsterdam / Philadelphia: John Benjamins.

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

Skehan, P. 1996. Second Language Acquisition Research and Task-based Instruction. Willis, J. and D. 1996. *Challenge and Change in Language Teaching*: 17-30. Oxford: Heinemann.

Varantola, K. 1984. On Noun Phrase Structures in Engineering English. Annales Universitatis Turkuensis, Series B, 168. Turku: University of Turku.

Zgusta, L. 1967. Multi-word Lexical Units. *Word* 23: 578-587.

Appendix I

As the **water_depths** are **marked** on **nautical_charts**, a further **navigational_aid** is **provided** by **soundings**, i.e., **systematic measurements** of the **depth_of_[the]_water** in which the **ship** is **sailing**. The traditional **device** for the purpose is the **lead-line**. **Patent sounding_machines** are **based** on the fact that the **pressure_of_[the]_water** on an **immersed body increases** with the **depth** to which it is **immersed**. A more modern **device** is the **echo_sounder**. In the **bottom** of the **vessel's hull** a **transmitting_oscillator** and a **receiving_oscillator** are so **mounted** that the latter **picks_up** the **echo reflected_back** from the **sea_bed** A **rotating_contact** causes a **condenser** to **discharge** through the **electromagnetic transmitting_oscillator**, so that a **sound_impulse** is **transmitted**. The **time-measuring_equipment comprises** a **neon_lamp rotating** in front of a **timing_scale**. When the **lamp passes** the **zero_position**, the **sound_impulse** is **transmitted**. The **time** it takes for this **impulse** to **reach** the **sea_bed** and return to the **receiving_oscillator** (which is essentially a **microphone** for **picking_up** the **sound**) is **marked**, or **read**, on the **scale** by a **flash emitted** by the **neon_lamp**, which has meanwhile **rotated** past the **zero_position**. The **current** from the **receiving_oscillator** has to be **amplified** in order to cause the **lamp** to **light_up**. **Echo_sounders** usually operate with **ultrasonic frequencies** (20,000 **cycles/sec.**), but **sounders operating** with an **audible sound_frequency** (3600 **cycles/sec.**) are used for **deep-sea soundings**. The **principle** of the **transmitting_oscillator** is really quite simple. The **coil** is **energised** by an **alternating_current** with a **frequency** of 3600 **cycles/sec.**, so that the **laminated armature**, attached to the **diaphragm**, is **alternately attracted** and **released**. The **diaphragm emits sound_waves** of this same **frequency**. Other types of **transmitter** make use of the **principle** of **magneto-striction** or the **piezo-electric_effect**.

Appendix II

95% confidence intervals
absence of menstruation
active group odds
additive adverse effects
adverse neurodevelopmental outcome
adverse neurodevelopmental sequelae
alternative metabolic substrates
analysis of subgroups
antepartum breech presentation
antepartum cardiac disease
artificial insemination clinic
asymptomatic-neonatal hypoglycaemia
Bayley mental scale
blood glucose concentration
blood glucose concentrations
breech multiple pregnancy
cardiac antepartum haemorrhage
cardiac breech presentation
central delivery suite
centre d failed
computerised collection system
continuous dependent variables
controlled cord traction
cycles of treatment
days of ventilation
degree of hypoglycaemia
degrees of hypoglycaemia
Department of Health
detailed neurological examination
development of hypoglycaemia
diagnosis of postpartum
diastolic blood pressure
dichotomous dependent variables
distribution of conceptions
distributions of insemination
early breast feeding
early childbearing years
early feeding practices
elective intravenous nutrition
empirical cumulated distribution
enzyme immunosorbent assay
fatness mass index
fetal growth retardation
frequency of inseminations
frequency of intercourse
frequent moderate hypoglycaemia
full term infants
full term neonates
gamete intrafallopian transfer
glucose oxidase method
Health Ethics Committee
higher susceptibility rates
highly significant increase
hypoglycaemia glucose concentration
hypoglycaemia plasma glucose
hypoglycaemia recurrent apnoea
hypoglycaemic neurological impairment
iatrogenically controlled ovulation
immediate postpartum period
incidence of hypoglycaemia
incidence of postpartum

incidence of vomiting
insemination donor clinic
late maternal refusal
local awareness campaigns
major congenital abnormalities
maternal blood loss
maximum regression coefficient
maximum sperm counts
mean birth weight
mean cell volume
mean gestational age
mean monthly temperatures
measure of non-compliance
mental development scores
mental developmental scores
method of testing
methods of testing
middle postpartum haemorrhage
minimal enteral feeding
minimum safe plasma
mothers' educational level
motor development scores
multiple intrauterine death
neonatal cell volume
neonatal packed cell
non-parametric sample test
number of conceptions
occurrence of hypoglycaemia
onset of hypoglycaemia
packed cell volume
persistence of hypoglycaemia
plasma glucose concentrations
postpartum cell volume
proportion of infants
proportion of susceptible
quality of semen
raised blood pressure
randomised controlled trial
rate of conception
rate of ovulation
rate of postpartum
rate of susceptibility
rate of uptake
rates of conception
rates of uptake
reduced developmental scores
regional health authorities
Registrar General's Classification
release of gonadotrophin
risk of hypertension
risk of postpartum
senior house officers
severe neonatal hypoglycaemia
severe neurological damage
somatosensory evoked potentials
special care nursery
treatment of infertility
waste of eggs
World Health Organisation

[121 of 883 triads retained]

Appendix III

```
        16 APR 92 / Observer: A mucky    business
hich has a staff of 8, transact          business
encounters: there is unfinished          business
r. And all at the bottom of the          business
hey will lead to more lucrative          business
hion houses are reporting brisk          business
oup's managed pubs and property          business
          of its US-based healthcare    business Glasrock for Dollars 72m (Pounds 3
                    having run the        business a bit cavalierly.'
                                          business activities, Andrew Jack writes. No
                                          business activities.
                                          business activities. Sales of snacks rose 2
ot of time casting doubt on the          business acumen of British
        the DTI's efforts to make        business advisory services more user-friend
t not according to Bill Emmott,          business affairs editor of
eeking to transfer its non-life          business after two years of heavy
                    tax' on              business aircraft mandated by Congress. Gam
for domestic mortgage indemnity          business amounted to
                    However,             business analysts welcomed Mr Patten's blue
              were a family              business and not a public company'.
          sex, women, big                business and the government, you get a poli
      Dalgety regarded agri              business as a core activity, it saw foods a
Giovanni Di Stefano, a former            business associate of Mr Giancarlo Parretti
              ailing retail              business back on track. Ahead of this year'
                    in                   business circles.
                                          business climate in the country.
                                          business climate.
                                          business climate. There were fears in Hong
      The average rate of                business closures (as distinct from insolve
                                          business closures are inevitable and will r
      proportion of                      business closures that results in legal ins
                                          business colleagues outside HunterPrint hav
                    the                  business community and the governor'.
HE traditional split in Genoa's          business community between a public sector
                    The                  business community may be more grateful. 'I
        sectors of the                   business community to feel that the council
nd significant support within a          business community which
jobs, some members of the local          business community,
            the city's                   business community, responds with scepticis
                    The                  business community, which overwhelmingly en
            Canada's                     business community.
    and the bulk of Canada's             business community.
shifted from politicians to the          business community.
      DALGETY, the foods and agri-       business company, raised pre-tax profits by
              some                       business concerns. The announcement of a mo
l contracting marketplace where          business conditions
        which indicated that             business conditions are improving, writes P
          improvement in                 business conditions by September.
                    on                   business conditions by Staffordshire County
The closely-watched index of            business conditions for manufacturing
                                          business conditions in the second half.
profitability in spite of weak          business conditions.
th the Hoar brothers over their          business conduct.
        FT  13 JUN 92 / Japanese         business confidence at 5-year low
the Bank of Japan reported that          business confidence in May had plunged
the Bank of Japan reported that          business confidence in May had plunged
                    decline in           business confidence in May, which hit the 1
to the 333 companies surveyed,           business confidence in many UK
        FT  13 JUN 92 / Japanese         business confidence slumps
Lybrand - records an upturn in           business confidence.
r Dean, aged 53, is a freelance          business consultant and was an executive
sation of profits, there was no          business criteria [!!!]in
```

```
                       similar    business culture and language. In 1980-88,
d result is a terribly immature    business culture in which discipline
        He argues that Japanese    business culture will change to allow other
                  many of their    business customers - a question highlighted
   street banks' charges to small- business customers are
re optic circuits to link local    business customers directly to
ochester says some Lloyds small-    business customers would save at least 20
        combination of a reviving   business cycle and fiscal stimuli from a ne
hat United was trying to link a     business deal with
                            and     business debts do not together make a stabl
                   Mr Tim Baker,    business development manager with Norwich U
                           The      business development unit of Manchester Bus
1 months before he died and his     business empire
mi, inherited almost the entire     business empire except for Seibu
                        largest     business empires in Italy (Agnelli's Juvent
jet, expensive cars and various     business enterprises. Only Pounds
how to respond to the difficult     business environment has
     warned that the Japanese       business environment would become increasin
                  political and     business ethics being called into question.
              mainly targets the    business executive and conference market.
                         that      business failures early this year were runn
because there have been so many     business failures in the sector) to
m committing one of the biggest     business frauds of the
                       attract      business from Taiwan.
night stay in a Marne la Vallee     business hotel.
  trippers or staying at nearby     business hotels at cut-price weekend rates.
                          its       business hours to encourage more trading. B
US and Wellworth food retailing     business in Northern Ireland.
d standards and with its own 39     business indicators. 'Unless an
/ Turkish fears grow on foreign     business influx
I might place some of its small     business initiatives with the shops if
te and the postponement of many     business initiatives.
be said that many, if not most      business insolvencies
                           of       business insolvencies throughout the period
he company along with other agri    business interests in the under-developed r
9, has resigned to follow other     business interests to concentrate on core
     production and flat February   business interests, writes Terry Hall in
d yesterday showed the ratio of     business inventories - were in line with
terday showed that the ratio of     business inventories to
        business rates impose on    business inventories to
                     'Our core      business investment in buildings and the co
Brothers of Belfast, and the US     business is the development, production and
to power the new Gulfstream GV      business jet
                          GV        business jet aircraft. The order
             flight of the new      business jet at Farnborough to meet the nee
              It is true the        business jet is due in late 1995.
and billed as the fastest-ever      business jet market held up rather better t
ion last year. However, the 186     business jet.
o find an education pundit or a     business jets sold, although 10.7 per
                     However,       business leader who agrees
       Mr Hata was responding to    business leaders and conservative oppositio
A timely combination of new pro-    business leaders' disquiet as the Nikkei in
                        though      business legislation and important
                          the       business levels remained well below normal.
company that once had a booming     business lobby, such as Mr Jacques Calvet,
een told that his parliamentary     business making costly, top-quality sports
              'The government's     business managers
                       GERMAN       business managers will be desperate to bull
       will help government's       business managers, unlike their national re
tactics by banks in the small       business managers. However, it seems the st
ple is commonplace in the small     business market, and constant rumbles of
            fares may benefit the   business market, though. Other
asset. They are steeped in the      business market.
              it will be the        business of
           in monitoring the        business of David Mellor, the new minister;
                                    business of government. He promised more op
```

months, net profit from normal	**business operations**
for Unilever's **agri**	**business operations** recently put up for sal
profits from normal	**business operations** would be about half the
	business opportunity.
industrial wasteland into a	**business park** for smaller industrial and se
permission for a large	**business park** in partnership with the Churc
a	**business park**, industrial areas and housing
permission for a large	**business park**.
eisure district, most of the 22	**business parks** are on the periphery.
promoting	**business parks** in the north-east, are not p
promoting	**business parks** in the north-east, are not p
t company specialising in large	**business parks**,
in large	**business parks**, has asked National Westmins
at least seven	**business parks**, there are hundreds of acres
Mr Jaffe is an old friend and	**business partner** of Ms Lansing's.
One	**business partner** reporting on the ripeness
customer orders with their own	**business partners**.
Britain's battered **small**	**business people** need is panic action to ref
s that 'there are many Japanese	**business people** who
they are	**business people**, judges or politicians.
ative fundraising efforts among	**business people**, led by its deputy
on	**business performance** confirmed the broad we
	business pessimism, and the political hiatu
ompany News: R. H. Macy **unveils**	**business plan**
urgently **reviewing its**	**business plan** and may decide to sell core o
	business plan for the company, says that th
at the	**business plan** once it was complete and had
new	**business plan** to lenders later this month a
look at the project when the	**business plan** was complete.
ould look at the project if the	**business plan** were
nformation we will **finalise our**	**business plan** which
part of the airline's **five-year**	**business plan** which it updates
unveiled a five-year	**business plan** which, it claims, will 'retur
reached Pounds 1.65bn. The	**business plan** written in spring last year h
iery have yet to **complete their**	**business plan**, in spite of
As part of JAL's	**business plan**, the airline plans to open at
imoto, general manager of Asian	**business planning** at Itochu,
	business planning has been based on very co
'put together 15 of the last 16	**business plans**
standing commission on laws and	**business practices**
ftware company who says: 'Their	**business practices** are reminiscent of
To **counter the**	**business problems**, the company has plans to
proved itself a useful	**business product** with a finite constituency
their lending margins. So basic	**business profits** for the year
ough trading conditions **cut agri**	**business profits** to Pounds 11.9m from
domestic and	**business property** costs such as rent levels
vey says that uncertainty about	**business prospects** is still holding
acco and insurance group's good	**business prospects** prompted profit
uniform	**business rate** in 1990, brought in an estima
The	**business rate** is set by the government, but
Councils **pay their**	**business rate revenue** into the national non
ear 95 per cent of the **assessed**	**business rate** was collected
government grant and the	**business rate**. However, the Department of t
government grant and the	**business rate**. This year, though, The Depar
	business rates - and council tax when that
	business rates - and council tax when that
	business rates could lead to unexpected inc
pointing out that	**business rates** had been cut by Pounds 1.25b
	business rates impose on business invest
whatsoever for	**business rates non-payment**'.
stripped out to avoid	**business rates** were 'the tip of a nasty ice
capital-intensive	**business** rather than one which generates a
when **customer retail**	**business** reached Pounds 1.154bn.
decisions and would not **affect**	**business relationships** between the two
stacle in **building constructive**	**business relationships**'.
t the well-publicised series of	**business scandals** in Ireland pointed

rsity of Strathclyde's graduate **business school.**
provided **business schools** across the world with a cl
ach, both geographically and in **business sector** terms, to take
As for the **business sector,** 'which has earned a massiv
lucrative business sector.
the apparent opposition of the **business sector.**
business sentiment, out next week, before m
The tankan, reflecting poor **business sentiment,** prompted a fall in the
Banking, insurance and **business services**   13.4%
BET, the **business services conglomerate,** is likely t
PRE-TAX profits at BET, the **business services group,** could have been al
business services have both recovered lost
FT   16 APR 92 / **Boost in business spending forecast**
and **business strategy** for years to come - is no
The latest review of **business strategy** goes further than a simil
dy progress on all parts of our **business strategy.'**
over **business strategy.**
department of business studies,
y important role in **galvanising business support** in recent weeks.
cant upper floors can now **grant business tenancies** to housing
lord and Tenant Act which gives **business tenants** the right to
anything **in business terms,'** he reckons. Meanwhile he i
things are profitable **in business terms.'**
general **fall in business travel** and tourism.
a **decline in business travel** caused by the Gulf war and
business travellers more choice.
ds to attract more tourists and **business travellers.**
In 1989 Dan-Air's **charter business** turned over about twice as much as
The new **business unit** consists of three product ope
e US has formed a new worldwide **business unit** focused
business unit to sell its new Canadair Regi
'The new **business unit's** mission is to build, on top
GPT has **set up a business unit,** GPT Cablecom, to tackle the
many **business units** beneath those groups in its
'strategic business units' (SBUs) in Vevey for each Ne
business units.
business volume in the insurance industry,
In the life insurance sector, **business volumes** are still depressed and
There was also a slight **rise in business volumes** during the quarter, the
increase in business volumes in the next three months.
d today. However, it warns that **business volumes** remain well below
ncial services companies expect **business volumes** to
business was holding up, helped Great Unive
A sizeable amount of the **day's business** was reported to have come from the
core of the business was smokers' accessories. Today, p
ales at the Darchem engineering **business were hard hit.** They fell from
side of the **health business will attract competition,** but is c
ating profit for that **sector of business** will be lower. Last year 54 per
traditional **way of doing business** will be threatened by the new law.
Women hoping **to set up in business** will have access to female   busi
mphasising that any **recovery in business** will start from
embargo on business with Hanoi, the sensitive issue ma
lient has tried **to carry on the business** with a view to
time they are **chasing us to do business with us,** not
in its extraordinary losses was **business writedowns** and closure
rritory might remains 'the most **business-friendly location** in
Wirtschaft in Cologne, a **business-orientated** research institute, say
The proportion of **business-related bankruptcies** to total indi
and **business-related individual bankruptcies** ac
exclude **business-related individual bankruptcies,** w
Spencer's **niche foods business.**
clear signs of an upturn in business.
from its **mainstream business.**
full, and doing a **thriving business.**
involved in lending to a **small business.**
group was looking to buy a **core business.**

```
or from the Freemans mail order      business.
      nearby to handle the volume of  business.
                           small       business.
pending is put the way of black       business. 'You don't
    that only gangsters are in the    business. But the success of Mr
              they are still in       business. But there are too few of them lef
ers were informed, was open for       business. Foreign investors were
kely to remain in charge of the       business. He will
y no comprehension of running a       business. So
        this week, had we been in     business. To illuminate my purpose and demo
unning a town is like running a       business. Unless you continue to
in turning round the UK banking       business. When he joined Midland in 1987
ince 'they are part of our core       business.'
everyone gets burnt in the film       business.'
      are always soliciting new       business.'
                    go out of         business.'
omers) is something rare in our       business.'
    in this country to sustain the    business.'
```

[278 citations retained out of 984]

Appendix IV


```
ws: Harland Simon may sell core       businesses
istoric buildings with low tech       businesses
cretionary portfolio management       businesses
                    for small         businesses and Midland the cheapest.
                   peripheral         businesses and adding new ones to our core
          closure of loss-making      businesses in Senegal and the Ivory Coast.
on which has sent so many small       businesses to the [wall]
        Small scale family-run        businesses which constituted the bulk of It
              owner-occupier          businesses will look elsewhere.'
rticultural and consumer health       businesses, Mr Scroggs said. Fisons
           information technology      businesses, Quality Software Products and S
         As one of the two core       businesses, William Hill is an essential so
interested in selling non-core        businesses, but
elling some of the private side       businesses, but promised the hole
          health insurance            businesses, is again meeting claims from po
                portfolio of          businesses, some problem areas such as foot
oup noticed a pick-up in its US       businesses, the relative
lised approach to running their)      businesses.
```

[18 citations retained out of 233]

Appendix V

SYNOPTIC CHART FOR 'BUSINESS'

51	the	71	the	91	the	business	59	and	53	the	45	the
42	of	62	of	45	of	business	36	in	27	in	21	to
32	to	30	a	27	a	business	29	is	23	to	17	a
23	in	29	in	26	in	business	25	of	21	and	16	of
14	its	20	to	19	its	business	22	which	18	a	14	and
13	a	19	and	17	and	business	22	was	13	not	10	its
11	is	17	its	15	new	business	18	has	13	has	10	be
10	and	10	for	13	that	business	16	to	12	for	8	it
8	for	8	on	13	The	business	14	with	10	of	8	in
7	that	7	or	12	small	business	13	plan	10	is	8	as
6	with	6	is	11	for	business	11	for	9	which	7	on
6	as	6	his	10	core	business	11	community	9	be	7	by
6	The	6	UK	7	their	business	10	will	8	been	6	with
5	on	5	up	7	our	business	8	would	8	are	6	is
5	core	4	out	7	on	business	8	rates	7	said	6	been
5	at	4	more	6	to	business	8	conditions	6	would	5	not
5	are	4	general	6	retail	business	8	as	6	at	5	has
4	who	4	at	6	hotel	business	7	confidence	5	he	5	Pounds
4	up	4	an	6	banking	business	7	It	5	had	4	will
4	may	3	writing	5	main	business	6	that	5	by	4	this
4	had	3	very	5	investment	business	6	rate	5	Mr	4	one
4	been	3	that	5	insurance	business	6	parks	4	with	4	for
3	will	3	specialist	5	his	business	6	jet	4	will	4	US
3	which	3	said	5	doing	business	6	had	4	only	3	years
3	we	3	political	5	clothing	business	6	by	4	from	3	year
3	was	3	own	5	agri	business	5	volumes	3	was	3	were
3	side	3	new	5	Japanese	business	5	the	3	up	3	up
3	set	3	link	5	British	business	5	says	3	there	3	says
3	part	3	growing	4	this	business	5	district	3	still	3	new
3	can	3	group's	4	some	business	5	customers	3	should	3	leaders
3	between	3	companies	4	pro-	business	5	The	3	on	3	into
3	an	3	The	4	most	business	4	unit	3	now	3	competition
3	UK	3	/	4	large	business	4	services	3	made	3	are
3	Pounds	2	within	4	family	business	4	sector	3	lower	3	an
3	But	2	win	4	engineering	business	4	people	3	have	3	after
3	92	2	which	4	derivative	business	4	park	3	an	3	May

2	years	2	weather-proof	4	customer	business	4	managers	2	you	2	would
2	when	2	vinyl	3	tile	business	4	it	2	writes	2	well
2	way	2	their	3	retailing	business	4	he	2	who	2	was
2	Australian	2	take	3	reinsurance	business	4	from	2	what	2	total
2	there	2	still	3	property	business	4	at	2	were	2	that
2	stopped	2	sport	3	own	business	4	a	2	well	2	still
2	some	2	shrink	3	many	business	4	He	2	want	2	start
2	project	2	running	3	management	business	3	until	2	two	2	sell
2	parts	2	run	3	local	business	3	though	2	their	2	sales
2	or	2	rise	3	helicopter	business	3	support	2	that	2	recent
2	more	2	reported	3	good	business	3	strategy	2	six	2	offset
2	margins	2	refrigeration	3	food	business	3	should	2	selling	2	much
2	loss-making	2	ratio	3	equipment	business	3	remained	2	restrictions	2	more
2	it	2	proportion	3	distribution	business	3	population	2	residential	2	manufacturing
2	into	2	paper	3	central	business	3	operations	2	reported	2	if
2	have	2	our	3	big	business	3	on	2	reorganise	2	housing
2	growth	2	other	3	battery	business	3	market	2	renewing	2	high
2	grant	2	not	3	aggregates	business	3	like	2	prompted	2	from
2	gradual	2	means	3	US	business	3	inventories	2	profits	2	existing
2	go	2	mail	3	UK	business	3	interests	2	over	2	end
2	from	2	lose	3	Irish	business	3	difficulties	2	outside	2	done
2	expansion	2	last	3	However	business	3	development	2	or	2	customers
2	domestic	2	into	2	wine	business	3	culture	2	once	2	continuing
2	development	2	international	2	way	business	3	could	2	manager	2	companies
2	closely-watched	2	interest	2	various	business	3	closures	2	make	2	but
2	build	2	insurance	2	up	business	3	climate	2	long	2	badly
2	assessment	2	index	2	transact	business	3	are	2	led	2	analysts
2	any	2	increase	2	tourist	business	3	activities	2	largely	2	always
2	all	2	help	2	systems	business	3	But	2	its	2	The
2	While	2	from	2	steel	business	2	year	2	it	2	Ms
2	Mr	2	floor	2	snacks	business	2	worth	2	industrial	2	Mr
2	Japan	2	existing	2	small-	business	2	world	2	increased	2	Ireland
2	I	2	drift	2	securities	business	2	while	2	impose	2	He
2	Halstead's	2	domestic	2	reviving	business	2	via	2	help	2	German
2	European	2	decline	2	protection	business	2	users	2	growing	2	Dollars
2	BDI	2	core	2	promoting	business	2	units	2	government		
2	'In	2	commercial	2	profitable	business	2	under	2	fell		
2	close	2	products	2	travellers	business	2	expected				
2	cement	2	private	2	travel	business	2	consumer				
2	car	2	pensions	2	terms'	business	2	computer				
2	by	2	order	2	survey	business	2	closed				

2	between	2	normal	2	space	business	2	between
2	automotive	2	no	2	she	business	2	being
2	as	2	music	2	sentiment	business	2	bankruptcies
2	US	2	manufacture	2	saw	business	2	as
2	Time	2	lucrative	2	said	business	2	about
2	Germany's	2	loans	2	prospects	business	2	The
2	Belstaff	2	lending	2	profits	business	2	Pounds
2	BET	2	leading	2	practices	business	2	However
2	The	2	instrument	2	plans	business	2	Britain
2	indemnity	2	planning	2	Unless	business		
2	holiday	2	partner			business		
2	health-care	2	owners			business		
2	government	2	overseas			business		
2	generics	2	out			business		
2	foods	2	last			business		
2	flooring	2	insolvencies			business		
2	five-year	2	individual			business		
2	film	2	increased			business		
2	export	2	failures			business		
2	drive	2	faces			business		
2	dough	2	ethics			business		
2	develop	2	environment			business		
2	construction	2	empire			business		
2	building	2	does			business		
2	black	2	director			business		
2	another	2	days			business		
2	among	2	but			business		
2	after-care	2	body			business		
2	One	2	back			business		
2	German	2	away			business		
2	GV	2	advisory			business		
2	Canada's	2	acquired			business		

457   510   546   396   333

Appendix VI

0 1/4 , Chrysler in continued **active trading**. slightly from Thursday's 38.938m s
yet another day of extremely **active trading**. Elsewhere, Peugeot was knocked bac
ore than Dollars 3 1/2 1/2 in **active trading** after the company said it backed Pr
s less than Dollars 22 1/4 in **active trading**. Merck also saw heavy turnover, clo
dipped 0.29 Dollars 49 3/4 in **active trading** after Merrill Lynch cut its long-te
isregarding down 1/8 at 44 in **active trading**. Citicorp was up 7/8 at 21 5/8 in a
icorp was up 7/8 at 21 5/8 in **active trading**. Canada ACTIVITY was limited, with
ose 9.7 to TAIWAN advanced in **active trading** with the weighted index putting on
g OSLO jumped 2.3 per cent in **active trading** as the economic outlook appeared na
erage climbed 4.9 per cent in **active trading** on rumours that the last minutes of
RONTO ended little changed in **active trading** boosted by a large block Nikkei in
m A HELSINKI closed higher in **active trading**, as the HEX index closed 13.1 or SH
ocks ended slightly higher in **active trading**. Based on preliminary Advances led
t close off intraday highs in **active trading**. The STOCKHOLM gained in moderate t
Toronto share prices lower in **active trading**. 1,889.5 in thin trading. Dealers a
continued its recent rally in **active trading** on the first day of the Other state
2.1 per cent and saw its most **active trading** this month on before closing FFr13
and arbitrageurs, In spite of **active trading** by domestic investors, traders said
ion. The TSE-300 fell 15.3 to **active trading**, writes Emiko Terazono in Tokyo. tr
ding day, the longest winning **active trading** by public funds was prompted by fin

se Dollars 7/8 to Dollars 10 in **busy trading** after the company TORONTO stocks fi
ed Dollars 7/8 to Dollars 17 in **busy trading** after FRANKFURT saw a flurry of act
lars 3 3/4 to Dollars 55 3/4 in **busy trading** after the TORONTO stocks finished l
Dollars 1/2 to Dollars 8 3/4 in **busy trading** ACTIVITY surged in the morning sess
lars 1 3/8 to Dollars 32 1/8 in **busy trading** after Mr PARIS eased on profit-taki
ollars 1/2 to Dollars 65 3/8 in **busy trading** after Merrill Lynch, TORONTO share
vy trading session with most in **busy trading** TORONTO stocks ended slightly highe
US SHARE prices held steady in **busy trading** yesterday as financial markets Hone

y at the end of a quiet and **cautious trading** adjusted positions and there has be
December options chairman's **cautious trading** statement on Monday. of trading was
g SINGAPORE closed lower in **cautious trading** as sentiment was dampened by Hong C
ing TAIWAN remained weak in **cautious trading** ahead of Saturday's parliamentary B

ng demand for Astra made up for **dull trading** in other issues. The Affarsvarlden
index slipped 1.09 to 279.36 in **dull trading**. Astra was the most in yet another
Jax, JOHANNESBURG was firmer in **dull trading** with the overall index 14 higher at

cross-shareholding, issued a **gloomy trading** US share prices were flat-to-lower
ted the market. The company's **gloomy trading** trading days it has lost more than
PARIS dropped 2.9 per cent in **gloomy trading** on the last day of the account L6,0

of the per cent lower after a **heavy trading** session. The TSE-300 fell 15.3 to a
stocks finished lower after a **heavy trading** session with most in busy trading T
stocks finished lower after a **heavy trading** session marked by the afternoon tra
rices ended mixed in continued **heavy trading**. Nikkei average gained marginally a
months before the in extremely **heavy trading**. Although much of that rise was con
d Wednesday's gains in further **heavy trading** as the JOHANNESBURG revived in late
ram trading combined 24 3/4 in **heavy trading** on disappointing fourth-quarter ear
ollars 3/8 to Dollars 9 5/8 in **heavy trading** after the A late selloff sent Toron
5 shares STOCKHOLM advanced in **heavy trading** as strong demand for Astra made up
CKHOLM rallied 4.1 per cent in **heavy trading** on speculation, confirmed market. F
e prices were flat-to-lower in **heavy trading** yesterday as the market trading. Lo
sing FFr13 better at FFr506 in **heavy trading** of 129,325 shares STOCKHOLM advance
-CHIP stocks jumped sharply in **heavy trading** on US stockmarkets Because of the '
ck markets rallied strongly in **heavy trading** on unexpectedly good trading patter
f the last trading day for the **Heavy trading** in Westpac's new shares and options

HANNESBURG ended mixed in **lacklustre trading** ahead of today's public US SHARE pr
days. The NEW ZEALAND saw **lacklustre trading** as the NZSE-40 index moved up just

omposite rose 2.79 to 637.16 in **late trading**. Shares in Tele-Communications led
trading day before to L1,500 in **late trading**. The telecommunications company had

URT saw a flurry of activity in **late trading,** having moved in a narrow activity
er cent. HONG KONG fell back in **late trading** after the Sino-British Joint Liaiso
re, Peugeot was knocked back in **late trading** to close PFr15 or 2.8 being delayed
.25 to AUSTRALIA lost ground in **late trading** as the local currency weakened agai
ors. SINGAPORE closed higher in **late trading** as foreign institutions came into M
ing by overseas institutions in **late trading.** BOMBAY fell to a low for the 1992/
as the JOHANNESBURG revived in **late trading,** helped by a better performance fro
d the but recovered slightly in **late trading** to finish down 18.47 at 1,476.01, a

NG ended lower in volatile but **light trading.** The Hang Seng Index of the day's t
er the recent MANILA dipped in **light trading** as the composite index fell 7.35 to
in 20m shares. MANILA slid in **light trading** but brokers still said that they ex
ces remained in easier vein in **light trading** yesterday as losses, and took its d
technical rebound led by Kepco **light trading** yesterday, writes Patrick Harverson
53, while the KUALA LUMPUR saw **light trading** continuing as the composite index d

into MANILA closed lower in **listless trading** as the composite index shed 10.92 t
in trading and the lower in **listless trading:** the all-share index shed 8 to 3,25

x dropped 7.52 to 348.75 in **moderate trading** PARIS dropped 2.9 per cent in gloom
otest SEOUL ended firmer in **moderate trading** on a technical rebound led by Kepco
ng. The STOCKHOLM gained in **moderate trading** as domestic interest rates fell bac
ity. Declines SEOUL rose in **moderate trading** but sentiment remained nervous ahea

US SHARE prices were mixed in **modest trading** yesterday as the Thanksgiving activ

ars 5/8 to Dollars 63 1/4 in **nervous trading** ahead of retailer's trading and imp
he SINGAPORE fell sharply in **nervous trading** on talk, confirmed after the tradin

ing on AUSTRALIA was firmer in **quiet trading.** The All Ordinaries index rose 9.7
ices moved marginally lower in **quiet trading** ahead of the release of low of 17,1
xed after drifting sideways in **quiet trading.** Gains in there was no discernible

PORE ended marginally lower in **slack trading** ahead of the holidays. The NEW ZEAL

X contract MILAN ended lower **in slow trading** and the Comit index fell 4.17 to 44

Y56 to Y695 on increased **speculative trading,** after rumours of at 142.9m shares

luation of the crown 3,134 in **steady trading.** The gold index added 24 or 3 per c

in the morning session on **technical trading** related to the hands in the first h

gen was also boosted by a **short-term trading** buy The Tokyo Stock Exchange suspen

ency index rose 0.6 to 103.2 in **thin trading.** BUYING by public funds and dealers
r in active trading. 1,889.5 in **thin trading.** Dealers again blamed weekend polls
by some institutional buying in **thin trading** and the lower in listless trading:
PUR drifted to a lower close in **thin trading** as investors remained market-makers
ines to a fifth straight day in **thin trading.** The trading after a one-day halt,
owed PARIS continued to fall in **thin trading,** and some sell orders from the UK V
eased just 0.05 to OSLO fell in **thin trading** as speculation about a devaluation
d by SINGAPORE closed firmer in **thin trading** as the Straits Times Industrial KUA
firm but off the day's highs in **thin trading.** The thin trading in very thin trad
to SKr33.5 MILAN ended mixed in **thin trading** and the Comit index eased just 0.05
sted by a large block Nikkei in **thin trading** HONG KONG closed sharply lower on p
that China TAIWAN retreated in **thin trading** as the weighted index lost 28.25 to
ort policies. AMSTERDAM rose in **thin trading** with the CBS Tendency index gaining
e fourth consecutive session in **thin trading** as many BOMBAY fell again in tradin
the HONG KONG eased slightly in **thin trading,** as modest early gains were erased
rading, genuine NEW ZEALAND saw **thin trading** as the NZSE-40 index eased 13.78 to
ay's highs in thin trading. The **thin trading** in very thin trading, Reuter report
four weeks. SEOUL slid in very **thin trading** in a technical adjustment after the
ading. The thin trading in very **thin trading,** Reuter reports from Tokyo. the New

[108 citations retained out of 265]

Language Learners' Use of a Bilingual Dictionary: A Comparative Study of Dictionary Use and Needs

James Salehe Mdee, *Institute of Kiswahili Research,*
University of Dar es Salaam, Tanzania

Abstract: This paper compares and contrasts dictionary use and needs of language learners at the University of York in the United Kingdom and at the University of Dar es Salaam in Tanzania. Five aspects are discussed in this study viz. dictionaries used, instructions and guidance on dictionary use, the functions for which students use a dictionary, information categories sought or lacking, metalanguage in the dictionaries and how it has been used to explicate the information categories.

These five aspects are discussed in relation to the observations made by the subjects as reflected in the data collected. From the data analysis are made some generalizations on how the language students of the two universities use their dictionaries and the lexicographical needs of the language learners in general and of each group in particular.

It is concluded that language learners do not look up the same information in a dictionary because each has specific language problems.

Opsomming: Die gebruik van 'n tweetalige woordeboek deur taalaanleerders: 'n Vergelykende studie van woordeboekgebruik en -behoeftes. Hierdie artikel vergelyk en kontrasteer woordeboekgebruik en -behoeftes van taalaanleerders by die Universiteit van York in die Verenigde Koninkryk en by die Universiteit van Dar-es-Salaam in Tanzanië. Vyf aspekte word in hierdie studie bespreek, nl. die woordeboeke wat gebruik word, instruksies en riglyne vir woordeboekgebruik, die funksies waarvoor studente 'n woordeboek gebruik, inligtingskategorieë wat verlang word of ontbreek, die metataal in die woordeboeke en hoe dit benut word om die inligtingskategorieë te verduidelik.

Hierdie vyf aspekte word bespreek met betrekking tot die waarnemings gemaak deur die studente soos weerspieël in die data wat versamel is. Uit die data-analise word sekere veralgemenings gemaak oor hoe die taalstudente van die twee universiteite hul woordeboeke gebruik en wat die leksikografiese behoeftes van die taalaanleerders in die algemeen en van elke groep in die besonder is.

Daar word tot die gevolgtrekking gekom dat taalaanleerders nie dieselfde inligting in 'n woordeboek naslaan nie aangesien elkeen spesifieke taalprobleme het.

Sleutelwoorde: TWEERIGTING- TWEETALIGE WOORDEBOEK, KOMMUNIKATIEWE BE-HOEFTES, WOORDEBOEKBEHOEFTES, WOORDEBOEKGEBRUIKSVAARDIGHEDE, WOOR-DEBOEKGEBRUIK, WOORDEBOEKGEBRUIKER, WOORDEBOEKMAKER, GEBRUIKERSVRIEN-DELIKE INLIGTINGSKATEGORIEË

Introduction

Dictionary use and needs are of great interest to both the lexicographer and the dictionary user. Research lexicographers seek to determine the needs of dictionary users in order to influence dictionary compilation accordingly. On the other hand, users of dictionaries want dictionaries which will meet their lexicographical needs.

The needs of dictionary users can be determined from the purposes for which the dictionary is used. Hartmann (1987) argues that although lexicographers have acknowledged a number of functions of the dictionary as guiding their work, until recently their understanding of these has not been informed by objective evidence on the real requirements of real users. According to Wiegand (1977), the functions of dictionaries may be determined more by the communicative needs of the dictionary users than the conventional information categories supplied by dictionary makers. The communicative needs of a dictionary user are: writing, speaking, reading and listening. A dictionary user needs a dictionary either for comprehension (reading and listening) or for production (writing and speaking).

The basic task of a bilingual dictionary is to provide L2 equivalents of L1 lexical items. A bilingual dictionary is either monodirectional or bidirectional. A two-language dictionary is monodirectional if it serves the needs of the native speakers of one of the two languages. It is bidirectional if it attends to the needs of the speakers of both languages (Tomaszczyk 1988). Theoretically, a bilingual dictionary serves speakers of both languages of the dictionary. An L1-L2 dictionary is a reading dictionary for the native speakers of L2, and a writing dictionary for the speakers of L1, and *vice versa*.

The objective of this paper is to analyze the dictionary use and needs of bilingual dictionary users as reflected in lexicographical data collected at the University of York in the United Kingdom and at the University of Dar es Salaam in Tanzania. The analysis will try (1) to establish dictionary users' needs and use of bilingual dictionaries in general, and of French and German bilingual dictionaries in particular, (2) to compare and contrast the dictionary use and needs of English speakers learning either French or German, and Swahili speakers learning French and (3) to determine ways of improving bilingual French-English, French-Swahili, and German-English dictionaries from the perspective of the foreign language learner.

The lexicographical research on language learners' use of bilingual dictionaries involved 56 students learning either French or German at the University of York (henceforth YU), and 14 students learning French at the University of Dar es Salaam (UDSM).

Research coverage

In this study, five aspects of dictionary use were investigated:

(1)   dictionaries consulted by language learners,
(2)   instructions and guidance on dictionary use,
(3)   the purposes for which a dictionary is consulted,
(4)   information categories sought or lacking in the dictionaries used, and
(5)   lexicographic metalanguage in the dictionaries and how it was inter-
       preted.

The results of these investigations are as follows:

Dictionaries used

The majority of the students at YU learning German or French mentioned the
Collins dictionaries as the dictionaries they use most. All 16 students of Ger-
man (100%) and 34 students (out of 40) learning French at YU (85%) used Col-
lins dictionaries. About 20% also consulted *Harrap's French-English Dictionary*.
Some 71% of the students of French at UDSM used *Collins French-English Dic-
tionary* and about 36% used CREDU *et al.*'s *Dictionnaire Français-Kiswahili*. Some
21% also consulted Cassell's English-French dictionaries, *Larousse's French-
English Dictionary* and Robert's *Petit Dictionnaire de la Langue Française*.
       Although the aim of the research was not to determine the reasons why a
particular dictionary was preferred to others, various reasons may be given for
this. The Collins dictionaries are more readily available to learners than many
other dictionaries, hence they are more frequently used by language students,
and most likely by their language teachers too. They are considered better than
their competitors in terms of the information rendered and the method in
which the information is presented. It is interesting to note here that a consid-
erable number of students of French at the UDSM used CREDU *et al*. It scored
the second highest percentage of students making use of it. Although this dic-
tionary is no more than a bilingual word-list, it has no competitor, therefore
students have no choice but to use it. It should be stressed here that bilingual
dictionaries are very important works of reference in translation classes, hence
where there is only one dictionary, as is the case with CREDU *et al.* for French
learners at the UDSM, students would naturally use it.

Instructions on dictionary use

Skills in dictionary use are rarely taught to language students. Many language
students know very little about the dictionaries they use, e.g. the information
the dictionary can offer and how such information can be retrieved. It therefore

takes them a lot of time to obtain any information they want from a dictionary, a factor which has attributed to the fact that dictionary use is seen by some students as boring and taking more of their time than they are willing to give. For a language learner to be able to use a dictionary successfully, requires specific competence because the dictionary contains a wide range of information (Herbst and Stein 1987). According to Herbst and Stein a dictionary is a reference book, hence its users require specific reference skills to be able to retrieve information from it. They argue further that although teachers use dictionaries as valuable and indispensable linguistic tools, they do not teach their students how to use them to their best advantage. Instructions offered on how to consult a dictionary are usually limited to how they can look up a lexical item. The dictionary user is asked to scan through the pages of the dictionary to the first letter of the lexeme he/she is looking up, and then to keep on scanning until he/she finally finds it. This information is not enough and will no doubt make dictionary use boring. Herbst and Stein note (from their experience in Germany) that when a monolingual English dictionary is first introduced in a class, it is accompanied by two to three lessons on skills in dictionary use. Some teachers rarely provide regular instruction in the use of dictionaries. They simply refer students to the dictionary without explaining in detail how the information could actually be retrieved.

Our hypothesis was therefore: Language learners are not taught how to use a dictionary. To test this hypothesis, two questions were asked:

(1)   Were you taught how to look up a lexical item or an idiomatic expression in a dictionary?
(2)   Have you ever been taught how to get the information you want in an entry such as n (noun), vt (transitive verb) or a derivative of a base word which is represented by an affix e.g. -ness?

The responses showed that 73% of the French and German students at YU were not taught how to look up a lexeme in a dictionary. About the same percentage (74%) of the French students at the UDSM indicated that they were instructed how to use a dictionary. The trend is repeated again in the second question above. 86% of the students at the YU indicated that they were not taught how to retrieve information encoded in alphanumerical symbols in an entry, while 57% of the UDSM students learning French said they were taught how to interpret the lexicographical information rendered in a dictionary entry.

The responses of YU students do not necessarily mean that they were never instructed on how to use a dictionary, otherwise they would not have been able to use a dictionary at all. One would argue that they got informal instruction either from friends or relatives at home, but not in school from a teacher. This is possible because most homes in the United Kingdom are literate and own and use dictionaries (Quirk 1973). A child is therefore exposed to dictionaries at an early age long before he/she needs to use them as aids to lan-

guage learning. On the other hand, many families in Tanzania are not literate and do not own dictionaries. Dictionaries are usually found in schools, hence students have to be taught how to use them when the need arises. This seems to explain why the instuction is still vivid in their minds while their counterparts in the United Kingdom are not even aware that they were taught how to use a dictionary, even if it was not done by their teachers.

Dictionary-using skills can be acquired through instructions given by the language teacher or through the students themselves reading the front matter which usually contains a wealth of information on how to retrieve lexicographical information from the dictionary. Whether one is instructed or not, it is very important for a dictionary user to read the front matter in order to understand the arrangement of the entries and the metalanguage used in that particular dictionary. Unfortunately, many language students do not read the front matter. Indeed, this is the main reason why many dictionary users do not know their dictionaries.

The subjects indicated that they do read the guide to the use of a dictionary. 89% of the YU students make reference to the guide, but only when they failed to interpret symbols used in an entry. This was in response to the question whether they first consult the instructions on how to use a dictionary whenever they acquire a new dictionary, or do so only when they cannot interpret the codes in an entry for a word they are looking up. 71% of French students at the UDSM first read the guide before using a dictionary for the first time. The responses from the students of the two universities show that because the YU students were not formally taught how to use a dictionary, they were not made to realize the need to know its structure, contents and metalanguage before consulting it.

On the other hand, the UDSM students were made aware of the importance of first reading the guide before using the dictionary. There is no doubt, therefore, that this was the reason why they studied it. Notwithstanding the fact that a dictionary user knows the significance of the front matter, its length is a factor that can deter him/her from reading it. Many dictionary users may not have the time to read through all the detailed explanations about the dictionary. Moulin (1983) was also of the same opinion when he observed that a dictionary is consulted when the user is not excessively pressed for time. A concise and clear introduction will encourage the reader to go through it.

A good dictionary is user-friendly. Such a dictionary is one which (1) enters all lexical items the expected user would want to look up, (2) puts every lexeme selected for the dictionary in its appropriate place where the user can easily find it and (3) provides at the point of entry the lexicographical information that the user needs. A dictionary which lacks some lexemes or information required by the user, or which cross-refers the user from one entry to another within the dictionary, is not user-friendly. Many language students need user-friendly dictionaries. 88% of the YU students and 64% of the UDSM students preferred all the necessary information of a headword to be clearly

shown in the entry. This information should be presented in simple codes which the user can easily decode without frequent reference to the front matter.

The guide on how to use a dictionary is only useful if the instructions are thoroughly explained and well elaborated on. Most of the YU students (71%) were of the opinion that the guide in their dictionaries was quite clear, that they were able to understand it, and that it helped them to retrieve information encoded in the symbols in an entry.

A few students were however not satisfied with the instructions in their dictionaries because these were not self-explanatory. Almost all the UDSM students (93%) were of the opinion that the guide in their dictionaries was very useful because it helped them interpret the codes and symbols in the entry.

The purposes for which a dictionary is used

Dictionary users have different reasons for making reference to a dictionary. As we have already noted, a dictionary may be used for encoding or decoding a language. Language learners use a dictionary most when they read or write, and less when they speak or listen to someone. Language learners consult the dictionary in order to spell words correctly or to check their meanings when they write. The research revealed that 96% of the respondents at YU consulted the dictionary when writing and 88% used it when reading. Some 30% used it when speaking and 23% referred to it when listening to a speech or lecture.

Similar results were observed at UDSM. 93% consulted the dictionary when writing and 79% used it when reading. None of them used a dictionary when listening to someone speaking. The findings confirm the purposes for which a language learner uses a dictionary: writing and reading.

Information categories sought in dictionaries

The information that French and German learners look for whenever they consult French-English, German-English or French-Swahili dictionaries, includes: gender, pronunciation, orthography, conjugation, collocation, comparative and superlative forms of adjectives, lexical equivalents, example sentences, irregular forms of lexical items and stylistic uses of lexical items. The frequency with which the information categories above were looked up in the dictionary, was tested. Using the codes (1) for most frequently, (2) for less frequently and (3) for never, the subjects were asked to indicate the information they made reference to by putting the code in the appropriate information category.

The results of the test showed that meaning or lexical equivalents of the headwords in L1 was recorded as the information most frequently sought by the language learners. 98% of the YU students and 79% of the UDSM students looked up this information. Other information sought frequently by the YU

students was: gender 80% and spelling 68%. This observation is consonant with the findings of Kipfer (1987) that dictionaries are used chiefly as a guide to meaning and spelling and occasionally pronunciation. Less frequently sought information which recorded high scores, was: collocations, stylistic labels, comparative and superlative forms of adjectives, irregular forms of lexemes, example sentences and pronunciation, in that order. Very interestingly, no information was categorized as never sought by the majority of the YU students. Almost every kind of information was looked up but with varying degrees of frequency.

Besides meaning, information most sought by the UDSM students, comprised the following categories: correct spelling of lexemes 93%, gender 79%, and irregular forms of lexemes 50%. In the category of less frequently sought information, the research results were as follows: pronunciation, example sentences, conjugation, and collocation of lexical items. Information which the UDSM students learning French did not look up very often in the dictionaries, was stylistic labels and comparative and superlative forms of adjectives.

Data analysis

An examination of the research findings shows some interesting observations:

(1)   Dictionary users do not look up the same information every time when they consult a dictionary. Moreover, some information categories which are important to a learner of one language, may not necessarily be important to a learner of another language.

(2)   The functions of a learner's dictionary are basically twofold:

(a)   to help the language student to decode, and
(b)   to encode the language which the dictionary user is learning.

A decoding dictionary provides the semantic information of the entry words. A bilingual dictionary is no doubt a decoding dictionary. This explains why semantic information (lexical equivalents) was rated by the German and French learners as the most frequently sought information. An illustrative example sentence is another kind of lexicographical information which helps the dictionary user to decode the headword of which the definition. is not clear. An example sentence complements the definition by implicitly stating some missing semantic features of the headword and by making the definition clearer.

The argument that a dictionary is a tool for language learning is proven by the encoding functions for which it is used. According to our data, gender and spelling were considered the most frequently sought information. Gender was put in the second position of the category of information most frequently sought by students of both universities. The importance of checking the correct gender of a noun in both French and German languages cannot be overempha-

sized. Knowledge of it is very necessary for a learner to be able to construct a grammatically acceptable sentence.

This is an area which language learners would want to be sure of before using a lexeme, failure of which would expose one's linguistic incompetence. In both languages, for example, a noun has an influence on the form of an article or a pronoun appearing together in one sentence. A change of gender would affect their forms in the same sentence, hence distort the cohesion of the lexemes in the sentence. For example, in German, the articles *der* (the masculine) and *das* (the neuter) change into *dem* (dative case). The feminine article *die*, changes into *der* (dative case) in a sentence or phrase. Failing to adhere to the appropriate gender of a noun would no doubt make the sentence ungrammatical.

A dictionary is a good reference work for checking the accepted or standard spelling of lexical items. Whenever a language learner is in doubt of the orthography of a lexeme, he checks it in a dictionary. The UDSM students rated it in the first position (93%) as the most frequently sought information, but the YU students placed it in the third position of this category. The difference is not surprising, because English and German belong to the family of Germanic languages, hence they have many lexical items with similar spelling. Likewise German, French and English belong to the Indo-European group of languages. Thus they have some common words as well. Many words from one of these languages have been adopted in the other. All these factors make the vocabulary of one language familiar and intelligible to speakers of another language. Therefore the problem of spelling for English students learning either French or German is not as great as that experienced by Swahili speakers learning French.

Swahili is a Bantu language and differs very much from French, because they have nothing in common except for a few French words which have been adopted in Swahili.

An illustrative sentence is very useful in encoding a language. A language learner can, through the example sentence, see the lexical items which are always found together with the headword, and identify the place of the headword in a sentence, and its relationship with other lexical items in a sentence.

An example sentence was rated in the third position as one of the most frequently sought types of lexicographical information by the YU students, and less frequently sought by the UDSM students. This disparity could be explained by the fact that the YU students use the dictionary more often as an aid for learning a language on their own than their counterparts at the UDSM.

Private learning of a language would require one to read the example sentences in order to find out the collocates of the entry words, and thus make one's sentences by analogy. This is possible because most of the YU students had been learning more than one foreign language, and some had been studying a foreign language for more than ten years. This is not the case with many of the UDSM students. They depend wholly on their language teachers and do not look up this information very often. Moreover, some of their dictionaries do

not provide elaborate example sentences (cf. CREDU *et al.*).

Pronunciation and verb conjugation were rated by all the students as less frequently sought information. CREDU *et al.*'s *Dictionnaire Français-Kiswahili* and *Collins German-English Dictionary* do not offer pronunciation. Unlike French dictionaries which generally provide pronunciation, German dictionaries have no provision for this information because it is claimed that the language is written as pronounced, hence there is no need for the information. The omission of this information has probably contributed to its low rating. Since languages have exceptions which do not follow the norm, it is important for German dictionaries to single out those lexical items whose pronunciations do not follow the rule, and indicate how they are pronounced. These exceptions cause the dictionary user to seek information on their pronunciation.

Verb conjugation is hardly indicated in French and German dictionaries because it is considered to be material for grammar books. (Sometimes it is inserted as an appendix to a dictionary). Each of the personal pronouns in these languages has its own verb form (present tense) conjugated from the base form. For example, the German verb *sein* (to be) has the following forms: singular, *bin, bist* and *ist,* and plural, *sind, seid* and *sind*. Although verb conjugation follows a pattern which is in most cases very consistent, language learners cannot be expected to remember the pattern or consult a grammar book whenever they want to use a conjugated form. A learner's dictionary should incorporate this information. Since English learner's dictionaries provide this information, e.g. *do, does* and also the forms *doing, did* and *done,* French or German dictionaries should also.

Stylistic labels in a dictionary was the information type which was categorized as not having been looked up by the UDSM students and which was rated as less frequently sought by the YU students. Style is an area of considerable complexity for the foreign language learner and an area in which errors are regularly made (Carter 1989). The explanation for this disparity of need could be the orientation one has with language usage. Students who are used to stylistic usage of lexical items are very cautious when using certain items, hence they would try to be sure that a word is used in its appropriate context. It seems the YU students are more cautious in this respect than their counterparts at the UDSM who do not realize its significance. Although Swahili words like those of other languages can be classified into different usage styles, this aspect is not seriously taught in Swahili classes. Hence no usage restrictions such as "dialect", "substandard", "formal", "derogatory", "offensive", "jocular", "archaic", "slang", "colloquial", etc. are marked in Swahili dictionaries. Dictionary users therefore cannot be expected to appreciate the stylistic use of words.

A collocation is a grouping of words which naturally go together through common usage, e.g. *heavy smoker, free of charge* (Summers 1987: 193). In these examples, the members of each set collocate with one another. None of the members can collocate with synonyms of their partners. *Smoker* cannot collocate with *big,* nor can *free of* collocate with *cost.* It is therefore important that a

language learner be familiar with words which collocate. A dictionary should help he/she get the right word which collocates with the word he/she wants to use.

Most of the subjects categorized collocations as less frequently sought information. In this category, 50% of the UDSM students and 71% of the YU students registered this opinion. The correct use of collocation is an indication of the extent to which one has mastered a language. For example, it is important for a speaker to be informed of the fixed collocations such as *breach of the peace, bring to a halt,* etc. and prepositions which collocate with certain lexical items: *graduate from, harmful to, reason for,* etc. This information is very important although it is not looked up as often as other information because not every word has fixed collocates. Although verbs more often collocates with prepositions than any other grammatical category, not all verbs collocate with prepositions in every context (cf. *go home* and *go to school,* or *look up, look at, look for,* etc.). Thus less frequently sought information does not necessarily imply that the information in question is insignificant, only that it is not looked up in a dictionary as often.

Tests on the verb, noun and adjective

Three categories of words, verb, noun and adjective, which usually have irregular forms, collocate with other words or have special characteristics, were tested to determine the type of information looked up in each of these categories.

The results showed that in a verb entry the language learners consult the dictionary to look up word class labels, phrasal verbs, verb conjugations and fixed expressions. The UDSM students less frequently looked up only the word class labels and phrasal verbs, while the YU students looked up all the information listed above, fixed expressions very often, and the rest, less frequently.

In noun entries, gender was the information that was rated as the most frequently looked up by all the students, a confirmation of their responses to question 3 above. Information about nouns less frequently sought by the YU students was collocations and irregular forms, and by the UDSM students collocations and countability and uncountability.

None of the information in adjective entries was looked up very often. The YU students looked up usage restrictions, collocations and irregular forms of adjectives less frequently, while the UDSM students only looked up usage restrictions less frequently.

The research further tested the frequency of looking up example sentences or illustrative phrases in entries of different word classes. The verb, adjective, adverb, noun and preposition was tested. The YU students indicated that they looked up illustrative sentences or phrases in verb (89%) and noun (77%) entries most often, and less frequently in adverb, adjective and preposition

entries. The UDSM students looked up example sentences under noun (57%) headwords most frequently, and less often under adverb and preposition headwords.

Illustrating a verb is very important because it collocates with other word categories: a noun when it functions as an object or adverbial noun, a preposition or prepositional phrase, and an adverb. In the verb entry a learner's dictionary should therefore indicate whether:

(1)   it is transitive, i.e. it takes a noun as a direct object or nouns as indirect and direct objects, e.g. *Moshi killed a lion; Juma bought Mary a golden ring*; and

(2)   it takes an adverb or a preposition to form verb phrases, e.g. *put down, put off, put on, look for, look up, look after*, etc.

Verbs form many of the fixed collocations which are either idiomatic expressions or fixed expressions, e.g. *breach of contract, look down, look back*, etc. Many verbs, especially those used quite often, take various complements of different categories. No other category is as heavily complemented as the verb. The research findings showed that not every lexical item entered in a dictionary needs to be illustrated. Example sentences in a dictionary have a purpose. It is therefore very important to emphasize here that every dictionary writer selects the information that will suit the needs of the users for whom the dictionary is written.

When the subjects were asked to state the information lacking in the dictionaries they use and which they think needs to be incorporated in their dictionaries, YU students recorded a sketch grammar of the target language (50%), idiomatic expressions (46%), verb conjugations (45%), usage notes (41%) and verb tenses (41%). The majority of the subjects did not express their views on this question. This may be interpreted to mean that none of the information tested was lacking in the dictionaries used by the majority of the YU students. Even the supplementary question that asked them to state any other information not given in the questionnaire prompted only a few to propose the inclusion of etymology. The UDSM students indicated the information lacking in their dictionaries as follows: usage notes (86%), verb conjugations (71%), idiomatic expressions (64%), verb tenses (57%), and pronunciation (43%).

A comparison of the views expressed by the UDSM and YU students shows that they all indicated idiomatic expressions, verb conjugations, usage notes and verb tenses as missing in their dictionaries. All this information is necessary for a good mastery of a language. Interestingly, the UDSM students indicated pronunciation, while the YU students did not rate it as significant. The YU students indicated that they needed a sketch grammar of the target language, which the UDSM students considered insignificant. This difference of opinion can be explained by the linguistic background and orientation of the two groups as already expounded above, i.e. the English students at YU learn

foreign languages from their teachers as much as on their own. A sketch grammar of the target language would be of greater importance to them than to students who depend wholly on their teachers. The UDSM students' first and second languages are Bantu or other African languages with nothing in common with French, a Romance language. Hence a French learner from Tanzania would want guidance on how to pronounce French words more than an English student learning French, because the two languages have some common lexemes, of which some may be spelt and pronounced more or less the same.

Conclusion

This paper has shown the significance of a learner's dictionary containing information which is consonant with the learners' needs. It also pointed out that language learners do not always look for the same information when they consult a dictionary. The various information categories discussed have indicated that certain information types are looked up more often by some language learners than by others. Further examination of the dictionary use and needs of language learners have shown that they have more problems with some categories of words, e.g. the verb, than with others. This implies that even at the level of the lexeme, all word categories do not get the same treatment. Some are rendered with very elaborate information while others are not.

References

Carter, R. 1989. Review of: *Longman Dictionary of Contemporary English*. Harlow: Longman, 1987 / *Collins COBUILD English Language Dictionary*. London and Glasgow: Collins, 1987. *International Journal of Lexicography* 2(1): 30-43.

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*. Tübingen: Max Niemeyer.

CREDU/TUKI/FLL. 1991. *Dictionnaire Français-Kiswahili*. Nairobi: CREDU.

Drosdowski, Günther, Helmut Henne and Herbert E. Wiegand (Eds.). 1977. *Nachdenken über Wörterbücher*. Mannheim / Wien / Zürich: Bibliographisches Institut.

Hartmann, R.R.K. (Ed.). 1983. *Lexicography: Principles and Practice*. London: Academic Press.

Hartmann, R.R.K. 1987. Four Perspectives on Dictionary Use: A Critical Review of Research Methods. Cowie, A.P. (Ed.). 1987: 11-28.

Herbst, T. and G. Stein. 1987. Dictionary-using Skills: A Plea for a New Orientation in Language Teaching. Cowie, A.P. (Ed.). 1987: 115-127.

Kipfer, B.A. 1987. Dictionaries and the Intermediate Student: Communicative Needs and the Development of User Reference Skills. Cowie, A.P. (Ed.). 1987: 44-54.

McDavid, Raven I., Jr. and Audrey R. Duckert (Eds.). 1973. *Lexicography in English*. New York: New York Academy of Sciences.

Mdee, J.S. 1990. *Theories and Methods of Lexicography in the Standard Swahili Dictionary.* Unpublished Ph.D. thesis. Leipzig: University of Leipzig.

Mdee, J.S. 1994. Swahili Learners' Views on the Need for a Monolingual Swahili Pedagogical Dictionary. *Lexikos* 4: 230-248.

Moulin, A. 1983. The Pedagogical/Learner's Dictionary II: LSP Dictionaries for EFL Learners. Hartmann, R.R.K. (Ed.). 1983: 144-152.

Quirk, R. 1973. The Social Impact of Dictionaries in the U.K. McDavid, Raven I., Jr. and Audrey R. Duckert (Eds.). 1973: 76-88.

Snell-Hornby, M. (Ed.). 1988. *ZüriLEX '86 Proceedings. Papers Read at the EURALEX International Congress, University of Zürich, 9-14 September 1986.* Tübingen: Francke Verlag.

Summers, D. (Editorial Director). 1987[2]. *Longman Dictionary of Contemporary English.* Harlow: Longman.

Tomaszczyk, Jerzy. 1988. The Bilingual Dictionary under Review. Snell-Hornby, M. (Ed.). 1988: 289-297.

Wiegand, Herbert E. 1977. Nachdenken über Wörterbücher: Aktuelle Probleme. Drosdowski, Günther, Helmut Henne and Herbert E. Wiegand (Eds.). 1977: 51-102.

The Utilization of Parallel Corpora for the Extension of Machine Translation Lexicons

Jeanne Pienaar and G.D. Oosthuizen,
Department of Computer Science, University of Pretoria,
Pretoria, South Africa

Abstract: There has recently been an increasing awareness of the importance of large collections of texts (corpora) used as resources in machine translation research. The process of creating or extending machine translation lexicons is time-consuming, difficult and costly in terms of human involvement. The contribution that corpora can make towards the reduction in cost, time and complexity has been explored by several research groups. This article describes a system that has been developed to identify word-pairs, utilizing an aligned bilingual (English-Afrikaans) corpus in order to extend a bilingual lexicon with the words and their translations that are not present in the lexicon. New translations for existing entries can be added and the system also applies grammar rules for the identification of the grammatical category of each word-pair. This system limits the involvement of the human translator and has a positive impact on the time, cost and effort needed to extend a bilingual lexicon.

Keywords: ALIGNMENT, BILINGUAL CORPORA, CORPUS, EXTENSION, LEXICON, MACHINE TRANSLATION, MONOLINGUAL CORPORA, PARALLEL CORPORA

Opsomming: Die benutting van parallelle korpusse vir die uitbreiding van masjienvertalingsleksikons. Onlangs was daar 'n toenemende bewustheid van die belangrikheid van groot versamelings tekste (korpusse) wat as bronne in die navorsing van masjienvertaling gebruik word. Die proses om masjienvertalingsleksikons te skep of uit te brei is tydrowend, kompleks en duur in terme van menslike betrokkenheid. Die bydrae wat korpusse kan maak tot die vermindering van koste, tyd en kompleksiteit is deur verskeie navorsingsgroepe ondersoek. Hierdie artikel beskryf die ontwikkeling van 'n stelsel wat gebruik maak van 'n afgepaarde tweetalige (Engels-Afrikaanse) korpus vir die identifisering van woordpare met die doel om 'n bestaande tweetalige leksikon uit te brei met hierdie woorde en hul vertalings wat nie in die leksikon voorkom nie of om nuwe vertalings vir bestaande inskrywings by te voeg. Die stelsel pas ook grammatikareëls toe vir die identifisering van die grammatikale kategorie van elke woordpaar. Die stelsel beperk die betrokkenheid van die menslike vertaler en het 'n positiewe impak op die vermindering van tyd, koste en moeite in die uitbreiding van 'n tweetalige leksikon.

Sleutelwoorde: AFPARING, EENTALIGE KORPUSSE, LEKSIKON, MASJIENVERTALING, PARALLELLE KORPUSSE, TWEETALIGE KORPUSSE, KORPUS, UITBREIDING

1.   Introduction

The term *corpus* originates from the Latin word "body" and can be defined as a sufficiently large collection of text samples of the written or spoken language. The utilization of a corpus as a source of information on the nature and understanding of language was delayed as a result of criticism by Noam Chomsky in the late 1950s. He viewed corpora as being inadequate in terms of size and representativeness. His view became the conformed opinion of the next generation of theoretical linguists. With the advance of computer technology and the availability of large collections of texts (corpora) in machine-readable form, the potential of using machine-readable corpora has been recognized in a variety of areas such as lexical knowledge acquisition, grammar-construction and machine translation. A large number of computerized corpora are available which vary in design, size and research purpose.

The purpose of this article is, firstly, to give an overview of the state of the art regarding the use of corpora in machine translation. Secondly, it describes the development of a system that utilizes an aligned bilingual (English-Afrikaans) corpus for the identification of word-pairs in order to extend a bilingual lexicon with the words and their translations that are not present in the lexicon, or for adding new translations for existing entries.

In the next section, various approaches which utilize the information contained in corpora, will be outlined. The purpose of section 3 is to identify some problems that are experienced in the use of corpora such as corpus size, corpus representativeness as well as vocabulary- and word frequency-related problems. The aim of section 4 is to give an overview of the alignment of bilingual corpora. In section 5, the alignment method utilized for aligning a sample of the South African Hansards is explained. The problems experienced during the preparation of the corpora for the project, as well as problems encountered with the alignment process, are described. The aligned bilingual corpus was used as a resource for the task of extending a bilingual lexicon. In section 6, the focus is on the method utilized to extend a bilingual lexicon with the aligned English-Afrikaans corpus as a resource. The development of the system is described and the results obtained from various experiments are discussed. The aim of the final section, section 7, is to give an idea of some future research topics utilizing bilingual corpora as resources, and to end with a few concluding remarks.

2.   Applications of Computer Corpora

There are several approaches to using computer corpora as components in machine translation systems. Corpora can broadly be divided into two types: monolingual and bilingual. In this section, an overview of these two types of

corpora will be given, and the various applications of both monolingual and bilingual corpora will be provided.

2.1   Monolingual Corpora

At the most basic level a monolingual corpus is an important resource for a linguist in the determination of language-usage in a given domain (Arnold *et al.* 1994). Monolingual corpora have proved to be excellent sources of lexical information as well as limited world knowledge and are used extensively in the production of monolingual machine translation lexicons. Large text corpora allow for a detailed study of how a word is used, thus enabling the evaluation of the accuracy of lexicon entries by comparing it with evidence of how the word is utilized in the real world. Very large corpora provide reliable statistics on occurrence and co-occurrence of lexical categories and word-senses. Terms or idioms in context can also be identified.

Some machine translation systems require the automatic grammatical analysis of texts (tagging[1] and parsing[2]) as a first stage of analysis and the monolingual corpus usually serves as input for the tagging process. The utilization of computer programs for automatic tagging and parsing have a positive impact on speed and consistency. An important use of annotated corpora is the provision of probability statistics for probabilistic language processing systems.

2.2   Bilingual Corpora

Recently, the interest of some research groups in machine translation and linguistics has shifted to bilingual corpora. There are currently very few large machine-readable bilingual corpora. The South African Hansards, proceedings of the South African parliamentary debates, which by law were published in English and Afrikaans are examples of bilingual corpora.

Bilingual corpora are especially useful if the user can view translation segments. In bitext (or multitext) the text is aligned in such a way that within each bilingual (or multilingual) segment the two texts are translations of each other. A properly aligned bilingual corpus is an exceedingly important and valuable source of information and can be used in several ways to contribute to direct

---

[1]   The tagging process associates with each word in a text or corpus of texts a tag identifying its lexical category and occasionally its syntactic features or subcategories (Garside *et al.* 1987).

[2]   Parsing (syntactic analysis) is the process of recognizing a sentence and simultaneously compiling a representation of its structure. Thus, a sentence will be partitioned into its constituent phrases, subphrases and lexical categories.

automatic machine translation with a degree of human assistance. Several alignment algorithms have been developed for aligning bilingual texts and attention will be paid in section 4 to different approaches to aligning sentences in bilingual corpora.

Bilingual corpora are important resources for the construction as well as the extension of bilingual lexicons, as they are rich repositories of information about actual language-usage. A lexicon can be created or updated by deriving lexical information from a corpus and lexical frequency lists can be constructed from raw (untagged) corpora. Bilingual corpora give detailed information on the properties of words and on the selection restrictions of verbs, which can provide a basis for the identification of the senses of the occurrences of a given word in a corpus. Examples of possible uses of a word are provided by a corpus and although a word's semantic type has been identified, the facts provided by a bilingual corpus can be utilized to at least confirm a classification according to a lexical definition and possibly also add supplementary attributes that are absent from a lexicon.

An annotated (tagged) aligned bilingual corpus together with a probabilistic model could be used to automatically provide equivalent terms in the two languages. These terms can then be automatically compiled into the relevant formalism for lexical entries in a machine translation system. Thus, an important and possibly complex channel of information transfer exists between corpora and lexicons.

Today, computer technology provides fast access to large memories and data sources. Therefore, methods based on the access of large machine-readable corpora can be investigated. Some of the corpus-based machine translation systems utilize the corpus as a database and examples of this type of corpus-based machine translation are sublanguage machine translation systems, example-based methods/systems and knowledge-based methods/systems.

Other corpus-based machine translation approaches use statistical and probabilistic techniques for the analysis of the source language text and the generation of the target language text. An example of this type of corpus-based machine translation is the statistics-based approach of Brown *et al.* (1990). Example-based translation and statistically-based translation are so-called "empirical" approaches which apply relatively low-level statistical or pattern-matching techniques. The term "empirical" is used to refer to the fact that whatever linguistic knowledge is used by the system, is derived empirically by examining the real texts and not by relying on the knowledge of linguists.

An example of the first type of corpus-based machine translation is a sublanguage machine translation system where the texts are written in a particular sublanguage for a specific subject domain. The sublanguage corpus is utilized as a database and is searched for a source language string which is similar or identical to the string to be translated.

Example-based and knowledge-based methods are not purely corpus-based approaches to machine translation as they are not limited to a particular

sublanguage or to specific corpora. However, large corpora of texts in source languages and target languages must be accessed to obtain the necessary knowledge. The acquired knowledge will be used in the translation of previously unseen texts. The problem of acquiring and managing required syntactic and semantic knowledge forces these approaches to be applied within specific domains. The incorporation of the large volume of knowledge tends to make the system large and complicated and, therefore, not cost-effective.

In 1949 Warren Weaver suggested the application of statistical techniques for the translation of text from one natural language into another (Weaver 1955). Unfortunately, efforts in this direction were soon abandoned for various theoretical and philosophical reasons. Over the last few years there has been a trend back towards the application of statistical techniques and methods in the analysis and generation of text. The growing availability of bilingual, machine-readable texts has also stimulated the interest in the utilization of these methods (Armstrong 1994).

Statistical methods have been successfully applied to lexicography and to natural language processing. However, the success of statistical-based approaches to speech research in recent years is the primary reason for the upsurge in applying statistical methods to machine translation. Statistically-based translation is entirely statistical and probabilistic. No grammatical information is incorporated as explicit rules for the analysis of source language texts and the generation of target language texts.

The essence of this approach is the alignment of the sentences in the two languages and the calculation of the probability that one word in a sentence of the source language text corresponds to two, one or zero words in the sentence of the target language text (Hutchins 1992). The knowledge-acquisition problem is eliminated, since linguistic information is not explicitly encoded, but the general suitability of this method may be in doubt as it requires very large amounts of good quality bilingual or multilingual data.

Bilingual corpora can play an important role in the evaluation of machine translation systems. The corpus can be used to test the experimental system on real data. The tests on the corpus will help to uncover errors, limitations as well as potential areas which need improvement. The corpus can be utilized to evaluate the linguistic quality of the translation by comparing it to the target language part of the corpus. With the help of the corpus it will be possible to make sure that phenomena which appear in theoretical linguistics do indeed occur in the real world.

However, certain words and syntactic structures will not be found in the corpus. The vocabulary of the corpus is limited and the translation provided for the source text of the corpus may not be the one and only correct option. Therefore, the corpus must only be used as a guide for the evaluation of machine translation systems. The role of the corpus can thus be seen as aid in the definition of research goals and for the provision of material for system testing.

112   Jeanne Pienaar and G.D. Oosthuizen

In this section, the applications of monolingual and bilingual computer corpora in machine translation were outlined. The utilization of corpora in machine translation is not problem-free and unrestricted, and in the next section some of the problems will be identified.

3.   Problems with Corpora

In contrast to the era of Chomsky, the late 1950s, the computers of today are more powerful, faster and have hundreds and millions of bytes of storage. It is therefore possible to utilize very large corpora for research purposes. The question of the importance of corpus size inevitably arises. As the corpus size increases, the number of new types of words decreases and less effort is needed to search larger and larger corpora. However, a large portion of all word-types encountered occur only once. The problem is how large the corpus must be to capture all the words of a language as a corpus of more or less 100 million words will only produce less than half the theoretical total of word-types (Sebba 1991).

It is important for the corpus to be representative of the totality of texts from which it is drawn. At present, no statistical or other models exist for the determination of the representativeness of a corpus. One of the problems in selecting a representative sample is that in order to make valid conclusions, the sampling must be random. It is required that the procedure of composing a random sample should be objective and it would seem impossible to obtain a representative sample from a corpus. However, presently no actual criteria exist for the selection of a "representative" corpus and thus far corpus linguists assembling large machine-readable corpora have made intuitively-guided decisions about what to include and in what proportions (Sebba 1991).

The decision about what types of text to include or exclude is also difficult. The Brown corpus, for example, contains samples of text which were drawn from sources such as newspaper reports, government documents and popular fiction, but excludes poetry. It is, therefore, not possible to generalize results without difficulty.

Bilingual corpora only exist in restricted fields. Although corpora may be viewed as translations of each other, the human translators do not usually translate sentence by sentence. The translations are strictly suited to the context in which the individual source language sentence occurs. Therefore, there are several problems regarding the utilization of automated methods to translate new sentences on the basis of existing translations, as one source language sentence can translate to zero, one or two target language sentences. Thus, it seems that the short-term value of these systems, if they go beyond the experimental stage, would be as (possibly interactive) aids to human translators.

4.   The Alignment of Bilingual Corpora

In section 2 it was mentioned that a bilingual corpus is best utilized if the texts are aligned in such a way that they are translations of each other. The production of alignments by hand is extremely time-consuming and requires the skill of individuals with knowledge of both languages. Recently researchers in bilingual lexicography as well as machine translation have shown interest in the study of parallel texts and done some work on the alignment of sentences (Simard *et al.* 1992, Brown *et al.* 1991, and Gale and Church 1991). Alignment does not have to stop at sentence level and research has also been conducted to find alignments between syntactic structures, noun phrases, collocations and words.

The most common form of alignment takes the sentence to be the organizing unit and techniques exist for performing this alignment of bitext automatically with a high level of accuracy (Arnold *et al.* 1994). The alignment of paragraphs and sentences is only the first step towards the identification of word-correspondences, the construction of a probabilistic dictionary for the utilization in the alignment of words in machine translation and for the construction of a bilingual concordance for use in lexicography. Alignment can be defined as follows (Simard *et al.* 1992):

> Given a text and its translation, an alignment is a segmentation of the two texts such that the *n*th segment of one text is the translation of the *n*th of the other (empty segments are allowed as the result of additions or omissions).

The extraction of pairs of sentences from corpora that are translations of one another remains a problem as the alignment algorithm must cater for several scenarios, namely (Simard *et al.* 1992):

1.   A single sentence in one language translates to one sentence in the other language.
2.   A single sentence in one language may give rise to two or more translated sentences in the other language.
3.   Two sentences can translate into one.
4.   Two sentences in one language translate to two sentences in the other language.
5.   A sentence may not be translated at all.
6.   A new sentence may have no equivalent in the source text.

The possibility also exists that sentences, paragraphs or even passages can be missing from the corpora. These obstacles prevent many potential users from taking advantage of the many benefits of bilingual corpora as the solutions to these problems are computationally prohibitive, and/or unreliable (Simard *et*

al. 1992). An aligned bilingual corpus provides several advantages. Some of these advantages are (Simard *et al*. 1992):

1.   A valuable source of information is given.
2.   A text and its translation can be viewed side by side, with explicit connections between individual components.
3.   An alignment may form the basis of deeper automatic analysis of translation. For example, it could be utilized to indicate possible omissions in a translation. It could also be utilized for the detection of errors, for example to identify common translation mistakes.

Until recently, the South African parliamentary debates were by law published in both Afrikaans and English and is known as the South African Hansards. A subset of the Hansards was used to construct an Afrikaans-English corpus for the purpose of the project that will be described in detail in section 6. A method was developed for the alignment of the bilingual corpus at sentence level and the objective of the next section is to provide a description of the steps followed to align the English-Afrikaans corpus.

5.   Alignment of the English-Afrikaans Hansard Corpus

The Hansards of the 1990 parliamentary sessions were used for the project and it was necessary that the corpus was in an aligned format. For each parliamentary session, there were two files: one English and one Afrikaans. The text was in XYWrite format and had to be converted to ASCII format. The next step was to combine all the English files into one large English corpus and all the Afrikaans files into one large Afrikaans corpus.

The corpora were not exact translations, since sentences, paragraphs or even passages were missing from the corpora. The duplication of sentences and paragraphs was also a phenomenon that had to be taken into account. It was, therefore, necessary to compare the texts and to manually remove the unmatched sentences, paragraphs and passages.

In total, between 10% and 15% of the data in each corpus were rejected. After the completion of this exercise the number of files in each corpus were the same, no unmatched passages existed, but about 10% unmatched paragraphs and about 20% unmatched sentences were still present. The process of removing the unmatched paragraphs and sentences is described in the next section.

The sizes of the two corpora made it impractical to obtain a complete set of alignments by hand. Therefore, some method had to be used for the alignment process to be done automatically. The method employed is outlined in the next section.

5.1   The Process of Alignment

The alignment of the South African Hansard corpora was done by implementing a simple algorithm. The alignment was done first at paragraph level and then the sentences within the paragraphs were aligned. The correctness of the method was checked by hand. A program was written to identify the start and the end of the proceedings.

The next step of the alignment process was to mark the end of each paragraph. It was found that the paragraphs in the South African Hansards were already within certain regions. Therefore, the process of marking the paragraphs was not too complicated. The paragraphs were aligned automatically by applying the alignment algorithm to the corpora and each paragraph was assigned a number. About 90% of the paragraphs were aligned correctly. The unmatched paragraphs occurred as a result of missing paragraphs or where the paragraphs did not correspond one to one. The texts were inspected for possible errors, and mismatches still present were removed. The alignment algorithm was applied again and at this stage the paragraphs corresponded 100%.

The paragraphs were aligned automatically and each paragraph was assigned a number. The texts were again inspected for possible errors, and mismatches still present were removed. The next step in the process was the implementation of the sentence-alignment program. Determining sentence boundaries was a problem. It was found that most of the sentences ended with ".", "]", "!", "?", or ">", and by using these symbols as beacons, each sentence was moved to a new line. The program then automatically numbered each sentence and a success rate of about 80% was achieved in aligning the sentences. The texts were inspected for possible errors and mismatched sentences still present were removed. The process of sentence alignment and checking was repeated and this time the success rate was higher. This process was repeated until all the sentences corresponded. Although the alignment of the corpus was a rather difficult and at times a tedious process and the correction of errors time-consuming, the lack of quality data and a roughly aligned corpus would have had a negative impact on the usefulness and accuracy of the system for this project. Tables 1 and 2 illustrate the format utilized for the numbering of the files, paragraphs and sentences in the Afrikaans and English corpora.

This alignment process forms the basis of the research project and the aligned corpora will serve as input to the program for the extension of a bilingual lexicon.

```
1 «SOF AFR1.TXT»
1.1
1.1.1
WOENSDAG, 2 MEI 1990«EOP»
1.2
1.2.1
VERRIGTINGS VAN DIE UITGEBREIDE OPENBARE KOMITEE-VOLKSRAAD«EOP»
1.3
```

1.3.1
Lede van die Uitgebreide Openbare Komitee kom om 15:30 in die Raadsaal van die Volks-
raad byeen.«EOP» ...
1.9
1.9.1
Die MINISTER VAN LANDBOU: Mnr die Voorsitter, terwyl ons vanmiddag die gebed ge-
doen het, het dit my deur die gemoed gegaan hoe gepas dit is dat ons die teenwoordigheid
van die Almagtige Vader ook afbid op die dinge wat vandag begin.
1.9.2
Die gesprek oor gesprekke is 'n begin en ons kan bid dat dit goed sal verloop. [Tussenwerp-
sels.]
1.9.3
Ek wil netnou daarby uitkom en bietjie oor onderhandelinge gesels, maar ek wil dit sterk af-
ets dat dit nie dieselfde is as wat vandag aan die gang is nie.«EOP» ...
1.302
1.302.1
«EOF AFR1.TXT»

Table 1: Alignment of paragraphs and sentences — Afrikaans section

1 «SOF ENG1.TXT»
1.1
1.1.1
WEDNESDAY, 2 MAY 1990«EOP»
1.2
1.2.1
PROCEEDINGS OF EXTENDED PUBLIC COMMITTEE - ASSEMBLY«EOP»
1.3
1.3.1
Members of the Extended Public Committee met in the Chamber of the House of Assembly
at 15:30.«EOP» ...
1.9
1.9.1
The MINISTER OF AGRICULTURE: Mr Chairman, while prayers were being read this
afternoon, I was struck by how fitting it is that we call upon the Almighty to bless what we
are starting to do today.
1.9.2
The talks about talks are a start, and we can only pray that they will go well. [Interjections.]
1.9.3
In a moment I want to come back to that and say a few words about negotiations, but I
clearly want to outline the fact that this is not the same as what is going on today.«EOP» ...
1.302
1.302.1
«EOF ENG1.TXT»

Table 2: Alignment of paragraphs and sentences — English section

6.   The Implementation of a System for the Extension of a Bilingual Lexicon

As mentioned before, the purpose of this research project was to examine the viability of utilizing aligned parallel corpora to extend existing bilingual lexicons. To reach this goal the system that is developed relies on the existence of a sentence-aligned corpus as well as a bilingual lexicon. The English-Afrikaans corpus was constructed from a subset of the South African Hansards. The Machine Translation Research Group at the University of Pretoria constructed an English-Afrikaans lexicon as well as an Afrikaans-English lexicon. In the English-Afrikaans lexicon the source words are in English and the target words in Afrikaans and vice versa in the Afrikaans-English lexicon. The availability of both lexicons made it possible to use either the English text or the Afrikaans text as the source language text depending on which lexicon the user wants to extend.

Following the advice of a computer linguist it was decided to start with a simple domain and gradually move towards a more complex domain. The first experiments were conducted by only using one source language sentence and one target language sentence. These sentences are sentence-pairs that have been extracted from the aligned corpus. The other experiments used samples consisting of several sentences extracted from the aligned corpora. The program had to take the markers added by the alignment process into account when reading the source language file and the target language file. The markers indicated the start and end of each file, the end of paragraphs as well as the file numbers, paragraph numbers and sentence numbers.

The aim of the next section is to give some background information of the system. An overview of the methodology that was used to implement and test the system is also provided. Various experiments were conducted and the results obtained from these experiments will be discussed.

6.1   Background

This system provides a mechanism for the automatic identification of word-pairs that are present in a sample consisting of bilingual texts, but are absent from the bilingual lexicon (Pienaar 1996). It is also possible that a word in the source language text does have an entry in the lexicon, but that the target word entry could not be found in the target language text and that another translation has been used. Depending on which lexicon is chosen to be updated, the Afrikaans-English lexicon or the English-Afrikaans lexicon, the English corpus will be used as the source language text and the Afrikaans corpus will be used as the target language text and vice versa. One source and one target language sentence are read at a time. The source sentence is parsed and for each source language word the lexicon is checked if an entry exists. If the source language word is present in the lexicon, the target language sentence is parsed to deter-

mine if the target language word as specified in the lexicon appears in the target language sentence. An indicator is used to show if the target language word has been found or not.

By taking the positions of the unmatched source and the target language words in the sentences as well as the sentence-lengths into account, the likelihood that certain words are translations of each other, is calculated. The formulas used to determine the likelihood factor are:

$$difference = abs(pos_{source} - pos_{target})$$
$$sentence\text{-}length = (sentence\text{-}length_{source} + sentence\text{-}length_{target}) / 2$$
$$likelihood\ factor = (sentence\text{-}length - difference) / sentence\text{-}length$$

The difference value is determined by calculating the absolute difference in position of each unmatched source and target word in the source and target language sentence. Unfortunately it is not possible to utilize the positions of the words and the sentence-lengths only in calculating the likelihood factor. Some reasons are:

• The lengths of the target language sentence and the source language sentence often differ.
• Two or more words in the source language sentence translate to zero or one word in the target language sentence.
• One word in the source language sentence translates to zero, one, two or more words in the target language sentence.
• The positions of the source words in the source language sentence do not correspond with the positions of the target words in the target language sentence.

To determine more accurately if (a) source and target word(s) are translations of each other, the incorporation of linguistic knowledge was inevitable. For each entry in the lexicon, the grammatical category (type) of the word is specified. By using this syntactic information, it is possible to determine the grammatical types of the neighbouring word(s) of the unmatched word. A few basic grammar rules were implemented to determine the possible type of each word-pair. At present, the rules can give an indication whether the grammatical category of the word-pair is a noun, verb, adverb, adjective or a preposition.

The grammatical category of the preceding word as well as of the following word is queried, and the simple grammar rules are applied by taking these known types into consideration. In the case of insufficient information, for example, if the type of the preceding or following word is also unknown, or if the grammatical information does not conform to the rules, the grammatical category cannot be ascertained and is indicated as *unknown*. If more than one rule is satisfied, the possible types will be provided and it will be the responsibility of the user to select the correct type of the word-pair. The possibility also exists that two words following each other can translate to one, two or

more words or that one word can translate to two or more words. The system caters for these scenarios in so far that it can identify a translation pair consisting of a compound noun written as two words (such as *water tariff*) and a compound noun written as one word (such as *watertarief*) as well as a verb-preposition pair (such as *gaan oor*) and a verb (such as *crosses*). Although the simplified grammar rules currently implemented give an idea of what the possible types of the unmatched words are, more intensive investigation is required for the refinement, improvement and extension of these rules.

The output of the system is a list of unmatched source words and possible translations. A likelihood factor indicating the numeric possibility that the word-pairs are translations of each other, as well as the possible grammatical type(s) are provided. The human translator or the user has to inspect the results and identify the correct translation pairs. The grammatical type of each word-pair must be verified. In the case of the type not being known the translator has to rely on his/her linguistic knowledge to identify the correct grammatical type of the word-pair. The word-pair and the grammatical type can then be added to the lexicon as a new entry. The objective of the next section is to give a brief overview of the methodology that the system follows to extend a bilingual lexicon.

6.2   Methodology

The process of identifying word-pairs, ascertaining the grammatical category and calculating the likelihood factor can be divided into the following stages:

- **Input**

The source and target words as well as the grammatical types are extracted from the lexicon and recorded. The source and target language sentences are read from the corpus.

- **Lexicon lookup**

The words from the source sentence are checked against the extracted lexicon entries. The target language sentence is searched for the translation specified in the lexicon. If the translations are located, an indicator for each target word is set to show that the searching process was successful. The source words not found are recorded.

- **Calculation of likelihood factor**

For each unmatched word-pair the likelihood factor is determined by taking the positions of the source and target words as well as the sentence lengths into account.

120   Jeanne Pienaar and G.D. Oosthuizen

- **Assignment of grammatical categories**

For each unmatched word-pair, grammatical rules are applied to attempt to recognize the grammatical type. For example, the grammar rule constructed for the identification of the adjective *<tall>* in the sentence *The very tall trees* takes into account that the adjective is preceded by the adverb *<very>* and followed by the noun *<trees>*. Thus the rule which will satisfy this specific example has the following format:

> < *(det) the (adverb) very (?) tall (noun) trees* > →
> < *(det) the (adverb) very (adj) tall (noun) trees* >

If the recognition process is successful, the grammatical type is assigned to the word-pair. In the case of an unsuccessful recognition process the assigned type is *unknown*.

- **Identification of compounds**

Compound nouns (such as *election manifesto* and *verkiesingsmanifes*) and verb-preposition pairs (such as *oorneem* and *take over*) are identified.

- **Output**

The result of the program is a list containing an entry for each identified word-pair, the possible grammatical type(s) and the likelihood factor.

- **Inspection**

The list must be inspected by the human translator or the user for correctness. The correct word-pairs and their grammatical type must be extracted and added to the lexicon. The purpose of the next section is to describe some of the experiments that were conducted.

6.3   Experiment

Purpose

The aim of the experiment was to evaluate the performance of the system by utilizing several sentences differing in terms of:

- complexity (for example, the tense of the sentence, the number of noun and verb phrases, positions of the verbs and presence of compound prepositions),
- sentence length and
- number of unmatched words.

Method

Twenty sentences were randomly chosen from the Afrikaans-English corpus and used as input to the system. The sentences satisfied the requirements as stated in the purpose of this experiment, since they differed in complexity, sentence length and the number of unmatched words. The sample extracted from the Afrikaans corpus containing the source language sentences is shown in table 3.

---

1. Ek sê dit, want ek was betrokke by 'n verkiesing.
2. Die skrapping van hierdie betrokke wet het my volle ondersteuning.
3. Geen deel van die samelewing bly onaangeraak deur hierdie probleme nie.
4. As ons praat van spesifieke behoeftes, beteken dit nie 'n swak standaard van dienslewering nie.
5. Ek dink ons moet 'n duidelike onderskeid tref tussen die vlakke van dienslewering.
6. Ek bedank agb lede vir hul steun aan hierdie wetgewing.
7. Die begrotingswetsontwerp is aanvaar.
8. Een van die probleme ten opsigte van die solvensie van die pesioenfonds was die geweldige las wat die stelsel op die fonds geplaas het.
9. Ek wil net vir die agb Minister vra of sy departement betrokke is by die opstel van sulke programme en of hy enige onderwysdepartemente genader het.
10. Daar is 'n behoefte aan formele en informele voorligtingsprogramme.
11. Die natuurlewe dra by tot hierdie gehalte.
12. Na 'n baie deeglike ondersoek is 'n omvattende gewysigde skema daargestel.
13. Die beweging van mense in die wêreld stimuleer groei.
14. 'n Private monopolie tree dikwels op teen die belang van die verbruiker.
15. Kruissubsidiëring word dus 'n werklikheid.
16. Ek het begrip vir die feit dat ons die bronne moet beskerm en dat weersomstandighede, seisoene en al daardie dinge 'n invloed het op ons bronne aan die kus.
17. Ek dink dat ons 'n tydperk binnegaan waar verwagtinge geskep gaan word.
18. Die departement behou die bevoegdheid oor die toewysing en toedeling van water uit die hoofbron, en hierdie water sal ook aan die raad teen 'n tarief beskikbaar gestel word.
19. Ek wil graag begin deur 'n woord van hulde te spreek soos dit al die gebruik is wanneer 'n senior amptenaar aftree na lang jare van diens.
20. Ons weet dat beperkte toegang tot die rekeninge en beleggings van vermeende handelaars dikwels ondersoeke kortwiek.

Table 3: Sample of the Afrikaans corpus

The sample extracted from the English corpus containing the target language sentences is shown in table 4.

1. I say this, because I was involved in an election.
2. The abolition of this specific act has my full support.
3. No segment of the society stays untouched by this problem.
4. When we talk about specific requirements, this does not entail a poor standard of service.
5. I think we should draw a clear distinction between the levels of service.
6. I thank the hon members for their support for this legislation.
7. The appropriation bill was adopted.
8. One of the problems in respect of the solvency of the pension fund was the tremendous strain which the system had placed on the fund.
9. I simply want to ask the hon Minister whether his department is involved in the installation of such programmes and whether he has approached any education departments.
10. There is a need for formal and informal education programmes.
11. The wildlife adds to this quality.
12. After a very exhaustive investigation, a comprehensive amended scheme was established.
13. The movement of people around the world stimulates growth.
14. A private monopoly frequently acts against the interests of the consumers.
15. Thus cross-subsidisation becomes a reality.
16. I have understanding for the fact that we have to protect all the resources and that weather conditions, seasons and those things have an influence on our coastal resources.
17. I think that we are entering a period where expectations will be created again.
18. The department retains its powers in regard to the allocation and appointment of water from the main source, and this water will also be made available to the board at a specific tariff.
19. I should like to begin by paying tribute, as is customary, to a senior official who has retired after many years of service.
20. We know that limited access to the accounts and investments of suspected dealers frequently hampers investigations.

Table 4:   Sample of the English corpus

The performance of the system was measured in terms of the number of the translation pairs correctly identified. For each word-pair the likelihood factor was calculated and the grammar rules were applied to assign a grammatical type to each translation pair.

Results and Discussion

The performance of the system was dependent on the sentence length, the number of unmatched words present in the sentence as well as the complexity of the sentence. The results varied according to the circumstances. The various performance results are:

• The system could identify the translation pairs and the possible grammatical categories to which each of them belonged, if the sentences contained only a few (three or less words out of ten words per sentence) unmatched words, irrespective of the sentence length. For example, for the sentence pair

> *Ek wil net vir die agb Minister vra of sy departement betrokke is by die opstel van sulke programme en of hy enige onderwysdepartemente genader het.*

and

> *I simply want to ask the hon Minister whether his department is involved in the installation of such programmes and whether he has approached any education departments.*

five words were unmatched and the target sentence length was twenty-seven words. Thus 18% of the words were unmatched. The word-pair *<net, simply>* was identified as a translation pair purely on the value of the likelihood factor since the grammatical type of the pair could not be determined by the grammar rules. The following word-pairs as well as their grammatical type were correctly identified:

♦ *<agb, hon>* and grammatical type of *adjective*,
♦ *<opstel, installation>* and grammatical type of *noun*,
♦ *<programme, programmes>* and grammatical type of *noun* and
♦ *<onderwysdepartemente, education departments>* and grammatical type of *noun + noun*.

Thus, a success rate of 80% was achieved for this specific example.

• As was to be expected, the system did not perform very well when the source and target sentences were very long and contained many unmatched words. For example, for the sentence pair

> *Ek wil graag begin deur 'n woord van hulde te spreek soos dit al die gebruik is wanneer 'n senior amptenaar aftree na lang jare van diens.*

and

> *I should like to begin by paying tribute, as is customary, to a senior official who has retired after many years of service.*

fourteen words were unmatched and the target sentence length was twenty-three words. Thus 57% of the words were unmatched. The system was able only to identify five (23%) translation pairs and their grammatical types correctly. The grammar rules were not very successful in determining the grammatical types, since most of the grammatical types of the neighbouring word(s) of the unmatched word were unknown.

- The system performed rather well if the grammatical types of the neighbouring words of the unmatched word were known. Unfortunately the grammar rules are less successful if the neighbouring words are also unmatched. For example, for the sentence pair

> *Ons weet dat beperkte toegang tot die rekeninge en beleggings van vermeende handelaars dikwels ondersoeke kortwiek.*

and

> *We know that limited access to the accounts and investments of suspected dealers frequently hampers investigations.*

the grammatical types of all of the neighbouring words of the unmatched words were known. The unmatched word-pairs for this specific example were: *<beperkte, limited>*, *<rekeninge, accounts>*, *<vermeende, suspected>* and *<kortwiek, hampers>*. The system was successful in identifying the correct grammatical types for the four word-pairs.

For the sentence pair

> *Die skrapping van hierdie betrokke wet het my volle ondersteuning.*

and

> *The abolition of this specific act has my full support.*

the grammatical types of all of the neighbouring words of the unmatched words were not known. The unmatched word-pairs for this specific example were: *<skrapping, abolition>*, *<betrokke, specific>*, *<wet, act>* and *<volle, full>*. The grammar rules were unable to determine the grammatical types of the unmatched word-pairs *<betrokke, specific>* and *<wet, act>* which follow each other. However, the system could identify the correct grammatical types for the other two word-pairs which were positioned elsewhere in the sentence and were surrounded by words of known grammatical types.

Table 5 is a graphical representation of the results obtained from the experiment as described above. The x-axis (horizontal axis) shows the percentage success rate that was achieved. The formula for the x-axis (horizontal axis) is as follows:

$$\% \ correct = (x \ / \ y) * 100$$

where x = *number of word-pairs correctly identified* and y = *number of unmatched words.*

The y-axis (vertical axis) indicates the percentage unmatched words per sentence. The formula for the y-axis (vertical axis) is as follows:

$$\% \ unknown/\ length = (y\ /\ l)\ *\ 100$$

where y = *number of unmatched words* and l = *length of sentence.*

The graph confirmed the expectation that the system is less successful in handling very long or very short sentences containing many unmatched words. The graph also implies a higher success rate if few unmatched words were present irrespective of the sentence length.
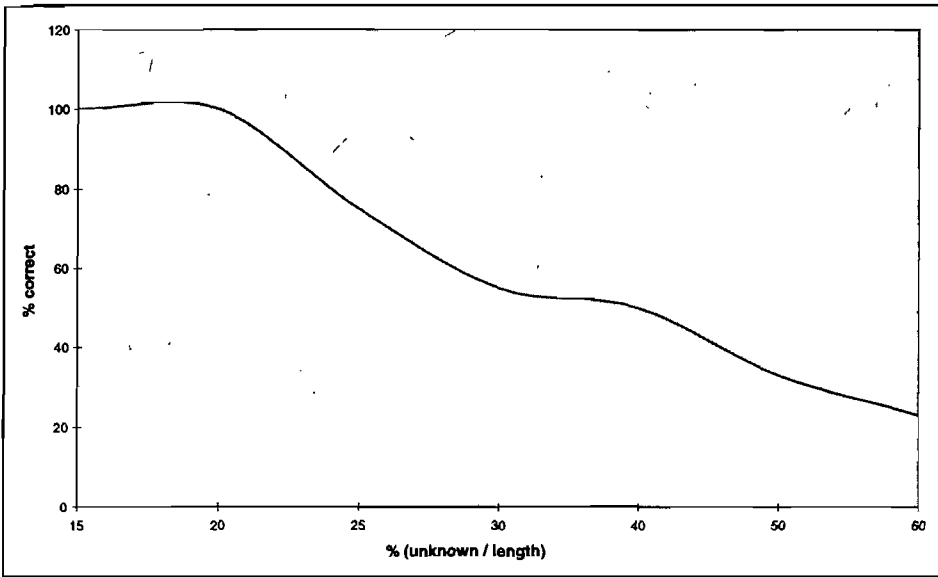
Table 5: Statistics for identification of word pairs

Conclusions

The results were encouraging and it would seem that the system is more successful in identifying unmatched words and determining their grammatical type if there are not too many unmatched words per sentence. Since the purpose of the system is to extend an existing lexicon and not to produce a new lexicon from scratch, the system will perform better if most of the words in the sentences already have entries in the lexicon and the system only has to identify a few new words per sentence.

7.   Future Research and Concluding Remarks

The object of this project was to examine the hypothesis that parallel corpora, in this case the English-Afrikaans corpus, can be utilized to extend an existing bilingual lexicon, in this case either the English-Afrikaans or the Afrikaans-English lexicon. To investigate and demonstrate the validity of this argument, a simple system was developed which was applied to sentence-aligned samples of parallel texts extracted from the South African Hansards.

Encouraging results were obtained from the experiments conducted. The results confirmed the expectation that sentences, irrespective of length, containing many unmatched words, would yield a lower success rate than sentences containing only a few unmatched words. Several problems were identified, but further research is necessary to clarify these issues and to propose possible solutions. Some of these issues as well as other possible future research topics are discussed in the next section.

7.1   Future Research

The alignment of the English and Afrikaans corpora was done at sentence level. The alignment process does not have to stop at this level and it is possible to determine the most probable word-pair alignments. Phrasal alignment is another possibility and can be achieved by phrasal parsing and the utilization of phrasal information. Although the success rate of these probabilistic techniques is dependent on the size and the quality of the corpus, the English-Afrikaans corpus satisfies these dependencies. It is difficult to align translations on the basis of words and to achieve this goal, a tagged corpus will be a valuable resource. Some specific issues identified that require further investigation, are:

- The simple grammar rules that were implemented for the identification of the grammatical categories can be refined in order to achieve more accurate results.
- Problems exist in determining the grammatical category for the unmatched word-pair if the neighbouring source and target words are also unmatched and the grammatical categories are not related. For example, for the sentence *Daar is 'n behoefte aan informele programme* the simple grammar rules were unable to assign the grammatical type of *adjective* to the unmatched word, *informele*, and *noun* to the unmatched word, *programme*.
- The treatment of verbs in terms of the difference in position in the sentence for different languages, for example in Afrikaans the verb often occurs to the end of the sentence while this is not usually the case in English.
- Difficulties with split verbs, for example *vat ... saam*.
- The treatment of compound prepositions, for example *out of*.
- The treatment of complex prepositions, for example *in place of*.

A tagged corpus would also add value to the system, since the linguistic information provided by the tags can be used to refine the process of determining the likelihood that word-pairs in parallel sentences are translations of each other and to ascertain the grammatical category of the word-pair. A tagged corpus can be utilized for deriving linguistic rules which can be incorporated into this system or into an existing machine translation system and thereby contribute to the improvement of the quality of the product. As the tags provide part-of-speech information, the tagged corpus can be used as a resource to study actual language-usage in English and Afrikaans.

7.2   Concluding Remarks

Although the system is very simple and some problems still exist, the results were encouraging. The refinement of the grammar rules and the incorporation of more linguistic knowledge should improve the results, while simultaneously reducing the cost of human inspection of the newly identified translation pairs. It would also be interesting to investigate the use of tagging and the effect it will have on the results. The system proved valuable in that it supported the extension of an existing bilingual lexicon with reduced human effort.

Bibliography

Armstrong, S. (Ed.). 1994. *Using Large Corpora*. Cambridge, Massachusetts: MIT Press.

Arnold, D., L. Balkan, R.L. Humphreys, S. Meijer and L. Sadler. 1994. *Machine Translation — An Introductory Guide*. Manchester: Blackwell Publishers.

Brown, P.F., J.C. Lai and R.L. Mercer. 1991. Aligning Sentences in Parallel Corpora. *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics (Berkeley)*. Berkeley: University of California.

Brown, P.F., J. Cocke, S.A. Della Pietra, V.J. Della Pietra, F. Jelinek, J.D. Lafferty, R.L. Mercer and P.S. Roosin. 1990. A Statistical Approach to Machine Translation. *Computational Linguistics* 16.

Brown, P.F., S.A. Della Pietra, V.J. Della Pietra, J.D. Lafferty and R.L. Mercer. 1992. Analysis, Statistical Transfer and Synthesis in Machine Translation. *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages, Montréal*.

Carbonell, J.G., T. Mitamura and E.H. Nyberg. 1992. The KANT Perspective: A Critique of Pure Transfer. *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages, Montréal*.

Catizone, R., G. Russel and S. Warwick. 1989. Deriving Translation Data from Bilingual Texts. *Proceedings of the First International Acquisition Workshop, Detroit*.

Chang, J-C. and H. Chen. 1995. Using Partially Aligned Parallel Text and Part-of-Speech Information in Word Alignment. *Proceedings of the First Conference of the Association for Machine Translation in the Americas, Columbia*.

Chomsky, N. 1962. A Transformational Approach to Syntax. Hill, A.A. (Ed.). *Proceedings of the 1958 Conference on Problems of Linguistic Analysis in English*. Austin: University of Texas.

Gale, W.A. and K.W. Church. 1991. A Program for Aligning Sentences in Bilingual Corpora. *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics (Berkeley)*. Berkeley: University of California.

Gale, W.A., K.W. Church and D. Yarowsky. 1992. Using Bilingual Materials to Develop Word Sense Disambiguation Methods. *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages, Montréal*.

Gale, W.A., K.W. Church and D. Yarowsky. 1993. A Method for Disambiguating Word Senses in a Large Corpus. *Computers and Humanities* 26.

Garside, R., G. Leech and G. Sampson. 1987. *The Computational Analysis of English — A Corpus-based Approach*. London: Longman.

Hutchins, W.J. 1986. *Machine Translation: Past, Present, Future*. Chichester: Ellis Horwood.

Hutchins, W.J. and H.L. Somers. 1992. *An Introduction to Machine Translation*. London: Academic Press.

Klavans, J. and E. Tzoukermann. 1996. Combining Corpus and Machine-Readable Dictionary Data for Building Bilingual Lexicons. *Machine Translation* 10.

Maclovitch, E. 1992. Where the Tagger Falters. *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages, Montréal*.

Mitamura, T. and E.H. Nyberg. 1995. Controlled English for Knowledge-Based MT: Experience with the KANT System. *Proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation, Leuven*.

Pienaar, J. 1996. The Utilization of Parallel Corpora for the Extension of Machine Translation Lexicons. Unpublished M.Sc. thesis. Pretoria: University of Pretoria.

Renouf, A. 1987. Corpus Development. Sinclair, J.M. (Ed.). *Looking Up: An Account of the COBUILD Project in Lexical Computing*. London: Collins.

Sato, S. and M. Nagao. 1990. Toward Memory-Based Translation. *COLING-90: Papers presented to the 13th International Conference on Computational Linguistics (Helsinki)*. Helsinki: Helsinki University.

Sebba, M. 1991. The Adequacy of Corpora in Machine Translation. *Applied Computer Translation* 1.

Simard, M., G.F. Foster and P. Isabelle. 1992. Using Cognates to Align Sentences in Bilingual Corpora. *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages, Montréal*.

Smadja, F. 1992. How to Compile a Bilingual Collocation Lexicon Automatically. *Proceedings of the AAAI Workshop on Statistically-Based NLP Techniques, San Jose, CA*. San Jose, California: AAAI Press.

Somers, H.L. 1993. Current Research in Machine Translation. *Machine Translation* 7.

Su, K. and J. Chang. 1990. Some Key Issues in Designing MT Systems. *Machine Translation* 5.

Su, K. and J. Chang. 1992. Why Corpus-Based Statistics-Oriented Machine Translation? *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages, Montréal*.

Sumita, E., H. Iida and H. Kohyama. 1990. Translating with Examples: A New Approach to Machine Translation. *Proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages (Austin TX).* Austin: University of Texas.

Weaver, W. 1955. Translation. Booth, A.D. and W.D. Locke (Eds.). *Machine Translation of Languages.* Cambridge, Massachusetts: MIT Press.

Kind, Considerate, Thoughtful:
A Semantic Analysis[1]

Catherine Travis, Department of Linguistics,
La Trobe University, Victoria, Australia

Abstract: This paper presents a semantic analysis of three English words denoting positive character traits, namely *kind*, *considerate* and *thoughtful*. These three words are closely related, and the differences (and similarities) in their meanings can be very difficult to pinpoint. It shall be shown that modern dictionaries demonstrate a great deal of circularity in their definitions of these words, reflecting the closeness of their meanings. An analysis of usage examples provides evidence on the basis of which their differences can be understood. Some of the differences that shall be noted are the following: *kind* and *thoughtful* necessarily involve doing something for another person, while *considerate* does not; *kind* involves not wanting anyone to feel bad, *considerate* involves not wanting anyone to feel bad as a result of one's actions, and *thoughtful* focuses on a specific situation another is in, and not wanting them to feel bad in that situation; *kind* involves wanting to do something for another's good, while *considerate* involves wanting to avoid something that may harm another; and both *considerate* and *thoughtful* imply some kind of thinking about another before one acts, as reflected in their morphology. The Natural Semantic Metalanguage approach, as developed by Wierzbicka (1972, 1980, 1996) and colleagues, is used to propose definitions for these words, with the aim of exhaustively capturing their meaning, and clearly delineating their range of use. Such definitions can be particularly valuable to second language learners, and can provide a basis for cross-linguistic, and cross-cultural, comparisons of related concepts.

Keywords: KIND, CONSIDERATE, THOUGHTFUL, SEMANTICS, LEXICOGRAPHY, NATURAL SEMANTIC METALANGUAGE, DEFINITIONS, CHARACTER TRAITS, PERSONALITY, DICTIONARY, WIERZBICKA, CIRCULARITY, CROSS-CULTURAL STUDIES

Opsomming: *Kind, considerate, thoughtful:* 'n Semantiese analise. In hierdie artikel word 'n semantiese analise aangebied van drie Engelse woorde wat positiewe karaktertrekke te kenne gee, naamlik *kind*, *considerate* en *thoughtful*. Hierdie drie woorde is nou verwant, en die verskille (en ooreenkomste) in hul betekenisse kan moeilik uitgewys word. Daar sal aangetoon word dat moderne woordeboeke 'n groot mate van sirkelbeskrywing in hul definisies van hierdie woorde bevat, wat die noue verband in hulle betekenisse weerspieël. 'n Ontleding van gebruiksvoorbeelde lewer bewyse op grond waarvan hulle verskille verstaan kan word. Sommige van die verskille wat uitgewys sal word, is die volgende: *kind* en *thoughtful* hou noodsaaklikerwyse in om iets vir iemand anders te doen, terwyl dit nie die geval met *considerate* is nie; *kind* hou in om enigiemand nie sleg te wil laat voel nie, *considerate* hou in om iemand nie sleg te wil laat voel as gevolg van 'n mens se optrede nie, en *thoughtful* fokus op 'n spesifieke situasie waarin iemand anders is, en om hom/haar nie te laat sleg voel in daardie situasie nie; *kind* hou in dat iets tot iemand anders se voordeel gedoen wil word, terwyl *considerate* inhou dat iets vermy wil word wat ander kan bena-

deel; en beide *considerate* en *thoughtful* impliseer 'n mate van inagneming van 'n ander voor 'n mens optree, soos weerspieël in hul morfologie. Die Natuurlike Semantiese Metataal-benadering, soos ontwikkel deur Wierzbicka (1972, 1980, 1996) en kollegas, word benut om definisies vir hierdie woorde voor te stel, met die doel om hul betekenisse volledig vas te lê, en om hul gebruiksgebied duidelik af te baken. Sulke definisies kan veral waardevol wees vir tweedetaalaanleerders, en kan as basis dien vir kruislinguistiese, en kruiskulturele, vergelykings van verwante begrippe.

Sleutelwoorde: *KIND, CONSIDERATE, THOUGHTFUL*, SEMANTIEK, LEKSIKOGRAFIE, NATUURLIKE SEMANTIESE METATAAL, DEFINISIES, KARAKTEREIENSKAPPE, PERSOON-LIKHEID, WOORDEBOEK, WIERZBICKA, SIRKELBESKRYWING, KRUISKULTURELE STU-DIES

1.   Introduction

Words denoting character traits are a little researched but interesting area in the field of lexicography. It is an area which seems to include many words referring to closely related concepts, the differences between which are often very difficult to capture. Such is the case for *kind, considerate* and *thoughtful*, which are commonly defined via each other in modern dictionaries. In this paper, an empirical analysis of these three words will be used to address the issue of how to define such concepts. A brief look at some dictionary defini-tions of these words will highlight some of the problems with these definitions, and an alternative analysis — the Natural Semantic Metalanguage (NSM) approach — shall be proposed. On the basis of examination of usage examples, definitions for *kind, considerate* and *thoughtful* constructed in accordance with this approach shall be put forward, and these definitions will be used to demonstrate the ability of NSM to accurately and exhaustively capture and explicate meaning.

2.   Treatment of *kind, considerate* and *thoughtful* in modern dictionaries

The closeness of meaning of *kind, considerate* and *thoughtful* can be seen by the circularity of definitions found in modern dictionaries. The following defini-tions from the *Australian Concise Oxford Dictionary* (1987) of *considerate* and *thoughtful* serve as a prime example of this.

> *considerate*: "thoughtful for others"
> *thoughtful*: "considerate (*of*), not haphazard or unfeeling"

Any dictionary user who does not know the meaning of both of these words would be no better off in the case of *considerate*, and still very much in the dark regarding *thoughtful*. The element added to "considerate" in the definition of *thoughtful* ("not haphazard or unfeeling") is so broad that it could cover any

range of words, and does little to delineate the meaning of *thoughtful*. Further-more, what does it mean to be "not haphazard *or* unfeeling"? Does it mean that a *thoughtful* person may exhibit either one or the other of these properties, and if so, are both truly part of the meaning of *thoughtful*? How do these properties relate to each other, and to "considerate", the first element proposed?

The *Collins English Dictionary* (1991), cited below, also provides circular definitions for *considerate* and *thoughtful*, and includes some circularity in its definition of *kind*.

kind: "1. having a friendly or generous attitude. 2. helpful to others or another: *a kind deed*. 3. considerate or humane. 4. cordial; courteous (esp. in the phrase *kind regards*). 5. pleasant, agreeable, mild: *a kind climate*."

considerate: "1. thoughtful towards other people; kind. 2. *Rare.* carefully thought out; considered."

thoughtful: "1. considerate in the treatment of other people. 2. showing careful thought. 3. pensive; reflective."

Thus, *considerate* is defined via *thoughtful* and *kind*, and both *thoughtful* and *kind* are defined via *considerate*. Multiple glosses are proposed for each word, which may be justified in the case of *considerate* and *thoughtful*,[2] but is questionable in the case of *kind*. This exemplifies another problem prevalent in dictionaries — that of positing polysemy where perhaps it does not exist (cf. Wierzbicka 1996: 270 ff). In the case of *kind*, for example, surely we would not want to say that "having a friendly or generous attitude", being "helpful to others or another", "considerate or humane" and "cordial; courteous" are all distinct meanings in the same way that "considerate in the treatment of other people" and "pensive, reflective" are for *thoughtful*. Polysemy is a feature of language, and cannot be ignored, but nor should it be posited without careful analysis that proves its existence.

In order to deal with problems of circularity, polysemy, multiple aspects of meaning and the many other difficulties encountered in attempting to cap-ture meaning, a rigorous lexicographic theory that specifically addresses these issues must be utilised. In the following section I shall discuss such a theory, proposed by Wierzbicka and colleagues, and will then go on to demonstrate the practical application of this theory in relation to *kind*, *considerate* and *thoughtful*.

3.   The Natural Semantic Metalanguage Approach

The Natural Semantic Metalanguage, or NSM, is a metalanguage used for defining words and concepts that has been compiled over nearly 30 years of extensive research by Wierzbicka (see especially 1972, 1980, 1996 and refer-ences therein) and colleagues (see especially Goddard 1989; Goddard and Wierzbicka (Eds.) 1994; Boguslawski 1970). The basic tenets underlying this

approach are as follows:[3]

(1)   semantic analysis should be carried out using paraphrase based on natural language as opposed to artificial symbols, features or markers;

(2)   semantic analysis must follow a reductive approach, defining complex concepts in terms of simpler ones;

(3)   there is a finite set of words, the meanings of which are so basic that they cannot be broken down any further;

(4)   this set of words (the so-called "semantic primitives") represents innate concepts that are fundamental to human thought, partly evidenced by their expression (be it by a word, morpheme or phrase) in all languages of the world;

(5)   these primitives can be used as "building blocks" to define all words and expressions;

(6)   there are distinct grammatical patterns that govern the combinability of these primitives, which represents the "innate grammar" of human cognition.

NSM makes use of the "primitives" as its lexicon, and the "innate" grammatical rules as its syntax to construct culture- and language-independent natural language definitions of words and concepts. Although the metalanguage is semi-artificial, the fact that it comprises a subset of natural language (principle (1)) means that the definitions can be understood on their own, without having to decode symbols and features.

The principle of defining complex notions in terms of simpler ones (principle (2)) solves the problem of circularity discussed above, and is perhaps the only way to guarantee resolution of this problem. It also ensures that definitions do not include scientific or other knowledge, such as, for example, when *salt* is defined as "sodium chloride" (cf. *Collins English Dictionary* 1991, among others). It is hard to imagine that someone who did not know the meaning of the word *salt* would know what *sodium chloride* is. This definition provides scientific information about *salt*; it does not capture the everyday meaning of *salt*, as used to refer to something we add to food to give it flavour.[4] If the language of the definition is always simpler than the word being defined, then this problem of turning to scientific information can readily be avoided. NSM follows a reductive approach, until the set of indefinable words is reached, and this set can be used to define all other words (principles (3), (4) and (5)). The proposed set of primitives currently numbers around 60 words, and the English version includes words such as *I, you, someone, something, good, bad, want, know, say, do, think, this, can* etc. The "universality hypothesis" (that these concepts can be expressed in some way in all languages of the world) has been empirically tested in a number of other languages, and has been strongly backed up by this testing.

The NSM lexicon is used to form sentences in accordance with its own

syntactic rules, which are maximally simple and believed to be language-independent (principle 6)).

The NSM approach allows the analyst to rigorously test and account for each component of meaning that is to be posited in a definition, as well as how it is to be presented. NSM definitions that clearly demarcate meaning can show precisely where related words differ and where they coincide. This is extremely useful for anyone in search of better understanding of meaning. It is particularly useful, perhaps, for second-language learners, and can also play an important role in cross-cultural understanding, facilitating comparisons of related concepts across languages, that can then offer insights into cultural values and attitudes.[5]

4.   Defining character-traits

An NSM definition consists of a series of components, each component representing different elements of the meaning of the word under consideration. In the analysis presented here, I shall first consider each element (or set of elements) individually, before compiling them to provide the completed definition.

The three words to be looked at here refer to character-traits that are realised in interpersonal relations. They refer to an attitude one takes to other people, or the way one thinks about others in interacting with them. As used in the frame "X is *kind* / *considerate* / *thoughtful*" these words all refer to an enduring quality, as opposed to something manifested in a one-off incident. That is, a "*kind* person", for example, is not someone who once did something *kind* for someone, but someone who regularly acts in a "*kind* way". This must be captured in the first component of the definition, setting the context for the components to follow. This can be done in the following way:

> "X is *kind* / *considerate* / *thoughtful*"
>   X **often thinks something like this about people:**

The use of "often" reflects the "enduring" nature of the quality being defined, and the expression "thinking about people" reflects the fact that these words refer to one's attitude in interpersonal relations. Including X thinks "something like this" gives the definitions a margin of leeway in their interpretation, and thus allows for metaphorical and other extensions. I shall now look at, and attempt to explicate, the attitude entailed in these concepts.

4.1   Kind

4.1.1   Not wanting others to feel bad

Perhaps the most obvious element of the meaning of *kind* is that it implies that

one would not want to do something that may hurt another. This is reflected in the expression:

(1)   He's so *kind*, he wouldn't hurt a flea.

In the light of this expression, it is important to consider the following, which is in apparent contradiction to this:

(2)   You have to be cruel to be *kind*.

What this implies is that sometimes one has to do things that may hurt another, for some eventual good to come of that. It seems, then, that *kind* does not imply not wanting to do something that may harm another, but a more general feeling that one does not want others to feel bad, be that as a result of one's own actions, or independently of them.

Consider the following example of an encounter between three children, one of whom (the one telling the story) has been searched by the other two for something he had been accused of stealing. The two children searching him protect him from further trouble, by pretending not to find the stolen object, and as they leave, one of them smiles at him.

(3)   The girl, before she went with him, gave me a look of *kindness*,[6] and I remembered that it was she who on my first day at this school had come up to me in the school yard to offer me a sweet from a bag of sweets she was carrying.

UE: 315[7]

Presumably the children do not report the stolen object because they don't want the narrator to get into trouble, and this is punctuated by a "look of *kindness*", which reminds the narrator of what he, we can assume, sees as another *kind* act, that of offering him sweets on his first day at school. The implication here is not so much that the girl doesn't want to do something that may harm the narrator, but that she doesn't want him to be harmed in any way. It is for this reason that she doesn't report the stolen object (which would have resulted in him being punished), and that she gave him sweets at school (to make him feel welcome on his first day).

I therefore propose that the following component be included in the definition of *kind* to capture this notion.

X is kind:
  X often thinks something like this about people:
    I don't want anyone to feel something bad

4.1.2   Doing something for the benefit of others

Closely related to this notion of not wanting others to feel bad is an implication that one wants to do things for their benefit, and that one does actually do something because of this. This must be stated independently of the component proposed in 4.1.1, because it is possible that one does not want others to feel bad, but doesn't actually do anything to cause them to feel good. This would be the case, for example, for *sensitive*, which implies avoiding saying or doing things that may hurt others, but not necessarily doing something that would specifically benefit them (Travis 1992). That *kind* must involve doing something for others can be seen in the unacceptability of the following sentence:

(4)   *He won't do anything for you, but he is *kind*.

Note also the following expression, used for asking a favour, which shows that being *kind* is associated with doing something.

(5)   Would you be *kind* enough to / so *kind* as to ...

At this stage, I propose the following components to capture this notion, which will be subject to modification with further analysis.

X often thinks something like this about people:
 I want to do something good for this person
X does something because of this

This reflects the fact that being *kind* implies a desire to do things for another's benefit, and the performing of some action because of this desire, but it does not specify that what one does is actually to the benefit of that person. This is important because it is possible that what is intended as an act of *kindness* does not in fact benefit the person to whom it is directed. This can be shown by the fact that we could say the following, of someone who is very generous in offering their help to another:

(6)   She's very *kind*, and she means well, but I'd rather she'd just let me do it myself.

Another example of this is found in the 1992 acceptance speech of then US President George Bush, in which he promised a "*kinder*, gentler America", one of the features of which was to be no new taxes. With the budget deficit at a record high at that time, whether this would in fact benefit the country was questioned (*Canberra Times* 1992: 10).

For this reason, *kind* cannot be defined as "doing something good for another", but must be defined as simply "doing something". Whether what one does, is to the other's benefit exists in the intentions behind the *kind* act, rather than in the result of it.

4.1.3   Action performed is not a "big" thing

The components proposed in 4.1.2 need qualification, because they imply that any action performed with the intention of benefiting another would be described as *kind* and this is far from true. Consider, for example, the case of aid workers who devote their lives to helping the starving or poverty-stricken in developing countries. Certainly these actions would not be described as *kind*, as is shown in the unacceptability of (7).

> (7)   *It was very *kind* of her to devote her life to helping the starving in Africa.

Rather than *kind*, "self-sacrificing", "selfless", or "humanitarian" would be more likely to be used to describe such people. Another example of something one could do for another's good that could not be described as *kind* would be cutting off another's hand (for example, to save their life by allowing them to avoid military service). It would be impossible to say of such an act:

> (8)   *He *kindly* chopped off his friend's hand.

It seems that both of these examples are unacceptable because of the magnitude of the act being performed, and that *kind* is reserved for relatively simple tasks, that perhaps do not require a great amount of effort, or do not have serious consequences for either party involved (such as would devoting one's life to helping others, or having one's hand cut off).

That *kind* acts are not "big" things is reflected in the expression "little deeds of *kindness*", used below in the poem "Little Things" by Julia Fletcher Carney:

> (9)   Little deed of *kindness*, little words of love
>   Help to make the Earth happy, like the heaven above.[8]

It is also reflected in the following Wordsworth quote.

> (10) That best portion of a good man's life,
>   His little, nameless, unremembered acts
>   Of *kindness* and of love.[9]

This can be captured in the definition by modifying those components pro-

posed in 4.1.2 above to indicate that what one wishes to do for another is not a "big" thing. This is, of course, "big" in a metaphorical sense, but this seems to accurately capture the notion implied here.

> X often thinks something like this about people:
>   I want to do something good for this person
>   this is not a big thing

4.1.4   Doing something which one is not obliged or expected to do

Another qualification that needs to be made is that something done out of *kindness* cannot be something one is expected or obliged to do. It implies that one does something for another simply because one wants to. Thus, *kind* would not tend to be used in the following context:

(11) ?The bus driver *kindly* sold me a ticket.

Selling tickets is part of the driver's job, something they are obliged to do, and therefore something that would not normally require any *kindness* on their part. Note that example (11) could be used in a context where there was some reason why the driver should not sell the speaker a ticket. Compare this with the following sentence:

(12) The bus driver *kindly* helped me load my bags onto the bus.

This would be natural, because helping passengers load their bags onto a bus is not one of the driver's prescribed duties, but something they may choose to do if they wish to.

The following is one final example, to illustrate this point, taken from a scene in a novel in which a woman has fallen through her seat in the cinema, and cannot get herself out.

(13) A lady *kindly* went to fetch the manager, ...
> MS: 37

There is clearly no obligation for a patron in the cinema to help another, and thus doing so can be described as *kind*.

Related to this is the fact that we do not usually use *kind* to describe the actions of people with whom one is in a close relationship. Thus, we do not usually say that spouses are *kind* to each other (?My husband is very *kind* to me), or that parents are *kind* to their children (?She's a very *kind* mother). I submit that this is because implicit in such close relationships are a number of obligations and expectations, and this then renders the use of *kind* strange in such a context.

The notion of a *kind* act being performed because one wants to do something for the benefit of another, and not because one feels obliged in any way to do so, can be captured in the definition of *kind* with the following components:

> X often thinks something like this about people:
> I think I can do something good for this person
> I know I don't have to do it
> I want to do it

4.1.5   Complete definition of *kind*

The components so far presented can now be combined to form the complete definition of *kind*.

> X is *kind*:
> X often thinks something like this about people:
> I don't want anyone to feel something bad
> I think I can do something good for this person
>   this is not a big thing
> I know I don't have to do it
> I want to do it
> X does something because of this

Being *kind* entails performing some action, that one perceives, and hopes, will be to the benefit of another. It implies that one does not want others to suffer in any way, and so, by implication, that one would not do anything to cause another to feel bad (unless it was for their ultimate good, as discussed above in reference to the expression "to be cruel to be *kind*"). One believes one can do some small thing for the benefit of another person, is aware that there is no obligation to do this, but still wishes to do it, and because of this, does actually do something.

I shall now go on to look at *considerate* and *thoughtful*, carrying out a similar analysis of these words to establish their meanings, and to highlight some of the ways in which these three words differ from each other, and how they are similar.

4.2   Considerate

4.2.1   Not wanting one's actions to cause others to feel bad; No notion of doing something for the benefit of others

Considerate is similar to *kind* in that both imply a wish that others do not feel

bad. It differs from *kind*, however, in that this is specifically related to not wanting others to feel bad as a result of one's own actions. Related to this is the fact that, unlike *kind*, *considerate* carries no implication that one wants to do something for others. Below are some examples to illustrate these points.

> (14) Mary is very *considerate* of her neighbours, and if she has a party she always makes sure it's not too noisy after about 11, so as not to disturb them.

This implies that Mary thinks of her neighbours, and of how her actions may negatively affect them, and she tries to avoid these negative effects. She takes the feelings of her neighbours into "consideration", and acts so as not to harm them in any way. Note that *kind* would be strange in this context, which I believe is because *kind* implies specifically doing something for the benefit of another, while the focus of *considerate is* on avoiding something that may harm another. This can be demonstrated by the following example where *kind* and not *considerate* would be natural.

> (15) Mary is very ?*considerate* of (*kind* to) her neighbours, and always waters their garden, and feeds and walks their dog when they go away.

Consider also the following example, which is from a sign seen in a restaurant in Australia.

> (16) We do not ban smoking here, but please be *considerate* of other diners.

This is asking smokers to think about the effect their smoking will have on other diners, and not to smoke if they judge this effect to be negative. It does not suggest that they do something for the benefit of others, but rather that they refrain from doing something that could cause discomfort to others. Thus, the focus is on not doing something bad, rather than on doing something good. Note that neither *kind* nor *thoughtful* could be used here, which can be explained by the fact that both of these words refer to wanting to do something good for others.

Having shown that *considerate* implies not wanting to do anything that may harm others, I shall now look more closely at whether or not it involves doing something for others. It appears so far that the focus of *considerate* is on *not* doing something, rather than on actually doing something. The fact that it need not involve specifically doing something can be shown clearly with the following example, which was discussed in the preceding section, and shown to be unnatural with *kind* (example (4)), but which works well with *considerate*.

(17) He won't do anything for you, but he is *considerate*.

Thus, doing something for others is not an essential element of the meaning of *considerate*. That is not to say that doing something for others is incompatible with *considerate*, and certainly not that being *considerate* means one does not do anything for others. A *considerate* person may well do things to benefit others, but this is not part of the invariant meaning of this word itself.

While actually doing something is not an essential element of the meaning of *considerate*, not doing something does seem to be essential. Thus if one does not want to do anything that may harm another (such as making loud noise at night, or smoking while others are eating), but for whatever reason does do this anyway, this cannot be described as being *considerate*. This can be seen in the unacceptability of the following example.

(18) ?Mary is very *considerate* of her neighbours, and she doesn't like to disturb them with loud music at night, but when she has parties, they're always noisy till all hours of the morning.

To be *considerate*, one must specifically refrain from doing what one has perceived would be harmful to another, not merely be aware that it could be harmful.

On the basis of the discussion so far, it seems the definition of *considerate* must include something along the following lines:

X is *considerate*:
  X often thinks something like this about people:
    if I do something (W), this person may feel something bad
    I don't want this
    if this person could feel something bad
      I will not do this (W)

These components capture the notion that a *considerate* person is someone who thinks about the potential negative effects their actions may have on others (or rather, the effects of a specific action, denoted here with (W)), and avoids doing that thing.

4.2.2 Not wanting others to feel bad thinking that one doesn't care about them

While it is evident that *considerate* implies not wanting to do something that may cause another to feel "something bad", there needs to be some kind of specification of what kind of bad feelings could fit into this context. For example, acting in a certain way so as not to cause others to become angry or indig-

142   Catherine Travis

nant (two emotions that certainly classify as "feeling bad") would not be described as being *considerate*. Thus being careful not to break something that belongs to another person, would not be described as being *considerate*, and nor would keeping a secret someone has told you. We have already seen that being *considerate* can be not disturbing one's neighbours with loud noise, or refraining from smoking while others are eating. Another example would be being quiet in an area where silence is for whatever reason desired. Thus, a librarian could say to people making noise in a library:

> (19) Others are trying to read. Could you please be more *considerate*, and quiet down a little.

It could also mean letting your partner know if you're going to be late home, when they may be expecting you.

> (20) She's *considerate* enough to let me know if she won't be home for dinner, so that I don't worry about cooking for her.

All of these contexts imply that a person could be made to feel something bad if the actor had not thought about them before doing something (smoking, making noise etc.). Specifically, it seems to be the case that the kind of bad feelings they may experience are related to what they would think that person has thought about them. That is, they may feel bad thinking that their own feelings have not been taken into "consideration". Thus, neighbours who have to listen to loud parties at night suffer because they cannot sleep, but also because those having the party have not thought about how they would be affected by their actions. Diners in a restaurant suffer the smoke of others' cigarettes, but they also suffer because the smoker has not thought about them. People trying to read where others are talking, suffer in that their concentration is interrupted, but also because the talkers have not thought about how they would feel. And finally, someone who prepares dinner for another who does not turn up, has perhaps gone to unnecessary trouble, but they also suffer because that person has acted without thinking about them. It seems that the element each of these examples has in common is a bad feeling caused by the fact that another has acted without thinking about them, in a situation where they should have done so.

If we now consider the examples mentioned above where *considerate* is unacceptable, we can find further evidence for this. Being careful not to break something that belongs to another, would not be described as being *considerate* because we would assume that the reason one is careful with others' possessions is not because one doesn't want that person to think that one has not thought about their feelings, but because of an understanding that people don't like to have their things broken. Similarly, people do not keep secrets others tell them because they have thought about how that person would feel if they told

others, but out of respect for the notion of a secret, or for the friendship they have with the person who told them the secret. On the other hand, if someone told you something without telling you that it was a secret, and you told it to someone else who wasn't meant to know, then presumably that could be described as inconsiderate, with the implication that you did not think about how that person would feel if you passed on what they had told you.

Note also that we do not use *considerate* to describe the way we interact with animals. Acting in a way so as not to harm animals is described as being *kind* to them, not as being *considerate* of them. This can be explained by this element of meaning I propose for *considerate*, namely that one does not think of what the animals will *think* if one behaves in a certain way. Even Dr. Doolittle, who certainly made an effort to understand the animals, and would not have wanted to cause them harm, is not described as being *considerate* of them.

It seems, then, that thinking about what another will think about oneself, and consequently feel, as a result of one's actions is inherent in the meaning of *considerate*. What a *considerate* person thinks is that they don't want others to feel bad thinking that they have done something without thinking about how that would make them feel. The components tentatively proposed in 4.2.1, must be modified to incorporate this, and this can be done in the following way.

> **X often thinks something like this about people:**
> **if I do something (W)**
> **this person may think: X doesn't think about what I could feel**
> **because X does W**
> **because of this, this person may feel something bad**
> **I don't want this**

4.2.3   Notion of "consideration"

As is reflected in the stem "consider", being *considerate* of others implies that one reaches an understanding of them through thinking about them. A *considerate* person is not someone who is intuitively aware of the potential negative effects of their actions, but someone who makes a conscious effort to think of, and understand, how others may feel.

A person who is not *considerate*, is someone who does not think of how they may make others feel through their own actions. One could therefore ask someone to be more *considerate*, by asking that they think about one before they do something, as shown in the following example.

> (21) Just be a bit more *considerate* — think about how I feel when you do things like that.

Note also that being *considerate* is incompatible with not thinking of others, as is illustrated by the unacceptability of the example (22) below.

(22) *He doesn't think about how I'm affected by what he does, but he is *considerate*.

We therefore need to include in the definition a notion that being *considerate* implies some conscious thought-process to achieve an understanding of another. This can be done with the following components.

X often thinks something like this:
   before I do something (W)
   I want to think about this person
   I want to know if this person will feel something bad
      because I do this (W)

4.2.4   Complete definition of *considerate*

We can now compile the complete definition of *considerate* based on the three notions thus far looked at individually.

X is *considerate*:
   X often thinks something like this about people:
   if I do something (W), this person may think:
      X doesn't think about what I could feel
         because X does W
   because of this, this person may feel something bad
   I don't want this
   because of this, before I do this (W)
   I want to think about this person
   I want to know if this person could feel something bad
      because I do this (W)
   if this person could feel something bad
      I will not do this (W)

Considerate implies thinking about one's actions and the potential effects they have on others, in terms of what one performing that action will make that person think about one's attitude towards them. That is, whether they will think that one has acted without thinking about how they would be caused to feel as a result of that action. The main focus of *considerate* is on not doing something that may harm another. The notion of wanting to do something for the benefit of another is not an essential element of its meaning. This is the most obvious difference between this word and both *kind*, as has been shown in the preceding discussion, and *thoughtful* (although not the only one). I shall now look at the meaning of *thoughtful*, and will attempt to show that, similar to *kind*, it is focussed on what one can do for the good of another, and unlike *considerate*, not on the potential negative effects of one's actions.

4.3   Thoughtful

4.3.1   Polysemy of *thoughtful*

Before looking at the meaning of *thoughtful*, the polysemy of this word must be considered. *Thoughtful* appears to have two meanings: one (that which concerns us here) is that used to describe a character trait that is manifested in interpersonal relations; the other refers not to one's attitude to people, but, for example, the approach one takes to an issue, or an argument. Thus, a newspaper might produce a "*thoughtful* editorial", or a speaker may present a "*thoughtful* discussion of an issue". There are a number of points that justify this polysemy, which I shall briefly discuss below.

Firstly, *thoughtful* as used to describe a way of relating to others (which for the moment I shall call *thoughtful* 1) contrasts with the adjective *selfish*, while the other use of *thoughtful* (which I shall call *thoughtful* 2) does not. Thus a "*thoughtful* 1 person" cannot be selfish, but a "*thoughtful* 2 writer", for example, can be. Secondly, *thoughtful* 1 refers to a person's character, and thus implies a degree of permanence, while *thoughtful* 2 can be used to refer to a temporary state. Thus, one can be *thoughtful* 2 for a moment, but one cannot be *thoughtful* 1 for a moment. And thirdly, the syntactic frames in which these words are used provide further evidence for their polysemy. Only *thoughtful* 1 can be used in the frame: "It was *thoughtful* of him to do X". If we were to describe the writing of a letter as *thoughtful* (a context in which both *thoughtful* 1 and *thoughtful* 2 would be possible), in this syntactic frame, this would have to be understood as *thoughtful* 1. Thus, "It was *thoughtful* of her to write the letter" could only imply that the letter was written with consideration of how that could benefit someone, and was written for that purpose. It cannot mean that the writer put a lot of thought into the ideas that made up her letter.

Both meanings imply some kind of "deep thinking", but the object of that thinking is quite different in each case. Having established that polysemy does indeed exist in this case, I shall now disregard *thoughtful* 2, and will focus only on *thoughtful* 1, which I shall call *thoughtful*.

4.3.2   Not wanting others to feel bad;
   Believing one can do something to prevent them from feeling bad

I have argued that *kind* implies a general wish for others not to feel bad, and that *considerate* implies not wanting others to feel bad specifically as a result of one's actions. *Thoughtful* also implies that one doesn't want others to feel bad, but this seems to be related to specific situations, and includes a notion that one can do something to help a person who is in a potentially harmful situation. Thinking about someone in such a situation, and wanting to do something to help them, or something to prevent that situation from developing, is described

as being *thoughtful*. Consider the following example, from a book about how to handle grief, and what one can do for those who are grieving.

> (23) This is a time when *thoughtful* friends should rally in a co-ordinated way in the areas of food, errands, child care, hospitality for relatives and chores.
>
> MM: 44

This implies that friends of the bereaved should think of the bereaved's situation, and try to improve it, determining how they are able to help that person, and doing this. Note that *considerate* would be unnatural in such a context. This can be explained by the fact that there is no implication here of concern over the potential negative effects of one's actions on another, but rather over what one can do for the good of another. *Kind*, on the other hand, could be used here, because, as we have seen, *kind* does imply doing something to benefit another. Consider also the following example, again about how to help people who are grieving.

> (24) Obviously there is no single dramatic gesture or pearl of wisdom that will dissolve the ache, but there are many acts of *thoughtfulness* that can convey your concern and help to soften the blow that a friend or loved one has suffered.
>
> GM: 72

"Acts of *thoughtfulness*" implies things that one can do to help alleviate another's suffering. Note that we could not talk of "acts of *consideration*", which is natural in view of the fact that *considerate* does not imply doing something for others, while we could talk of "acts of *kindness*", as *kind* does involve doing something for others.

The notion of thinking of the situation others are in, not wanting that situation to cause them to feel bad, and wanting to do something to help them can be captured with the following components.

> X is *thoughtful*:
>   **X often thinks something like this about people:**
>   **if something like this happens to a person**
>   **this person can feel something bad**
>   **I don't want this**
>   **I think this person will not feel it if I do something (W)**
>   **I want to do something (W) because of this**

This clearly marks one distinction between *thoughtful* and *kind*, the latter being defined as thinking "I don't want anyone to feel something bad; I think I can do something good for this person", which is much more general than what has

been proposed here for *thoughtful*. It also shows one way it differs from *considerate*, which refers specifically to consideration of the potential negative effects of one's actions on another, and avoiding doing anything that may have such effects.

4.3.3   Wanting others to feel good

We have seen that *thoughtful* implies wanting to alleviate another's pain or suffering. It can also be used to mean wanting to do something for the benefit of another.

This element of the meaning of *thoughtful* is evident from its use in modifying some nouns. For example, it can be used to describe a present ("a *thoughtful* present"), or a letter ("a *thoughtful* letter"). The components proposed above do not account for such a use, where there is no implication of not wanting others to feel bad. A "*thoughtful* present" is not one given because one doesn't want others to feel bad, but because one wants them to feel something good. It is one into which the giver has put a lot of thought, and as a result of this, has found something they believe the recipient will like, or that will be of benefit to them in some way. Something similar is implied by "a *thoughtful* letter".

Consider also the following example, where the writer is describing a woman's impression of a new friend of hers in relating to various people.

> (25) She was impressed with the *thoughtfulness* and hospitality Sam showed his guests, ...
>
> KA: 84

This does not imply that Sam treated his guests in a way to minimise any potential hardship they might undergo, but treated them so as to maximise their comfort, to make them feel good.

It is important, then, to include in the definition that not only does a *thoughtful* person not want others to feel bad, they want to do something to cause them to feel good. This can be captured with the following components.

> **X often thinks something like this about people:**
> **I think this person will feel something good**
> **if I do something (W)**
> **I want to do something (W) because of this**

This is similar to what has been proposed in the definition of *kind*, namely that one thinks one can do something for the good of another, and one wants to do it. It differs from *considerate* which, as has been discussed, does not include a notion of wanting to do something good for another.

4.3.4   Doing something for others

I have argued that performing some action is an essential element of the meaning of *kind*, and that not performing some action is essential to *considerate*. *Thoughtful* seems to be similar to *kind* in this regard, in that it does involve doing something. The fact that we can talk of "acts of *thoughtfulness*" (example (24)), is one piece of evidence supporting this, while, as mentioned, we cannot talk of "acts of *consideration*". Also consider the following sentence, shown to be unacceptable for *kind* (example (4)), but acceptable for *considerate* (example (17)), which is questionable for *thoughtful*.

(26) ?He won't do anything for you, but he is *thoughtful*.

It seems that being *thoughtful* means that not only does one think about what would be good for others, or what one can do to help others in a potentially harmful situation, but one does do something for another as a result of what one has thought about them. Remembering to give a present on someone's birthday and choosing something particularly appropriate may be described as *thoughtful*, but simply thinking of a wonderful present for someone, but not actually giving it to them, would not be. The following component must therefore be added to those proposed above (which I shall not reproduce here).

X does something because of this

4.3.5   Not a "big" thing

Another element that remains to be captured in the definition of *thoughtful* is that, like *kind*, what one wants to do for another is not a "big" thing; it is rather a small act that is fairly easy to perform. Thus, while helping a sick friend with errands, child care etc. can be described as *thoughtful* (example (23)), taking a week of work to help them would not be.

(27) *He *thoughtfully* took the week off work to look after his sick friend.

It seems that this is because taking a week off work requires too much effort to be *thoughtful*, and is therefore too "big" a thing. *Thoughtful* is reserved for "smaller" things, such as remembering people's birthdays, calling people on special occasions, keeping people informed of news they may want to hear, and so on.

The component given below must be included in the definition, following those components outlining that one wants to do something to prevent another from feeling bad in a given situation, and that one believes one can do something good for them (not reproduced here).

X often thinks something like this about people:
  I want to do something because of this
    this is not a big thing

4.3.6   Notion of "thought"

The notion of a *thoughtful* act being something into which one has put some "thought" is reflected in the examples given above, of a "*thoughtful* letter" or a "*thoughtful* present". These imply things that one has spent some time thinking about (as opposed to an idea that automatically popped into one's head) and because of that, has been able to work out what would be good for the recipient.

This notion of thinking involved here is also reflected in the morphology of the word, being built on the stem *thought*, as was also discussed for *considerate*.

This notion of "hard thinking" can be captured in the definition with the following component.

X often thinks something like this about people:
  I want to think about this person before I do something

Note that this is similar to what was proposed for *considerate*.

4.3.7   Complete definition of *thoughtful*

On the basis of this discussion, I propose the following definition of *thoughtful*.

X is *thoughtful*:
  X often thinks something like this about people:
    If something like this happens to a person
      this person will feel something
    I don't want this person to feel something bad
    I think this person will not feel it if I do something (W)
    I think this person will feel something good
      if I do something (W)
    I want to do this (W) because of this
      this is not a big thing
    I want to think about this person before I do it
  X does something (W) because of this

A *thoughtful* person is someone who thinks of the situation others are in, and of what they can do for that person, either to minimise their discomfort, or to

150   Catherine Travis

maximise their comfort in some way. It is someone who goes to some effort to think of small things they can, and want to do for another, and who does do those things.

5.   Conclusion

I have presented here definitions for three common and closely related English words referring to positive character traits. Although these three words appear on the surface to mean something very similar, a detailed analysis of usage examples has revealed fine differences between them. I have argued that *kind* is focussed on not wanting others to feel bad, wanting to do good things for them, and on actually doing something. This explains its use in expressions such as "So *kind* he wouldn't hurt a flea", and "Would you be so *kind* as to ...", where neither *considerate* nor *thoughtful* are used. *Considerate* is focussed on not doing something that could hurt another by demonstrating that one has not thought about that person. Thus, it is *considerate* to not play loud music at night, or to not make noise while others are reading, contexts where neither *kind* nor *thoughtful* are used. *Thoughtful* is focussed on thinking of the specific situation that another is in, and what one can do for them in that situation, and involves doing something for that person. Thus, it is *thoughtful* to help sick friends in small ways, and to remember people's birthdays, for example, while such acts would not be described as *kind* or *considerate*.

I hope to have shown, in the course of this analysis, that the Natural Semantic Metalanguage is an extremely useful theoretical framework for defining words. It facilitates the exhaustive explication of word meaning, and thus allows for differences and similarities between related words to be clearly and accurately stated. As well as helping native speakers better understand their own use of words, this theory has wide-reaching applications for second language learners, and for cross-cultural analysts, and thus has a great deal to offer the field of lexicography.

Notes

1.   I would like to express my gratitude to Anna Wierzbicka, for her inspiration and support with this paper, from its inception to the final stage. I would also like to thank Nick Enfield for his helpful comments and thought-provoking discussion.
2.   The polysemy posited for *considerate*, however, seems to apply more to the noun *consideration* and the verb *consider* than it does to the adjective *considerate*. In the case of *thoughtful*, the validity of (2), and whether this really is a separate meaning from (1) and (3), is unclear. The polysemy evident in *thoughtful* is discussed in section 4.3.1.
3.   See Goddard (1994: 7-14) for a more detailed and theoretical discussion of the principles underlying NSM.

4.   See Wierzbicka (1985: 193 ff) for a discussion of the meaning of *salt*.

5.   See Travis (to appear), Wierzbicka (1991a, 1991b, among others) for examples of such analyses.

6.   I do not wish to suggest that *kind* means the same as its derivations, such as *kindness* and *kindly*, aside from belonging to a different word-class. I do believe, however, that there is a common core to these derivations, and thus will not restrict my data to *kind* alone. It is the common core that I am attempting to capture in my definition of *kind*, and also for *considerate* and *thoughtful*.

7.   I have tried, where possible, to use naturally occurring examples, and have indicated the source and page numbers of such examples as done here. All come from written sources; the initials are the author's initials and the numbers represent the page number of the quote. The full reference is given in "Sources Cited" under "References". Those that are not marked in this way are constructed examples.

8.   Quoted in Stevenson (1946: 1036).

9.   Quoted in Stevenson (1946: 1037). From Wordsworth: "Lines composed a few miles above Tintern Abbey".

References

Australian Concise Oxford Dictionary. 1987. Melbourne: Oxford University Press.

Boguslawski, A. 1970. On Semantic Primitives and Meaningfulness. Greimas, A.J., R. Jakobsen and M.R. Mayenowa (Eds.). 1970: 143-152.

Canberra Times, 17 August 1992: 10.

Collins English Dictionary (Australian Edition). 1991. Sydney, Australia: Harper Collins Publishers.

Goddard, C. 1989. Issues in Natural Semantic Metalanguage. *Quaderni di Semantica* 10(1): 51-64.

Goddard, C. 1994. Semantic Theory and Lexical Universals. Goddard, C. and A. Wierzbicka (Eds.). 1994: 7-29.

Goddard, C. and A. Wierzbicka (Eds.). 1994. *Semantic and Lexical Universals.* Amsterdam: John Benjamins.

Greimas, A.J., R. Jakobsen and M.R. Mayenowa (Eds.). 1970. *Sign, Language and Culture.* The Hague: Mouton de Gruyter.

Stevenson, B. 1946. *Stevenson's Book of Quotations.* London: Cassell's Publishing.

Travis, C. 1992. *How to be Kind and Considerate in Japanese.* Unpublished BA Honors Thesis. Australian National University.

Travis, C. To appear. *Omoiyari* as a Core Japanese Value: Japanese Style Empathy? Athanasiadou, A. and E. Tabakowska (Eds.). *Speaking of Emotions: Their Conceptualization Expression.* CLR Series. Mouton de Gruyter.

Wierzbicka, A. 1972. Semantic *Primitives.* Frankfurt: Athenaum.

Wierzbicka, A. 1980. *Lingua Mentalis: The Semantics of Natural Language.* Sydney, Australia: Academic Press.

Wierzbicka, A. 1985. *Lexicography and Conceptual Analysis.* Ann Arbor: Karoma Publishers, Inc.

Wierzbicka, A. 1991a. *Cross-Cultural Pragmatics: The Semantics of Human Interaction.* Berlin: Mouton de Gruyter.

Wierzbicka, A. 1991b. Japanese Key Words and Core Cultural Values. *Language in Society* 14(4): 491-513 .

Wierzbicka, A. 1996. *Semantics: Primes and Universals.* Oxford: Oxford University Press.

Sources Cited

GM   **Gordon, M.** 1988. How to Comfort Those Who Grieve. *Readers Digest*, March: 68-64.

KA   **Kiiman, A.** 1978. *Crisis.* New York: Holt Rinehart and Winston.

MM   **McKissock, M.** 1985. *Coping with Grief.* Sydney, Australia: Australian Broadcasting Commission.

MS   **Morgan, S.** 1987. *My Place.* Perth, Australia: Fremantle Arts Centre Press.

UE   **Upward, E.** 1990. The Theft. Lefanu, S. and S. Hayward (Eds.). 1990. *Colours of a New Day: Writing for South Africa:* 313-328. London: Penguin Books.

Afrikaans, American and British Models for South African English Lexicography: Racial Label Usage[*]

M. Lynne Murphy, *Department of Linguistics,*
University of the Witwatersrand, Johannesburg, South Africa
and Department of English, Baylor University, Waco, USA

Abstract: This article examines the treatment of racial labels in monolingual English diction-
aries of South Africa. Considering past controversies regarding racist language in Afrikaans dic-
tionaries and considering the changing role of English in democratic South Africa, we can expect
that English dictionaries will be more carefully scrutinized in future for potential offence. With the
boom in the South African lexicography industry, now is a prime time to reflect on the issues
involved and suggest courses of action. This article places South African English dictionary tradi-
tions in a national and international context, by comparing the traditions and the roles of the lan-
guages in Afrikaans, British and American lexicographical traditions. While South African English
lexicography is rooted in the British tradition, its possible evolution on the American model is
demonstrated, and thus the role of the dictionary in a postcolonial community is briefly discussed.

Keywords: LEXICOGRAPHY, DICTIONARY, DEFINITION, SOCIAL LABELING, RACE,
RACISM, ETHNICITY, ENGLISH, SOUTH AFRICA, AFRIKAANS, EPITHETS, USAGE LABELS

**Opsomming: Afrikaanse, Amerikaanse en Britse modelle vir die Suid-
Afrikaanse Engelse leksikografie: Gebruik van rasse-etikette.** Hierdie artikel
ondersoek die hantering van rasse-etikette in eentalige Engelse woordeboeke in Suid-Afrika. As
polemieke van die verlede oor rassistiese taal in Afrikaanse woordeboeke oorweeg word, sowel as
die veranderende rol van Engels in demokratiese Suid-Afrika, kan verwag word dat Engelse
woordeboeke in die toekoms noukeuriger ondersoek sal word vir moontlike kwetsing. Met die
ontploffing in die Suid-Afrikaanse leksikografiese bedryf is dit nou die regte tyd om die betrokke
vraagstukke te oordink en om handelswyses voor te stel. Hierdie artikel plaas die Suid-Afrikaanse
Engelse woordeboektradisies binne 'n nasionale en internasionale konteks deur die tradisies en
rolle van die onderskeie tale in die Afrikaanse, Britse en Amerikaanse leksikografiese tradisies te

---

vergelyk. Terwyl die Suid-Afrikaanse Engelse leksikografie gewortel is in die Britse tradisie, word sy moontlike ontwikkeling volgens die Amerikaanse model aangetoon, en sodoende word die rol van die woordeboek in 'n postkoloniale gemeenskap kortliks bespreek.

Sleutelwoorde: LEKSIKOGRAFIE, WOORDEBOEK, DEFINISIE, SOSIALE ETIKETTE-RING, RAS, RASSISME, ETNISITEIT, ENGELS, SUID-AFRIKA, AFRIKAANS, BENOEMINGS, GEBRUIKSETIKETTE

This essay contrasts the practices of and attitudes toward racial label treatment in South African English, Afrikaans, and other English dictionaries. As I discuss below, the comparison of these dictionary traditions not only lends insight into the peculiarities of South African English lexicography, but also indicates new directions that South African English lexicography may take, given similarities in the social situations of Afrikaans, American English, and South African English.

The treatment of racial labels such as **black**, **African**, **white**, **Coloured**, and **Asian** in English dictionaries for the South African market is discussed elsewhere (Murphy, forthcoming). In general, it was found that dictionaries of South Africanisms, including Branford and Branford's *A Dictionary of South African English* and the new *Dictionary of South African English on Historical Principles*, capture the South African senses of racial labels and give the most insightful commentary on their usage. However, these dictionaries are not widely used by average dictionary users. Instead, the South African English market depends upon dictionaries that are either produced for the British market (such as the Oxford and Collins concise dictionaries) or derivatives of these dictionaries, such as *The South African Pocket Oxford Dictionary*. These dictionaries typically fail to represent South African senses and usage of general English words like **black** and **Bantu**.

The changing linguistic, social, and educational situations in South Africa predict a boom in English lexicography for a South African audience, as does the increasing computerization of lexicographical databases. Due to computerization, dictionaries for specific regions or purposes can be extracted and edited from the databases of major international publishers, and we can already see this trend with the publication of *Chambers-Macmillan South African Dictionary for Junior Primary* and *The South African Oxford School Dictionary* and the foreseen preparation of South African editions of general English dictionaries by the Dictionary Unit at Rhodes University (Penny Silva, personal communication). As a lexicographical tradition for South African English emerges, the question is whether it will mimic traditions available in South Africa or the traditions of other Englishes. Such traditions provide stark contrasts in the treatment of racially sensitive words and possible models for dictionary making in postapartheid South Africa.

This paper relies on basic work on South African English racial labels

(reported in Murphy, forthcoming) in order to compare South African English dictionary treatments of racial terminology to that in Afrikaans, British and American English dictionaries. The focus of comparison is on the inclusion and usage descriptions of racial labels. Three South African English dictionaries for adults could represent the nascent South African English lexicographical tradition: *A Dictionary of South African English*, 4th edition (henceforth DSAE, 1991), and *A Dictionary of South African English on Historical Principles* (DSAE-HP, 1996), and *The South African Pocket Oxford Dictionary* (SAPOD, 1987). However, the first two of these are not dictionaries of English, but rather dictionaries of South Africanisms in English. The last is an adaptation of a British pocket dictionary, and, as a pocket dictionary, does not necessarily provide a model for standard desk dictionaries. In contrast, the dictionaries that South African English speakers actually use, are represented by the *Collins Concise English Dictionary*, 3rd edition (CCED, 1992), and the *Concise Oxford Dictionary of Current English*, 8th edition (COD, 1990). These two dictionaries are written for a primarily British audience, but are the two most popular English desk dictionaries in South Africa. Thus, they are part of the standard against which English-speaking South African consumers will measure South African English dictionaries. The remainder of this essay looks first at the indigenous model for lexicography provided by Afrikaans and then the models provided from abroad, by the British and American lexicographical traditions.

A South African Model: Afrikaans Lexicography

Compared to Afrikaans lexicography, South African English lexicography is a new field. While monolingual Afrikaans dictionaries have been published in South Africa since at least 1926 (Gouws 1986), the first major dictionary of South African English (DSAE, first edition, 1978) was limited to South Africanisms, and a general English dictionary (SAPOD) was not produced until 1987 (Béjoint 1994). Lexicographical practice in Afrikaans differs from that in English to the extent that the languages hold different social positions in (and out of) South Africa, and to the extent that they reflect very different cultural and communicative norms.

Afrikaans lexicographers have been at the forefront of developing a new South African model for lexicographical policy-making. Evidence for a South African (Afrikaans) model for English lexicography comes from two recent sources: developments in AFRILEX, the relatively new association for lexicography in Africa, and an articulated policy strategy for treating taboo items in volumes of the *Woordeboek van die Afrikaanse Taal* (WAT). Lexicographical planning and policy-making are high priorities in this new model.

AFRILEX is considered relevant to the Afrikaans model since its founding sixteen-member board included no English lexicographers, and an over-representation of Afrikaans lexicographers.[1] The first item in the call for interest in

the formation of this association is the establishment of "a national policy for lexicography" (letter, Mariëtta Alberts and Daan Prinsloo, 10 March 1995) and in its first year of existence, the AFRILEX board pushed forward a National Lexicography Bill, whose purpose is to set up official dictionary units for the eleven official languages of South Africa.[2] This is in contrast to similar organizations elsewhere (such as the European Association for Lexicography and the Dictionary Society of North America), whose foremost stated aim is scholarly exchange. It is tempting to interpret the perceived need for planning as a reflection of an Afrikaner cultural rejection of ambiguity and preference for hierarchical organization, which is commonly contrasted to South African English interactional styles in the South African social psychology and cross-cultural communication literature (see, e.g., Kinloch 1985, Louw and Foster 1992).

This combination of needs for decisive and authoritative practice can be seen in a recent controversy concerning the status of racial insults in the largest Afrikaans dictionary project, the *Woordeboek van die Afrikaanse Taal* (WAT). After heated discussion at a multiracial seminar convened on the topic, the editors of the WAT released for comment an eight-paged "Policy for the Treatment of Insulting and Sensitive Lexical Items in the *WAT*" that was intended to represent, through compromise, the interests of ingroup, outgroup, and academic interests (Bureau of the WAT 1994). This document proposed that "racist" headwords be listed but not defined or otherwise discussed in the printed version of the WAT. (People interested in such definitions would have to contact the Bureau of the WAT for access to the unpublished electronic version of the manuscript.) The editors maintained that although the racial climate in South Africa is changing, "the inclusion of racist lexical items in the WAT would be undesirable in this period of reconciliation. Such a move would not only hinder reconciliation, but would at the same time result in greater alienation" (Bureau of the WAT 1994: 1, Harteveld and Van Niekerk 1995: 252). (The editors fail to make clear whether it is the lexicographers or the referents of the racist terminology who would be alienated by this move.) The editors received a large number of responses from local and international commentators, some of whom protested that the policy ran counter to basic lexicographical practice and that such censorship might hamper, rather than aid, reconciliation. The seventeen-paged revised policy (Harteveld and Van Niekerk 1995) differs in its treatment of "racist" words, in that a nonracist synonym is given in the printed version. Thus "wholly racist lexical items" are treated differently from other lexical items in that the definition is limited to a single near-synonym, with no semantic oppositions or citations given. Compounds or idioms containing "wholly racist" lexical items are explained, but not permitted synonyms, antonyms, references, or illustrations. The "wholly racist" words are contrasted to "partially racist" words, which are treated in greater depth. So, for example, while **kaffir** is considered to be "wholly racist", **meidjie** is considered "partially racist", since it is often used as a term of endearment (in which case, it seems that it might also be partially sexist). Throughout the policy, the editors write

as if they consider the words themselves to be racist, rather than the use (or users) of those words.[3] While a word may very well be taboo, as an inanimate, arbitrary thing, it cannot have racist intentions. Thus, the decision as to whether or not to include an item that is potentially insulting, is based upon outgroup usage of the term, since ingroup usage is not necessarily racist.[4]

Contrast with South African English lexicographical practices is evident. None of the five English dictionaries considered here have any blanket policies against printing or defining potentially offensive words, although the number of such items in any dictionary depends in large part upon the dictionary's purpose. Thus, English racial labels are included as dictionary headwords if they suit the dictionary's general criteria for inclusion (relative frequency, non-slang usage, etc.), whereas Afrikaans racial labels are required to fit an extra criterion, being noninsulting. While South African English dictionaries do label potentially insulting material as "*derogatory*" and/or "*offensive*", such measures have been deemed insufficient for the WAT.

The WAT has good reason for concern. Racially provocative language in Afrikaans dictionaries has historically caused many problems for their publishers, resulting in boycotts and book burnings (Hauptfleisch 1993) and protests from academics (e.g., Links 1991). Historically, the WAT used racist metalanguage in its definitions, for example, using the offensive term **kaffer** in definitions of compound terms like **Kafferhond**, which was defined in a 1968 volume of the WAT as "Dog belonging to a Kaffir" (Hauptfleisch 1993: 126). However, although this problem has been attended to, the continued furor and defensiveness over racist language in Afrikaans is no doubt related to the perception (in some circles) that Afrikaans is a "racist language". This sentiment stems mostly from the salient rôle of Afrikaner individuals in forming racist policies and the position of Afrikaans in the maintenance of such policies. Thus, Kinloch (1985) reports a number of studies that have shown that Black, Indian, and increasing numbers of Coloured South Africans have far more negative attitudes toward Afrikaners than toward English-speaking White South Africans. Since the Afrikaans language serves as a symbol of the Afrikaner people, non-Afrikaners sometimes consider it an inherently racist language or "the language of the oppressor" (Benjamin 1994). Resentment toward the Afrikaans language has greatly contributed to protests against racially insulting language in Afrikaans dictionaries, and sensitivity to these resentments and past and potential protests has led Afrikaans lexicographers to react supercautiously by usual lexicographical standards.

This very South African lexicographical phenomenon leads to the question: Does the Afrikaans experience make predictions for the future of South African English lexicography? Whereas Afrikaans was, until recently, the language of political power, its power is diminished in the new *status quo*, and English's status as the language of economic power only increases with the reintegration of South Africa into the world community. As the institutions associated with the Afrikaans language lose power or switch to other lan-

guages (especially English), English-medium institutions become the objects of protest. For example, since South Africa's first democratic elections, the traditionally liberal, English-medium universities were the first targets of loud and violent protests on matters including the nature of the curriculum and the racial composition of the student body and staff. Furthermore, white English-speaking liberals have repeatedly been called "racist" in the opinion pages of the popular press, call-in radio, and television talk shows. While there are still small conservative Afrikaner organizations and individuals pushing for racial separation and other racist policies, it is the English-speaking liberals whose politics are now questioned publically. As English becomes more and more entrenched as a language of education, commerce, and politics, we can expect its dictionaries to be scrutinized more carefully for words, passages, or sentiments that are potentially racially or politically offensive.

However, it seems unlikely that South African English dictionaries will be subject to the same degree of (self-)censorship as Afrikaans dictionaries. Since South Africans recognize that English is an international language (rather than a South African artifact), perceptions of English will not necessarily be based upon perceptions of South African English speakers and the institutions with which they are associated. However, with more widespread use of English dictionaries by Black people, we can expect that racial language in English dictionaries will attract more notice. And as South African English dictionaries are developed that may replace the currently popular British dictionaries, more South African racial terminology will be included in the dictionaries people use. The trend in South African English lexicography, if the (South African-edited) dictionaries of South Africanisms are indicative, is toward more explicit cautionary labeling. This signifies a move from the British style toward a more American model of usage description, to which I turn next.

The English Models: British and American Traditions

South Africa and the United States are similar in the complexity of their racial relations, but different in the nature of that complexity. Both provide challenges for lexicographers, who need both to accurately reflect semantically and socially complicated words and to satisfy a dictionary-using public that is sensitive to the affective power of the words defined and the metalanguage used to define them. However, since South African English lexicography is derived from (and often situated in) the British lexicographical tradition, we can expect that dictionary treatments of racial terms will be rather different in the two cultures.

Murphy (1991: 61) found three types of problems in American racial term definitions and usage treatment:

(a)   the polysemy of racial labels is underrepresented, often to the point of contradiction in cross-referring definitions;

(b)   a white norm is sometimes implicitly assumed in dictionary definitions of racially-charged terms; and

(c)   usage notes for labels for Black Americans do not inform the reader of differences in usage among Black and white users.

As shown in Murphy (forthcoming), criticisms (a) and (b) hold for the dictionaries used by South African English speakers as well. This essay is concerned with point (c), which notes that usage information presented in dictionaries tends to assume a White user. This criticism is less apt for the South African English dictionaries and British imports used in South Africa. This relative success is caused by a relative absence of usage information and by the fact that South African English is largely a language of White people. Taking the latter point first, arguments about what people prefer to be called in English have not assumed the importance in South Africa that they have in the United States. Thus, it is not at all common in South Africa for public discourse to revolve around whether people of African ancestry should be called **African** or **Black**. For those items for Black Africans that do require cautionary labels (**Native**, **Bantu**), little differentiates White and Black usage of the terms, since they are not widely used by Black Africans. Lack of Black African interest in English language auto-ethnonyms is in part explained by the fact that those labeled usually do not speak English as a first language, but further explained by the differences in "Black" racial identity in South Africa and the United States (Greenstein 1993). In the United States, a relatively coherent "Black" group identity has formed in the past two centuries, such that **Black** (or a near-synonym thereof) is the primary means of racial/ethnic identity among and for a well-delimited portion of the population, which has a common language and its own set of cultural, religious, and political traditions. In South Africa, on the other hand, the "Black" group has been internally and externally divided on the basis of ethnic/language groups, such that in many contexts the Black/African identity is subsidiary to a Zulu or Xhosa or Sotho identity. A Black/African identity was important to anti-apartheid organizing and will continue to be important in the "Africanization" of business, government, education, *et cetera*. However, at a personal, intra-African or political level, ethnic identity is often primary. The contrast between the heterogeneity of the Black/African group in South Africa and the homogeneity of the Black (African-American) group in America may account for the fact that **Black** is more easily used in South Africa in the "non-White" sense. In the United States, where **Black** is strongly attached to a coherent social group, this label carries too much specific denotative and connotative information for it to be more generally used in most contexts (i.e., to include Asian Americans, Latinos, Native Americans, *et cetera*).

So, while racial items like **Black** may not be as controversial in South Africa as in the United States (and thus not require so much usage notation), the situation described above leads to the hypothesis that ethnic terminology

for Black/African groups (e.g., **(ama)Zulu, Basotho**) assumes greater impor-
tance in South Africa and South African English dictionaries. However, since
all of the African ethnic terminology is borrowed into English from cultures
with much shorter written histories, the possibility of usage conflict (and conse-
quent offense) is diffused. Firstly, the referents of the ethnonyms have tradi-
tionally had little stake in English, its orthography, and especially its diction-
aries. Secondly, the orthographic history of the Southern Bantu languages is
quite short and varies in standardization. Literacy (and literariness) in these
languages is not as well-established as in a language like English. Thus, the
cultural attachment to orthographical forms (especially in terms of capitaliza-
tion) is slighter than in a language with a longer history of standardized ortho-
graphy, literary language, and general literacy. As language policies, general
literacy, and postapartheid racial/ethnic identities evolve, mismatches between
Bantu language labels and their borrowed counterparts in English may lead to
offense and protest, and thus the English language dictionaries will be required
to take a more consciously prescriptive role in ethnic label orthography. This
may already be under way. On the University of the Witwatersrand campus,
posters advertising matches between the Wits and AmaZulu football clubs
have been defaced in order to "correct" the capitalization of **AmaZulu** in
various ways (with attendant commentary on white peoples' ignorance of Afri-
can languages). On a larger scale, proposals by Neville Alexander that the
orthographies of the Nguni and Sotho groups of languages be regularized (in
order to aid in the administration of 11 official languages), have been met by
uproar that "the government wants to take away our languages and their indi-
vidual characters". This is fairly ironic, considering that the diverse ortho-
graphies of the indigenous South African languages were largely the work of
white missionaries and academics and therefore represent appropriation of the
languages for the benefit of Europeans' goals, not indigenous (or necessarily
linguistically logical) aims.

General English dictionaries for the South African market, coming from
the British lexicographical tradition, include fewer prescriptions regarding
usage and less encyclopedic information than American dictionaries (Béjoint
1994). Similarly, their audiences differ, in that Americans are widely perceived
to grant dictionaries more influence and greater status than the British do
(Quirk 1973). So for this reason as well, bias in usage labeling is less pernicious
in South African English dictionaries than in American dictionaries, simply be-
cause they contain less usage labeling and less is expected of their treatment of
usage. Read (1986) and Algeo (1989) suggest that the American tradition's
deviations from the British tradition reflect the United States' postcolonial iden-
tity. Depending upon one's perspective, prescriptivism and encyclopedism in
American dictionaries can reflect either Americans' linguistic and intellectual
insecurities or their belief in (and practice of) socio-economic mobility (or, per-
haps a combination of these). British dictionaries, in a sense, have fewer pre-
scriptive controversies to arbitrate, since the "standard" language in Britain is

better defined and highly accessible and recognizable. The United States, how-
ever, tolerates and supports a wider range of "standard" forms, including re-
gional "standards". So, when attempting to speak "standardly", Americans are
subject to more contradictory standards than the British, and thus may require
a dictionary to settle these controversies. The prescriptive tradition in Ameri-
can dictionaries can also be traced to the postcolonial need to assert a national
identity through a nationally distinguishable language.

While South African English lexicography hails from the British tradition,
the similarities between the South African and American situations are enough
to predict that South African English lexicography will move toward a more
prescriptive or at least cautionary style. This prediction is based on several
facts. First, a "standard" South African English is increasingly recognized.
Thus, South African English dictionaries may take the rôle that American dic-
tionaries have had in promoting a new "standard". Furthermore, South African
English stands to have more than one standard, based on ethnic varieties of
English, particularly the forms Black South African English used in the political
domain. (See De Klerk 1996 and particularly Wright 1996.) Second, South
Africa is a nation in social transition, and thus its communicative styles are
changing. As Chick (1991) discusses, South African English conversational
styles are shifting from the deference-based system of British culture to a soli-
darity-based system, the communicative style most closely associated with
Americans. The solidarity system is an escalating and unstable type of system.
Thus, underlying assumptions about proper linguistic behavior may be shifting
such that these assumptions will be reflected in dictionary style. Third, oppor-
tunities for social mobility in South Africa are shifting rapidly in the postapart-
heid climate. Greater mobility is especially afforded to nonnative speakers of
English, who can be expected to increasingly demand dictionaries that serve
their needs in mastering "standard" South African English. These needs may
be served by the advanced learner's dictionaries that are due to recent innova-
tions in the British lexicographical industry, but these needs might also be
served (especially in integrated educational institutions) by general use diction-
aries with a more prescriptive mission. As suggested above, nonnative
speakers' increased access to South African English and its dictionaries will
likely create more interest among these speakers in the ethnonyms that
describe them in South African English.

Conclusions

In conclusion, South African English and its lexicography are in a state of flux,
and thus new demands will be made of South African English dictionaries and
their treatments of racial labels in the near future. Changes in racial label
presentation, definition, and usage labeling are required in large part because
of the changing demography of English speakers. As South African educa-

tional, political, and commercial spheres become more and more integrated, the racial/ethnic make-up of (educated) English speakers in South Africa becomes more diverse. This diversity, in turn, has two relevant implications. Firstly, as more of the referents of the various racial labels become speakers of English, they will have more interest and stake in the words that English uses to describe them (and which they, then, use to describe themselves). Secondly, the diversity of the English-speaking population will result in a more diverse dictionary-using population. Thus, South African English dictionaries will be expected to reflect a wider range of perspectives toward the words and language therein, or else risk economic (and social-political) consequences. However, the predictions made here, that racial labels in English (and their treatment in dictionaries) will become more controversial as English is more widely used, may be overstated. The political and economic situation in South Africa may not support active linguistic controversies when so many other controversies must be resolved in a new (and economically troubled) democracy. However, naming controversies do exist in South Africa (viz. **Bantu/Sintu** and the capitalization of **Coloured**), and shifts in identity, such as are forced by current efforts toward nation-building, will continue as the new South Africa and new South Africans reconcile with the past and move toward the future. Since labeling is a (if not *the*) crucial step in identity formation, it seems that it will only increase in importance in the coming years.

Notes

1.   Although AFRILEX is an international body, the 1996 board included six South African scholars and publishers of Afrikaans, five scholars of African languages from South Africa and three from neighboring states, and two employees of the National Terminology Services (responsible for all 11 official languages). No lexicographers working primarily on English are included, and few of the South African board members work at English-medium institutions.

2.   The South African nationalistic focus of what purports to be an African organization has caused some consternation among non-South African (and even some South African) members of AFRILEX, but much of this frustration is not expressed to the Board, which seems to assume that South African interests are the interests of the organization as a whole.

3.   Thus, the policy document (which is largely composed of justifications for the policy) makes such invalid arguments as "Highly advanced technical language can not insult anybody, therefore nobody needs to be warned against its use. Racist terms are always hurtful, therefore dictionary users should indeed be warned against its use." (Harteveld and Van Niekerk 1995: 254). People can indeed be insulted by technical language (the history of **Bantu** provides evidence — see Khumalo 1984), just as people can be affectionately referred to by "racist" words.

4.   This policy, then, is subject to the same criticism as usage labels in American English dictionaries, which, as Murphy (1991) points out, are based on the assumption that the language

user is White. Thus, the treatment of words like **kaffer** in the WAT are made without consideration for their ironic and sometimes affectionate ingroup use observed by the author in Johannesburg. The same phenomenon is seen for items such as **nigger** in American English, and while the reclamation of such taboo terms is more prevalent in the United States now, the opportunity for such use may increase in South Africa as the distance from apartheid times increases. Certainly, while the "racial" sense of **Bantu** is listed as *"offensive"* in many dictionaries, ironic and jocular use by ingroup members does occur, especially when poking fun at the attitudes or policies of White South Africa. While ingroup users of these words are taking advantage of their taboo status, they are not identifying themselves as racist by using the words.

Dictionaries cited

Chambers-Macmillan South African Dictionary for Junior Primary. 1996. Manzini, Swaziland: Macmillan Boleswa.

Collins Concise English Dictionary³. 1992. Glasgow: HarperCollins.

The Concise Oxford Dictionary of Current English⁸. 1990. Oxford: Clarendon.

A Dictionary of South African English⁴. 1991. Cape Town: Oxford University Press.

A Dictionary of South African English on Historical Principles. 1996. Cape Town: Oxford University Press.

The South African Oxford School Dictionary. 1996. Cape Town: Oxford University Press.

The South African Pocket Oxford Dictionary. 1987. Cape Town: Oxford University Press.

Woordeboek van die Afrikaanse Taal X: M. 1996. Stellenbosch: Bureau of the WAT.

References

Algeo, John. 1989. Dictionaries as Seen by the Educated Public in Great Britain and the USA. Hausmann, Franz Josef, Oskar Reichmann, Herman Ernst Wiegand, and Ladislav Zgusta (Eds.). 1989. *Dictionaries: International Encyclopedia of Lexicography* 1: 28-34. Berlin: Walter de Gruyter.

Béjoint, Henri. 1994. *Tradition and Innovation in Modern English Dictionaries. (Oxford Studies in Lexicography and Lexicology.)* Oxford: Clarendon.

Benjamin, Jean. 1994. Language and the Struggle for Racial Equality in the Development of a Non-racial Southern African Nation. Fardon, Richard and Graham Furniss (Eds.). 1994. *African Languages, Development and the State*: 97-110. London: Routledge.

Bureau of the WAT. 1994. Policy for Handling Insulting and Sensitive Lexical Items in the *Woordeboek van die Afrikaanse Taal (WAT)*. Photocopied document, Stellenbosch, South Africa.

Chick, Keith. 1991. Sources and Consequences of Miscommunication in Afrikaans English — South African English Encounters. Cheshire, Jenny (Ed.). 1991. *English Around the World: Social Contexts*: 446-461. Cambridge: Cambridge University Press.

De Klerk, Vivian (Ed.). 1996. *Focus on South Africa. (Varieties of English Around the World* G15.) Amsterdam: John Benjamins.

Gouws, Rufus. 1986. Afrikaans Lexicography. R.R.K. Hartmann (Ed.). 1986. *The History of Lexicography: Papers from the Dictionary Research Centre Seminar at Exeter, March 1986*: 75-83. (*Amsterdam Studies in the History of Linguistic Science* 40.) Amsterdam: John Benjamins.

Greenstein, Ran. 1993. Racial Formation: Towards a Comparative Study of Collective Identities in South Africa and the United States. *Social Dynamics* 19(2): 1-29.

Harteveld, P. and A.E. van Niekerk. 1995. Policy for the Treatment of Insulting and Sensitive Lexical Items in the *Woordeboek van die Afrikaanse Taal (WAT)*. *Lexikos* 5: 249-266.

Hauptfleisch, D.C. 1993. Racist Language in Society and in Dictionaries: A Pragmatic Perspective. *Lexikos* 3: 83-139.

Khumalo, J.S.M. 1984. New Term for "Bantu" in Linguistics Studies. *South African Journal of African Languages* Supplement 1984(1): 111-120.

Kinloch, Graham C. 1985. Racial Attitudes in South Africa: A Review. *Genetic, Social, and General Psychology Monographs* 111(3): 261-281.

Links, T.H. 1991. Ideology ("Apartheid") and Afrikaans Dictionaries. Paper presented at the 8th Biennial Meeting of the Dictionary Society of North America, University of Missouri at Columbia, 9-11 August.

Louw, Johann and Don Foster. 1992. Intergroup Relations and South African Social Psychology: Historical Ties. *Canadian Psychology* 33(3): 651-656.

Murphy, M. Lynne. 1991. Defining Racial Labels: Problems and Promise in American Dictionaries. *Dictionaries* 13: 43-64.

Murphy, M. Lynne. Forthcoming. Defining People: Race in South African English Dictionaries. Submitted for publication to *International Journal of Lexicography*.

Quirk, Randolph. 1973. The Social Impact of Dictionaries in the UK. McDavid, Raven I., Jr. and Audrey R. Duckert (Eds.). 1973. *Lexicography in English*: 76-88. New York: New York Academy of Sciences.

Read, Allen Walker. 1986. Competing Lexicographical Traditions in America. R.R.K. Hartmann (Ed.). 1986. *The History of Lexicography: Papers from the Dictionary Research Centre Seminar at Exeter, March 1986*: 197-206. (*Amsterdam Studies in the History of Linguistic Science* 40.) Amsterdam: John Benjamins.

Wright, Laurence. 1996. The Standardisation Question in Black South African English. De Klerk, Vivian. 1996: 149-162.

'n Kontrastiewe beskouing van tweetalige woordeboeke in Suid-Afrika en Japan[*]

Takashi Sakurai, *Departement Japannees,*
Meikai Universiteit, Chiba, Japan en
Ernst F. Kotzé, *Departement Afrikaans en Nederlands,*
Universiteit van Port Elizabeth, Suid-Afrika

Abstract: A Contrasting View on Bilingual Dictionaries in South Africa and Japan. Both dictionaries that combine Afrikaans and African languages and dictionaries that combine Japanese and Western languages incorporate unrelated languages. In this respect lexicographers in Japan and South Africa have the potential for fruitful cooperation. Nevertheless, major differences between bilingual dictionaries in South Africa and Japan came to the fore during the transforming of the *Afrikaans/Zoeloe-woordeboek (Afrikaans/Zulu Dictionary)* of Kotzé and Wela into an Afrikaans-Japanese dictionary. Differences in microstructure include the use of phonetic transcriptions, indication of the distinction between transitive and intransitive verbs, the ordering of example-sentences and the use of certain symbols as aids. The difference in social context is reflected in the choice of sensitive lexical items and the size of the dictionaries. These differences had to be accounted for in the remaking.

The remaking of existing dictionaries is an approved method in Japanese lexicography and there are some historical examples of this practice. However, this does not merely ential a shift of target language but also adapting the structure of the dictionary to the conventional format acceptable to the users (in this case Japanese speakers). Bilingual dictionaries in South Africa that fall under the same typological category as the Afrikaans-Japanese dictionary (in treating genealogically unrelated languages) can benefit from the comparison made in this article.

Keywords: ABUSIVE LEXICAL ITEMS, AFRIKAANS, BILINGUAL DICTIONARY, EXAMPLE-SENTENCES, JAPANESE, LABELLING, LEARNER'S DICTIONARY, MACROSTRUCTURE, MICROSTRUCTURE, OSTENSIVE DICTIONARY, PHONETIC TRANSCRIPTION, TRANSCULTURALISATION, TRANSITIVITY, TRANSLATION EQUIVALENT

Opsomming: Sowel tweetalige woordeboeke wat Afrikaans en Afrikatale kombineer as dié wat Japannees en Westerse tale kombineer, hanteer tale wat geen onderlinge taalkundige verwantskap het nie. In hierdie opsig het leksikograwe van Japan en Suid-Afrika 'n vrugbare samewerkingsterrein. By die verwerking van die *Afrikaans/Zoeloe-woordeboek* van Kotzé en Wela tot 'n Afri-

---

[*]   Hierdie studie vorm 'n deel van die projek genaamd "Vertalende brugwoordeboeke met Afrikaans as bron- of teikentaal" in die Departement Afrikaans en Nederlands, Universiteit van Port Elizabeth, Port Elizabeth.

kaans-Japannese woordeboek het groot verskille tussen die tweetalige woordeboeke in Suid-Afrika en dié in Japan na vore gekom. Mikrostrukturele verskille sluit in die gebruik van fonetiese transkripsies, die aanduiding van die onderskeid tussen oorganklike en onoorganklike werkwoorde, die volgorde van voorbeeldsinne en die gebruik van bepaalde simbole as hulpmiddels. Deur die hantering van aanstootlike leksikale items en die grootte van die woordeboeke word die verskil in sosiale agtergrond gereflekteer. Hierdie verskille moes by die verwerking verdiskonteer word.

Die heropstelling van bestaande woordeboeke is 'n aanvaarde metode waarvolgens die leksikografie in Japan bedryf word, en daar is 'n aantal voorbeelde daarvan uit die geskiedenis. Dit beteken egter nie 'n eenvoudige vervanging van die teikentaal nie, maar die hervorming van die woordeboekstruktuur en die aanpassing daarvan by die gebruikers (in hierdie geval Japanneessprekendes). Tweetalige woordeboeke in Suid-Afrika wat tipologies by die Japannees-Afrikaanse woordeboek aansluit (m.a.w. deur die bewerking van genealogies onverwante tale) sou kon baat by die vergelyking wat in hierdie artikel getref word.

Sleutelwoorde: AANLEERDERSWOORDEBOEK, AANSTOOTLIKE LEKSIKALE ITEMS, AFRIKAANS, ETIKETTERING, FONETIESE TRANSKRIPSIE, JAPANNEES, MAKROSTRUKTUUR, MIKROSTRUKTUUR, OORGANKLIKHEID, OSTENSIEWE WOORDEBOEK, TRANSKULTURALISASIE, TWEETALIGE WOORDEBOEK, VERTAALEKWIVALENT, VOORBEELDSINNE

0.   Inleiding

In hierdie artikel word daar gewys op beduidende verskille tussen Afrikaanse en Japannese woordeboeke waarvan daar by die opstel van 'n Afrikaans-Japannese woordeboek kennis geneem is. Die bespreking is gerig op tweetalige woordeboeke, maar verklarende woordeboeke, veral handwoordeboeke, word ook betrek waar nodig.

Vergelykings kan ook getref word met elektroniese woordeboeke wat onder andere in CD-ROM-formaat beskikbaar is. In hierdie artikel gaan dit egter om die strukturering van geskrewe leksikons.

1.   Twee genealogies onverwante tale

Tweetalige woordeboeke kan ten opsigte van die behandelde tale onder andere in twee tipes onderskei word, te wete woordeboeke wat tale van dieselfde taalfamilie kombineer en woordeboeke waarin tale van verskillende taalfamilies figureer. Hierdie subklassifikasie verteenwoordig 'n wesenlike onderskeid tussen hierdie tipes woordeboeke. Dit is byvoorbeeld nagenoeg onmoontlik om by twee tale van uiteenlopende taalfamilies, soos Afrikaans en Japannees, onderlinge *faux amis* ("vals vriende") aan te dui.

As die twee tale tot dieselfde taalfamilie behoort, soos Afrikaans en Duits, (en as dié twee tale boonop kulturele verwantskap vertoon), sal die opstel van

die woordeboek minder probleme oplewer, want vertaalekwivalente lê voor die hand. Maar as dié twee tale geen taalkundige verwantskap besit nie, sal die opsteller dikwels moeite ondervind om selfs vir die eenvoudigste woord 'n ekwivalent te vind.

Watter een van hierdie woordeboektipes dominant is, hang af van die taal-situasie in die betrokke land. Japan is 'n feitlik eentalige land. Alhoewel min-derhede voorkeur gee aan die gebruik van hulle eie tale, Ainu, Koreaans en Chinees, word dié tale haas glad nie gebruik buite hulle klein kringe van moe-dertaalsprekers nie. In hierdie situasie het die meeste Japannese slegs vir die studie van buitelandse tale tweetalige woordeboeke nodig.

Byna alle Japannese leer Engels op middelbare en hoërskool (st. 5 tot 10 in Suid-Afrika). Aan die universiteite is Engels ook die eerste buitelandse taal en as tweede buitelandse taal word tradisioneel Duits of Frans gekies. Gevolglik is die merendeel van die tweetalige woordeboeke wat in Japan gepubliseer word, dié wat Japannees en Westerse tale kombineer, tale wat geen onderlinge genea-logiese of tipologiese verwantskap vertoon nie.

Die tweetalige woordeboeke wat Afrikaans as bron- of teikentaal het, word ook deur die taalsituasie in die land bepaal. Vanselfsprekend is Afri-kaans-Engelse/Engels-Afrikaanse woordeboeke die belangrikste in Suid-Afri-ka. Maar dit is opvallend dat daar toenemend woordeboeke gepubliseer word wat Afrikaans en die verskillende Afrikatale kombineer. Dat die woordeboeke waarin tipologies uiteenlopende tale gekombineer word, so belangrik is, is 'n punt wat Afrikaanssprekendes en Japannese gemeen het. In hierdie opsig het die leksikograwe van die twee lande 'n potensieel vrugbare samewerkings-terrein. In Europa kom 'n mens ook daagliks in kontak met ander tale en die leksikograwe daar het lang ervaring van tweetalige woordeboeke. Maar *tweetalig* beteken hier gewoonlik 'n kombinasie van twee Westerse tale, byvoorbeeld Duits en Frans. Woordeboeke wat een van die Westerse tale met, byvoorbeeld, 'n Asiatiese taal kombineer, is net vir diegene met spesifieke oog-merke. Goeie tweetalige woordeboeke met 'n heterogene taalsamestelling is eerder die uitsondering as die reël.

Die Verenigde State van Amerika publiseer baie woordeboeke wat Engels met tale van inheemse Amerikaners of met tale in die Stille Oseaan kombineer. Maar hierdie tweetalige woordeboeke is eerder gemik op akademiese beskry-wing as vir praktiese gebruik. (Amerikaners wil gewoonlik nie self ander tale aanleer nie, en beoog eerder die onderrig van Engels aan ander.) As 'n woorde-boek eers klaar opgestel is, glo die navorsers dat die taal gered is, en die woor-deboek sal in die meeste gevalle nie gereeld hersien word vir 'n verbeterde uitgawe nie.

Japannese leksikograwe sou, in die lig van bogenoemde oorwegings, medewerkers nie in Europa of in die Verenigde State van Amerika nie, maar in Suid-Afrika vind.

2.   'n Afrikaans-Japannese woordeboek

Die poging om 'n Afrikaans-Japannese woordeboek op te stel waarvan in hierdie artikel verslag gedoen word, is in wese 'n verwerking van Kotzé en Wela (1992), waar Zoeloe as doeltaal deur Japannees vervang word.

Om by die skryf van 'n woordeboek van die makrostruktuur van 'n ander woordeboek gebruik te maak, is in Japan nie ongewoon nie, maar word as 'n uitweg aanvaar indien 'n mens gou die buitelandse taal moet invoer. Sommige voorbeelde uit die geskiedenis kan aangetoon word.

Die oudste bestaande Japannese woordeboek, *Tenrei-banshou-myougi*, wat in die eerste helfte van die 9de eeu deur die Boeddhistiese priester Kuukai opgestel is, is die verjapanisering van die Chinese woordeboek *Yù-piang*. Die eerste groot Nederlands-Japannese woordeboek *Haruma-wage* is in 1796 opgestel op die basis van François Halma se Nederlands-Franse woordeboek. Heropstelling is vandag nog steeds 'n uitweg in geval van nood. Die jongste Nederlands-Japannese woordeboek *Oranda-go-jiten* wat in 1994 gepubliseer is, is die verwerking van *Kramers' Nederlands-Engels Woordenboek*.

'n Woordeboek behoort ideaal gesproke *ab initio* volgens die opsteller se eie plan ontwerp te word. Die praktyk vereis normaalweg 'n lang periode van samewerking tussen verskeie leksikograwe. Hierdie gewone werkwyse is in die huidige omstandighede onmoontlik vir 'n Afrikaans-Japannese woordeboek. Die enigste moontlike en realistiese keuse in hierdie geval is leksikografiese herbewerking.

Verwerking is nie die eenvoudige vervanging van die teikentaal nie, maar behels die hervorming van die woordeboekstruktuur. Hierdie aspek word hieronder verder bespreek.

Die redes waarom die *Afrikaans/Zoeloe-woordeboek* as basis gekies is, is dat dié woordeboek

(a)   die tale van uiteenlopende taalfamilies kombineer en
(b)   o.a. op die behoeftes van die aanleerder op universiteitsvlak gerig is.

Die Afrikaans-Japannese woordeboek in wording het ook hierdie kenmerke en het in dié woordeboek 'n goeie model gevind.

3.   Verskille tussen die woordeboeke in die twee lande

In hierdie afdeling word die praktyk van die leksikografie in beide Japan en Suid-Afrika onder die loep geneem, en kenmerke wat Afrikaanse woordeboeke gemeen het, vergelyk met dié van Japannese woordeboeke.

3.1   Fonetiese transkripsie

Dit is opmerklik dat geen tweetalige woordeboek met Afrikaans as een van die tale fonetiese transkripsies bevat nie. Dit geld ook Afrikaanse handwoorde-

boeke. Die rede is voor die hand liggend. Dit word as onnodig beskou vir die woordeboekgebruikers, almal Suid-Afrikaners wat elke dag ouditief aan Afrikaans blootgestel word. Buitelandse gebruikers was tot dusver nie 'n faktor vir Afrikaanse leksikograwe nie. Maar vir sulke gebruikers wat geen geleentheid tot blootstelling aan Afrikaans het nie, is fonetiese transkripsies 'n *sine qua non*.

Daar is 'n soortgelyke probleem in die geval van Japannese woordeboeke. Japannese woordeboeke met 'n vreemde taal as brontaal bevat in die reël fonetiese transkripsies, maar by woordeboeke met Japannees as brontaal ontbreek sulke transkripsies deurgaans. Een beduidende rede is dat Japannese leksikograwe nooit situasies voorsien het waar so baie buitelanders Japannees bestudeer nie. Nog 'n rede is die fonologiese eenvoud van die Japannese taal. Die artikulatoriese konsekwentheid en relatiewe afwesigheid van allofonie gee aanleiding daartoe dat transkripsies in die Romeinse alfabet feitlik as fonetiese transkripsies funksioneer. Die jongste Japannese woordeboeke vir buitelanders wat as gevolg van die onlangse toename in die studie van Japannees gepubliseer is, toon slegs die uitspraak met behulp van die gewone Romeinse transliterasie aan.

Daarteenoor is die Afrikaanse uitspraak ingewikkelder as die Japannese, byvoorbeeld ten opsigte van verskillende diftonge en allofone van die letter "e" ([ə][e][ɛ]). In 'n tweetalige Afrikaanse woordeboek behoort hierdie variasie duidelik aangetoon te word deur middel van fonetiese transkripsies.

Die gebruik van fonetiese transkripsies sou ook inpas by Japannese onderrigmetodes ten opsigte van buitelandse tale. Omdat 'n mens buite die groot stede in Japan vroeër selde met moedertaalsprekers van byvoorbeeld Engels kontak kon maak, het die tradisionele konvensie by Engelsonderrig ontstaan dat fonetiese transkripsies van die begin af ingevoer is om die uitspraak te leer. Dit wil sê, saam met Engels word die Japannese verplig om ook fonetiese transkripsie te leer. Dié kennis gebruik hulle dan verder vir die studie van die ander buitelandse tale.

In Suid-Afrika, hierteenoor, word fonetiese transkripsie feitlik nêrens as hulpmiddel vir taalonderrig benut nie. Kennis daarvan kan die Afrikaanse leksikograwe nie van die gemiddelde woordeboekgebruiker verwag nie.

Maar in sowel Afrikaanse as Japannese woordeboeke behoort die uitspraak met behulp van fonetiese transkripsies duidelik in ooreenstemming met internasionale gebruik weergegee te word.

3.2   Die onderskeid oorganklik / onoorganklik by werkwoorde

Sintaktiese inskrywingskategorieë wat by Afrikaanse woordeboeke gebrekkige aandag ontvang, is die onderskeid tussen oorganklike en onoorganklike werkwoorde. Dit is by Japannese woordeboeke een van die basiese inligtingskategorieë. As hierdie onderskeid nie aangedui word nie, sal so 'n woordeboek nie as standaard beskou word nie. Selfs 'n klein sakwoordeboek wat net die ver-

taalekwivalente sonder voorbeeldsinne opneem, dui die onderskeid tussen die twee soorte werkwoorde aan. 'n Goeie voorbeeld daarvan is die artikel **propose** in *Sanseido's Daily Concise English-Japanese Dictionary* wat duidelik tussen *vt.* en *vi.* onderskei.

> **‡pro·pose** [prəpóuz] *vt.* 提案する；申し
> 込む；推薦〔指名〕する；企てる. — *vi.* 計
> 画する；結婚を申し込む. ***pro·pós·al** *n.*
> 申込み；提案，計画；結婚の申込み. **-pós-**
> **er** *n.* 申込人；提案者.

Hierdie benadering sou Afrikaanse leksikograwe miskien vreemd vind, maar dit is 'n grammaties belangrike eienskap van die Japannese taal.

In Afrikaans (en ook Engels) kan 'n werkwoord in baie gevalle sowel oorganklik as onoorganklik gebruik word. In Japannees bring die twee werkwoordkategorieë egter totaal verskillende woordvorme mee. Vergelyk die volgende voorbeelde:

Afrikaans	Japannees
brand (oorg./onoorg.ww.):	*yaku* (oorg.ww.) *yakeru* (onoorg.ww.)
hang (oorg./onoorg.ww.):	*kakeru* (oorg.ww.) *kakaru* (onoorg.ww.)

Oorganklike en onoorganklike werkwoorde in Japannees word gevolglik as onderskeie lemmas gehanteer. Dié verskil is so wesenlik dat dit in enige woordeboek vindbaar behoort te wees. Ook in 'n Afrikaanse woordeboek sou Japanneessprekende gebruikers 'n aanduiding van hierdie basiese onderskeid by werkwoorde verwag.

Dit is een voorbeeld van die opvallende grammatikale verskille wat 'n leksikograaf in ag moet neem by die opstel van 'n woordeboek wat tipologies ongelyksoortige tale kombineer.

Afgesien van hierdie eienskap van Japannees wat dié inligtingstipe noodsaaklik maak, is dit gewoon gebruikersvriendelik as 'n woordeboek oorganklikheid of onoorganklikheid by werkwoorde aandui.

3.3   Aanstootlike leksikale items

3.3.1   Woorde en uitdrukkings

Soos in Afrikaans is daar ook in Japannees items wat die opsteller van 'n woordeboek versigtig moet hanteer. Maar oor wat as sensitief beskou word, verskil

Japan van Suid-Afrika. Dit verteenwoordig 'n verdere probleem by 'n Afri-kaans-Japannese woordeboek. Vir Afrikaanse leksikograwe sou die meeste pro-bleme die woorde en uitdrukkings wees wat 'n rassistiese konnotasie het, byvoorbeeld **ou meid**. Die letterlike betekenis daarvan is gewoon *bejaarde huis-hulp* en sy vertaalekwivalent in Japannees, *toshi-totta kasei-fu* het hoege-naamd geen slegte konnotasie nie. Sulke Afrikaanse woorde of uitdrukkings kan buitelandse aanleerders onbewus verkeerd gebruik. Om dit te vermy, behoort die opstellers van 'n tweetalige woordeboek, indien hulle sulke leksi-kale items opneem, met behulp van 'n etiket aan te dui dat dié items rassistiese konnotasies het.

Wat Japannese woordeboeke betref, is sensitiewe items hoofsaaklik diskri-minerende woorde en uitdrukkings wat vroue, fisies en/of geestelik gestrem-des, asook mense van 'n onderdrukte kaste (*hi-sabetsu buraku-min*) beskryf. As 'n woordeboek die items foutief hanteer, word van die redakteur verwag om in die openbaar verskoning te vra en die inskrywing by die volgende uitgawe reg te stel. In die ergste geval sal die woordeboek dadelik aan alle boekwinkels onttrek word en die gebruik daarvan deur biblioteke streng beperk word.

Hoe leksikograwe dié tipe items hanteer, illustreer *Kenkyusha's New College Japanese-English Dictionary* (1995⁴), 'n gewilde Japannees-Engelse handwoorde-boek. In hierdie woordeboek word 'n "x" as etiket ná die lemma gebruik om die gewraakte lemma te merk. Nog 'n hanteringswyse is om die item weg te laat. Byvoorbeeld *buraku*, 'n soortnaamwoord wat *dorpie* beteken, kom nie daarin voor nie, vermoedelik uit vrees vir assosiasie met *hi-sabetsu <u>buraku</u>-min* (letterlik vertaal: mense van 'n miskende dorpie). *Die opstellers sou vir die gebrekkige opname van die basiese woordeskat gekritiseer kan word, maar hulle sou kan antwoord dat 'n handwoordeboek nie volledig hoef te wees nie.*

In die sensitiewe taalomstandighede in Japan veroorsaak ander gewone Afrikaanse woorde ook 'n probleem. 'n Goeie voorbeeld is gek. Die naaste Japannese vertaalekwivalent van die woord is *ki-chigai*, maar dit word as 'n diskriminerende woord vir 'n psigoot beskou en die gebruik daarvan in die skriftelike en elektroniese media is feitlik verbode.

By die lemma **crazy** in die bogenoemde *Kenkyusha's New College Japanese-English Dictionary* staan die uitdrukkings *ki ga kurutta* en *kyou-ki no* in plaas van *ki-chigai* as vertaalekwivalente. Dit is wel sosiaal korrek, maar hierdie alternatiewe gee nie die grammatikale verhouding en die register van **crazy** weer nie, want *ki ga kurutta* is 'n onhandige parafrase en *kyo-ki no* is 'n skryf-taalvorm.

Die inskrywing vir gek toon blyke van dieselfde problematiek. Geen finale oplossing lê voor die hand nie. In dié stadium lyk die enigste moontlikheid die gebruik van alternatiewe leksikale items. Leksikograwe moet soms aan maat-skaplike eise voldoen met opoffering van die gepaste vertaalekwivalente.

3.3.2 Voorbeeldsinne

Voorbeeldsinne moet ook baie versigtig gehanteer word. Alhoewel nie elke woord in die sin problematies is nie, kan die idee wat die sin oordra as diskriminerend beskou word. Die volgende sin in die *Afrikaans/Zoeloe-woordeboek* staan op die grens:

Die vlek op haar gesig doen afbreuk aan haar skoonheid.

As dieselfde voorbeeldsin in Japannees vertaal word, sal daar in Japan sterk protes van vroue verwag kan word. Die implikasie sou wees: Deur die waarde van vroue gelyk te stel met oppervlakkige skoonheid, word hul inherente menslikheid geïgnoreer. En so 'n verklaring met betrekking tot vroue met vlekke op hulle gesigte, is 'n belediging vir sulke vroue, en vir persone met fisiese gebreke in die algemeen.

Sulke interpretasies deur gebruikers van die teikentaal moet deur die opsteller van 'n tweetalige woordeboek in ag geneem word. Kennis daarvan strek egter oor die grense van die leksikografie en val eerder binne die gebied van die kulturele antropologie.

Desnieteenstaande word dié voorbeeldsin sonder verandering in die Afrikaans-Japannese woordeboek gebruik, omdat dit tog die kulturele verwysingsraamwerk en die taalgebruikskundige konteks van Afrikaans verteenwoordig.

3.4. Ordening van die inskrywings in 'n artikel

3.4.1 Leksikale items

'n Belangrike onderwerp wat normaalweg nie so intensief bespreek word nie, is die volgorde van die inskrywings in 'n artikel.

In Afrikaanse tweetalige woordeboeke word al die vertaalekwivalente ná die grammatiese inligting gelys en daarna volg die voorbeeldsinne. Dié volgorde het die voordeel dat die struktuur van die artikel deursigtig is en dat die gebruiker die betekenis van die lemma gou kan begryp met behulp van die lys vertaalekwivalente wat reeds aan die begin van die artikel gegee word.

Maar so 'n volgorde is in Japan ongewild omdat die verband tussen 'n vertaalekwivalent en 'n voorbeeldsin waarin die vertaalekwivalent gebruik word, nie onmiddellik duidelik is nie.

Die tipiese volgorde binne die lemma van Japannese woordeboeke is: 1. vertaalekwivalent, voorbeeldsin(ne)   2. vertaalekwivalent, voorbeeldsin(ne) ... soos die volgende voorbeeld uit *Kenkyusha's New English-Japanese Dictionary* (1980[5]), die standaard Engels-Japannese woordeboek.

greet¹ [gríːt] 〔OE *grētan* to approach, greet <
(WGmc) *ʒrōtjan* to cry out, address (G *grüssen*) ←
IE *ghrēd-, *gher- to call out〕 ── vt. **1** (口頭・身ぶ
り・書面などで)…に挨拶を述べる, 敬礼する (hail,
salute): ～ a person *with* a cheerful "Good morning!"
元気よく「おはよう」と人に挨拶する. **2** (親愛・尊
敬・歓喜, 時には悪意・ののしりなどをもって)迎える
〔*with*〕: …～ a person *with* cheers, an embrace, a smile,
etc. / ～ a person's return *with* jeers [a scowl, a vol-
ley of arrows] やじりながら[にがい顔をして, いっせ
いに矢を放って]人の帰りを迎える / His remark was
～*ed with* hisses. 彼がそういったらやじが飛んだ /
Jeers ～*ed* the candidate. あざけりの笑声が候補者を
迎えた. **3** 〈光景などが〉…に現れる (appear to), 〈目・
耳などに〉触れる, はいる (strike): A wide expanse of
sea ～*ed* the eye. 広い海が目にはいった / A terrible
noise ～*s* our ears [us]. すごい物音が聞こえてきた /
A sweet smell ～*s* the nose. 甘い香りがただよってく
る. ── vi. 《廃》挨拶をかわす.

Aangesien die subkategorisering van mikrostrukture siklies herhaal word vir
elke betekenisonderskeiding, kan dit tot verwarring lei. Daarbenewens is die
ordening ongerieflik omdat die gebruiker deur die artikel moet lees om al die
vertaalekwivalente te kry. Maar die bogenoemde ordening in Japannese woor-
deboeke moet anders begryp word. Die onderliggende beginsel is die rangskik-
king van betekenisse, naamlik 1. betekenis 1 2. betekenis 2 ... By elke betekenis
word die vertaalekwivalent(e) aangedui en direk daarna word die voorbeeld-
sin(ne) gegee waarin die vertaalekwivalent(e) gebruik word. En met behulp
van die nommers kan die gebruiker maklik sy/haar pad tussen die vertaal-
ekwivalente vind. Hierdie stelsel word naastenby ook in Afrikaanse verkla-
rende handwoordeboeke gevolg.

3.4.2   Voorbeeldsinne

Ook die rangskikking van die voorbeeldsinne in 'n Afrikaanse woordeboekarti-
kel kom vir 'n Japanneessprekende vreemd voor. So byvoorbeeld bevat die arti-
kel van die byvoeglike naamwoord gesamentlik in die *Afrikaans/Zoeloe-woor-*
deboek twee voorbeeldsinne wat soos volg aangebied word:

gesamentlik (～e) b.nw. ⸗ngokuhlanganyela *(joint, to⸗*
tal) **1. hulle het ～ opgetree** benze ngokuhlanganye⸗
la **2. die gesamentlike bedrag is . . .** isamba sesi⸗
hlangene ngu⸗ . . .

In die eerste sin bepaal **gesamentlik** die werkwoord **op(ge)tree** en in die laaste bepaal dit die selfstandige naamwoord **bedrag**. Hierdie volgorde veroorsaak geen probleem nie omdat Afrikaanse byvoeglike naamwoorde sowel werkwoorde as naamwoorde kan bepaal. Maar Japannese byvoeglike naamwoorde se primêre funksie is die bepaling van 'n selfstandige naamwoord. As 'n byvoeglike naamwoord 'n werkwoord bepaal, moet die byvoeglike naamwoord verbuig word, en gevolglik word die bepaling van 'n werkwoord as 'n sekondêre funksie beskou. Vergelyk die volgende voorbeeld:

hayai (infinitief) vinnig   *hayai kisha* vinnige trein
                                   hayaku hashiru vinnig hardloop

(Die vertaalekwivalent van *hayai* is eintlik *vinnig wees*. Adjektiewe in Japannees verbuig soos werkwoorde.)

Vir die Japanneessprekende woordeboekgebruikers is dit aanvaarbaarder indien voorbeeldsinne waarin die attributiewe gebruik van byvoeglike naamwoorde geïllustreer word, eerste staan. Die ordening van die voorbeelde is vir die Japanneessprekende meer logies soos in die volgende artikel in die Afrikaans-Japannese woordeboek:

gesamentlik [xəsá:məntlək] (~e) b.nw. **1.** 合計の *goukei no* (total); **Die gesamentlike bedrag is ...** 合計の金額は... *Goukei no kingaku wa...* **2.** 総合した *sougou-shita* (joint); **Hulle het ~ opgetree.** 彼らは皆集まって現れた. *Kare-ra wa mina atsumatte arawareta.*

'n Afrikaanse leksikograaf sou die verandering miskien as onnodig beskou, maar die motivering hier is om die inskrywings sodanig te rangskik dat hulle so min as moontlik teen die taalgevoel van die gebruikers indruis, veral in die geval van aanleerderswoordeboeke.

In die Afrikaans-Japannese woordeboek onder bespreking is talle veranderinge ingevoer. Die makrostruktuur daarvan stem enigsins ooreen met dié van die *Afrikaans/Zoeloe-woordeboek* waarop dit gebaseer is, maar die mikrostruktuur is heeltemal hervorm. Waarskynlik is hier sprake van die "transkulturalisasie" van die aanbiedingswyse van 'n bestaande woordeboek.

3.5   Simbole en hakies

Terwyl die gebruik van tekens in Afrikaanse woordeboeke nog relatief beperk is, benut Japannese woordeboeke 'n verskeidenheid simbole en hakies as hulpmiddel by etikettering. *Sanseido's Daily Concise English-Japanese Dictionary* het byvoorbeeld ses tipes hakies met verskillende funksies wat in die toeligtingsmateriaal soos volg gebruik word:

17	「ハ ロ ン」 「メティエ」	**furlong, métier** などの訳語がそれぞれ引用符 (「 」)に包まれているのは専門語・職域語としてはその 訳語ないし音訳語が普通であることを示す.
18	[　　　]	この括弧は発音表記を示す. 発音については発音略 解 26〜37 及び発音記号一覧表を参照のこと.
19	〔　　　〕	この括弧は前出語の言い換えを示す: *be on* 〔*off*〕 *one's game* 調子がよい〔悪い〕=*be on one's* *game* 調子がよい; *be off one's game* 調子が悪 い.
20	(　　　)	この括弧は括弧内の部分が省略または添加できること を示す: **last'** の項の *at* 〈*long*〉〜=*at last* or *at* *long last.* また, **held** の項の, 「過去〈分詞〉」は 「過去及び過去分詞」を示す.
21	《　　　》	この括弧は言語区分, 注意すべき用法, 補足的説 明, 学習上必要と思われる解説, 著作者の作品中 特に有名なものを示す場合などに用いる: 《英》, 《米》, 《俗》, **will** の項の《単純未来》, **cress** の項の《食 用》, **Milton** の項の《*Paradise Lost*》など.
22	【　　　】	この括弧は術語, 専門語の表示に用いる: 【医】, 【ク リケ】など. なお, 本書中の略語表記については別に略 語解を参照のこと.

Deur die gebruik van verskillende hakies word die kategorie van die etiket dui-
delik geïdentifiseer en tegelykertyd word ruimte bespaar. Hierdie benadering
kan maklik in Afrikaanse woordeboeke gevolg word. Die volgende gebruik van
die etiket (**bet. ...**) "betekenis" in die *Woordeboek van die Afrikaanse Taal* demon-
streer die tipiese indruk van oorvloedigheid wat deur die gebruik van afkor-
tings geskep en deur die benutting van verskillende hakies voorkom kan word:

koeëlafsluiter Afsluiter (bet. 1, 2 , 4) wat 'n koeël (bet.1, 2a) is ...

As byvoorbeeld die hakies [ ] vir betekenissubklassifikasies ingevoer word, kan
die inskrywing met behoud van ondubbelsinnigheid verkort word. Vergelyk
die volgende voorbeeld:

koeëlafsluiter Afsluiter [1, 2, 4] wat 'n koeël [1, 2a] is ...

In hierdie geval moet die funksie van dié hakies natuurlik in die toeligtende
aantekeninge verduidelik word.

Uiteraard kan leksikograwe nie oormatig met hakies te werk gaan in 'n
woordeboek nie, want dit kan verwarring veroorsaak. Die gebruik van verskil-
lende simbole en hakies is 'n moontlike verswarende faktor by leesbaarheid,

omdat dit teksdigtheid verhoog. Dit is egter een van die opvallende kenmerke wat Afrikaanse en Japannese woordeboeke van mekaar onderskei.

3.6   Illustrasies

Nog 'n onderskeidende kenmerk van woordeboeke in die twee lande is die gebruik van illustrasies. Daar bestaan baie min Afrikaanse ostensiewe woordeboeke, terwyl daar in die meeste Japannese woordeboeke talle afbeeldings vir die doel van betekenisverklaring aangewend word.

Illustrasies is 'n noodsaaklike element, veral in die geval van 'n tweetalige woordeboek wat die tale van verskillende kulture en nature kombineer. Heelwat begripsprobleme by items in die Afrikaans-Japannese woordeboek wat nou ongeïllustreer is, sou sonder meer opgelos kon word deur afbeeldings te gebruik.

Die springbok byvoorbeeld is die nasionale dier van Suid-Afrika, en dié naamwoord het (nog) 'n relatief hoë gebruiksfrekwensie in Afrikaans. Maar die dier is in Japan onbekend en het gevolglik geen Japannese naam nie. Die enigste en maklikste manier om dit in die beperkte spasie van 'n woordeboekartikel eksplisiet te verklaar, is om 'n afbeelding daarvan te laat sien.

Een nuwe Japannese verklarende handwoordeboek, *Nihon-go-dai-jiten* (1994), toon in sy alfabetiese deel selfs heelwat *kleur*foto's. Dit skep weliswaar 'n indruk van kommersialisering, maar is beslis gebruikersvriendelik.

3.7   Formaat

3. 7.1 Grootte

Daar is ook belangrike verskille tussen die grootte van woordeboeke in Suid-Afrika en dié in Japan.

Sanseido's Daily Concise Kanji Dictionary (1995) bevat 664 bladsye (plus 'n bylae van 61 bladsye) maar is net 1.5 cm dik. Japannese uitgewers beywer hulle om woordeboeke dun te maak deur sover as moontlik dun papier te gebruik. Daarteenoor is Jan Kromhout se *Skoolwoordeboek / School Dictionary* (1995[21]) met byna dieselfde aantal bladsye (658) 4 cm dik.

In Japan is dit nie vreemd dat 'n persoon byvoorbeeld buite die kantoor 'n woordeboek met hom/haar saamdra nie. Handwoordeboeke moet dus draagbaar wees, met ander woorde, hulle moet so klein en dun wees dat hulle in 'n aktetas of 'n handsak pas. Aan die ander kant is dit in Suid-Afrika vermoedelik ongewoon vir persone om woordeboeke saam met hulle te dra. Draagbaarheid word dus nie in aanmerking geneem by die beoordeling van die woordeboekmark nie. Dit sou een rede wees waarom Afrikaanse woordeboeke dik en groot is in vergelyking met Japannese woordeboeke.

3.7.2   Omslag

Die feit dat die meeste kleiner Afrikaanse handwoordeboeke in sagteband is, is vir Japannese gebruikers onvanpas. 'n Woordeboek is in Japan 'n boek met status, en 'n soliede omslag word as gegewe aanvaar. 'n Sagtebandwoordeboek sal in Japan as 'n verbruiksartikel beskou word, en nie die vertroue van die gebruikers wen nie.

Die grootte en die omslag van woordeboeke word deur die houding teenoor woordeboeke in 'n land bepaal. Dit kan wees dat dieselfde woordeboek in twee lande in heeltemal verskillende formate gepubliseer word. Die kulturele agtergronde wat woordeboeke betref, veral dié van die tale wat in tweetalige woordeboeke verteenwoordig word, vorm deel van die ondersoekgebied van die leksikografie.

4.   Ten slotte

Die leksikografie is in Japan nie as 'n akademiese vak gevestig nie. Die praktiese kennis van woordeboekmaak word deur uitgewers as 'n soort bedryfsgeheim bewaar. Die enigste taalinstituut van die Japannese regering, *The National Language Research Institute*, beplan om 'n omvattende verklarende woordeboek op te stel en is op die oomblik besig om 'n databasis op te bou. Dié inrigting het egter geen ervaring van die leksikografie nie. In hierdie opsig sou instansies in Suid-Afrika wat met die opstel van woordeboeke gemoeid is, byvoorbeeld die Buro van die Woordeboek van die Afrikaanse Taal, deur samewerking goeie leiding kan bied.

Bibliografie

Collick, R.M.V., K. Hirata en S. Tanabe. 1995⁴. *Kenkyusha's New College Japanese-English Dictionary.* Tokio: Kenkyusha.

Gù, Y.-W. 543. *Yù-piang.* Jiànkang (Nánjing).

Halma, François. 1710. *Woordenboek der Nederduitsche en Fransche Taalen / Dictionnaire Flamand et François.* Amsterdam.

Inamura, S., G. Udagawa e.a. 1796. *Haruma-wage.* Edo (Tokio).

Jockin-La Bastide, J.A. en G. van Kooten. 1993. *Kramer's Nederlands-Engels Woordenboek.* Amsterdam: Elsevier.

Koine, Yoshio e.a. 1980⁵. *Kenkyusha's New English-Japanese Dictionary.* Tokio: Kenkyusha.

Kotzé, Ernst en Patrick Wela. 1991. *Afrikaans/Zoeloe-woordeboek.* Kaapstad: Tafelberg.

Kromhout, J. 1995²¹. *Skoolwoordeboek/School Dictionary.* Pretoria: J.L. van Schaik.

Kuukai. 9de eeu. *Tenrei-banshou-myougi.* Kioto.

Sanseido Jisho-henshuujo. 1990. *Sanseido's Daily Concise English-Japanese/Japanese-English Diction-ary.* Tokio: Sanseido.

Satake, H., T. Sakurai en S. Sugito. 1995. *Sanseido's Daily Concise Kanji Dictionary.* Tokio: San-seido.

Umesso, T., H. Kindaichi, A. Sakakura en S. Hinohara. 1989. *Nihon-go-dai-jiten.* Tokio: Kodansha.

Snijman, F.J. (Hoofred.). 1976. *Woordeboek van die Afrikaanse Taal.* Deel VI. Pretoria: Staatsdrukker.

Van Sterkenburg, P.G.J. en W.J. Boot. 1994. *Oranda-go-jiten.* Tokio: Kodansha.

Legal Terminology in African Languages[*]

Mariëtta Alberts, *National Terminology Services, Pretoria, South Africa*

Abstract:. This article deals with the project on legal terminology in the African languages. It focuses on terminology aspects relating to the coining of terms for the legal profession. Terminology development in South Africa has been hampered by a number of sociolinguistic factors. During recent years South Africa has seen considerable changes regarding the function of the indigenous languages, and this situation gave rise to an urgent need for terminology in various subject fields and knowledge domains. A demand arose for legal terminology in African languages. This need for legal terminology gave rise to the establishment of the Centre for Legal Terminology in African Languages. The aim of the Centre is to make legal terminology more accessible to the local indigenous population. The background of the legal terminology project is given and the various principles on which the working procedure is based, are discussed. Various aspects regarding the present project (such as financing, time-schedule, training and terminological problems encountered) are treated.

Keywords: LEGAL TERMINOLOGY, SOCIOLINGUISTIC FACTORS, TERMINOLOGY DEVELOPMENT, AFRICAN LANGUAGES, INDIGENOUS LANGUAGES, MULTILINGUALISM, SUBJECT FIELDS, TERMINOLOGY, TRANSLATION, INTERPRETATION, CRIMINAL LAW, CRIMINAL PROCEDURAL LAW, TERM

Opsomming: Regsterminologie in Afrikatale. Hierdie artikel handel oor die projek oor regsterminologie in die Afrikatale. Dit fokus op terminologiese aspekte wat verband hou met die skepping van terme vir die regsprofessie. Terminologie-ontwikkeling in Suid-Afrika is deur 'n aantal sosiolinguistiese faktore belemmer. Suid-Afrika het die afgelope jare drastiese veranderinge beleef ten opsigte van die funksie van die inheemse tale, en hierdie situasie het aanleiding gegee tot 'n dringende behoefte aan terminologie op verskeie vakgebiede en kennisterreine. 'n Behoefte aan regsterminologie het in die Afrikatale ontstaan. Hierdie behoefte het tot die stigting van die Sentrum vir Regsterminologie in Afrikatale gelei. Die doel van die Sentrum is om regsterminologie meer toeganklik vir die plaaslike inheemse bevolking te maak. Die agtergrond van die regsterminologieprojek word verskaf en die verskillende beginsels waarop die werkswyse gebaseer is, word bespreek. Verskeie aspekte in verband met die huidige projek (soos finansiering, tydsbepaling, opleiding en terminologiese probleme wat ervaar is) word behandel.

Sleutelwoorde: REGSTERMINOLOGIE, SOSIOLINGUISTIESE FAKTORE, TERMINOLOGIE-ONTWIKKELING, AFRIKATALE, INHEEMSE TALE, MEERTALIGHEID / MULTILIN-

---

GUALISME / VEELTALIGHEID, VAKGEBIEDE, TERMINOLOGIE, VERTALING, TOLKING, STRAFREG, STRAFPROSESREG, TERM / VAKWOORD

Introduction

There are several factors influencing the terminological development of South African languages. Most of these are sociolinguistic factors. Terminology development in South Africa has been retarded by a number of ideological, historical, and educational factors, "the most fundamental of which are the language policies adopted in the Republic of South Africa (RSA)" (Mtintsilana and Morris 1988: 109).

With English and Afrikaans being the official languages in the RSA up to 1994, the State actively supported terminology development in Afrikaans to try to ensure parity with English (from which the international terms usually originate or enter South Africa). The African languages enjoyed neither official status nor any officially recognised functional status until the introduction of mother-tongue primary school education in 1953.

Jansen (1992: 1) states: "Recent years in South Africa have seen considerable changes. There has been a complete restructuring of the labour force and a greater involvement of the different population groups in matters of own and general interest. The fields of science and technology have become more accessible and various careers which were traditionally dominated by certain population groups are now infiltrated by people from other groups. All these factors have made increasing demands on the various indigenous languages spoken by members of the different population groups in South Africa." More than ever, this situation gave rise to an urgent need for terminology in various subject fields and knowledge domains.

In the new democratic South Africa, the Constitution provides for multilingualism and the development of the linguistic heritage. South Africa now has eleven official languages.

Multilingualism in South Africa is a sociolinguistic fact to be taken seriously. A large proportion of the indigenous population of South Africa can only be reached by means of indigenous languages. Information flow is prevented from being established by factors such as low literacy rates and terminologically poorly developed African languages (Fourie 1993: 132).

Political and societal changes have largely transformed the functional role of the African languages. The national and self-governing states that were created by the previous government led to the adoption of certain African languages either as the official language or as the *de facto* dominant language of these states. A substantial demand for terminology creation arose from this, since various business matters (civil service departments, local administrative bodies, courts of law, etc.) were conducted in the vernacular (Mtintsilana and Morris 1988: 109).

Origin of the Project

During an annual general meeting of the South African Institute for Translators (SATI) in 1985, translators, interpreters and other officials of the then self-governing states in South Africa indicated that they encountered numerous problems when dealing with legal terminology. They came across problems when translating legal documents, when compiling legislation for their various states, and when interpreting in court. These problems occurred since legal terminology in African languages either did not exist or were inadequate.

It was realised that, although English and Afrikaans were at that stage the official languages of South Africa, "the different indigenous languages could not be ignored but had an important role to play in legal procedure" (Jansen 1992: 2). The problem, however, was that the indigenous languages could not meet the demands placed on them, as legal terminology in these languages was either altogether lacking or, where present, not sufficiently complex to deal with the legal concepts of modern developments.

The South African Translators' Institute requested the National Terminology Services (NTS) to investigate the situation. The needs assessment done by NTS indicated a need for terminology in criminal law and related domains.

In 1987 a working group was formed. This working group later developed into the Committee for Legal Terminology in African Languages. Since 1996 this Committee is known as the Centre for Legal Terminology in African Languages. The Centre has its own constitution and is managed by an Executive.

The Objectives of the Centre

The objectives of the Centre are to:

* compile legal terminology lists, legal dictionaries and other related products in African languages, Afrikaans and English;
* undertake research on legal concepts and determine how these can at best be designated and systematised, to promote research and further terminology work and to make the results thereof available; and
* obtain the co-operation of experts and interested persons and promote common interests.

The main objective of this Centre is thus to make legal terminology more accessible to the local indigenous population; "not only to make it more available but also to make it more comprehensible" (Jansen 1992: 2).

The aims of the Centre are revealed in its coat of arms. The **bridge** symbolises an attempt to reach out to other fellow citizens by means of translation, the **stars** bring the light of knowledge to the existing darkness and the **scale** refers both to the justice and to the function of an interpreter to ensure that the original (source language) and the translation (target language) convey the

same message. The translation must be a mirror image of the language from which the translation is done. Gold and blue, the colours of the coat of arms, are two typical South African colours and the counterbalance in colour rein-force the source- and target-language image.

Coat of arms of the Centre for Legal Terminology in African Languages

Working Procedure

Background

The project started as a result of the needs expressed by Sepedi interpreters who experienced difficulty in interpreting in court due to a lack of legal termi-nology in Sepedi. Although this project was initially aimed at meeting the needs of the Sepedi interpreters, the initial working group intended from the start to incorporate all the indigenous languages in the project. This is one of the reasons why the name changed from Committee to Centre. Various Com-mittees dealing with the various indigenous languages can now be created under the auspices of the Centre. The Centre caters for any language group that feels the need for legal terminology in a specific language.

The members of the Centre and interested persons meet quarterly to dis-cuss general matters. Before the commencement of the main meeting, the committee for Sepedi usually convenes to discuss terminological problems encountered in the intermediate period.

The Centre for Legal Terminology in African Languages consists of repre-sentatives from different disciplines. The members of the Centre are all volun-tary workers. At present there are several members of the legal professions (e.g. lawyers, magistrates, court interpreters, etc.), members of the Justice Training College, interpreters, translators, academics (of the Departments of Law and African Languages of various universities), Language Board members, mem-bers of the Provincial Language Committee, linguists, terminologists / termi-

nographers, researchers, anthropologists, etc. There are also representatives from the South African Broadcasting Corporation (SABC), and the Departments of Justice and of Education.

"Although the Centre may seem overburdened with academics," alleges Jansen (1992: 4), "the idea is nevertheless to compile terminology lists that can be used in everyday dealings with law." It is only possible to communicate concepts clearly when the subject is known properly. Only then can the message be conveyed to laypersons.

The target group, for which this terminology is intended, includes legal practitioners, interpreters, translators, compilers of legislation, students of law and even the man in the street. The Centre intends this dictionary to be used not only by speakers of the African languages, but also by English- and Afrikaans-speaking South Africans — either to clarify the meaning of a term in their own language, or to communicate more successfully with someone speaking an African language.

In a recent letter from the Department of Justice, the Director-General stressed the value of the project for the legal profession and emphasised the importance of compiling and publishing this legal dictionary in the eleven official languages. The Centre enjoys the blessings of the Departments of Justice and of Education of the central government and of the Northern Province government.

In 1993 the Centre became a member of the International Academy of Language Law situated in Montreal, Canada. The Centre is also a member of CoHSSSA — the Consortium of Human Sciences Societies of Southern Africa. The Centre has close co-operation with the African Association for Lexicography (Afrilex), since various members of the Centre are also Afrilex members.

Principles

At this stage, the Centre for Legal Terminology in African Languages only concentrates on providing African language equivalents in the subdomains of *criminal law* and *criminal procedural law* because of the dire need in these legal fields.

At the outset the undertaking seemed boundless, so the Centre decided to demarcate it by limiting the scope of the work to those legal terms relating to criminal law and criminal procedural law, as these legal fields are traditionally best known to the African population groups as a result of their own tribal penal organisation. Because the Government of Lebowa was the first to draw attention to the deficiencies already mentioned, Sepedi (Northern Sotho) was chosen as the first indigenous language into which terms would be translated. The Centre is presumptuous enough to plan terminology lists for other African languages (Jansen 1992: 3). Any interested language group can use the Centre as the umbrella organisation when working on criminal law and criminal pro-

cedural law in any of the other official languages. The Centre will make the data already collected available to other language groups. It also plans to cover other legal fields when the first project is completed.

Procedure

The Centre tries to fulfil the role it has undertaken by working in different phases:

First Phase:

* The first phase comprises the compilation of lists of legal terms in English in the specified domains.
* The legal expert of the Centre excerpts these legal terms using English as the source language.
* The English terms are provided with fairly simple definitions and examples of how the terms should or could be used. The definitions must be kept simple yet should clearly explain the legal concepts.
* The English terms, with their definitions and example sentences, are then rendered into Afrikaans (that is, the Afrikaans equivalents, definitions and example sentences are provided).

Second Phase:

* During the second phase of the work, the English / Afrikaans term list is edited to conform with terminological and terminographical standards as set by modern dictionary compilers. The know-how for this work comes from the terminologists / terminographers who are members of the Centre.
* Thereafter, the information is entered into a computer and adapted even further in accordance with the requirements of the database system.

Third Phase:

* The third phase of the process involves the translation of the various English and Afrikaans terms, definitions and examples into an indigenous African language.
* At this stage, the terms with their definitions and / or examples are given to the Sepedi Committee, who coins / designates / denotes Sepedi terms and other relevant information according to the concepts. Problematic concepts or terms are dealt with at the quarterly meetings.

Fourth and further Phases:

*   As soon as other language groups (that is, other than Sepedi) decide to form their own committees the English / Afrikaans / Sepedi terms, definitions and / or examples will be made available to them to render the terms and other information into the relevant language(s).

The Present Project and Problems Encountered

Financing

In 1993 the Centre received a donation from XB Brokers. This donation kept the Centre going and enabled it to purchase a computer which assisted the progress. In the same year, the Centre received a complimentary *Trilingual Legal Dictionary* by V.G. Hiemstra and H.L. Gonin, donated by Juta & Co Ltd. The project was stimulated by this donation, since the dictionary forms, to a certain extent, the basis of the work.

Fortunately, the Department of Arts, Culture, Science and Technology (DACST, previously Department of National Education) also regards the project of the Centre as a worthwhile effort which deserves financial backing. In 1995 DACST presented the Centre with a grant-in-aid that enabled the Committee, *inter alia*, to appoint assistants and thus make better progress with the project. When presenting the cheque, the Director-General of DACST expressed the hope that the Centre will in future be able to branch out and add the equivalents of other African languages.

Computerisation

The computerisation process was unfortunately hampered by the working procedure. The legal expert at UNISA used to excerpt the legal terms from the *Trilingual Legal Dictionary*. These terms were keyed into her computer at UNISA and a printout was given to the Sepedi Committee. The Sepedi Committee added the Sepedi definitions and term equivalents on a separate computer or typewriter. After the terms were discussed at a quarterly meeting the English / Afrikaans and Sepedi terms were again keyed into a computer at the Department of African Languages at the University of Pretoria, where an assistant was available to do the job.

With the grant from DACST, it was possible to buy another computer. The computers are located at UNISA. As was mentioned before, the Centre is now also in a position to employ assistants to excerpt terms. The legal terms and definitions are taken from various sources such as dictionaries, textbooks, study material, etc. The assistants enter all relevant information directly into the

computer. The senior researcher (a legal expert) now writes simple English and Afrikaans definitions by taking all the information on a specific concept into account. The printouts are then given to the Sepedi researcher to add the Sepedi definitions and equivalents. The latter liaises with the Sepedi Committee.

At present the data is entered into the *Multiterm* database management system of the NTS. This system will also be used to enter terminology in the other languages as soon as the committees are established and they have started with the compilation of their lists. The reason is that the *Multiterm* system can handle multilingual data and it is necessary to link the various languages by means of a divisor language (or a bridging language) such as English or Afrikaans. While every committee can work on its own in its decentralised geolinguistic area, it will be possible to link various languages if need be, in order to eventually publish a multilingual legal dictionary.

Time-schedule

Although the Centre is at this stage only dealing with a small portion of the legal domain, it is still a huge task. The compilation of a dictionary is time-consuming and can last for years — even if work on it is done on a full-time basis. The people working on this dictionary do it on a voluntary, part-time basis. The Centre would, however, like to expedite the process. Thus it was decided at the meeting in May 1996 that another *modus operandi* will be followed. Only the core terms of *criminal law* and *criminal procedural law* would be excerpted. By doing this, an overview of the subject field can be gained, the problematic terminology can be dealt with first, and a list of terms from A-Z can be provided. This basic list can be distributed among subject specialists and linguists to get their input. As soon as this list is evaluated and edited, it can be given to the other language groups to start off with the basic terms. The assistants can immediately start revising of this basic list by adding the rest of the excerpted terms. The supplementary terms are not necessarily core terms, but they are just as important to the subject field as a whole. The various language committees can then immediately start work on the supplements and provide them with term equivalents and definitions in the relevant languages.

Training

The Centre also devotes some time to training. Before every quarterly meeting of the Centre, the Committee for legal terminology in Sepedi convenes to discuss problems they encountered while denoting legal concepts in the months preceding the meeting. The Committee members discuss a specific concept to outline the exact meaning and then they coin a Sepedi equivalent.

At the moment, African language terminologists of the National Termi-
nology Services (NTS) attend these working sessions. As part of their training
and with a view to elaborating the languages, NTS terminologists coin equiva-
lents in their mother-tongue when the Committee discusses problematic cases
in Sepedi. (NTS employs mother-tongue speakers of all the African languages.)
This is a very valuable exercise for all concerned. During these sessions, termi-
nologists learn in a practical environment how important it is to know and
make sure of the meaning of concepts before even thinking of providing
equivalents in the relevant language.

Terminology

One should realise that there are existing legal terms in all the African lan-
guages. Every language already has several legal terms at its disposal, since
every one of the language groups has a comprehensive legal and court system.
The various languages make a clear distinction between different legal aspects.
Several legal terms are already documented, explained and published by either
lexicographers, researchers in the legal field, or anthropologists.

The Centre can, however, not totally rely on existing documentation. One
problem is that existing terms in different languages do not necessarily denote
the same concept and are therefore not term equivalents. This gives rise to con-
fusion. Terms that are merely translated forms, sometimes appear to be more
unintelligible to African language users than the original Afrikaans or English.

Indigenous law contains elements of both *criminal law* and *civil law*. There is
no separate criminal law and criminal procedural law. To unite these principles
and procedures with that of the Roman Dutch Law and British Common Law is
very difficult. It is of vital importance to write simple definitions to explain the
legal concepts. It is of equal importance to ensure exact communication by
retaining the legal meaning. A further problem facing the terminologist is that
there are no explanatory legal dictionaries on the South African legal system
available to consult.

The Centre has to ensure that the definitions will be acceptable to the legal
professions, yet simple enough to be understood by people at grassroots level.
If not, it will give rise to controversy. In court, the interpreter, who does not
have a thorough knowledge of the source language, can easily misinterpret the
Afrikaans term "breekgoed" (English "crockery") for "gebreekte goed" (English
"broken goods"). The Sepedi exclamation "Sebatakgomo!" (English "Help!") can
be interpreted by the Sesotho interpreter as "Ke batla kgomo" (English "I want
an ox"). There was a an actual case where a Sepedi woman who was raped,
pleaded that she cried for help, but the rapist was acquitted because of the mis-
interpretation by the court interpreter. The terminologist designating concepts
by means of terms must be sure of the underlying meaning of the concept
before coining a term.

The terminologist working on this project must have a sound background in legal approach and court procedure, as well as a good command of Afrikaans and/or English and the language of the relevant ethnic group. It is important to realise that this dictionary will be the channel through which legal and linguistic information will flow. This dictionary should also contain extra information on usage. It is not only a glossary of legal terms, but should guide the user on language and style usage, on the difference in the interpretation and translation of certain concepts, misinterpretation, misstatements and misrepresentations.

The Committee is careful not to incorporate offensive and/or sensitive items in the dictionary. Terminology is in nature abstract and exact, and it is therefore reasonably secure against emotive connotations that can be attached to words. To a certain extent, the domain overlaps with social and political subject fields. Therefore the Centre has to beware of incorporating potentially offensive and sensitive terminology, such as:

gewoonteverbintenis or *gebruiklike verbinding* (English "customary union")

> <u>degrading term</u> for:

gewoonteregtelike huwelik or **inheemsregtelike huwelik** (English "customary marriage" or "indigenous marriage") (cf. Act 38 of 1927 and Act 76 of 1963)

deelgenoot van 'n gebruiklike verbinding (English "partner of a customary union")

> instead of:

eggenoot van 'n inheemsregtelike huwelik (English "spouse of a customary marriage")

toordokter (English "witchdoctor")

> <u>degrading term</u> for:

tradisionele dokter (English "traditional doctor")

> <u>wrong term</u> for:

towenaar (Engels "sorcerer")

The Centre has to deal with various aspects concerning the denotation of concepts. There are various ways of coining or creating terms in the target language(s). It is in some cases relevant to coin a new legal term in the relevant African language. Sometimes the meaning of an old (existing) word can be extended to incorporate the meaning of a specific concept. It is sometimes necessary to make use of transliterations where the Greek or Latin (that is, the classic) stem of the term is being used to create a target-language equivalent. When dealing with the various African languages, it is interesting to look at the harmonisation or unification of the languages when coining legal terms. The terminologists have to look into the possibilities of harmonisation in dealing

with the terminology of a project such as this, for which such an urgent need exists.

The Centre for Legal Terminology in African Languages encounters interesting problems while compiling this dictionary. Although Roman Dutch Law forms the basis of the South African legal system the legal system is also greatly influenced by, *inter alia*, British Common Law.

According to South African law (traditional Roman Dutch Law) the term **defamation** (Afrikaans "laster") is used for the *criminal offence*. No distinction is made whether *written* or *by word of mouth*. According to British Common Law two terms are being used:

> **libel** written
> **slander** by word of mouth

The difference in meaning between these legal systems and the lack of explanatory dictionaries pose a big problem when determining the exact meaning of a concept. It is therefore necessary to excerpt legal terms from existing documentation and firstly to define these concepts accurately in the South African context before it is possible to denote them with equivalents in any of the African languages. A term with an exact meaning will promote exact communication. By supplying definitions for the various concepts, the Centre for Legal Terminology in African Languages tries to standardise the legal terminology of *criminal procedural law* and of *criminal law*.

It is necessary to remember that in South Africa, Afrikaans and the other indigenous languages are all minority languages compared to a dominant language like English that is used world-wide. Minority languages like Afrikaans and the African languages find it very difficult to create a scientific and technical terminology next to the existing terminology of the dominant English language. There are, as previously discussed, certain political and economic factors inhibiting or impeding the coinage and penetration of terms in the minority languages, as it is very difficult for minority languages to compete with a dominant language. One should try to understand that it is not only difficult to communicate without proper knowledge of the various indigenous languages and lack of training in terminological principles. There are also several other difficulties and obstacles that have to be dealt with by terminologists if they want to provide acceptable multilingual terminology.

Terminologists who have to supply term equivalents for English terms, have a variety of possibilities, namely

* *loan words* (from English, Dutch, Flemish, German, French ... mainly European languages) (affidavit / affidavit / afidavite; abortus / abortus / abortuse)
* *transliteration* (using the Greek or Latin stem to coin an equivalent) (**addendum** L neutral of *addendus*: addendum / toevoeging / adentamo

190   Mariëtta Alberts

(tlalletso); **bail** ME *bail, baille,* from MF *bail,* from *bailler* to give, deliver, from L *bajulare* to bear a burden, keep in custody, from *bajulus porter,* load carrier: bail / borg / peile)

* *extension of meaning* (using existing words but broadening the meaning to encompass the new concept) (abet / aanmoedig / go thusana le go thusa bosenying; abductor / ontvoerder / motshabisamongwe)

* *neologisms* (claim / kleim / kleime; adopt (child) / aanneem (kind) / adopta (ngwana))

In all these cases, the principles of harmonisation (e.g. search for an existing word in the various dialects or related languages) can be applied. It should also, however, be established whether this is what the users want.

As was mentioned before, the terminologist has to apply specific terminological principles when denoting concepts. He/she also has to apply certain linguistic principles. No terminologist can coin a term if he/she does not know the basic word-formation principles of a language. Unfortunately, the basic word-formation principles for all African languages have not yet been established. There is a dire need for such principles. The National Terminology Services would like to form a working relationship with any person in any language group who has the linguistic background and knowledge to assist the office with this research.

The work of the terminologist is furthermore closely linked to language planning and language policy. The terminologist is always consulting, be it with subject specialists, linguists or dictionary users, on all registers of terminology usage. It is expected of the terminologist to be aware of changes that may occur across a wide spectrum of the linguistic field — from scientific and technological areas to grassroots level. In the new dispensation, terminology must also be supplied at grassroots level, especially for education purposes in order to facilitate exact communication. The terminologist is dependent on the comments and collaboration of specialists in all spheres of life. Terminologists would like language planners to consult with them when deciding on language policies. All South Africans are experiencing exciting times regarding language, and now is the time for all kinds of language professionals to consult one another before major decisions are taken. This is one of the reasons why the Centre aims to consult with subject specialists, linguists and mother-tongue speakers before publishing any dictionary. The approval of all new terms by a Committee of mother-tongue speakers is vital and fully appreciated by the Centre.

Close co-operation between language planners and terminologists is vital because one of the objectives of language planning is terminology unification. Nahir (1984: 308) defines terminology unification as "establishing unified terminologies, mostly technical, by clarifying and defining them, in order to reduce communicative ambiguity ..." According to Cluver (1989: 247) this objective coincides with the standard view of the objectives of terminographical

work on developed standard languages. The focus is here on collecting terms and standardising them by means of a definition, rather than on creating new terms.

Conclusion

The Centre for Legal Terminology in African Languages aims not only to compile and publish a multilingual explanatory legal dictionary, but also to train terminologists. The Centre aims to be instrumental in the coining of legal terms to promote exact communication in all spheres of the legal profession and in all languages.

At the moment the Centre has only one Committee working in Sepedi. Interest was, however, shown by the Xhosa Dictionary Project at Fort Hare, and the Centre hopes to establish a Committee for Legal terminology in isiXhosa. Representatives of the other languages, such as Sesotho, isiZulu and Tshivenda already attended meetings and it is hoped that they will also be able to start with projects.

At every meeting of the Centre the chairperson declares the aim of the Centre to extend the current English / Afrikaans / Sepedi project to incorporate the other official languages. The Centre can, however, only extend an invitation to the other language groups — it does not intend to force a legal terminology project on any language group. The stakeholders of each language should decide for themselves whether they want to collaborate in this project.

The Centre is a young but dynamic institution and the members are working hard to fulfil their mission. Although the members of the Centre are very proud of the work already done, they would like to accelerate the pace, but for that they need the input and collaboration of the various stakeholders.

References

Cluver, A.D. de V. 1989. *A Manual of Terminography*. Pretoria: HSRC.

Fourie, D.J. 1993.The Rejection of Terminology: Observations from the African Languages. *South African Journal for African Languages* 12(1): 11-15.

Jansen, Marthie. 1992. *Double, Double Toil and Trouble — The Problems Encountered when Compiling a Legal Dictionary in African Languages*. Paper read at the ILLA Conference at the HSRC, Pretoria.

Mtintsilana, Priscilla N. and Rose Morris. 1988. Terminography in African Languages in South Africa. *South African Journal of African Languages* 8(4): 109-113.

Nahir, M. 1984. Language Planning Goals: A Classification. *Language Problems and Language Planning* 8(3): 294-327.

Lexicography and Sign Language Engineering: The Zambian Experience

Vincent M. Chanda, *Department of Literature and Languages, University of Zambia, Zambia*

Abstract: Sign language as used by deaf communities, is a real and fully-fledged human language, not based on any spoken language, and not universal in the sense of there being only one sign language worldwide. A deaf community is a linguistic minority, but a linguistic minority with special linguistic needs because of the very nature of sign language. In Zambia, like in the vast majority of other Third World countries, the linguistic needs of deaf and hard-of-hearing people have been ignored. This article examines the genesis and implementation of a dictionary project for sign language, the Zambian Sign Language Dictionary Project, regarded as a first step towards the development of a Zambian National Sign Language. The article highlights the specificity of sign language lexicography.

Keywords: AMERICAN SIGN LANGUAGE, ARTICULATED LANGUAGE, BORROWING, DEAF, HAND SHAPE, HARD-OF-HEARING, ICONICITY, INDIGENOUS SIGN, LOCATION, MOVEMENT, ORIENTATION, SIGN, SIGN LANGUAGE, SIGN-WORD SEARCH SYSTEM, WORD-SIGN SEARCH SYSTEM

Opsomming: **Leksikografie en gebaretaalontwikkeling: Die Zambiese ervaring.** Gebaretaal, soos deur dowe gemeenskappe gebruik, is 'n ware en volwaardige mens-like taal wat nie op enige gesproke taal gebaseer is nie, en nie universeel is in die sin dat daar slegs een gebaretaal ter wêreld is nie. 'n Dowe gemeenskap is 'n taalkundige minderheid, maar 'n min-derheid met spesiale taalbehoeftes juis weens die besondere aard van gebaretaal. In Zambië, soos in die oorgrote meerderheid van ander Derdewêreldlande, is die taalkundige behoeftes van dowe en hardhorende persone geïgnoreer. Hierdie artikel ondersoek die ontstaan en verwesenliking van 'n woordeboekprojek vir gebaretaal, die Woordeboekprojek vir Zambiese Gebaretaal, wat beskou word as die eerste stap na die ontwikkeling van 'n Zambiese Nasionale Gebaretaal. Die artikel beklemtoon die spesifiekheid van gebaretaalleksikografie.

Sleutelwoorde: AMERIKAANSE GEBARETAAL, GESPROKE TAAL, ONTLENING, DOOF, HANDVORM, HARDHOREND, IKONISITEIT, INHEEMSE GEBAAR, POSISIE, BEWE-GING, ORIËNTASIE, GEBAAR, GEBARETAAL, GEBAAR-WOORD-SOEKSISTEEM, WOORD-GEBAAR-SOEKSISTEEM

1. Introduction

The present article examines the genesis and implementation of a sign lan-

guage lexicographic project, the Zambian Sign Language Dictionary Project, a project of the Zambian National Association of the Deaf (ZNAD). The editorial work was supervised by the writer and the product, i.e. the dictionary, is currently in the press (Co-op Printing, P.O. Box 50208, Lusaka, Zambia).

A deaf community in any country is a linguistic community with special linguistic needs including lexicographic needs which the country must endeavour to satisfy since all citizens have the same linguistic rights. The project was launched as a result of the realization that the linguistic rights of the Zambian deaf people, which include the development of a national sign language (since the Zambian deaf cannot use the national official language, English, in its spoken form), have been neglected or ignored.

Where there is a deaf community — through the establishment of a deaf school, deaf club, deaf centre, etc. — an elaborate manual system of communication emerges and eventually develops into a fully-fledged language comparable to articulated language in both scope and complexity (Chanda 1997: I, Serpell and Mbewe 1990: 281). Such a language is known as sign language (SL) in the linguistic literature.

Contrary to one widespread belief, SL language is not universal in the sense of there being only one SL worldwide. This belief is based on the fact that many SL signs are iconic, or pictorial, in that they somehow describe depictively what they mean or stand for. That SL is not universal in the sense defined above is evidenced by the existence of several SLs (American Sign Language, British Sign Language, Australian Sign Language, Chinese Sign Language, French Sign Language, Kenyan Sign Language, etc.) most of which are not mutually intelligible. At this juncture, three points must be made regarding iconicity in human language as a way of explaining why there exist several SLs which are not mutually intelligible. Firstly, many artefacts (e.g. houses) do not have the same set of features worldwide nor are all actions (e.g. building a house) performed in the same way worldwide. Consequently, given some artefact or action X with some feature F such that X is [+F] in culture C1, but [-F] in some other culture C2, a deaf community belonging to C1, but not to C2, may "mimic" [+F] to refer to the artefact or action X, in which case a deaf person belonging to culture C2 will not understand the sign. Secondly, even when some artefact or action X has the same set of features in two different cultures, C1 and C2, deaf people belonging to C1 and C2 will not necessarily mimic the same feature from X. The latter point reminds us of the fact that articulated languages do not necessarily onomatopoeically depict the same sound in the same way. Lastly, within the same community a feature F which is mimicked to refer to some artefact or action X may be abandoned due to cultural change while the sign is retained so that the sign is iconic only diachronically.

It is also worth noting that LSs are not based on spoken languages, as is evidenced by the fact that, while American English and British English are mutually intelligible, American Sign Language (ASL) and British Sign Language (BSL) are not.

Minimally an SL sign consists of four components, whose functions are the same as the function of distinctive features of phonemes in articulated language, namely:

(a)   **hand shape**, i.e. the configuration of the active hand during the production of the sign,

(b)   **location**, i.e. the place of articulation of the sign (e.g. the chin),

(c)   **movement**, i.e. the movement of the hand(s) during the production of the sign, and

(d)   **orientation**, i.e. the direction in which the palm(s) face(s) (e.g. downward, upward, left, right).

2.   Historical Background to the Zambian Sign Language

The first school for deaf children in Zambia was opened in 1955 by the Dutch Reformed Church at their Magwero Mission in the Eastern Province. Since then several deaf schools, deaf school units, deaf clubs and deaf centres have been established countrywide and a national deaf association, the Zambian National Association of the Deaf (ZNAD), has been formed..

At the Magwero Mission School for the Deaf, the missionaries were essentially "oralists", that is to say, they were teaching the deaf somehow through speech and speechreading and writing, but the children soon developed an indigenous SL through, in the words of Serpell and Mbewe (1990: 283), cross-cultural borrowing. Nowadays, the medium of instruction for the deaf in Zambia is to a large extent what is known as "total communication", i.e. a system whereby any means of communication (signing, gesturing, writing, pictures, etc.) is used, but American Sign Language (ASL) has undoubtedly become the most important medium of instruction and means of communication in deaf communities in Zambia. This is due to a number of teaching and promotional activities initiated by Mr Mackenzie Mbewe, a deaf person who holds a B.A.(Ed.) (University of Zambia, 1978) and a B.Phil. in Special Education (University of Birmingham, 1984). After studying at the Ibadan School for the Deaf, Nigeria, for two years (1968-1970), he immediately started teaching and promoting ASL in Zambia. Mr Mbewe is currently the Executive Director of the ZNAD.

The increasing role played by SL in deaf education in Zambia and the realization that deaf people in the country use indigenous signs, most of which are similar, led to the decision to compile a *Zambian Sign Language Dictionary*, the core of which would be made of standardized indigenous signs while other signs would be borrowed from ASL. Such a dictionary would be the basis for the development of a Zambian Sign Language (ZAMSL), an idea which was the result of the rejection of "oralism" in favour of "manualism", because oralism

limits access to knowledge, and impairs the relationship between deaf children and their parents while SL gives deaf children a normal development (Wallin 1993: 3).

3.   The Zambian Sign Language Dictionary

3.0   General

As already stated, the *Zambian Sign Language Dictionary* is a product of the Zambian National Association for the Deaf (ZNAD). The lexicographical work was wholly funded by the Finnish Association of the Deaf (FAD).

3.1   Objective

The object of the Zambian Sign Language Dictionary Project was to compile an SL dictionary including:

(a)   as many standardized signs as possible, and
(b)   a sizeable ASL number of signs needed in deaf education in Zambia.

3.2   Target Users

The dictionary is primarily targeted at deaf schools and deaf school units but also at the clergy and government ministries such as the Ministry of Health, whose services are essential.

3.3   Personnel

To implement the project, a Sign Language Committee was set up comprising:

(a)   two academic members of staff of the University of Zambia, including the writer,
(b)   three hearing officials from the Ministry of Education, and
(c)   six deaf members from the Zambian National Association of the Deaf (ZNAD).

The project was managed by the ZNAD and benefitted from services provided from time to time, but was aided throughout the duration of the project by SL experts from the Finnish Association of the Deaf (FAD).

3.4   Special Equipment

Because of the nature of the language and work involved, the following pieces of equipment were used:

(a)   a video camera,
(b)   a still picture camera,
(c)   a heavy-duty tripod,
(d)   a colour television set, and
(e)   drawing sets.

3.5   Methodology

3.5.1   Data Collection

All the signs were collected from deaf and hard-of-hearing people whose competence in written English was adequate for the purposes of the project. Most informants were at deaf schools, deaf school units, deaf clubs and deaf centres. The signs were collected during field work in several districts throughout the country with the use of a video camera and a still picture camera.

To collect signs, the field workers used a list of English words and sometimes phrases and the informants were asked to supply the equivalent indigenous signs. However, some informants were not always able to say whether the signs they gave, were indigenous or borrowed from ASL but this was not problematic since the researchers know ASL. The result was that three types of signs were collected:

(a)   indigenous signs,
(b)   ASL signs, and
(c)   modified ASL signs.

Most signs were indigenous signs. It is important to note that, for the purposes of the project insofar as it is an aspect of language corpus planning, ASL signs and modified ASL signs were retained and regarded as ZAMSL signs borrowed from ASL.

3.5.2   Lemmatization

To decide which signs should be lemmatized, the signs collected during fieldwork had first to be ascertained at national conferences attended by the members of the Zambian Sign Language Committee as well as deaf and hard-of-

hearing delegates from all over the country. The inclusion policy was as follows:

(a)   where there was only one sign for a given English item retained, the sign had to be lemmatized, and

(b)   where there was more than one sign retained for a given English item (which was generally the case for indigenous signs), it was agreed to lemmatize the sign which displayed the greatest geographical distribution except when two signs were thought to have practically equal geographical distribution, in which case both signs were lemmatized as synonyms.

3.5.3   Processing

The lemmas, i.e. those signs which had been selected for inclusion in the dictionary, were processed manually. The processing consisted in:

(a)   drawing every sign,
(b)   giving the English gloss,
(c)   giving grammatical information about the English gloss,
(d)   giving a verbal description (in English) of how the sign is to be produced, and, where necessary,
(e)   giving cross-references.

3.5.4   Evaluation and Editing

All the processed signs had to be evaluated at national conferences which, as explained above, were attended by the members of the Zambian Sign Language Committee and delegates from all over the country. The final work, i.e. editing, was supervised by the writer.

3.6   Type of Dictionary

With regard to the number of languages involved, dictionaries are generally either monolingual or bilingual. However, there is no monolingual SL dictionary in which signs belonging to a given SL (e.g. ASL), are explained in signs belonging to the same SL. All published monolingual SL dictionaries involve an articulated language (e.g. English). In such dictionaries, the lemma is a lexeme or, sometimes, a phrase or a morpheme in an articulated language and the lemma is followed by a verbal description (in the same articulated language) of how the equivalent sign is to be produced. However, the verbal description is

always followed by a photograph or, more often, a drawing of the sign, but such photographs or drawings are to be regarded only as "illustrations". Articles from a monolingual SL dictionary from Sternberg (1990: 34) are shown in Appendix A.

The *Zambian Sign Language Dictionary* is a bilingual dictionary with the following features:

(a)   the lemma consists of a drawing representing a sign;
(b)   the lemma is followed by:
 (i)   an English gloss,
 (ii)   the literal English gloss of a ZAMSL sentence containing the lemma (Exag, i.e. example of SL grammar),
 (iii) the actual English gloss of the example in (ii) (Enlat, i.e. English language translation), and
 (iv)   a verbal description in English of how the sign is performed.

Among pieces of information which are found optionally, it is important to mention facial expression (FE). Examples of articles from the *Zambian Sign Language Dictionary* are shown in Appendix B.

3.7   The Outer Access of the *Zambian Sign Language Dictionary*

3.7.1   General

With reference to Hausmann and Wiegand (1989), Gouws (1996: 19) writes:

> The dictionary user has to get to the information he needs, and the structure of the dictionary should help him to reach that information. This component of the dictionary is known as the *access structure* ... The outer access structure guides the user up to the lemma, whereas the inner access structure helps the user to reach the specific information category within an article. The access structure can also be described as the *search path* for the dictionary user.

For dictionaries in articulated language, the outer access structure is generally the alphabetical ordering of the lemmas. Likewise, in the type of monolingual SL dictionaries described in the previous section, the outer access structure is the alphabetical ordering of the lemmas.

The *Zambian Sign Language Dictionary* has been designed with two search systems, namely:

(a)   the **sign-word search system**, which enables the user to identify the English equivalent(s) of a sign, and

(b)   the **word-sign search system**, which enables the user to locate the sign for a given English item.

3.7.2   The Sign-Word Search System (SWSS)

As explained above, the sign-word search system enables the dictionary user to identify the English equivalent of the sign he/she has seen. The SWSS is based on handshapes and locations (see 1.2) in that lemmas (signs) are arranged according to:

(a)   the hand formation (handshape) at the beginning of the production of the sign, and
(b)   location, i.e. on/near which part of the body the sign is made.

The first classification is by handshapes. Signs with the same initial handshape are classified in the same section, an arrangement which is reminiscent of alphabetical ordering. In the *Zambian Sign Language Dictionary* handshapes are ranked according to the number of lemmas (signs) they include and are arranged in such a way that a more frequently used handshape comes before a less frequently used one. In Appendix C which shows the handshapes used in the *Zambian Sign Language Dictionary* S, E, A, etc., refer to the signs representing the ASL alphabet, which has been adopted in ZAMSL.

The second classification of the lemmas is by location. Within each section, containing all signs with the same initial handshape, signs are classified according to their locations, starting from above the head down to the main natural space as shown in Appendix D.

It is clear that with the above dual classification of signs, coupled with the fact that the table of contents gives the pages where the categories of handshapes and locations are found, any sign contained in the dictionary, can be easily located, provided that the user's observation of the sign is correct. The English gloss of the sign is given after the lemma.

3.7.3   The Word-Sign Search System (WSSS)

Let us suppose that a dictionary user wants to know the sign for "boy". The first thing to do is to locate "boy" in the English alphabetical index at the end of the dictionary. In this index, every item is followed by at least two numbers, the first of which is always the page number of the lemma and the other(s) is/are the number(s) of the lemma(s) (lemmas are numbered). Therefore, after locating "boy" in the index, the user will go to the indicated page number where he/she will see the sign with the number indicated after the page number in the index.

4.   Conclusion

As it displays all the essential design features of human language (duality of structure, creativity, discreteness, spatial and temporal displacement, etc.), sign language as used by deaf communities is a real human language. Human language need not involve speech.

Sign language being a fully-fledged language and deaf people being a minority in any country, a deaf community is a minority linguistic community. However, by virtue of the nature of its language, which is its first language (L1), a deaf community has special linguistic needs which are difficult to meet but must be met somehow in order for the deaf and hard-of-hearing to be able to enjoy their rights as citizens and participate in national development.

In Zambia, like in most other countries in the Third World, deaf people have been neglected since their natural language, or first language (L1), that is to say, sign language, has been ignored. The Zambian Sign Language Dictionary Project was planned to partly remedy this serious deficit.

The study of sign language can greatly contribute to linguistic theory (Fischer 1979: 51-64) and sign language lexicography can contribute to the metalexicographical debate. Perhaps the most important lesson to be drawn from the Zambian Sign Language Dictionary Project is that in lexicography the methodology as well as the human and other resources to be used depend on the language involved.

References

Chanda, V.M. (Ed.). In the press. *Zambian Sign Language Dictionary*. Lusaka: Co-op Printing.

Fischer, Susan D. 1979. Twixt the Hand and the Lip: The Application of Linguistic Theory to Non-oral Languages. *Application of Linguistic Theory in the Human Sciences*: 45-75. East Lansing: Michigan State University, Department of Linguistics.

Gouws, R.H. 1996. Bilingual Dictionaries and Communicative Equivalence in a Multilingual Society. *Lexikos* 6: 14-31.

Hausmann, F.J. and H.E. Wiegand. 1989. Component Parts and Structures of Monolingual Dictionaries. Hausmann, F.J. *et al.* (Eds.). 1989-1991. *Wörterbücher: Ein internationales Handbuch zur Lexikographie / Dictionaries: An International Encyclopedia of Lexicography / Dictionnaires: Encyclopédie internationale de lexicographie*: 328-360. Berlin: Walter de Gruyter.

Serpell, R. and M. Mbewe. 1990. Dialectal Flexibility in Sign Language in Africa. Lucas, Ceil (Ed.). 1990. *Sign Language Research: Theoretical Issues*: 273-287. Washington, D.C.: Gallaudet University Press.

Sternberg, M.L.A. 1990. *American Sign Language Concise Dictionary*. New York: Harper and Row.

Wallin, L. 1993. Deaf People and Bilingualism. A paper presented at the International Conference on Bilingualism in Deaf Education, Stockholm, Sweden, August 16-20, 1993.

Appendix A: Examples of ASL Signs (Sternberg 1990)

EYEGLASSES (ī' gläs' əs), *n. pl.* (The shape.) The thumb and index finger of the right hand, placed flat against the right temple, move back toward the right ear, tracing the line formed by the eyeglass frame. See also GLASSES.

HEAD (hĕd), *n.* (The head is indicated.) The tips of the fingers of the right right-angle hand are placed at the right temple, and then move down in an arc to the right jaw.

Appendix B: Examples of ZAMSL Signs
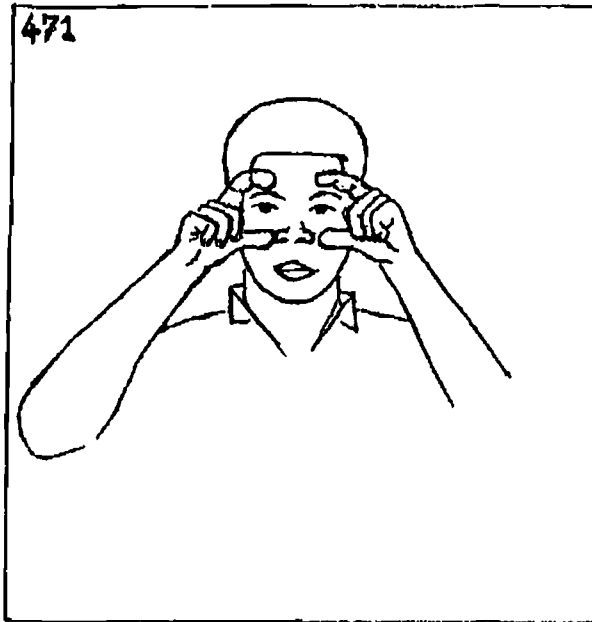
EYEGLASSES (n pl)
SPECTACLES (n pl)

Usage:

Exag:	Problems eye/ eyeglasses/ expensive.
Enlat:	Spectacles for sight problems are expensive.
Articulation:	Double-hand performance. With both hands in the 'bent L' hand position, palms facing each other, place them slightly to the front of each eye so that each eye is in between the thumb and index finger. This indicates the shape of the lenses.

1004   HEAD (n)
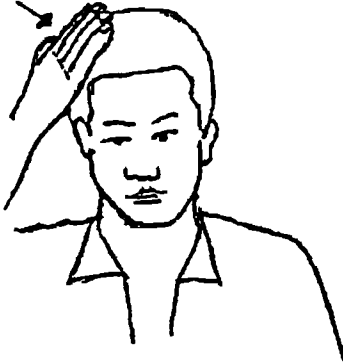
Usage:

Exag:	Butchery/me go/pig head/buy.
Enlat:	I am going to the butchery to buy the head of a pig.
Articulation:	One-hand performance. Just touch the right temple using the palm of the dominant flat hand. (Tapping can be done once or repeatedly "twice").

Appendix C: Ordering of Handshapes in the *Zambian Sign Language Dictionary*

Handshape	Name	Handshape	Name
	S-handshape		Bent L-handshape
	E-handshape		L-handshape
	A-handshape		
	Flat E-handshape		N-handshape
	Common 'good' handshape		Bent 2-handshape
	Closed Q-handshape		R-handshape
	Q-handshape		U-handshape
	Open Q-handshape		V-handshape
	G-handshape		Combined 3-handshape
	X-handshape		Bent U-handshape
	I-handshape		Bent 3-handshape
	K-handshape		3-handshape
	P-handshape		
	D-handshape		

Handshape	Name	Handshape	Name
	M-handshape		O-handshape
	Closed W-handshape		Combined finger handshape
	W-handshape		Flat C-handshape
			Cupped handshape
	F-handshape		C-handshape
	Open F-handshape		Invented claw handshape
	9-handshape		Flathand handshape
	Open 9-handshape		Flathand (thumb out) handshape
			5-handshape or spreadhand handshape
	B-handshape		Y-handshape
	4-handshape		Love-handshape
	Bent flathand handshape		Middle finger handshape
	Bent diagonal flathand handshape		T-handshape
	Bent flathand (thumb out) handshape		I-handshape

Appendix D: Body Locations in Sign Verbal Description

Right hand side   Left hand side

Location A (Above the head)

Location B

Location C

Location D

Location E

Location F

Location G
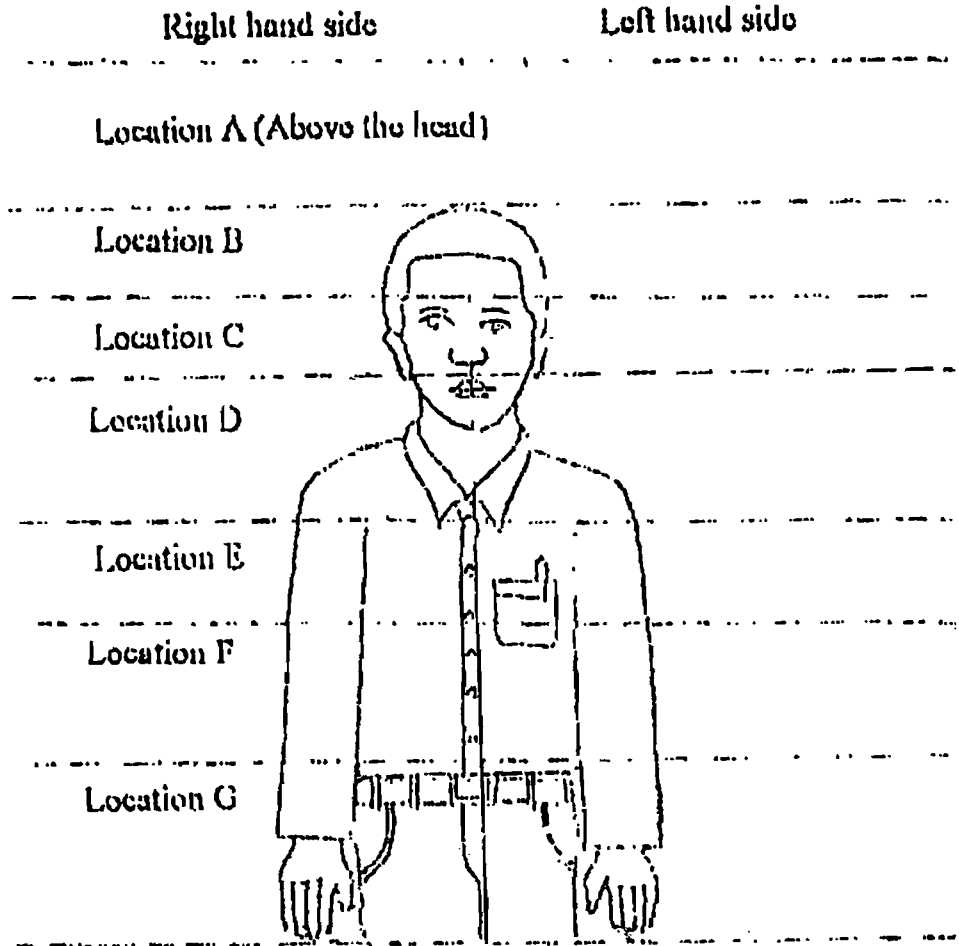
The Southern Dutch
Dialect Dictionaries

Joep Kruijsen, *co-editor of the* Woordenboek van de Limburgse Dialecten,
University of Nijmegen and
Jacques van Keymeulen, *co-editor of the* Woordenboek van
de Vlaamse Dialecten, *University of Ghent*

Abstract: The article on the three dictionaries of the southern Dutch dialects is divided into two parts. In the first part Joep Kruijsen (University of Nijmegen, The Netherlands) treats the history of the *Woordenboek van de Brabantse Dialecten* (*Dictionary of the Brabant Dialects*), the *Woordenboek van de Limburgse Dialecten* (*Dictionary of the Limburg Dialects*) and the *Woordenboek van de Vlaamse Dialecten* (*Dictionary of the Flemish Dialects*), three dictionaries which together record the vocabulary of the southern Dutch dialects. He describes the plan and method of collection and presentation. Because of the systematic arrangement (introduced by Weijnen) the three dictionaries are unique in Dutch and international lexicography. They combine a dictionary with a word atlas.

In the second part of the article Jacques van Keymeulen (University of Ghent, Belgium) deals with a number of new methodological developments which were introduced following the start of the publication of part III General Vocabulary (apart from Part I Agricultural Terminology and Part II Nonagricultural Terminologies). After the institution of REWO (Regionale Woordenboeken), a coordinating body for the three dictionaries within the Dutch Language Union, and because of the introduction of sophisticated software, the databases can be combined to give a survey of the whole of the southern Dutch language area.

Keywords: LEXICOGRAPHY, ONOMASIOLOGY, DIALECT GEOGRAPHY, LANGUAGE VARIATION, PHONOLOGY, HISTORICAL LINGUISTICS, DICTIONARY, REGIONAL DICTIONARY, WORD ATLAS, CARTOGRAPHY, GENERAL VOCABULARY, AGRARIAN TERMINOLOGY, TERMINOLOGY OF TRADITIONAL CRAFTS, DIALECT, TRADITIONAL DIALECT, FLEMISH, BRABANT DIALECT, LIMBURG DIALECT, DUTCH, THE NETHERLANDS, BELGIUM, FRENCH FLANDERS, METHODOLOGY, SYSTEMATICAL ARRANGEMENT, AUTOMATION, DATABASE, USER GROUPS, USEFULNESS

Samenvatting: Dialectwoordenboeken van het zuidelijke Nederlands. Het artikel over de drie dialectwoordenboeken van het zuidelijke Nederlands valt uiteen in twee delen. In een eerste deel heeft Joep Kruijsen (Katholieke Universiteit Nijmegen, Nederland) het over de ontstaansgeschiedenis van het *Woordenboek van de Brabantse Dialecten*, het *Woordenboek van de Limburgse Dialecten* en het *Woordenboek van de Vlaamse Dialecten*, drie woordenboeken die samen de woordenschat van de zuidelijk-Nederlandse dialecten beschrijven. Hij schetst opzet en methode van verzamelen en presenteren. Door het systematische ordeningsprincipe (geïntroduceerd door Weijnen) nemen de woordenboeken een unieke plaats in in de Nederlandse en internationale dialectlexicografie. Ze zijn een combinatie van een woordenboek en een woordatlas.

In een tweede deel heeft Jacques van Keymeulen (Universiteit Gent, België) het over een aantal methodologische vernieuwingen die naar aanleiding van het begin van de publicatie van deel III Algemene Woordenschat (naast deel I Landbouwwoordenschat en deel II Niet-agrarische Vaktalen) werden geïntroduceerd. Door de installatie van het overlegorgaan REWO (Regionale Woordenboeken) binnen de Nederlandse Taalunie en door de introductie van gesofistikeerde software kunnen de gegevens van de drie databases samengevoegd worden tot overzichten voor het hele zuidelijk-Nederlandse taalgebied.

Sleutelwoorden: LEXICOGRAFIE, ONOMASIOLOGIE, DIALECTGEOGRAFIE, TAAL-VARIATIE, FONOLOGIE, HISTORISCHE TAALKUNDE, WOORDENBOEK, REGIONAAL WOORDENBOEK, WOORDATLAS, CARTOGRAFIE, ALGEMENE WOORDENSCHAT, AGRARISCHE TERMINOLOGIE, TRADITIONELE VAKTERMINOLOGIE, DIALECT, TRADITIONEEL DIALECT, VLAAMS, BRABANTS, LIMBURGS, NEDERLANDS, NEDERLAND, BELGIË, FRANS-VLAANDEREN, METHODOLOGIE, SYSTEMATISCHE ORDENING, AUTOMATISERING, DATABASE, GEBRUIKERSGROEPEN, BRUIKBAARHEID

I.   THREE LEXICOGRAPHICAL PROJECTS

1.   The origin of the southern Dutch dictionaries

In 1958 Toon Weijnen took over the chair of Dutch and Indo-European Linguistics at the Nijmegen Faculty of Arts. Weijnen was more or less the personification of research on the Brabant dialects and at that moment he must already have had very explicit ideas on at least one of his later projects. The roots of the *Woordenboek van de Brabantse Dialecten (Dictionary of the Brabant Dialects)* lie in the enquiries that he himself had undertaken in the thirties and that in 1937 resulted in a Ph.D. dissertation on dialect borders in Brabant. The dissertation was written under the supervision of Jacques van Ginneken. Between 1937 and 1958 he published many books and articles on lexical and phonological topics in the Brabant dialects in the Netherlands and in Belgium. Very often Weijnen looked at the neighbouring dialects of Zeeland and Limburg for elucidatory facts. Just as the Leiden specialists under Kloeke were primarily oriented towards the role of the dialects of Holland in the Dutch language area and as Heeroma in Groningen started his observations on Dutch in Lower Saxony, so the starting-point for Weijnen's Dutch philology was to be found in the dialectological situation in Brabant. This special interest of his has been noticeable in all his later work, whether in Dutch or in European dialectology.

In 1960 the Netherlands Organisation for Pure Research made it possible to start new fieldwork in Brabant. Jan van Bakel became involved in the project; written enquiries were started and within five years the project developed into a real institution, the Nijmeegse Centrale voor Dialect- en Naamkunde (NCDN). In 1967 the "Preliminary Introduction" of the *Woordenboek van de Brabantse Dialecten (WBD)* appeared, together with the first fascicle on "the farm".

In the meantime, at the beginning of the sixties, preparations for the *Woordenboek van de Limburgse Dialecten (WLD)* started as well. The first impetus for the Limburg project was research by means of an extended questionnaire, undertaken in 1914 by Schrijnen and Van Ginneken, together with the school inspector Verbeeten. Schrijnen and Van Ginneken were later appointed professors in Nijmegen. They collected, by correspondence, a large amount of precious dialect material at about 90 localities in Limburg. At the beginning of the thirties the folklorist Winand Roukens expounded a plan for a dictionary of the Limburg dialects in the journal *Veldeke*. He contacted the dialect centre of the University of Leuven where Grootaers and Pauwels were making recordings of the dialects in the Belgian province of Limburg. To Roukens we owe the idea of a dictionary of Limburg dialects, covering the two provinces of Limburg in Belgium and in the Netherlands.

The final step towards the realisation of this Limburg dictionary does not come from Roukens though, but from Weijnen. The Brabant dictionary had made a good start and Weijnen had a central role in the Faculty of Arts of the University of Nijmegen. Therefore he was in a good position to realise the Limburg pendant of the Brabant dictionary. Frans Peters and later Pieter Goossens gathered an enormous amount of new data on the Limburg dialects. In the beginning they were financed by the Netherlands Organisation for Pure Research and later by the faculty itself. They also received a copy of a similarly extensive amount of data on the agricultural lexicon, collected orally by Jan Goossens in Belgian Limburg. In 1983 the first fascicle, with an introduction, appeared. At about the same time the cultural departments of the provincial governments decided to participate in the project, first in the Netherlands, later in Belgium too.

Well over ten years later than Weijnen, Willem Pée started the preparation of the *Woordenboek van de Vlaamse Dialecten (WVD)* at the University of Ghent. Pée was inspired by the two projects of Weijnen with regard to the plan and the microstructure of the Flemish dictionary. This dictionary is issued in two parallel publications. A "scientific text" which contains entry forms, phonetic data and place code numbers for every article, is meant for specialists. The parallel "dictionary text" aimed at the general public, is based on the mapping of the data in the scientific text. In the dictionary text the information of the scientific text is rewritten in a globalised form; place code numbers are replaced by indications with regard to frequency and region. In 1979 the introduction and the first fascicle appeared. With this project the whole southern Dutch language area is covered by dictionaries with the same lexicographical approach (see Annex A for the map).

In 1990, the Dutch Language Union (Nederlandse Taalunie), which fosters linguistic cooperation between the Netherlands and Belgium, instituted a coordinating committee for the regional dictionaries covering the two countries. Around this time the Belgian provinces involved in the two Nijmegen projects made the appointment of two editors for these projects possible at the

University of Leuven. Within the framework of the Dutch Language Union the three projects, in Nijmegen, Leuven and Ghent are coordinated more and more. This is important, as work on the so-called general part of the lexicon is about to begin. There will be a common database, fed by the three projects. It has also been agreed upon that three or four editors for each dictionary will try to finish the general lexicon within the foreseeable future.

2.   Methods of collecting and treating the data

What do the dictionaries look like, what do they describe and how is it done? Three matters have to be raised in this respect: the material treated, the systematic design, i.e. the macrostructure of the dictionaries, and the organisation of the articles, i.e. the microstructure of the dictionaries.

2.1   The data

The dictionaries take stock of the lexicons of the dialects in the regions involved. Dialects are geographically differentiated diasystems.

In the second half of the twentieth century the use of dialects is restricted to certain social classes, though different for each region, and to specific situations. The traditional dialect lexicon described here is to be conceived as the relatively stable, natural colloquial language of the majority of the language community in the first half of this century. The oldest generation still has knowledge of this standard. In the course of the second half of this century this traditional dialect-vocabulary has become an historical one. The dialect dictionaries are to be seen as historical dictionaries of a special kind: they record an actual lexicon by describing almost exclusively the *knowledge*, not the *usage* of a language.

Knowledge or usage, the situation differs from region to region. In the Netherlands use of the traditional dialect occurs only together with the standard language. This has already been the case for two generations. Knowledge of the dialect is still to be found among the oldest generation, among the middle generation it is already rare. In the Dutch province of Brabant, Northern Brabant, the use of dialect is restricted to very specific situations and here too, it is in danger of becoming a regiolect, a variety between standard and dialect without the relative stability of the traditional dialects. In Limburg, especially in the south, the use and *a fortiori* the knowledge of dialects, is more widespread and accepted than in Brabant, but here too the pressure of the standard language in situations beyond regional issues is high.

In Belgium the position of the dialect differs from that in the Netherlands. The Dutch standard language has gained prestige during the last hundred

years. Adoption of the (northern) standard language was and still is not a matter of course. The emancipatory language conflict was not only directed against French as the dominating language, but also against dialects that were thought to keep the Flemish people from speaking the (northern) standard language correctly. This antidialect attitude is still noticeable. In the Netherlands, Brabant and Limburg dialects are still held in esteem, whereas in Belgium, the appreciation of dialects is only growing slowly.

In French Flanders and in the north-eastern part of the province of Liège, the situation of the Dutch dialects is even more complicated. In the northernmost part of France, the Dutch standard language disappeared three centuries ago. The only variety of Dutch which speakers from this area understand, is their own dialect; they have no knowledge of standard Dutch any more. In the province of Liège, between Voeren (Furon) and Eupen, this situation has now existed for two centuries. Dutch dialect speakers are wedged in between speakers of the dominant French language and the neighbouring German language. Here too Dutch is unknown as a written language. In these two areas the alarm for the Dutch dialects should be sounded.

2.2   The macrostructure, systematic ordering in three parts

One of the most salient features of the three southern Dutch dictionaries is the nonalphabetical ordering of the entries; they are presented in systematic order, by word-field, or rather (because word-field should suggest a semantic ordering) by the field in which they are used. They are arranged according to activities and objects around a certain occupation, such as work in the fields, flailing, cooperage or mining, the names of birds, etc.

At the beginning of the sixties, a theoretical discussion on the most suitable ordering of the lexicon of a large and internally differentiated area was conducted in dialect lexicography. Weijnen and Van Bakel often argued in favour of abandoning the formal criterion of the written word, i.e. the graphic registration of speech in dialect lexicographic work. Yet the publication in 1967 of the first fascicle of the dictionary without an alphabetical ordering was regarded as something new. No doubt, practical reasons influenced this choice to have it processed completely. It was no longer necessary to collect all the data, and to start with the treatment of A words before the first fascicle could be published. The dictionary-user will find all words concerning a specific field concentrated in one place. In an alphabetical dictionary, however, the dictionary-user will have to glean the sought-for information from diverse entries.

Alphabetical indexes on keywords in a fascicle and cumulative indexes for a volume make it possible to consult the material in other ways, e.g. by word-form, in order to get an idea of the polysemy of a given word.

By ordering the dictionaries in such a systematic way, Weijnen and Van Bakel joined the geolinguistic rather than the lexicographic tradition. The Ley-

den linguistic atlas treated limited word-fields and also the French regional atlases were ordered according to word-fields. In the southern Dutch dictionaries the geolinguistic design of the language map and atlas on which the geographically scattered heteronyms of a concept are given, is combined with the completeness of a dictionary. Because the areas covered by the three dictionaries are rather extensive, the number of heteronyms for a concept can be high. For a concrete and familiar object like a scoop-shovel no less than seven heteronyms are found in the Limburg dialects, each with its own geographical area. For the names of plants or animals this number will be much higher. The question for a dictionary like the one under discussion will not be whether a word like *eren* in the meaning of "to plough" (from Lat. *arare*) exists in southern Dutch dialects, but rather which geographical position the word *eren* takes among other words for "to plough" and what the dialectal pronunciation of this word is.

An article in the dictionary gives all current words for a certain concept in one of the three areas, together with their pronunciation and their location. The order of the concepts is as close as possible to the reality of the dialect-speaker. It is obvious that word-fields can be ordered in more than one way. The concept "clover" could be treated as part of the plant-names in the general vocabulary, as part of the section on crops in the volume on agriculture, or as part of the section on fodder in the volume on cattle breeding. The editor makes the final choice and provides the necessary references, but this will be influenced by the field which the dialect-speaker himself assigns to the word.

The general division of the dictionaries is based on the thematic ordering according to fields.

Each regional dictionary consists of three parts, and each part of a series of fascicles in which a separate item is dealt with. Roughly speaking, each part will comprise about 2,000 pages.

The first two parts treat the vocabulary related to occupations. Part I deals with the agricultural vocabulary. Until the middle of this century, the greater part of the dialect-speaking population was active in agriculture. Part II covers other technical and craft terminologies: those of the baker and the butcher, of the brewer and the miller, the miner and the textile-worker, etc. Part III contains the general vocabulary (i.e. the lexicon which is not sociologically bound), supposed to be generally known among the dialect-speaking community. It includes the terminology of social life, of education, administration and religion, household words for clothing, cleaning, cooking, the names of plants and animals, of health and hygiene. The so-called closed word categories (adverbs, prepositions, conjunctions, etc.) will also be dealt with here.

2.3   The microstructure: the article

Representativeness and verifiability of the data become visible in the article itself, i.e. in the microstructure of a dictionary. All possible sources have been

described and are well-documented in the introductions. These sources consist in the first place of the enquiries made by the editors themselves. During the past 30 years approximately 200 correspondents of the institute in Nijmegen had more than 100 questionnaires completed by dialect spokespersons both in Brabant and Limburg. The data from these questionnaires was supplemented by other sources: lists from other dialect institutes, direct oral enquiries by the editors, local dictionaries and monographs, theses, etc. The collection of data is as complete as possible. Not only familiar articles are mentioned, but also those with mainly documentary value. Another keyword is precision: precision in phonetic documentation and precision in localisation. Many linguistic maps are inserted in the fascicles to give a clear view of the latter.

Thus an article consists of:

—   the title of the article, i.e. the keyword or the description of the concept to which it is easy to refer, e.g. *scoop-shovel*.
—   the acknowledgement of the sources: questionnaires and monographs or other sources consulted.
—   the elucidation of the title: the keyword is explained; the connection with other articles is often underlined. Special notes of *informants* are reported, encyclopedic peculiarities of a concept in a technical language can be given. In the example of the scoop-shovel an illustration is inserted, *details about its use are given, etc.*
—   the corpus of the article containing all the heteronyms, the terms given for the concept enquired about: a possibly constructed Dutch (or Dutchified) reference form and the phonetic variants noted down as exactly as possible. For each variant the exact geographical location is given by means of place-codes, often visually presented in the form of a map. This applies to parts I and II of the Brabant and Limburg projects in Nijmegen; the Flemish dictionary in Ghent presents its geographical information in a slightly different way, as can be seen in the annexes.

3.   The usefulness of dialect dictionaries

In what ways is such an extensive description of language useful?

The traditional dialect lexicon is disappearing fast and is in some regions, as we have seen, present in the memory of the oldest generation only. The collection of this lexicon therefore has a high priority. In spite of this fact, appeals like "save the dialects" or "preserve before it is too late" are dangerous and biased. These phrases could lead to the misunderstanding that the lexicographer nostalgically interests himself in an older and purer and therefore more highly esteemed language variety. On the contrary, he/she is interested in a rapid registration of a historical group of words and meanings, a vocabulary belonging to a society that has already disappeared.

The usefulness of dialectal dictionaries can be indicated as follows:

(a)   Such regional dialect dictionaries are linguistically useful, because they uncover part of the linguistic reality. The oral enquiries often make it possible to note down subtle distinctions in phonetic variations (as in *WBD* and *WLD*), but the main linguistic importance undoubtedly lies in the semantic and etymological domains. The dictionaries present the possibility to study fields of words and meanings onomasiologically and semasiologically and to trace the history of the distribution of word-forms and meanings.
       Weijnen himself has always stressed the importance of the study of dialects for general linguistics. On the very first page of his doctoral dissertation he notes: "The study of language without linguistic geography is almost impossible", and in his inaugural speech in 1958 he pointed out that "for the correct understanding of the standard language knowledge of the dialects is indispensable". Dialect-lexicography is therefore not a goal in itself, but knowledge of the dialects, made possible by a systematically ordered collection of data, is an important part of the linguistic knowledge in a broader sense.

(b)   The three dictionaries are culturally and historically useful, they uncover a culture-historical landscape. Because a word more often than the referent takes root, lexicological research is helpful to cultural history. The large dialect dictionaries reflect the daily life of the dialect-speaking population in the first half of the twentieth century. They also show in what way material and spiritual innovations have been assimilated in the language. In the specifications of the article, encyclopedic information is added, so that a description of the trade itself is also found in the fascicles in which the technical languages are treated.

(c)   As reference books, the dictionaries are practically useful. They are helpful to answer questions like: "What does word x signify, and where is it used?" "How does one express y?", etc. In this respect the dictionaries can be compared to other historical dictionaries like the *Middelnederlandsch Woordenboek* (*Middle Dutch Dictionary*) or the *NED* (*A New English Dictionary on Historical Principles*).

(d)   The large dialect dictionaries are also indispensable for the study of social variations in Dutch. Parts I and II (agricultural and other craft terminologies) treat languages with a socially restricted use, but in general the language of the dictionaries reflects a socially and situationally determined variety.

(e)   Dialect dictionaries can also have a political function in that they can be seen as a demonstration or as a symbol of the own identity of a group of language-users in contrast to the users of the standard language. Though not intentionally, they can be useful for the codification of a regional standard and therefore for the promotion of literature in the dialect. This

contribution towards a cultural identity of a certain region forms the main reason for the interest of and the participation by provincial governments in the projects.

4.   Prospects

Only a close cooperation between the universities, the national scientific organisations and the provincial governments can generate the funding of these three projects. The provincial authorities have discovered that the description of their dialects is a regional cultural issue and they have therefore taken the financial responsibility. The dictionaries are financed following a tripartite agreement: the two national scientific organisations coordinated by the Dutch Language Union, the three universities and the provinces involved (eight at the moment) share the funding, although for different reasons of interest. The procedure is successful. The description of the southern Dutch language area resembles a web in which many threads connect the institutes and the authorities, a web in which, through these threads, communication between the institutes, quality of the work and continuity are guaranteed.

II.   NEW METHODS IN THE TREATMENT OF THE GENERAL VOCABULARY

0.   Introduction

One of the most important principles in making a dictionary is not to change methods *en route*. In the case of the southern Dutch dialect dictionaries a change is however possible, due to the systematic arrangement of the dictionaries and their publication in three parts. All three projects are about to start with part III General Vocabulary, i.e. the vocabulary that is not restricted to professional activities.

Part III is of the utmost importance, not only because a larger public will undoubtedly be interested, but also because it documents the last phase of a long continuum of dialectal varieties of Dutch in the northern part of Belgium onto which the (relatively) recently introduced standard Dutch language is grafted. The three dictionaries are a frame of reference for sociolinguistics in Belgium.

Starting with part III makes it possible to evaluate and improve the methods used so far. Improvement is needed for a number of reasons:

(a)   Although the three projects basically have the same plan, they differ *inter alia* in the way the publications are issued.

(b)   The scope of the projects is not in accordance with the financial means, hence there is a need for speeding up the work.

(c)   The geographical representativeness of the investigation is not systematically guaranteed, i.e. it is uncertain whether every existing lexeme for a given concept has at least been recorded once.

(d)   Data collection by correspondence, involving questionnaires completed by dialect speakers without any phonetic training, does not guarantee accurate phonetic data. As a result, the phonetic component in the three dictionaries is sometimes secondary (i.e. reconstructed) since it is based on other sources or on specialist literature. This especially is the case for the *WVD* and for the investigation of the general vocabulary. Moreover, primary and secondary phonetic data cannot be distinguished unless the different sources are presented separately. The phonetic component is also highly redundant with regard to the regular sound phenomena.

(e)   The fact that there are two different user groups (namely the general public and specialists) is not always sufficiently taken into account.

(f)   The enterprise insufficiently makes use of the advantages of automation.

In his Ph.D. thesis Van Keymeulen (1992) evaluated the approaches and methods of the three dictionaries and tried to formulate new proposals regarding the methods of collection and publication. These new proposals were summarized in a report (Ryckeboer and Van Keymeulen 1992) presented to the REWO committee (Permanent Overleg Regionale Woordenboeken) (Permanent Consultation for Regional Dictionaries). This committee was instituted in 1990 by the Dutch Language Union in order to ensure the continuation and uniformity of the three projects. The institution of this committee already met the first two points of criticism (a) lack of cooperation and (b) financial difficulties above, since it aims at stimulating cooperation and encourages concerted action for funding.

In what follows there will be focused on the last four points of criticism: (c) geographical representativeness; (d) the reduction of the phonetic component; (e) the two user groups and (f) automation. A last paragraph will dwell on the presentation of the general vocabulary in a systematically arranged dictionary.

1.   Collection of lexical data

Since dialect vocabulary is characterized by its geographical differentiation, it is of the utmost importance that every existing dialect word is recorded at least once. A further objective which is harder to achieve because the voluntary cooperation of the dialect speakers has to be depended on, is the detection of word boundaries (isolexes), made possible by the abundance of the data.

The following proposals regarding the geographical representativeness of the data collection are especially important for the *WVD*, because, as Joep Kruijsen pointed out, this project started about ten years later than the others.

In Ghent, the field work for the general vocabulary has just started, whereas in Nijmegen the collection of the data is considered to be complete, although additional investigations may be needed to fill in geographical gaps. The proposals pertain to field work in different phases:

(a)   *Establishment of an inventory of concepts on the basis of:*
       (i)   the Hallig-Von Wartburg system (adapted by Corry Frissen);
       (ii)   a systematic reorganization of the existing alphabetical dialect dictionaries;
       (iii) systematic inventories of the standard language; and
       (iv)   older questionnaires.

(b)   *Oral investigation by the staff through a network of control points*
       The Flemish territory (in its dialectological sense, see annex A for the map) was divided into lexical areas on the basis of the existing knowledge of the geographical word patterns. One control point was selected for each of these areas (plus six urban points, plus some additional points in order to have a regular pattern). The inventory of concepts is tested in a discussion group of dialect speakers in order to discover the proportion of dialecticity (i.e. lexical contrast with the standard language) and heteronymity (i.e. the existence of geographically differentiated dialect lexemes). Emphasis is put on the lexical/semantic contrast with the Dutch standard language and on geographical differentiation. The field work aims at reducing the initial inventory and at guaranteeing the geographical representativeness. The dictionaries of the dialects of Brabant and Limburg try to ensure the geographical representativeness by incorporating the local amateur dictionaries also.

(c)   *Large-scale investigation by correspondence*
       The questionnaires for a large-scale investigation by correspondence are based on the results of the oral investigation. Only the concepts for which the dialectal lexemes show geographical differentiation, are accounted for in the large-scale investigation.

2.   Two user groups: the general public and specialists

Before the methods of lemmatizing and publishing the data are discussed, it is necessary to dwell on the users the dictionaries are aimed at. Joep Kruijsen pointed out that the dictionaries meet both a scientific and a social demand, which implies that the needs of both the general public and specialists are to be accounted for. In both cases it is taken for granted that both user groups have a good knowledge of standard Dutch.

3.   Lemmatizing the data

The main task of the editor is lemmatization, i.e. grouping together the dialect words that go with lexically relevant concepts. In the tradition of these three dictionaries the term lemma is used for the whole of a dictionary article, not for the headword. The main task for the lexicographer is to ascertain the way the dialect speakers categorize reality. This is achieved by field work and by close analysis (comparing) of the answers to the questions in the questionnaires. The focus of this paragraph, however, will be on the entry forms of the dialect words and on the consequences of the reduction of the phonetic component.

As Joep Kruijsen already indicated, the dictionaries use the "Dutchified" forms of the dialectal lexemes as entries. This means that all phonological and morphological particularities of the dialectal lexemes are replaced by their standard Dutch pendants. This technique is the only possible way of representing entry forms in regional dictionaries that cover many related but different dialects. This technique, however, may put a strain on the dialect-speaking user.

The lexicographer's hope that a dialect speaker will find back "his/her" word in the dictionary, is based on the assumption that a dialect speaker is able to abstract from his/her sound system to the standard Dutch sound system as represented in Dutch orthography. This implies that the dictionary user is considered to have absorbed two related, but different sound systems: the dialectal sound system, and the sound system (as represented in orthography) of the Dutch standard language. In the mind of the dialect speaker there are intuitive sound rules that relate to the two sound systems. They come into existence in that part of the vocabulary that shows no contrast between dialect and standard language, and can be transferred to the contrastive part of the vocabulary.

In standard Dutch, for example, the [œy] sound (written *ui*) occurs in many high-frequency words (the sound is so to speak "panlexical"), which also occur in the Western Flemish dialects, where the Dutch [œy] sound is represented by the [y] sound (written *uu*); e.g. Dutch [œyt] <uit> — West-Flemisch [yt] <uut>. Hence the intuitive rule "dialect [y] <uu> equals Dutch [œy] <ui>" guarantees that a speaker of West Flemish recognizes in the entry form *uit* "his/her" dialect word. The rule can be transferred to the contrastive part of the vocabulary. Thus the West Flemish word *uuvallig* "dirty" can be written in its Dutchified entry form *uivallig* (a word that does not exist in standard Dutch), without frustrating the dialect-speaking user too much. This method has its flaws, but the system works if only because all the phonological variants of a given dialect word are presented in the dictionary text. Fortunately, the Dutch orthography is highly phonological.

Parts I and II of the dictionaries contain the phonological documentation of all the words. Each of the three dictionaries uses its own home-made phonographic orthography; the Flemish dictionary the broadest one, the Limburg dictionary the narrowest one. This practice will in future be continued for parts I

and II. However, it often forces the lexicographer to reconstruct, at least partly, the phonological contours of the words according to the (written) clues given by the respondents and the specialist literature. In part III the phonological component will be reduced to the primary data, i.e. data orally collected by the editors themselves and data written down by respondents with a phonetic training. The phonetic data is entered into the database, but is not made available in printed form any more.

Because of this, a way had to be found to "rescue" the phonological component as much as possible and to find a solution for the fact that the "intuitive sound rules", a rather slippery notion, perhaps do not always lead to the identification of a dialect word with its Dutchified form, if only because the intuitive insights into the sound relation between dialect and standard Dutch may differ from person to person.

The reduction of the phonological component in the dictionary text will be compensated by the following measures:

(a)   Every dictionary will be preceded by a phonetic/phonological introduction in which the regular sound patterns of all the dialects under investigation will be described.

(b)   A set of lexical variants will be introduced next to the Dutchified entry forms whenever the lexicographer fears that the average user will not be able to retrieve a dialect word because the sound relation between dialect and standard Dutch is too obscure. The lack of transparency can be due to a number of reasons:
(i)   The dialect word shows an irregular sound pattern.
(ii)   A sound relation may be regular but exceptional.
(iii)   Several sound changes occur at the same time.
(iv)   A sound change affects the consonantic framework of the dialect word.
(v)   The affected dialect lexeme does not occur in standard Dutch.

Some examples:
(i)   In Eastern Flanders *duvel* "devil" contains the sound relic [y], hence: Dutchified entry form **duivel**, lexical variant *duvel*.
(ii)   *Kurre* "herd", with an intervocalic [d] becoming [r], is regular, but the words affected by the change are of low frequency; hence: Dutchified entry form **kudde**, lexical variant *kurre*.
(iii)   In *voreeuw* "harness" several regular sound changes occur at the same time, affecting the consonantic framework; hence: Dutchified entry form **gareel**, lexical variant *voreeuw*.

(c)   Primary phonetic data will be entered into the database.

4.   Database and publications

The published version of the dictionary distinguishes the lexical difference from the standard language. In principle only the data of the control points are taken into account for lexemes that do not show differentiation from standard Dutch or geographical differentiation. So we end up with a database, a scientific text and a dictionary text:

(a)   *Database*
All the data is stored in a database that contains sources, questionnaires, entry forms, lexical variants, phonological information (based on primary phonetic data) and exact location. The database contains interpreted material. The raw material, i.e. completed questionnaires, is kept in the archives.

(b)   *Scientific text* (see annex E)
With the database as basis a text file is generated that contains lemma titles, sources, entry forms, lexical variants, phonological information and exact location. In practice this text file differs from the database, as the data is further interpreted and sometimes even regrouped. This file is made available electronically (on floppy disk or otherwise) and is meant for specialists who want to have detailed information with regard to location and phonetics.

(c)   *Dictionary text* (see annex D)
The dictionary text meant for the general public is the only publication in printed form. The text is based on the mapping of the data of the scientific text. It contains lemma titles, semantic and encyclopedic information, sources, entry forms, lexical variants and globalized indications with regard to frequency and location. It also contains word maps and illustrations.

Thus each dictionary will be issued in two parallel publications: one meant for the general public, the other for specialists. The former serves as a kind of manual for the latter.

5.   Automation

A major result of the growing cooperation between the three projects is the development of a common computer program that will replace the existing software. The program can cope with all kinds of data that have a question-answer structure. It is interfaced with a word processor and a cartographic program (MapInfo), in such a way that texts (consisting of lexemes, source indications and place code numbers) and maps can be generated automatically. The databases of the three dictionaries can be combined in order to plot word maps of the whole area of the southern Dutch dialects.

The program is menu-driven, very flexible and allows for correction pro-
cedures in all stages of the input and the lemmatization. Lemmas are auto-
matically generated by putting together the data from different sources by
means of references to the questionnaires, question numbers and/or answers.

6.   Systematic arrangement of the General Vocabulary

The three dictionaries are systematically arranged. For part I Agricultural Ter-
minology and part II Nonagricultural Terminologies a systematic arrangement
is relatively easy, as it is possible to divide a production process in different
stages. For part III General Vocabulary an overall classification of all the con-
cepts of the dialect-speaking community is needed. The editorial boards of the
three dictionaries do not consider it their task to present the definitive structure
of the general vocabulary, but rather aim at presenting the lexemes in such a
way that it is possible for linguists to investigate conceptual structures. The
classification should serve a practical purpose and should in any case be easy
for the average user to handle.

In what follows, the principles for solving the assignment problems are
presented. The focus will be on the so-called open word classes (substantives,
adjectives, verbs, etc.). The closed word classes (adverbs, prepositions, conjunc-
tions, etc.) present specific problems, because the meaning of the lexemes of
these classes is more grammatical than representational. For these classes an
alphabetical (semasiological) arrangement will perhaps prove to be necessary.

Generally speaking, the dictionaries present the vocabulary of dialect
speakers living in agrarian surroundings during the first half of the 20th cen-
tury. The concepts should be classified in conformity with what Weijnen called
"the concrete coherence of things in daily life" as experienced by the dialect-
speaking community (Weijnen and Van Bakel 1967: 40). This means that so-
called "scientific taxonomies" that classify the external world on the basis of
inherent characteristics, can hardly be used. For example, a class "sounds" in
which all existing noises are brought together is only of relative value for the
dictionaries. Concepts like "to ring" or "to thunder" should no doubt be dealt
with in connection with "house" and "weather", if the arrangement in the dic-
tionary is to reflect the way the dialect speakers perceive the coherence in daily
life.

A second principle is that of referential coherence per fascicle, i.e. each
fascicle should contain the concepts that the user would expect to be there. For
instance, although the words for *baker* and *bread* are clearly part of the general
vocabulary, the lemmas "baker" and "bread" should also figure in the fascicle
on bakery, because it is hardly imaginable to issue a fascicle on the terminology
of a trade without the words for producer or main product.

Having said that, means have to be found to objectify subjective notions
like "concrete coherence" or "things in daily life". If "the concrete coherence of

things" is to be the guiding principle of the classification of the general vocabulary, the point of departure has to be man himself and his needs. A functional view (what is the use of this?) rather than an ontological view (what is this?) is held. For example, a tomato is a plant but it is eaten by man. Hence, the words for *tomato* are classified in a class "food" and not in a class "plants".

For the assignment of a given concept to a certain class, the notion of frequency has been introduced in an attempt to objectify intuitions about folk taxonomies. For example, in solving the problem whether *chicken* is to be an item in the class "birds", "food", or "hatchery" it is considered how frequently chickens are associated with birds, food or the hatchery in the daily life of the average dialect speaker. It turns out that a chicken is not considered a "bird", but belongs to the category "poultry". As there are more people eating chicken than breeding them, the words for *chicken* should be assigned to the class "food" rather than to "hatchery". However, because chickens are often bred in the countryside, one may consider a cross-reference or even the repetition of the lemma in a class "hatchery", if the referential coherence of a class demands it. Moreover, it turns out that in the same dialect different words are used for the same objects in different situations. In many Eastern Flemish dialects potatoes are called *erpels* when they are grown, but they become *patatten* when they are cooked and eaten.

Practice will show whether the above-mentioned guiding principles can cope with all cases. Not every problem can be solved by this functional-frequentative principle. In a number of cases it will be necessary to fall back on scientific taxonomies because no ordering principle from human experience can be given, e.g. wild plants and animals, qualities of character, etc. In any case, if the world of the dialect speaker is to be reflected in the classification of the general vocabulary, these principles seem a good point of departure. Problems can be solved up to some extent by repeating lemmas or by cross-references.

In practice, the classification of the general vocabulary is based on a rearrangement of the already existing classifications of Hallig-Von Wartburg (1952) and of the *Woordenboek van de Achterhoekse en Liemerse Dialecten (WALD)* (*Dictionary of the Achterhoek and Liemers Dialects*) (see annex F). Man is placed at the centre of things, and reality is assigned to him in ever broadening circles. Thus the classification of the general vocabulary consists of four major classes: 1 Man; 2 Domestic life; 3 Society; 4 External world. Within each class, the concepts are as much as possible clustered around human needs; hence, for example, in 2 Domestic life the subclasses 2a House; 2b Family life and 2c Food are distinguished. In any case, owing to the elaborate registers (systematic and alphabetical, both on lemma titles and entry forms/lexical variants), the user is in a position to control decisions and correct them if need be.

7.   Conclusion

The three dictionaries of the southern Dutch dialects inventory the vocabulary of the oral tradition of the Dutch language in the south of the Netherlands, in Belgium and in the north of France. The project as a whole is a major international and interuniversity undertaking. Owing to the ever closer cooperation between the three editing boards and to automation, an ever better idea of the geographical variety in the southern Dutch language area is obtained. The traditional dialects are disappearing rapidly because of the pressure of standard Dutch and because of the disappearance of the small-scale agrarian culture of which the traditional dialects bear witness to. Large parts of the centuries-old vocabularies of the traditional dialects will not survive into the third millennium. The three dictionaries must be written now or never.

Bibliography

Hallig, R. and W. von Wartburg. 1952. Begriffssysteem als Grundlage für die Lexikographie. *Abhandlungen der deutschen Akademie der Wissenschaften zu Berlin. Klasse für Sprachen, Literatur und Kunst, nr.* 4. Berlin: Akademie Verlag.

Ryckeboer, H. and J. van Keymeulen. 1992. Rapport over de lexicografie van de Nederlandse dialecten. *Publikatieblad van de Nederlandse Taalunie* 13: 46-53.

Schaars, A. 1984 -. *Woordenboek van de Achterhoekse en Liemerse Dialecten.* Doesburg: Staring Instituut Doetinchem.

Van Keymeulen, J. 1992. *De algemene woordenschat in de grote dialectwoordenboeken (WBD, WLD, WVD). Een methodologische reflectie.* Unpublished Ph.D. dissertation. University of Ghent.

Weijnen, A. and J. van Bakel. 1967. *Voorlopige inleiding op het* Woordenboek van de Brabantse Dialecten. Assen: Van Gorcum.

Annex A: Map of the southern Dutch language area

ALPic   Atlas Linguistique Picard
ALW   Atlas Linguistique de la Wallonie
DL   Dictionnaire Liégeois
RhWb   Rheinisches Wörterbuch
WALD   Woordenboek van de Achterhoekse
         en Liemerse Dialecten
WBD   Woordenboek van de Brabantse Dialecten
WfWb   Westfälisches Wörterbuch
WLD   Woordenboek van de Limburgse Dialecten
WVD   Woordenboek van de Vlaamse Dialekten
WZD   Woordenboek der Zeeuwse Dialecten
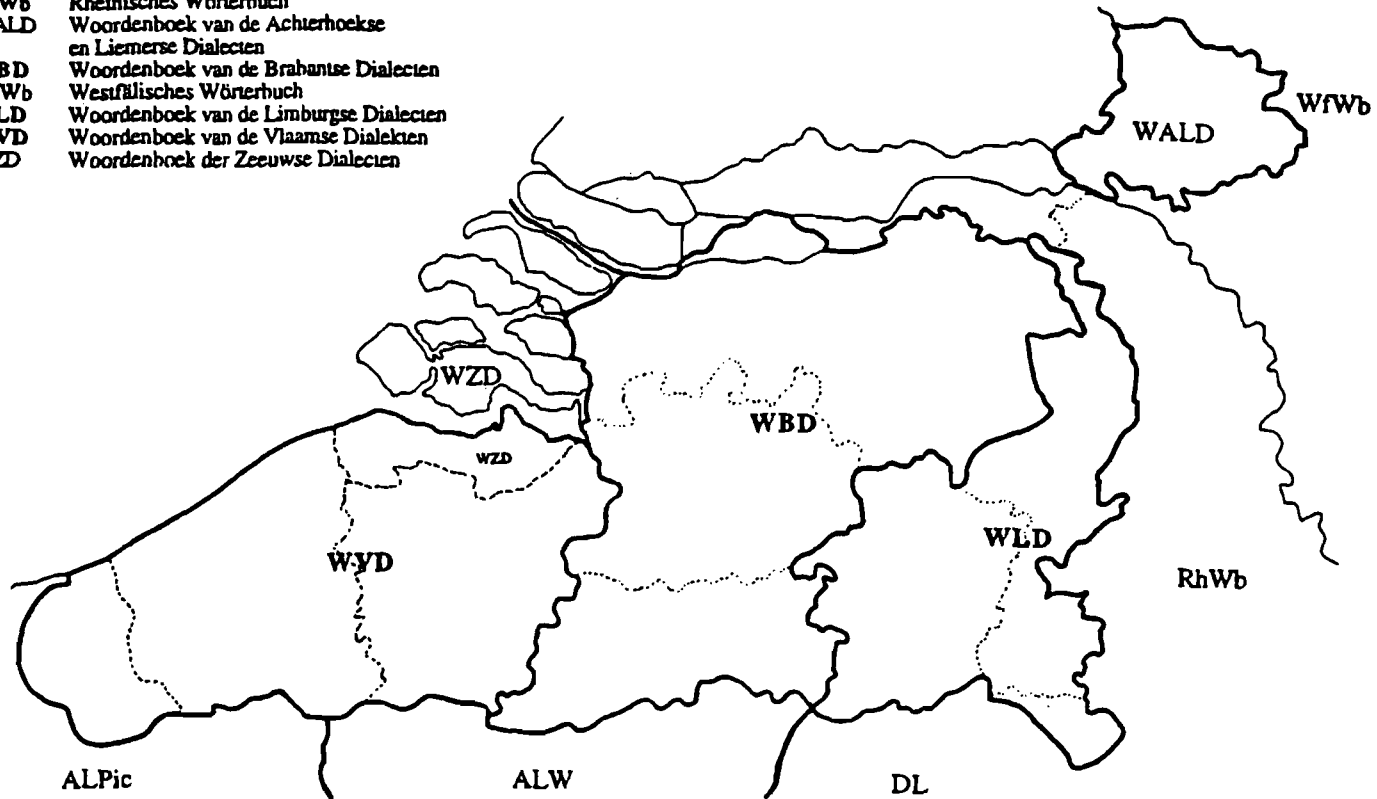
Annex B: A page from the *Woordenboek van de Brabantse Dialecten* (pt. I, fasc. 1, p. 153)
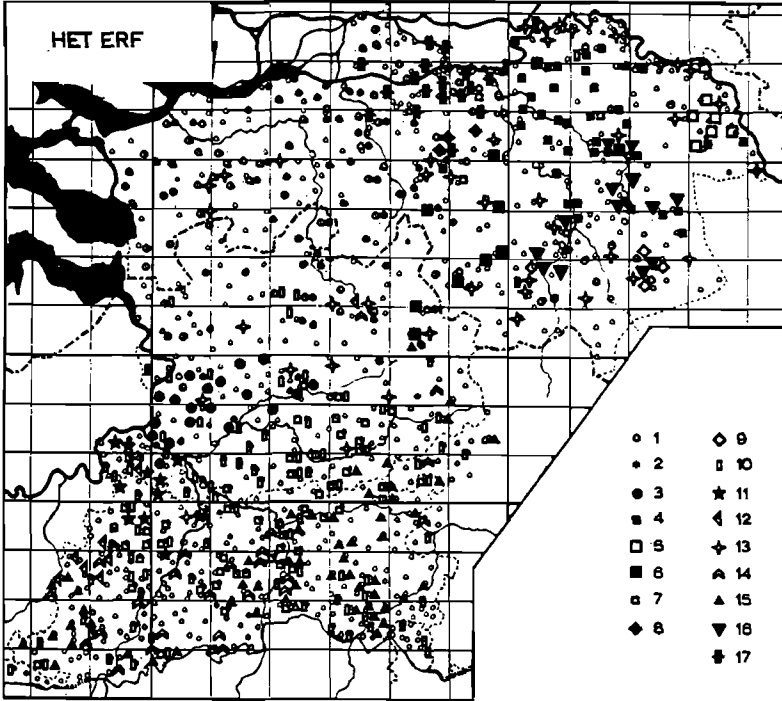
K 240; **boerenerf**: *boerənerf*, L 114a, 158, 239; **werf** (krt no. 1): *wérf*, K 189, 203, 209, 210, 237, 240, 241, 274, 317; *wèrf*, K 102, 149, 165, 191, 194, L 100, 150, 183; *wéjərf*, K 189; *werf*, I 57, 78, 79, 118, K 100a, 101, 101a, 102*, 131, 133a, 133c, 137, 146, 152, 158, 158b, 158c, 161a, 162a, 163c, 164a, 167, 168, 176, 177, 182, 182b, 185, 188a, 235, L 93, 152, 158, 239, 241, 242, 259, 260, 261; *wéèrf*, K 188a, 209, 212, 237, 240; *wèèrf*, K 216a, L 91b, 93, 99, 153, 264; *weerf*, K 215; *wurf*, K 151; *waarf*, I 78a, 79, K 102a, 145a, 153, 0153, 157, 173a, 173b, 174, 175a*, 188, 188a, 203, 224, 227, 248; *warf*, I 57, 78, K 151, 157, 173, 173*, 174, 175, 175a, 176, 177b, 188, 225; *waerf*, hs Renders L 226; *waaref*, Leopold, gegeven voor 's-Gravenmoer-Dongen; *wèèrf*, gegeven voor de lage Maaskant, Elemans; *wá.rəf, wârəf*, De Bont; *werf*, Goossenaerts; **werft** (krt no. 2): *wérft*, K 182, 183, 184a, 207, 210, 232, 240, 268, 288, 330; *wèrft*, K 184, 191, 192, 193a, 204, 291; *werft*, K 124a, 125, 125a, 126, 127, 127a, 129a, 130, 148, 149, 155,

158b, 161, 161a, 162, 163, 163c, 164, 179a, 180, 181, 182, 182a, 183b, 184, 184a, 193a, 194, 195, 197, 198, 207, 210, 211, 237, 246; *wéérft*, K 178*, 184a, 201, 211, 237, 273, 287; *wèèrft*, K 190, 191, L 232; *wéérft*, K 180; *waerft*, K 201, 208, 231; *wurft*, K 151, 152; *wörft*, K 268; *waarft*, K 176, 177b, 180, 226, 245; *warft*, K 245, 246; *werft*, Corn. Vervl.; **boerenwerf**: *boerewerf*, Goossenaerts; **boerenwerft**: *boerewurft*, K 152, Weijnen EV 8; **voorhoofd** (krt no. 3): *vurft*, K 252, 254, 256, 285, 287, 288, 291, 296; *vörft*, K 262; *veurft*, K 281; *verft*, K 252; *vóerft*, K 286; *vurrəft*, K 246, 249, 250, 255, 268, 282, 283, 286, 296, 320; *vörrəft*, K 267; *veuruft*, K 287; *vurf*, K 281; *vörft*, uitspr. *vörrəft*, de open plaats welke vóór de hoeve ligt (...) de achterplaats heet ook *vörft*, Corn. Vervl.; **misse** (krt no. 4): *missə*, K 149, L 91b, 99, 100, 144, 145, 145a, 146, 147, 148, 149, 150, 151, 152, 153, 157, 177, 177c, 178, 178a, 179, 180, 180a, 180b, 180**, 181, 182, 183, 185, 189, 200, 201, 202, 203b, 205, 206, 208, 208a, 233, 244, 256; *missə*, grond voor de

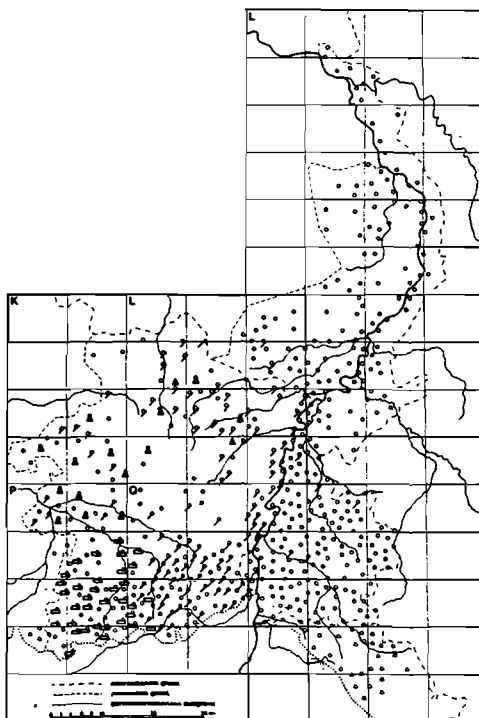## Annex C: A page from the *Woordenboek van de Limburgse Dialecten* (pt. I, fasc. 5, p. 131).

Kaart 71. BRAAM

(Rubus fruticosus L.)

⟋ braam
⇒ braamdoorn
▲ doorn

Afb. 64. Braam *(Rubus fruticosus* L.)

brō.ndzŏᵃ.n P 187, 188; *bryǫ.n–* P 121, 184; *brondōn* P 176, 178; *brondēᵃn* P 115, 173, 179, 180, 182, 218, 219; *brǫndyən* P 174, 175; *brǫn-dy.ęn* P 224, 227; *brǭ.ndyǫ.n* P 220, 223; **doornstruik:** *dērəstrÿ.k* L 423; *djǫ.nəstruk* P 186; **doorns:** *dōrəs* K 316; *dō.rəs* K 353, 357, 359, L 315; *dō.rəsə* L 354; *dęrə* Q 13; *dēⁱrə* L 316; *dēⁱrə* L 362, 367; *dēən* L 413; *dięn* K 361; *diᵉⁱǫ.n* P 57; *dīᵒn* P 51; *djō.n* P 49; *djǭ.n* P 53; *dyənə* P 45; **braamberenstruik:** *brǫmbęrəstruk* L 214, 214a, 215, 217, 245b, 246a; **brameltenstruik:** *brōməltəstryk* L 291; **bramenstruik:** *brōməstruək* L 318b; *briəmə–* L 318b; **braambuttenhout:** *brembytenhāt* K 278 *(hut* is struik); **berenstruik:** *bērəstruk* P 214; **tuinsteekselen:** *tunšteksələ* Q 113 *(tuin* is hier „heg"); **kretsdoorn:** *kręzjŏn* P 55; *krɛ.dzdjǫ.n* Q 2; **doornebraak:** P 46; **braakdoorns:** *brā.gdjǭ.n* P 46; **reefdoorn:** *rēvdžŏn* P 50; **reveldoorn:** *rēᵃ.valdjō.n* P 52.

5.1.24 BRAAMBESSEN (kaart 72)

(JG 1b (gedeeltelijk), 1c, 2c)
[Als aanvulling op de vraag die in het voorgaan-

de lemma is behandeld werd ook geïnformeerd naar de benamingen van de vrucht van de braamstruik.]

braamberen: *brǫmbērə* K 357, 358, P 44, 46, 49, 50, 51, 57; *brǭm–* L 288, 288a; *brom–* L 164, 192a, 210, 214, 214a, 216a, 217, 244d, 246b; *brǫm–* K 316, 318, 357, 359, P 45; *bręm–* P 48; *brǫmbēr* L 286, 314, 318, 318b, 353, 374, P 54; *brǫwm–* L 288a; *brombē.rə* K 361, L 282, 312, 313, 314, 315, 316, 318, 352, 353, 354, 355, 413, Q 1, 2a, 74; *brǭ.m–* K 360, P 51, 56, Q 2; *brum–* L 286; *bruəm–* L 414; *brämbērə* L 246, 246c, 249; *brǫm–* L 159a, 163, 164, 192, 192a, 209, 210, 212, 214, 214a, 215, 215a, 216, 217, 245, 245b, 246, 246a, P 108a, 115, 172; *brō.m–* Q 241; *bręm–* L 244c, P 45; *brum–* L 247, P 47; *brǫmbēᵃrə* L 214, 250, P 113; *brōmbęirə* P 176a; *brom–* P 113, 115, 172, 173, 176; *broəm–* P 181; *brum–* P 47, 48, 176; *bro.mbęi.rə* P 119, 120 121, 177a, 178, 184, Q 1, 78; *brǫmbęiərə* P 184; *brǫmbiərə* L 289; *brǫ.mbiə.rə* Q 152; *bro.m–* P 46, 117, 119, 188, Q 2a, 73, 156; *bręm–* K 316; *brum–* L 317, 417; *bro.mbiərn* L 416; *bro.mbi.rn* L 415; *brombię.rə* P 186; *brōmbiɛrə* P 187; *brǫ.m–* P 192, Q 159,

Annex D: A page from the *Woordenboek van de Vlaamse Dialecten*, the Dictionary Text (pt. III General Vocabulary, fasc. 1 Birds, p. 47)

ROODBORSTJE
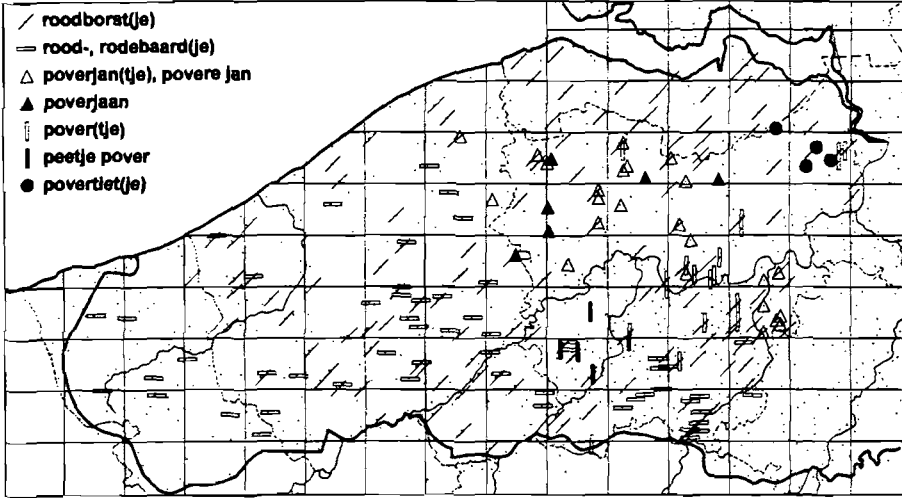
winterkneutje : ♦ Eeklo.

winterkoning(-kje) : s p o r . OV zuid beoosten Schelde, Waasl. en OZV ; z e l d z . Veurne-Ambacht, ook Stene, Gits, Zuidzande en Gent.

♦s p o r . WV, OZV en OV ; z e l d z . WZV. Wdb : Joos : winterkeuninksken. Ts : Bij-dragen Deinze : *winterkeuninkske*, omg. Deinze. Div : Avifauna : winterkeu-ninkske, OZV. Dialect OZV : winterkeu-ninkske(n), -koninkje(n).   Nederlandse Vogelnamen : winterkeuninkske, OZV.

winterpover : ♦ Kloosterzande.

wipstaartje [*wupstaartje*] : Poelkapelle.

ROODBORSTJE

Het roodborstje (Erithacus rubecula) is een klein, gedrongen vogeltje met oranje-rode borst. Men ziet het in parken en tui-nen. 's Winters komt het soms bij de men-sen om voedsel bedelen. Zie afb. 12.

WVD 66ᴍ (1995), 201 ; WVD 66 (1995), *102* ; *N 9 (1961), 24* ; *Materiaal Men-*schaert (1991), 83 ; Vandecasteele (1978), 84 ; Wielewaal (1952), 250 ; DC 6 (toegift)

(1938), 28 ; ZND 34 (1940), 72 (toegift). Volk en Taal 3 (1890), 170 ; Annalen Land van Waas 73 (1970), 114 ; Duumpje 13 (1987), 17 ; Bijdragen Deinze 55 (1988), 266. Dialect OZV (1982), 179 ; Neder-landse Vogelnamen (1995), 192.
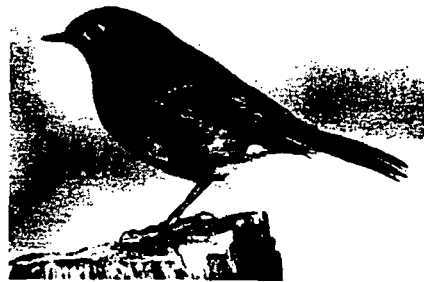
Afb. 12. Roodborstje.

arme jan : Ieper.

peetje pover : f r e q . OV zuid tussen Leie en Schelde, ook Dikkelvenne.

peetje povers [*peetje pofers*] : Kruishou-tem.

Annex E: A page from the *Woordenboek van de Vlaamse Dialecten*, the Scientific Text (pt. III General Vocabulary, fasc. 1 Birds, p. 47)

winterkneutje: ZND 34: I 158. **winterkoning:** WVD 66: I 166; N 9: I 179, N 82; *wintərkeuning:* DC 6: I 143; *wintərkeuningk:* DC 6: I 115. **winterkoninkje:** WVD 66: H 51, H 84, I 106, I 116, I 140, I 143b, I 208, N 33, O 199; N 9: O 46; Wielewaal: I 175, O 37, O 111, O 207; DC 6: I 107, I 109, I 112, I 116c, I 125a, I 138, I 139, I 142, I 144a, I 162; ZND 34: H 1, H 2, H 9a, H 12, H 21, H 29, H 69, H 84, H 110, H 113, I 155, I 175, I 183, I 193, I 213, I 241, I 250, I 252, I 258, N 28a, N 35, N 36, N 38, N 67, N 69, N 83, N 131, N 141, O 14, O 36, O 38, O 47, O 86, O 130, O 144, O 181, O 208; *wientərkeeningkskən:* ZND 34: O 150, I 264; *wientərkeuniengksjə:* ZND 34: N 74; *wientərkeuniengsjə:* ZND 34: H 24; *wintərkeuneingkskə:* ZND 34: O 82; *winterkeuningkskən:* ZND 34: O 216; *wintərkeuningsjə:* WVD 66M: H 18; N 9: N 26; *wintərkeuningskə:* ZND 34: N 38; *wintərkeuningskən:* WVD 66M: I 241; ZND 34: I 164, I 176; *wintərkeuningskie:* ZND 34: N 38; *wintərkeuningstjie:* ZND 34: H 36; *wintərkooningskən:* N 9: I 264; *wintərkûningskə:* ZND 34: I 218. **winterpover:** DC 6: I 116c. **wipstaartje** wupstaartje *wupstertjə:* Vandecasteele: N 31.

Annex F: Classification of the General Vocabulary

1.   Man
       1a Body
       1b Clothing
       1c Internal reality (mind, emotion ...)
2.   Domestic life
       2a House
       2b Family life
       2c Food
3.   Society
       3a Societal behaviour
       3b School
       3c Religion
       3d Amusement
4.   The external world
       4a The material world
       4b Plants and animals
       4c The a priori (time, space, form, ...)

A 38 Million Words
Dutch Text Corpus and its Users

J.G. Kruyt and M.W.F. Dutilh,
Instituut voor Nederlandse Lexicologie INL, Leiden, Nederland

Abstract: The use of text corpora has increased considerably in the past few years, not only in the field of lexicography but also in computational linguistics and language technology. Consequently, corpus data and expertise developed by lexicographical institutions have gained a broader scope of application. In the European context this has led to a revised view of corpus design. In line with these developments, the Institute for Dutch Lexicology (INL) has since 1994 been providing external access to steadily improving corpora via Internet. In August 1996, the *38 Million Words Corpus* was available for consultation by the international research community. The present paper reports on the characteristics of this corpus (design, text classification, linguistic annotation) and on its use, both in dictionary projects and in linguistic research. In spite of limitations with respect to corpus design, the INL corpora accessible via Internet have proved to meet external needs. By providing these facilities, the INL has acquired a much broader experience in corpus-building than before, which is essential for new, internal dictionary projects. Giving external access to corpus data which was developed primarily for internal purposes, may be profitable for all parties involved.

Keywords: LARGE ELECTRONIC DUTCH TEXT CORPUS, CORPUS DESIGN, TEXT CLASSIFICATION, TOPIC, PUBLICATION MEDIUM, LINGUISTIC ANNOTATION, ON-LINE ACCESS VIA INTERNET, CORPUS USERS

Samenvatting: Een tekstcorpus Nederlands (38 miljoen woorden) en de gebruikers ervan. Het gebruik van tekstcorpora is de laatste jaren aanzienlijk toegenomen, niet alleen op het gebied van de lexicografie maar ook in de computationele linguïstiek en de taaltechnologie. Ten gevolge daarvan kregen de corpusdata en de expertise opgebouwd door lexicografische instellingen een breder toepassingsdomein. Op Europees niveau leidde dit tot een herziene visie op corpussamenstelling. In overeenstemming met deze ontwikkelingen, geeft het Instituut voor Nederlandse Lexicologie (INL) sinds 1994 externe toegang via Internet tot steeds beter wordende corpora. In augustus 1996 was het *38 Miljoen Woorden Corpus* gereed voor consultatie door het internationale onderzoeksveld. Dit artikel beschrijft de karakteristieke kenmerken van dit corpus (corpussamenstelling, tekstclassificatie, linguïstische annotatie) en het gebruik in zowel woordenboekprojecten als in taalkundig onderzoek. Ondanks beperkingen ten aanzien van corpussamenstelling, is duidelijk gebleken dat de INL corpora die via Internet toegankelijk zijn, voorzien in een externe behoefte. Door deze faciliteiten aan te bieden, heeft het INL een veel bredere ervaring in corpusopbouw opgedaan dan voorheen. Deze is van essentieel belang voor nieuwe interne woordenboekprojecten. Het verlenen van externe toegang tot corpusdata die primair voor interne doeleinden ontwikkeld zijn, kan voor alle betrokken partijen profijt hebben.

Trefwoorden: GROOT ELEKTRONISCH NEDERLANDS TEKSTCORPUS, CORPUSSA-MENSTELLING, TEKSTCLASSIFICATIE, ONDERWERPSDOMEIN, PUBLICATIEMEDIUM, LINGUISTISCHE ANNOTATIE, ON-LINE TOEGANG VIA INTERNET, CORPUSGEBRUIKERS

1.   Introduction

In the early eighties, large electronic text corpora of national languages were developed mainly for lexicographical purposes (Zampolli and Cappelli 1983). Until the early nineties, however, a major problem was the management of the huge amounts of data stored in the computer, which caused lexicographers still to work with paper copies of concordances (cf. Clear 1987). Presently, more flexible access to large corpora is feasible. Corpora published on CD-ROM are distributed by the *Linguistic Data Consortium* (LDC) in the USA and the *European Language Resource Association* (ELRA) in Europe. Several institutions with a long-standing lexicographical background provide access to corpora via the Internet, for example *CobuildDirect Service* (Krishnamurthy 1996), the Italian *DBT* (Biagini and Picchi 1996) and the Dutch corpus services of the Institute for Dutch Lexicology INL (Kruyt 1995a, b, Kruyt *et al.* 1995).

The use of corpora has increased considerably in the past few years. Recent studies show the importance of corpus data for lexicography (e.g. Noël *et al.* 1995, several studies in Gellerstam *et al.* 1996 and Kiefer *et al.* 1996). Major publishers spend money on commercial corpus-based dictionaries, such as *Collins Cobuild English Language Dictionary* (1987) and *Longman Language Activator* (1993). Outside the field of lexicography, large corpora have become important for computational linguistics (Church and Mercer 1993) and language technology. From the perspective of a European infrastructure for language technology, the European Commission considered the corpus data and expertise developed by lexicographical institutions important enough to support projects in which the institutions contribute to the realization of the intended European infrastructure (cf. Kruyt 1995a, Teubert 1995, Zampolli 1996).

Corpora users have different attitudes towards corpus design. Lexicographers traditionally aim at a "representative" or "balanced" corpus, that is, the corpus should be appropriate as the basis for generalizations concerning the language as a whole. Corpus size (very large corpora) rather than corpus design is considered essential by many computational linguists using statistical methods of language analysis (cf. Church and Mercer 1993). Biber (1994) shows how complex it is to achieve "representativeness", even with the present computational methods for language analysis. Indeed, corpus practice demonstrates that lexicographical corpora for standard-language dictionaries may have very different corpus designs (Kruyt and Putter 1992, Kruyt and Van Sterkenburg 1996). The complexity of the notion "representativeness" (cf. Teubert 1995: 119), the different interests of corpus users and the costs of corpus development, have, at a European level, led to a shift of focus from building a separate, closed corpus for each project or application towards the development of

reusable, multifunctional and harmonized reference corpora for the European languages (Zampolli 1996). Flexible corpus use is ensured by the option of selecting user-defined subcorpora from a very large corpus with a composition as diversified as possible (Kruyt and Van Sterkenburg 1996).

In line with the interest in corpora and the European views on corpus development, the INL has broadened its scope in the past few years. Besides the ongoing compilation of the dictionary projects *Woordenboek der Nederlandsche Taal* (WNT) en *Vroegmiddelnederlands Woordenboek* (VMNW), the INL decided to participate in European corpus and lexicon projects. The INL also decided to make corpora accessible via Internet, so as to provide corpus facilities to the (inter)national research community. The INL opted for a phased approach, i.e. developing steadily improving corpora. Although representativeness was not aimed at, corpus design was well thought-out. In 1994, a *5 Million Words Corpus*, with a diversified composition and automatically annotated for lemma and part of speech, was made accessible via Internet (Kruyt 1995a, b). A *27 Million Words Newspaper Corpus*, with improved linguistic annotation and retrieval functionalities, followed in 1995 (Kruyt *et al.* 1995). At the end of August 1996, a *38 Million Words Corpus* with a diversified composition was made available in a similar way. This corpus is different from the former ones in various aspects: (a) size, (b) a broader coverage with respect to topic (subject domain), text types (with publication medium as parameter) and time span, (c) a more extended linguistic annotation, (d) the application of international standards for text classification and linguistic annotation, and (e) improved retrieval functionalities. The *38 Million Words Corpus* and its users will be characterized in the following sections. Where relevant, the use of the other INL corpora accessible via Internet will be discussed.

2.   INL 38 Million Words Corpus 1996

2.1   Composition

The INL has been acquiring electronic texts from several publishing houses since 1992. For reasons of copyright under Dutch law, the types of use permitted by the copyright holder are specified in a written contract between the INL and the copyright holder. Most text providers so far have given permission for internal use, as well as for external consultation by Internet for noncommercial research purposes. External use was particularly relevant to the *38 Million Words Corpus*. The required permission of the copyright holders limited the availability of texts to be incorporated into the corpus. Under this restriction, a corpus as diversified as possible was aimed at, so as to offer the research community an optimal opportunity to investigate language phenomena in different text types.

Corpus texts have been selected from the INL electronic text archive according to the following criteria. The language covered is standard Dutch

and Flemish (i.e. no dialect) as used in the Netherlands and Belgium. The corpus should preferably consist of components with a more or less equal size but with different contents. Broad coverage and balanced proportions were aimed at with respect to topic, publication medium and time span. However, coverage and balance were affected by availability or copyright restrictions, and in some cases, by inappropriate text formats. The resulting corpus consists of three main components: a component with varied composition (ca. 12,7 million words), a newspaper component (ca. 12,4 million words), and a component of legal texts (ca. 12,9 million words).

The varied component covers the period 1970-1995. It includes 18 single books and one title with 24 volumes, texts from issues of seven magazines, texts from 50 daily issues of the Belgian newspaper *De Standaard* (other newspapers being incorporated in other INL (sub)corpora), texts written to be read out in TV news broadcasts for adults and for youths, 18 *Queen's Speeches*, parliamentary reports over two months, and three issues of the *Law Gazette of the Kingdom of the Netherlands*. This subcorpus covers six topics (cf. section 2.2). For most text sources, all available text material for the purpose has been included. From some magazines with a large number of annual issues, half or a quarter of the issues have been selected for reasons of balance. For a more detailed survey, see the appendix.

The newspaper component consists of issues of the *Meppeler Courant*, dating from 1992-1995. Another newspaper available at the INL, *NRC*, was not selected for this (sub)corpus, as it forms part of the contents of the INL's *27 Million Words Newspaper Corpus 1995*. The selected newspaper is published three times a week. The INL receives a selection of articles per newspaper issue. All the material available up to 1996, grouped into monthly files, has been included. Two topics ("mixed" and "sports") are covered (cf. section 2.2).

The legal text component is a compilation of Dutch legal texts operative in 1989, including 5,875 laws, orders and decrees, protocols, agreements, treaties or conventions, etc., dating from 1814 up to 1989. This subcorpus has been derived from the *NLEX* database (the version without the text added by the publisher), with exclusion of texts undated or written in French.

2.2   Text classification

The corpus texts have been classified according to two parameters, viz. publication medium (in a broad sense) and topic (subject domain). For both parameters, a set of classification categories was distinguished on the basis of external, rather than linguistic criteria (cf. Biber 1994). The value of this (traditional) type of classification for corpus linguistics is criticized, particularly with regard to topic (Sinclair and Ball 1995). However, a new, commonly accepted, linguistically founded classification scheme has not been developed yet. In classifying the corpus texts, our sole intention was to assist the researcher in defining subcorpora from the whole corpus (cf. section 2.3).

The publication medium categories distinguished are: "book", "newspaper", "magazine", "written to be spoken", "reported speech", and "miscellaneous". Reference works (*Handboek van de Nederlandse pers en publiciteit*) or specific codes (ISBN, ISSN) assigned by the publisher have been used for classifying corpus texts as "newspaper" and "magazine" or "book" respectively. "Written to be spoken" refers to a text written beforehand, which is to be read out in public. In the corpus, this category is covered by the TV broadcast texts and the *Queen's Speeches*. "Reported speech" refers to a grammatically and stylistically corrected report of spoken language, rather than to transcribed spoken language. The parliamentary reports belong to this category. "Miscellaneous" includes texts that could not be classified in one of the former categories.

For topic classification, the two-level classification scheme proposed by Norling-Christensen (1996) for topic classification in the European *PP-PAROLE* project[1] is applied, based on the topic scheme used for the corpus underlying the Danish Dictionary edited by the Society for Danish Language and Literature (DSL) in Copenhagen.

Our corpus texts appeared to cover only part of the (sub)categories of the *PAROLE* scheme. The resulting topic categories are "HEALTH" with subcategories "health" and "psychology", "HUMANITIES" with subcategories "philosophy" and "language", "LEISURE" with subcategories "leisure" and "sports", "SCIENCE" with subcategories "astronomy" and "environment", "SOCIETY" with subcategories "social studies", "politics" and "law". A final category "MIXED" refers to texts covering a broad variety of topics, e.g. newspapers. For magazines, topic classification is based on branch-codes listed in the reference work *Handboek van de Nederlandse pers en publiciteit*, which have been translated into the *PAROLE* topic categories. For books published since 1980, so-called *CIP* (*Cataloguing in Publication*) data are available in the source. *CIP* data include up to three codes, which have their origin in different Dutch classification schemes (e.g. *UDC*, *NUGI*, *SISO*), as well as keyword terms. This data is reinterpreted in terms of the *PAROLE* topic scheme. Books without *CIP* data are classified on the basis of the title of the book or information in the front or back matter (cf. Dutilh and Kruyt 1992). Newspapers and TV news texts, covering many topics, have been classified as "mixed". However, the sports pages of the *Meppeler Courant* could be classified as "leisure / sports", based on the title of these pages as encoded in the electronic files. The classification of the remaining texts was based on general knowledge about the text.

2.3   Access to the corpus data

A retrieval (corpus query) system has been developed which enables the researcher to search for single words or for word patterns in the corpus, including some rather primitive, predefined word classes (e.g. past participle) and syntactic patterns (e.g. noun phrase NP, prepositional phrase PP) which can be customized and extended by the user. The result of a query is, in the end, a

series of concordances (keywords in context) meeting the query specifications. A major problem in information retrieval is the effectiveness of a search (recall and precision) (cf. Kruyt 1995b), in our case, the extent to which the query system retrieves the exact linguistic data the researcher needs from the corpus (no more and no less). Two functionalities of the corpus retrieval system reduce the overflow of data in the output of a query. One is linguistic annotation in terms of lemma (headword) and part of speech (POS). The other is the option to select a user-defined subcorpus from the whole corpus.

The researcher may address a query to a subcorpus selected from the perspective of his research purposes rather than to the whole corpus. Corpus composition (section 2.1), text classification (section 2.2), and text date enable the user to select subcorpora easily according to the parameters corpus component, topic, publication medium, and period. Selection of one, more or all of these options results in the display of text source surveys on the screen. The researcher has the opportunity to select individual text sources from these surveys. In this way, each researcher can define his own subcorpora, based on selection at the level of individual texts. This reduces an overflow of output caused by data meeting the query but coming from texts without relevance to the research purposes. For each defined subcorpus, its size can be displayed on the screen.

The other functionality reducing an overflow of irrelevant data is linguistic annotation, the explicit encoding of linguistic features in the electronic text (cf. Grefenstette 1996). The main function of linguistic annotation is that searches may be specified in terms of various linguistic features. The word-forms (tokens) in the corpus texts have automatically been annotated with lemma (headword) and two types of part of speech (POS)[2]. One POS scheme includes thirteen basic POS categories (Van der Voort van der Kleij *et al.* 1994). The other POS scheme is fine-grained, each POS being subcategorized in terms of type and/or characteristic features, conformant with the European *MECOLB* standard[3] (Raaijmakers and Dutilh 1995). For example, the *MECOLB* POS tag for the word-form "loopt" ("walks") is "VRB (intrans, indic, pres, sg, 2/3)", specifying it as an intransitive verb with its values for the features mood, tense, number and person. As a result of the linguistic annotation, a query may include references to specific word-forms, to specific basic parts of speech, to *MECOLB* parts of speech, *MECOLB* POS subtypes and features, and to headwords (lemmas), either separately or combined in one query definition. For example, the following searches may be expressed in the formal query language:

(1)   "Search the occurrences of the word-form 'werk' (work)". The first output is a list containing "werk" specified as noun and "werk" specified as verb. The user makes a selection and the relevant concordances appear on the screen.

(2)   "Search the occurrences of the word-form 'werk' under the condition that 'werk' is a noun". This query, with the double specification ("werk" plus

"as noun"), immediately results in the relevant concordances (i.e. without occurrences of the verb form).

(3)   "Search the occurrences of the lemma 'president' (president)". The output is a series of concordances with occurrences of singular "president" and plural "presidenten".

(4)   "Search the occurrences of the lemma 'president' followed by a prepositional phrase PP, within a distance of (say) 7 arbitrary word-forms". The output concordances show occurrences of word-forms of the lemma "president" (cf. example (3)) only if they are followed by a PP within the specified distance (and not all the other ones without the PP). For example: "president van Amerika" (president of America), "presidenten uit diverse landen" (presidents from various countries), instead of "president", "presidenten" not followed by the PP specification.

(5)   "Search the occurrences of the lemma 'werken' (to work) with *MECOLB* feature 'present tense'". The output is a series of concordances with occurrences of the verb "werken" as far as they have been annotated by the feature "present tense".

From the perspective of search effectiveness, the annotated corpus has an added value with respect to a "raw" (not annotated) corpus or "raw" texts available on CD-ROM, Internet, etc. "Raw" text can essentially be addressed at the level of word-form (token) only, whereas annotated text can be addressed at all linguistic features expressed by the encoding (including combinations). Due to the headword and POS annotation, the researcher does not need to specify the whole paradigm of a word. This is not only a matter of user-friendliness. For ambiguous word-forms such as "school" (ambiguous for noun "school" and verb form "sheltered") or "sleep" (ambiguous for noun "train", and verb forms "polished" and "drags"), the headword and POS encoding enables the search engine (the computer program) to discriminate between the different headwords. As a consequence, the researcher will only retrieve the occurrences of the headword he is interested in, and not all the others as well. The facility of combining search specifications or conditions in one query (examples (2), (4) and (5)) allows an additional curtailment of the output. As opposed to these positive effects on effectiveness, it should be noticed that incorrect encodings and encodings that could not be disambiguated, result in some overflow of data (found but not intended by the researcher) and/or some deficiency of data (intended but not found). For a more detailed description of retrieval facilities, see Kruyt *et al.* (1995).

2.4   Use of the 38 Million Words Corpus

The *38 Million Words Corpus* was ready for consultation via Internet at the end of August 1996[4]. It is consulted by lexicographers in dictionary and lexicon

projects as well as by individual users for various purposes. In the short term, it will be used in university courses in corpus linguistics. The INL keeps record of particular user data, not only so as to trace potential misuse, but also and more positively, to obtain insight into the needs of the corpus users. Where relevant, other INL corpora will be referred to, particularly the earlier *5 Million Words Corpus 1994* and the *27 Million Words Newspaper Corpus 1995* which also are accessible via Internet. Published work refers to these corpora rather than to the most recent one.

2.4.1 Use by lexicographers

The *38 Million Words Corpus* is used in the preparatory phase of a new dictionary project at the INL, which will start after completion of the *Woordenboek der Nederlandsche Taal (WNT)* in 1998. Depending on the concept for this dictionary, texts will be selected from the INL text archive and corpora for incorporation into the closed corpus for the dictionary. The assumption is that additional texts are to be selected and acquired.

In addition to this strictly internal use, the *38 Million Words Corpus 1996* is being consulted within the framework of several international corpus-based lexicon projects. The Dutch-Flemish project *Referentie Bestand Nederlands (RBN) (Reference Database of the Dutch Language)*, a project under the authority of *Commissie Lexicografische Vertaal Voorzieningen (CLVV) (Committee for Lexicographical Translation Facilities)* and supervised by Prof. dr. W. Martin, aims at the development of a lexical database for the purpose of noncommercial dictionaries with Dutch as either source or target language. The INL corpora were used for the composition of the entry list. Prof. Martin selected a subcorpus of ca. 10 million words from the *38 Million Words Corpus* to be consulted by the lexicographers for determining the contents of several fields of the microstructure (e.g. lexical and grammatical collocations, idioms). Lexicographers from several cities in the Netherlands and Belgium work on the INL computer system daily (cf. diagram on p. 238).

The EC-funded project *LE-PAROLE*[5] aims at the development of comparable corpora (each 20 million words) and lexica (each 20,000 entries) for 12 Western European languages, according to European standards with respect to linguistic background, contents, linguistic annotation schemes, text representation and access. The INL is responsible for the Dutch corpus and lexicon. The entry list for the Dutch lexicon has been determined on the basis of linguistically annotated INL corpora containing a total of ca. 54 million words: the *27 Million Words Newspaper Corpus 1995*, a *15 Million Words Corpus* with diversified composition and the varied component (ca. 12 million words) of the *38 Million Words Corpus*. The varied component and the newspaper component of the *38 Million Words Corpus* have been selected as subcorpora for determining syntactic complementation patterns for various types of POS.

Vliegen (1996) studied complementation patterns of Dutch verbs of visual, auditory, olfactory and gustatory perception and verbs of verbal communication within the framework of the EC-funded project *DELIS* (LRE 61.034). This project aimed at methods and tools to build lexical entries based on evidence extracted from textual corpora, combining a corpus-based lexicographical approach and frame-based semantic theory. In Vliegen (1996), the *5 Million Words Corpus 1994* is mentioned among the corpora he consulted for this study. The INL user records show that he also consulted the *27 Million Words Newspaper Corpus 1995* and the *38 Million Words Corpus 1996* for his research on verbs of perceiving and verbal communication.

2.4.2   Use by individuals

Since the *38 Million Words Corpus 1996* was developed rather recently, the individual use of this corpus will be considered against the background of the use of the earlier INL corpora accessible via Internet (the *5 Million Words Corpus 1994* and the *27 Million Words Newspaper Corpus 1995*). Only external (i.e. non-INL) use will be discussed. Note that the figures presented reflect the momentary status on a particular date; figures change daily.

By March 1, 1997, 175 external users signed a personal user agreement for one or more INL corpora: 138 have access to the *5 Million Words Corpus 1994*, 98 to the *27 Million Words Newspaper Corpus 1995*, and 58 to the *38 Million Words Corpus 1996*. It should be noticed that ca. 30 subscribers had not consulted the corpora yet, and that 16 user accounts are reserved for students of the Free University of Amsterdam, who will follow a short-term course in corpus linguistics. Monthly user records show that the number of users of each of the corpora is steadily growing (two new users per month on the average for the earlier corpora; 3,5 per month on the average for the most recent one). The users are mainly (over 80%) from the Netherlands and Belgium; ca. 14% comes from Germany, the USA, the United Kingdom and South Africa. Other users are from Norway, Denmark, Austria, Slovenia, Latvia, Malaysia and Korea.

By March 1, 1997, the corpora were accessed 3337 times in total. "Accessed" means that a user made contact with ("logged in to") the INL computer so as to address one or more queries to a corpus. For each of the corpora, the diagram (see p. 238) shows the number of log-ins over the period August 1996 up to March 1997. The rates for the *38 Million Words Corpus* increase very fast. This can mainly be explained by its use by the lexicographers of the *RBN* project (see under section 2.4.1). The number of log-ins by the other users is represented by the dotted line. If this line is taken into account, the three corpora show steadily rising curves, although somewhat flatter for the older ones than for the latest one. The highest rate (926) is still for the oldest, the *5 Million Words Corpus*. The average number of queries per consultation is very different for the corpora and for the individual users. As a rough indication, the proportion can be fixed at four to five queries per consultation.
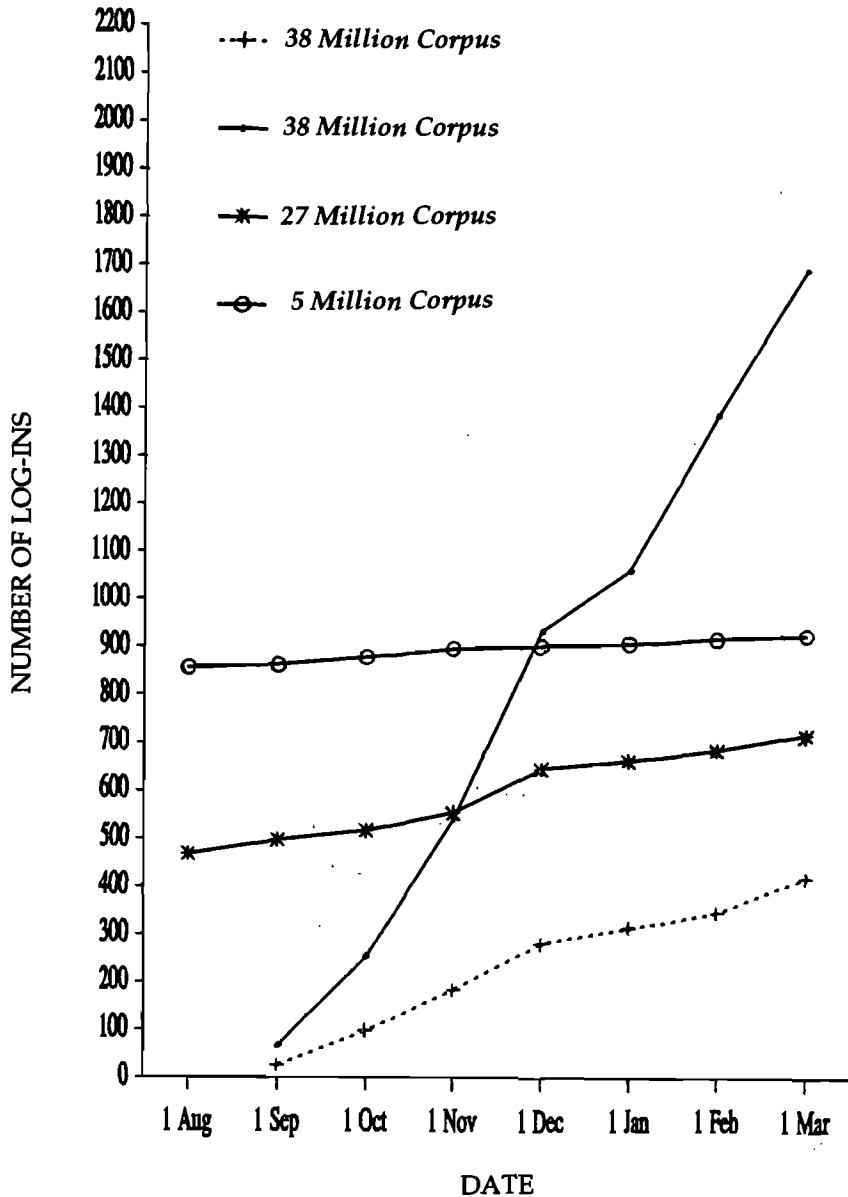
Number of log-ins in the period of August 1, 1996 up to March 1, 1997, for three INL corpora containing 5 million, 27 million and 38 million words, respectively. The dotted line shows the number of log-ins for the *38 Million Words Corpus*, excluding the users of the *RBN* project.

From publications and from an analysis of the queries over the past half-year (cf. Kruyt 1995a for earlier research with INL corpora), it can be concluded that the corpora are consulted for essentially two purposes: incidental looking up of particular words or phrases and research in the field of linguistics and social studies. Some examples of research are the following: Hoeksema and Klein (1996) investigated the usage of Dutch "even" (equally) as a comparative and as an adverb of degree within the framework of the *PIONIER* project *Reflections of Logical Patterns in Language Structure and Language Use*. Pollman (1996) used the historical distribution of dates in newspaper corpora for a psychological essay about memory and the systematics of collective historical consciousness. In her thesis, Cornelis (1997) investigated the passive construction in several Dutch text corpora. Corpus composition appeared to have an impact on the results (Cornelis 1997: 209). Other research topics appearing from the queries concern the orthography of geographical names, male and female variants of nouns, conjunctions, reciprocal pronouns, verbs with strong and weak inflection, inflection of separable verbs, specific verb and noun constructions, words with particular prefixes or suffixes, the vocabulary in the field of social legislation, fashionable words and neologisms.

3.   Conclusion and discussion

By providing easily accessible instruments for corpus-based research, the three INL corpora on Internet have proven to meet external needs from the (inter)-national research community. The function of the *38 Million Words Corpus* in international lexicon projects demonstrates its relevance for lexicographical purposes in spite of its shortcomings with respect to corpus design and text classification (cf. section 2).

Our conclusion is that a less ideal corpus is apparently better than no corpus. This may apply particularly to a minority language such as Dutch, as there exist no other Dutch corpora comparable in size, coverage, linguistic annotation and easy access. In English, for instance, several large corpora are available (e.g. the *British National Corpus* and the *Cobuild Bank of English*). Rather than a series of steadily improving corpora (cf. section 1), the INL might at once have opted for an ideally representative general-language corpus composed according to the principles outlined by Biber (1994). In order to achieve representativeness, Biber proposes a cyclical method consisting of four stages: (1) pilot empirical investigation / theoretical analysis, (2) corpus design, (3) compiling a portion of the corpus with grammatical tagging (pilot corpus), and (4) empirical investigation on the pilot corpus by automatic language analysis. The results of stage (4) are used to confirm or modify the design parameters of stage (2), and the process is repeated until representativeness is reached. A critical factor with respect to this method, though promising, seems to be feasibility due to, among other things, labour-intensiveness and, for Dutch, the lack of machine-readable texts covering the various registers and copyright restrictions. The INL could

not have met the needs for corpus data, as has been done since 1994, if a method like the one proposed by Biber had been applied.

The success of the INL corpora can be understood from the efforts needed for corpus-building. The development of the corpora required several man-years per corpus. Additionally, the INL has the technical infrastructure and the specialists in different disciplines (lexicographers/corpus linguists, computational linguists and information scientists) needed for large-scale, annotated corpora. For researchers, a corpus is an instrument rather than an end in itself. But even in the lexicographical projects referred to above, the use of an available, easily accessible corpus was preferred to building a corpus specifically for the purpose. In these projects particularly, the option of defining subcorpora has been applied.

From an internal point of view, the development of the three corpora, particularly the last one, has yielded much experience and insight with respect to the procedures to be followed, their routing, the time needed for the various phases and the problems to be solved. This experience will be very useful for the planning of new dictionary projects to be started at the INL after the completion of the current ongoing dictionary projects. The corpora developed so far may function as pilot corpora in the sense of Biber's stage 3.

For the near future (1997-1998), the INL aims at expanding and enhancing the research instruments, both for internal and external use. Within the framework of the *LE-PAROLE* project, a *corpus is being prepared in which text struc-tural elements are encoded in TEI* format, an international standard for the encoding and interchange of electronic text for research purposes, developed in the past years by the *Text Encoding Initiative (TEI)* (Sperberg-McQueen 1994). This corpus will be the basis for a large, syntactically annotated corpus, which will enhance retrieval functionalities. Furthermore, the INL intends to offer the research community on-line annotation facilities which enable the researcher to annotate his own texts by use of linguistic software developed by the INL. Due to the external importance of these facilities, the Netherlands Organization for Scientific Research NWO will cofinance the required hardware.

Until the early nineties, the INL developed corpora for lexicographical purposes only. By broadening its scope (cf. section 1), a much broader experi-ence in corpus-building has been acquired which is indispensable for new internal lexicographical projects and for the INL Integrated Language Database of 12th-21st Century Dutch (cf. Kruyt 1995b). In view of the need for corpus data and the efforts needed for building corpora, corpus builders may, from the outset of a new dictionary project, consider the possibility of giving access to their data to external users and establishing the legal conditions to realize this. For the INL, this has proven to be profitable for all parties.

Notes

1. The *LE-PAROLE* project (LE2-4017) is a project cofinanced by the European Commission. The aim of the project is the development of corpora and lexica for 12 Western European languages, which are comparable with respect to linguistic background, contents, text representation (in *TEI*) and access. They will be used for (multilingual) research and language technology products. The *PAROLE*-specifications for the contents of corpus and lexicon are based on European standards developed in *EAGLES* and related projects and have been specified in the earlier *PP-PAROLE* project (MLAP63-386). *PP-PAROLE* was preceded by the *NERC* project (Calzolari *et al.* 1996), a feasibility study into a Network of European Reference Corpora (cf. Kruyt 1995a).

2. Lemma and basic POS category have automatically been assigned by the lemmatizer / POS-tagger DutchTale, developed by the INL (Van der Voort van der Kleij *et al.* 1994) in the framework of the European *NERC* project (see note 1). Improved versions have been developed by S. Raaijmakers (Kruyt *et al.* 1995). These improvements particularly addressed the encoding of POS and headword for word-forms that were not found in the lexicon, and the disambiguation of word-forms that were assigned more than one POS and/or headword on the basis of the lexicon. For information on the lexicon component, see Van der Voort van der Kleij and Kruyt (1997).

3. The *MECOLB* standard has been developed in the framework of the European project *MECOLB* (MLAP93-21), sponsored by the European Commission and coordinated by R. Neumann, Institut für Deutsche Sprache, Mannheim. The *MECOLB*-tag set for Dutch morphosyntactic annotation was developed in cooperation with the TOSCA Research Group (University of Nymegen), under the direction of Prof. dr. J. Aarts. The *MECOLB* POS-encodings in the corpus have been partially disambiguated by use of a neural network which was trained on a corpus developed in cooperation with the TOSCA Research Group.

4. Access to the INL corpora is provided free of charge for noncommercial research purposes. For each corpus, a separate personal user agreement is to be signed. An electronic user agreement form can be obtained from the INL mail server

    Mailserv@Rulxho.LeidenUniv.NL

    or by request from the INL helpdesk

    Helpdesk@Rulxho.LeidenUniv.NL

    A hard copy of the agreement form must be made, a copy kept, and a signed copy returned to the Institute for Dutch Lexicology INL, P.O. Box 9515, 2300 RA Leiden, The Netherlands, fax. 31 71 527 2115. After receipt of the signed user agreement, the applicant will be informed of his/her user name and password.

5. See note 1.

References

Biagini, Lisa and Eugenio Picchi. 1996. INTERNET and DBT. Gellerstam, Martin, Jerker Järborg, Sven-Göran Malmgren, Kerstin Norén, Lena Rogström and Catarina Röjder Papmehl (Eds.). 1996: 47-53.

Biber, Douglas. 1994. Representativeness in Corpus Design. Zampolli, Antonio, Nicoletta Calzolari and Martha Palmer (Eds.). 1994: 377-407.

Calzolari, Nicoletta, Mona Baker and Johanna G. Kruyt (Eds.). 1996. *Towards a Network of European Reference Corpora, Report of the NERC Consortium Feasibility Study.* Pisa: Giardini Editori e Stampori.

Church, Kenneth W. and Robert L. Mercer. 1993. Introduction to the Special Issue on Computational Linguistics Using Large Corpora. *Computational Linguistics* 19(1): 1-24.

Clear, Jeremy. 1987. Computing. Sinclair, J.M. (Ed.). 1987: 41-61.

Cornelis, Louise H. 1997. Passive and Perspective. Van den Hoven, Paul and Wolfgang Herrlitz (Eds.). 1997: 1-295.

Dutilh, M.W.F. and J.G. Kruyt. 1992. *Feasibility Experiment Design Criteria. Investigation into Text Typological Classification Tools.* NERC Paper WP6-94. Unpublished report. Leiden: Instituut voor Nederlandse Lexicologie.

Gellerstam, Martin, Jerker Järborg, Sven-Göran Malmgren, Kerstin Norén, Lena Rogström and Catarina Röjder Papmehl (Eds.). 1996. *Euralex '96 Proceedings* I-II. Göteborg: Department of Swedish, Göteborg University.

Grefenstette, Gregory. 1996. Approximate Linguistics. Kiefer, Ferenc, Gábor Kiss and Júlia Pajzs (Eds.). 1996: 83-96.

Handboek van de Nederlandse pers en publiciteit. Gedrukte media. 1990. Schiedam: Nijgh Periodieken.

Hoeksema, Jack and Henny Klein. 1996. From Comparative to Adverb of Degree; Dutch *even.* Jonkers, Roel, Edith Kaan and Anko Wiegel (Eds.). 1996: 59-70.

Jonkers, Roel, Edith Kaan and Anko Wiegel (Eds.). 1996. *Language and Cognition* 5. Groningen: Centre for Language and Cognition, Department of General Linguistics, University of Groningen.

Kiefer, Ferenc, Gábor Kiss and Júlia Pajzs (Eds.). 1996. *Papers in Computational Lexicography COMPLEX '96.* Budapest: Linguistics Institute, Hungarian Academy of Sciences.

Krishnamurthy, Ramesh. 1996. The Data is the Dictionary: Corpus at the Cutting Edge of Lexicography. Kiefer, Ferenc, Gábor Kiss and Júlia Pajzs (Eds.). 1996: 117-144.

Kruyt, J.G. 1995a. Nationale tekstcorpora in internationaal perspectief. *Forum der Letteren* 36(1): 47-58.

Kruyt, J.G. 1995b. Technologies in Computerized Lexicography. *Lexikos* 5: 117-137.

Kruyt, J.G. and E. Putter. 1992. *Corpus Design Criteria.* NERC Paper WP6-129. Unpublished report. Leiden: Instituut voor Nederlandse Lexicologie.

Kruyt, J.G., S.A. Raaijmakers, P.H.J. van der Kamp and R.J. van Strien. 1995. On-line Access to Linguistically Annotated Text Corpora of Dutch via Internet. Rettig, Heike (Ed.). 1995: 173-178.

Kruyt, J.G. and P.G.J. van Sterkenburg. 1996. Corpus Design Criteria. Calzolari, Nicoletta, Mona Baker and Johanna G. Kruyt (Eds.). 1996: 57-72.

Neumann, Robert. 1995. *MLAP93-21 MECOLB Progress Report II.* WP5 Final report Quality Assessment. Unpublished report. Mannheim: Institut für Deutsche Sprache.

Noël Dirk, Bart Defrancq and Filip Devos. 1995. Considering Bilingual Dictionaries Against a Corpus. *Lexikos* 5: 57-81.

Noordman, L.G.M. and W.A.M. de Vroomen (Eds.). 1994. *Informatiewetenschap 1994, Wetenschappelijke Bijdragen aan de Derde StinfoN-conferentie.* Leiden: StinfoN.

Norling-Christensen, O. 1996. *Design and Composition of Reusable Harmonised Written Language Reference Corpora for European Languages.* Unpublished *PAROLE* paper. Copenhagen: Society for Danish and Literature (DSL).

Pollmann, Thijs. 1996. De wet van het uitdovend verleden. Wetenschapsbijlage, *NRC Handelsblad*, 11 January 1996: 1.

Raaijmakers, Stephan and Tilly Dutilh. 1995. Morphosyntactic Corpus Annotation for Dutch. INL Contribution to the *MECOLB* Project. Neumann, Robert. 1995: 10-21.

Rettig, Heike (Ed.). 1995. *Language Resources for Language Technology. Proceedings of the First European TELRI Seminar.* Mannheim: Institut für Deutsche Sprache.

Sinclair, J.M. (Ed.). 1987. *Looking Up. An Account of the Cobuild Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary.* London / Glasgow: Collins ELT.

Sinclair, John and Jackie Ball. 1995. *Text Typology.* Draft document EAGLES Text Corpora Working Group, submitted to EAGLES Workshop Issues in Corpus Linguistics, Madrid 1996.

Sperberg-McQueen, C.M. 1994. The Text Encoding Initiative. Zampolli, Antonio, Nicoletta Calzolari, Martha Palmer (Eds.). 1994: 409-427.

Teubert, Wolfgang. 1995. Language Resources: The Foundations of a Pan-European Information Society. Rettig, Heike (Ed.). 1995: 105-128.

Van den Hoven, Paul and Wolfgang Herrlitz (Eds.). 1997. *Utrecht Studies in Language and Communication* 10. Amsterdam / Atlanta: Rodopi B.V.

Van der Voort van der Kleij, John, Stephan Raaijmakers, Mathijs Panhuijsen, Martijn Meijering and Rogier van Sterkenburg. 1994. Een automatisch geanalyseerd corpus hedendaags Nederlands in een flexibel retrievalsysteem. Noordman, L.G.M. and W.A.M. de Vroomen (Eds.). 1994: 181-194.

Van der Voort van der Kleij, John and Truus Kruyt. 1997. Lexicon for Linguistic Annotation of Dutch Text. *TELRI Newsletter* 5: 32-35.

Vliegen, Maurice. 1996. Verbs of Perceiving and Verbal Communication in Dutch: Clausal Complementation. Gellerstam, Martin, Jerker Järborg, Sven-Göran Malmgren, Kerstin Norén, Lena Rogström and Catarina Röjder Papmehl (Eds.). 1996: 309-319.

Zampolli, A. 1996. Corpus Linguistics, the Use of Computers in the Humanities, Computational Linguistics. Calzolari, Nicoletta, Mona Baker and Johanna G. Kruyt (Eds.). 1996: XI-XXXIX.

Zampolli, A. and A. Cappelli (Eds.). 1983. *The Possibilities and Limits of the Computer in Producing and Publishing Dictionaries. Proceedings of the European Science Foundation Workshop, Pisa, 1981.* Pisa: Giardini Editori e Stampatori.

Zampolli, Antonio, Nicoletta Calzolari, Martha Palmer (Eds.). 1994. *Current Issues in Computational Linguistics: In Honour of Don Walker.* Pisa: Giardini editori e stampatori / Dordrecht: Kluwer Academic Publishers.

Appendix. Composition of INL 38 Million Words Corpus 1996

TOTAL CORPUS				38,164,250

VARIED SUBCORPUS

TOPIC	SUBTOPIC	MEDIUM	PERIOD	TOKENS
MIXED				6,380,508
	—	1 NEWSPAPER	1995	2,607,579
	—	2 WRITTEN-tb-SPOKEN	1991-1995	3,772,929
HEALTH				413,967
	Health	5 BOOKS	1992-1993	242,114
	Psychology	3 BOOKS	1993-1994	171,853
LEISURE				
	Leisure	2 MAGAZINES	1992-1995	1,017,881
HUMANITIES				1,030,033
	Languages	1 BOOK	1992	69,628
		1 MAGAZINE	1991-1995	515,921
	Philosophy	6 BOOKS	1993-1994	444,484
SCIENCE				1,040,131
	Environment	2 MAGAZINES	1989-1995	781,261
		1 BOOK	1989	61,559
	Astronomy	1 MAGAZINE	1992-1993	197,311
SOCIETY				2,874,744
	Politics	1 BOOK	1982	245,482
		2 MAGAZINES	1992-1995	628,026
		1 MISCELLANEOUS	1991	36,491
		1 WRITTEN-tb-SPOKEN	1970-1986, 1988	36,188
		1 REPORTED SPEECH	Nov-Dec 1995	1,869,099
	Social Studies	2 BOOKS	1990-1991	59,458
TOTAL				12,757,264

NEWSPAPER SUBCORPUS

TOPIC	SUBTOPIC	MEDIUM	PERIOD	TOKENS
MIXED				9,127,200
	—	1 NEWSPAPER	1992-1995	
LEISURE				3,305,237
	Sports	1 NEWSPAPER	1992-1995	
TOTAL				12,432,437

LEGAL SUBCORPUS

TOPIC	SUBTOPIC	MEDIUM	PERIOD	TOKENS
SOCIETY				12,974,549
	Law	1 MISCELLANEOUS	1814-1989	
TOTAL				12,974,549

On the Development of Scientific Terminology in African Languages: The Terminographer's Dilemma in a New Dispensation·

Louis Jacobus Louwrens, *Department of African Languages,*
University of South Africa, Pretoria, South Africa

Abstract: There currently exists a noticeable tension in South Africa between the political aim of one homogeneous South African nation on the one hand and the autonomy each language deserves in practice according to the constitution on the other. The real development of individual languages and the purposeful cultivation of language pride necessarily accentuate races and ethnical differences, which are contrary to the ideal of nation-building. Consequently, languages are subtly denied acknowledged constitutional rights in practice, which will impact negatively on the development of especially the African languages into technical and academic languages in their own right. The question thus arises whether it is sensible for the terminographer to develop scientific and technical terms for the African languages, while everything at this stage indicates that these terms will hardly, if ever, be used by subject specialists.

Keywords: TERMINOLOGY, AFRICAN LANGUAGES, LANGUAGE ATTITUDES, MEDIUM OF INSTRUCTION, LANGUAGE-TEACHING, LINGUISTIC AWARENESS, LANGUAGE PLANNING, LANGUAGE POLICY, MULTILINGUALISM, NATION-BUILDING, CONSTITUTIONAL RIGHTS, HARMONIZATION, STANDARD LANGUAGE, STANDARD SOTHO, STANDARD NGUNI, LANGUAGE BOARDS, LANGUAGE ACADEMIES, ETHNICITY, MONOLINGUAL SCHOOLS, RADIO SERVICES

Opsomming: Oor die ontwikkeling van wetenskaplike terminologie in Afrikatale: die terminograaf se dilemma onder 'n nuwe bedeling. Daar bestaan tans 'n merkbare spanning in Suid-Afrika tussen die politieke strewe na een homogene en ongedifferensieerde nasie enersyds, en die outonomie waarop elke taal konstitusioneel in die praktyk geregtig is andersyds. Die daadwerklike uitbouing van individuele tale en die doelgerigte kweek van 'n taaltrots aksentueer noodwendig rasse en etniese verskille, wat direk in stryd is met die ideaal van nasiebou. Gevolglik word tale in die praktyk op subtiele wyse vele konstitusioneel erkende regte ontsê wat veral die Afrikatale se ontwikkeling tot volwaardige tegniese en akademiese tale nadelig beïnvloed. Die vraag ontstaan dus of dit in so 'n klimaat vir die terminograaf sin maak om wetenskaplike en tegniese terme vir die Afrikatale te ontwikkel, terwyl alles in hierdie stadium daarop dui dat sodanige terme selde, indien ooit, deur vakspesialiste gebruik gaan word.

Sleutelwoorde: TERMINOLOGIE, AFRIKATALE, TAALHOUDINGS, ONDERRIGME-
DIUM, TAALONDERRIG, TAALBEWUSTHEID, TAALBEPLANNING, TAALBELEID, VEELTA-
LIGHEID, NASIEBOU, KONSTITUSIONELE REGTE, HARMONISERING, STANDAARDTAAL,
STANDAARDSOTHO, STANDAARDNGUNI, TAALRADE, TAALAKADEMIES, ETNISITEIT,
ENKELMEDIUMSKOLE, RADIODIENSTE

In the "New South Africa", the terminographer involved in the development of
scientific terminology in African languages is faced with one major uncertainty
brought about mainly by two factors, namely (a) existing attitudes regarding
the suitability of the indigenous languages for the purposes of scholarly dis-
course, and (b) the way in which the autonomy of these languages is affected
by the ideal of nation-building. As for the first point, two views can be distin-
guished: on the one hand, there are those who feel that the teaching of African
languages through a medium other than the vernacular itself is totally inexcus-
able; on the other, there are scholars who maintain that the indigenous lan-
guages are, at least for the present, not adequately equipped to be used as
media of instruction on a tertiary level. Currently the latter view predominates,
which explains the existing practice whereby African languages are taught
mainly, for example at tertiary institutions, through the medium of either
English or Afrikaans. This poses a serious problem for the terminographer,
which leads to the question whether it is worth the time and effort to embark
on ambitious projects for the development of terminology if there is no proof
that these languages will indeed in future reach a point where they will become
accepted as worthy means of scholarly discourse. These two issues are
addressed below and disquieting developments are identified which clearly
reveal how South Africa's indigenous languages are in practice denied the
autonomy they constitutionally deserve — a situation which calls for a
reassessment of the terminographer's position in a future dispensation.

Any effort to expand the scientific terminology of an African language
rests on the tacit assumption that, in future, the language will become a lan-
guage of learning. If this is not the case, what sense does it make to spend
manpower, time and money on term development if tuition at tertiary level is
certainly going to be conducted mainly through a medium other than the lan-
guage concerned?

Although the vision that the African languages can be developed into
appreciable languages of science and technology is not far-fetched, it is not
always met with the same degree of enthusiasm by scholars and university
teachers, and here both mother-tongue and non-mother-tongue speakers are
included. It will, therefore, be unwise for the terminographer to occupy him- or
herself with the development of terminology without assessing in advance how
the outcome of such an endeavour will be received.

The following remark by Nkondo (1987: 72) represents the viewpoint of
those who totally reject the idea of teaching African languages through a
medium other than the vernacular itself:

> For the native speaker no didactic principle would suggest anything other than the language itself for the medium of study. This is a very thorny issue, as we know, in spite of its obviousness. Arguments of internationalization or reaching a wider readership are indefensible.

Supporters of this view are often adamant that the *status quo* should be changed immediately, notwithstanding some almost insurmountable problems which would face the university teacher should this happen haphazardly. Furthermore, they claim that the poor performance of mother-tongue speakers at university level can mainly be attributed to students' poor proficiency in English due to ineffective schooling at pre-university level. With regard to mother-tongue instruction, the problem is aggravated by the fact that the teaching of African languages at secondary level is conducted through the medium of the vernacular, whereas at tertiary level the student is expected to switch to a foreign medium of instruction. Consequently, the questionable view is held that mother-tongue instruction through the medium of the mother tongue will lead to a dramatic increase in the pass rate of students in first language courses.

On the contrary, there are scholars who are strongly opposed to the view that African languages should be taught through the medium of the vernacular. One of the major arguments raised by this group is that the lack of terminology makes it impossible — at least for the moment — to achieve the standard of scientific reasoning which is expected at university level. One would then assume that, given the availability of an elegant and sophisticated terminology, this objection against the use of the vernaculars as media of instruction will fall away. That does not seem to be the case, however, since a second prominent argument raised by supporters of this view is that English is the international language through which scientific knowledge is accessed. It will hence serve very little purpose to present theses, dissertations, conference papers and scientific articles in an African language, because such contributions will deny international researchers as well as many fellow South Africans the privilege of sharing in newly acquired knowledge. Scholars and researchers therefore steer away form the vernaculars when it comes to academic work. So, for example, statistics reveal that during the period 1985-1995 only one article written in an African language was published in the *South African Journal of African Languages* despite the fact that the editorial policy of this journal makes provision for contributions in languages other than English. Even Afrikaans-speaking academics prefer to publish in English in order to reach a wider local and international readership. These observations reveal deeply rooted prejudices against the African languages, which will undoubtedly hamper their future development as languages of learning.

Against this background, the question arises whether it will be worthwhile to develop elaborate sets of scientific terms for African languages, given the very clear signals that a substantial number of scholars will probably

never, or only rarely, use them. This question can be approached in one of two possible ways — the one rather negative; the other more positive. Viewed negatively, the terminographer could adopt the view that since there are no guarantees that the existing prejudices towards the indigenous languages will disappear, terminographic work undertaken in advance would be premature. Is there any assurance that even the most comprehensive and scientifically sophisticated set of terms will, in the long run, change the minds of those who are currently in favour of English? Viewed more positively, it could be argued that the terminographer should not focus on the here and the now, but on the future. Attitudes may change for a variety of reasons, such as, for instance, the launching of vibrant language-awareness programmes as suggested in many of the subcommittee reports of the Language Plan Task Group discussed at a national conference held on 29 June 1996. It could therefore be argued that when a new linguistic awareness awakens, a properly developed terminology needs to be in place to serve a variety of language users at different levels.

The vision of African languages becoming accepted media of instruction at tertiary level is not so unrealistic, particularly if cognisance is taken of what happened elsewhere in the world. Discussing the advent of German as a literary language during the eighteenth century, Cluver (1987: 27) observes that a language's surge forward can be amazingly swift, provided that nationalism has, due to whatever contributing factors, reached that critical point which inspires a people to reassess its cultural heritage.

The question is, however, whether the present climate in South Africa is indeed conducive to the awakening of such a renewed linguistic nationalism. This does not seem to be the case, due to a disturbing conflict between the politically inspired vision of one homogeneous South African nation on the one hand, and the ideal of promoting multilingualism on the other. The active promotion of multilingualism, that is the promotion of the different languages as unique entities spoken by different racial and ethnic groups, directly opposes the ideal of establishing maximum cohesion and unity among the peoples of this country. The quest for a single undifferentiated "rainbow nation" requires, of necessity, the systematic eradication of everything that might distinguish one group from another. Linguistic identity, therefore, becomes a problem in as far as it enshrines racial and ethnic diversity. Despite what is written in the Constitution, the careful observer becomes aware of a number of subtly disguised initiatives which are all intended to obstruct whatever might inspire the awakening of a renewed linguistic awareness. The following can be cited as examples of endeavours aimed at maximising national unity:

(1)   At a workshop of the Language Plan Task Group's Subcommittee on the Development of South African Languages that was held at the University of South Africa on 28 March 1996, it became clear that this subcommittee has been assigned the task, *inter alia*, of advising the Ministry of

Arts, Culture, Science and Technology on the feasibility of the *harmoniza-tion* of the African languages. In its most extreme sense, harmonization amounts to the unification of languages which belong to the same lan-guage group into one so-called "standard language" of unpredictable composition. In practice, this scenario would entail the harmonization of Northern Sotho, Southern Sotho and Tswana into so-called "Standard Sotho", and the harmonization of Xhosa, Swati, Zulu and Ndebele into so-called "Standard Nguni". Fortunately, this proposal was met with severe opposition at the national conference of the Language Plan Task Group that was held on 29 June 1996, and the matter was postponed for the time being, pending further research. The question that remains unanswered, however, is why, in the first place, the Ministry regarded it necessary to put this point on the agenda? Given the agreements reached during constitutional negotiations regarding the autonomy of the coun-try's eleven languages, it becomes difficult — if not impossible — to reconcile the concept of *harmonization* with the undertaking given in the Constitution, namely:

> Recognising the historically diminished use and status of the indigenous languages of our people, the state must take practical and positive measures to elevate that status and advance the use of these languages.

(2)   A second controversial point discussed at some length at the said sub-committee workshop relates to the status of what has been labelled "Apartheid Language Boards". On the one hand, there were those who felt strongly that Language Boards are nothing but relics of the old apartheid system in which they were used as political tools to divide people who actually belonged together. Supporters of this view therefore proposed that these Boards be replaced by Language Academies of a *non-ethnic* nature by bundling together languages which belong to the same language group such as, for example, Northern Sotho, Southern Sotho and Tswana under one and the same Language Academy. Sup-porters of this view furthermore voiced the idea that in future the emphasis should be on the *similarity* between languages, and not on the *differences* which exist between them. On the other hand, there were those who were adamant that each language deserves its own academy, since it is constitutionally entitled to it — a view that fortunately enjoyed the support of the majority of delegates at the national conference of the Language Plan Task Group referred to above. Some supporters of the latter view did not hesitate to point out that the old Language Boards played a major role in the upliftment of South African languages to the extent that they are much further developed today than most of the lan-guages in the rest of Africa. But the mere fact that influential linguists,

language practitioners and government officials are divided on such an obvious point, namely whether languages should in practice be granted the privileges and autonomy which they are constitutionally entitled to, is symptomatic of a climate that will suffocate rather than stimulate the revival of a renewed linguistic awareness.

(3)   Thirdly, the severe opposition with which the principle of monolingual schools was met during recent constitutional negotiations can be re-called. This principle has not been enshrined in the Constitution, since it is viewed as being in direct conflict with the concept of nation-building. The ideal of one homogeneous nation cannot be attained if children are allowed to be educated in separate schools and in separate languages. Multilingual schools, on the contrary, are very useful and effective instruments in the equalising process, since they will, in the long run, gradually erode linguistic awareness which, in its turn, will ultimately lead to the demolition of linguistic boundaries between groups.

(4)   The fourth example of this equalising strategy pertains to the changing of names of radio services. The principle has been adopted that no radio station will in future use in its name the name of the language in which it communicates with its listeners. Names like Radio Tsonga, Radio Zulu, Afrikaans Stereo and the like are being replaced by more neutral labels, since the identification of a particular broadcasting service with a spe-cific language once again underscores the uniqueness of that language and the people *who speak it. Such a situation cannot be tolerated since it undermines the vision of one homogeneous and undifferentiated nation.*

The picture sketched above leads to a number of questions, for instance: How can the awakening of a renewed linguistic nationalism be envisaged if not only politicians, but also a number of very influential linguists, support efforts which are intended to destroy the identity and autonomy of the languages involved? How is the Language Plan Task Group's view that multilingualism is not a problem but an asset, to be reconciled with initiatives such as those men-tioned above, which are obviously aimed at suppressing the emergence of a renewed linguistic awareness? Given these developments it becomes very dif-ficult to accept that the African languages will ever reach a point where they become fully fledged languages of science and technology.

In the light of these observations, the terminographer will have to asses carefully the relevance of his or her present activities in terms of the possible future status of the African languages in this country. As one prominent ter-minographer puts it, the question that needs to be answered is analogous to one asked by a railway company intending to build a new railway line across a deserted country. Should it proceed with the building project in the hope that a new town will emerge at the end of the line after its completion, or should it rather wait for the town to develop first before it starts building the line?

Terminographic work can only be of value if language users, and particularly subject specialists, sincerely appreciate and fully utilise the terminographer's output. No matter how well-trained and resourceful terminographers are, and no matter how elegant and sophisticated the terms which they create might appear to be, terminographic work in the African languages stands in danger of becoming a wasteful endeavour if the language issue is politicised to the extent that linguistic diversity is subjected to the ideal of nation-building. Unless each and every language is in practice granted all the privileges it is constitutionally entitled to, and unless honest efforts are initiated that will create a climate which will stimulate the growth of a renewed language pride among speakers of all the different languages, it is doubtful whether the indigenous languages will ever develop much further than their present position.

References

Cluver, A. de V. 1987. A Sociolinguistic Approach to the Study of Technical Languages. *Logos* 7(2): 13-30.

Nkondo, C.P.N. 1987. Problems of Terminology in African Languages with Special Reference to Xitsonga. *Logos* 7(2): 69-78.

The African Dimension of the *Oxford English Dictionary*

Edmund Simon Christopher Weiner, *Oxford English Dictionary, England*

Abstract: African languages have contributed and continue to contribute a great many vocabulary items to English, both directly and via the intermediary of other non-African languages. The *OED Additions Series* contains 63 words of African origin, most of which have come into the language during the last years. The second edition of the *OED* contains about 275 words of African origin, drawn from 30 languages. The treatment of these words is somewhat uneven, owing to two factors: the unavailability to the editors of relevant information at the time of compilation of *OED*[1]; and supremacist attitudes, which caused entries for words of this kind to be shorter and less detailed, and affected their definition and description. In the third edition of the *OED*, words of all kinds should receive the same degree of attention, which implies that data collection from the reading of primary sources should include all varieties of World English; description and definition should be undertaken from a neutral standpoint; and etymological research and documentation must be as full as is practicable. All items of African origin will be sent to linguistic specialists (the bulk of this work has been done in the *DSAE* for items of South African origin). The checking of other etymologies will bring to light items whose African origin was not previously indicated (for example, borrowings from Brazilian Portuguese). In a sample from the letter M, there has been a gain in the numbers of African etymologies and the accuracy of treatment. Loanwords of all origins will be given as much attention as our resources allow, but there may be more ground to make up in the African sector of the lexicon.

Keywords: AFRICAN LANGUAGES, ENGLISH LANGUAGE, *OXFORD ENGLISH DICTIONARY*, *OED ADDITIONS SERIES*, LEXICOGRAPHICAL DOCUMENTATION, DEFINITION, ETYMOLOGY, LOANWORDS, SUPREMACIST ATTITUDE, THIRD EDITION OF THE *OED*, READING PROGRAMME, *DICTIONARY OF SOUTH AFRICAN ENGLISH ON HISTORICAL PRINCIPLES*

Opsomming: Die Afrikadimensie van die *Oxford English Dictionary*. Afrikatale het 'n groot aantal woordeskatitems tot Engels bygedra en doen dit steeds, sowel direk as via nie-Afrikatale. Die *OED Additions Series* bevat 63 woorde van Afrika-herkoms waarvan die meeste gedurende die afgelope paar jaar in Engels opgeneem is. Die tweede uitgawe van *OED* bevat ongeveer 275 woorde van Afrika-herkoms wat uit 30 tale ontleen is. Hierdie woorde word weens twee faktore enigsins inkonsekwent gehanteer: die onbeskikbaarheid van relevante inligting vir die redakteurs tydens die samestelling van *OED*[1]; en 'n meerderwaardige houding, wat veroorsaak het dat inskrywings vir hierdie tipe woorde korter en minder gedetailleerd is, en hul definisies en beskrywings beïnvloed het. In die derde uitgawe van *OED* behoort alle tipes woorde dieselfde mate van aandag te ontvang, wat impliseer dat dataversameling uit die lees van primêre bronne alle variëteite van Engels in die wêreld moet insluit; beskrywing en definisie moet vanuit

'n neutrale oogpunt onderneem word; en etimologiese navorsing en dokumentasie behoort so volledig te wees as wat prakties moontlik is. Alle items' van Afrika-herkoms sal aan taalkundige spesialiste gestuur word (die grootste deel van hierdie werk is in *DSAE* vir items van Suid-Afrikaanse oorsprong gedoen). Die nagaan van ander etimologieë sal items identifiseer waarvan die Afrikaherkoms nie voorheen aangedui is nie (byvoorbeeld ontlenings aan Brasiliaanse Portugees). In 'n steekproef van die letter M was daar 'n toename in die aantal Afrika-etimologieë en in die akkuraatheid van die hantering. Leenwoorde uit alle tale sal soveel aandag geniet as wat ons hulpbronne toelaat, maar daar mag 'n groter agterstand in die Afrika-deel van die woordeskat wees wat uitgewis moet word.

Sleutelwoorde: AFRIKATALE, ENGELS, *OXFORD ENGLISH DICTIONARY, OED ADDITIONS SERIES,* LEKSIKOGRAFIESE DOKUMENTASIE, DEFINISIE, ETIMOLOGIE, LEENWOORDE, MEERDERWAARDIGE HOUDING, DERDE UITGAWE VAN DIE OED, LEESPROGRAM, *DICTIONARY OF SOUTH AFRICAN ENGLISH ON HISTORICAL PRINCIPLES*

The languages of Africa have made an important contribution to the English language. Through creolization, grammatical features of African origin have entered various dialects of English. These may not be very influential on the various standard varieties of the present day, but for all we know they may one day assume a much greater importance. More importantly for the purpose of this paper, African languages have contributed a great many vocabulary items to the English lexis, both directly, and via the intermediary, of other non-African languages such as Portuguese, Spanish, Afrikaans, and so on.

1.   African sources in *OED*[1]

The first edition of the *OED* excluded the word *African*. This is what the editor, James Murray, wrote in his original Preface to Volume 1 (a Preface that was not reprinted in the completed *OED*[1]):

> The word *African* was one of the earliest instances in which the question of admission or exclusion arose with regard to an important adjective derived from a geographical proper name. After much careful consideration, and consultation with advisers, it was ·decided (perhaps by a too rigid application of first principles) to omit the word, as having really no more claims to inclusion than *Algerian, Austrian,* or *Bulgarian*. But, when *American* was reached, some months afterwards, it was seen that *Americanize* and *Americanism* must of necessity be included, and that these (with "the *Americanizing* of our institutions") could not be explained without treating *American,* and explaining its restricted application to the United States. *American* was accordingly admitted. Then the question arose, whether the exclusion of *African* was consistent with the inclusion of *American*: but the question came too late; *African* had been actually omit-

ted, on its own merits. And the inconsistency is only on the surface; *American* is included, not on its own account, but to help to the better explanation of derived words. (Murray 1884: ix.)

We would not now accept this principle, and of course the second edition of the *OED* (Simpson and Weiner 1989) contains *Algerian, Austrian,* and *Bulgarian,* as well as *African,* which were all added in the *Supplement* (Burchfield 1972). This rather astonishing omission is in a way symbolic of a tendency to underestimate the role of Africa in the English language at that time, of which I shall presently say more.

The contribution of the African countries to the English language continues to be made. In the *OED Additions Series* Oxford University Press publishes work in progress on words and senses since the completion of the Second Edition in 1989. So far, Volumes 1 and 2 (Simpson and Weiner 1993) have appeared. Volume 3 is in production, and further material is in preparation. In this body of edited material (which contain in all about 12,000 vocabulary items) there are 63 words of African origin. Most of them have entered English during the last 25 years. They are drawn from at least 30 languages. Not many of these words are part of the central vocabulary of World English. Some are the names of ethnic groups and languages and many are terms as yet restricted to a particular area of Africa. But as the use of English in some of these areas increases, we can expect such loanwords to assume a greater importance.

It is estimated that OED^2 contains 275 words of African origin, drawn from 80 languages. The treatment of these words in the dictionary is somewhat uneven, owing to two main factors that influenced the compilation of the first edition (OED^1) both of which are familiar in lexicography; namely, the unavailability of information and a supremacist attitude.

1.1   Unavailability of relevant information

Relevant information may have been unavailable to the editors of OED^1 because they did not have access to scholars and experts and their books. Recalling that the first edition was compiled between 1879 and 1928, we should not be surprised if (especially while the first half of the alphabet was being compiled, up till about 1905) the editors were unable to obtain the information they needed and ought to have deployed in their entries.

This had a particular effect on three parts of the entry; the documentation, the definition, and the etymology.

1.1.1   Effect on the documentation

Few quotations were available, giving the impression that the word was of infrequent occurrence and marginal to the language. For example, the word

impi (a Zulu regiment) had, in the first edition, only three quotations: 1879 *Daily Telegraph*, 1885 *Harper's Magazine*, and 1888 Rider Haggard *Maiwa's Revenge*, none of them African publications. This entry was published in 1900. The same entry has now been treated in appropriate detail in the new *Dictionary of South African English on Historical Principles* (Silva 1996). If we look at the *DSAE* to see what the potential documentation at that date was, we find five examples up to 1900, all different from those in *OED*[1], and all taken from people writing in Africa itself, not in Britain or America; moreover three of them are markedly older than *OED*'s examples, dating from 1836, 1846, and 1862.

Close by in *OED*[1] is the plant name **imphee** (now usually spelt **imfe**). The first edition had only two examples, one of 1880 from an Australian source, and one of 1893 from a book on Indian produce. By contrast, *DSAE* has no fewer than ten pre-1900 examples, the earliest dating from 1828.

1.1.2   Effect on the definition

Largely because of the scanty documentation, the definition was often very brief, sometimes combining into one section a range of meanings which, in an indigenous English word, would have merited a series of distinct sense sections. For example, in *OED*[1] the word **induna** is defined simply as "An officer under the king or chief of the Zulus, Matabele, and other South African tribes". The earliest example is *ante* 1875. When we turn to the *DSAE*, we find that there are three senses which date from before 1900: "1.a. A headman, councillor, or officer under a chief" (1835 onwards), "b. *transf.* A foreman, head servant, mine overseer, or policeman" (1857 onwards), and "2. *fig.* One in authority" (1897).

1.1.3   Effect on the etymology

In the absence of accurate linguistic information, etymologies often fall well below the normal standards in both completeness and accuracy. For example, the word **cola** (the West African nut-bearing tree) has the etymology "*Kola, Kolla, Goora*, in Negro languages of W. Africa". This is evidently unacceptable. The languages are not named; the region is far too large for the origin to be localized; and one has little confidence in the forms and their phonetic accuracy. "Negro languages" is also an objectionable expression, an example of the supremacism discussed below.

Another example is the word **macute** (a form of currency). In *OED*[1] nothing more was said than "origin unknown". Some addenda and corrigenda, published at the end of the completed *OED*[1], included a note on the origin of this word from "the Reverend W. Holman Bently, writing from the Congo Free State", which, though more informative, is still lacking in precision and reliability. This note was inserted into the etymology in *OED*[2], following the inten-

tion of the compilers of the addenda, but in the draft OED^3 text it has been moved to the quotation section and replaced by information taken from a much more recent scholarly study of Portuguese Africanisms.

1.2   A supremacist attitude

Supremacism is undesirable in a dictionary; but, to be fair to OED^1's editors, such an attitude must have been largely unconscious, and shared with the vast majority of their British contemporaries. Moreover, I would see it rather as the (slightly) more excusable implicit assumption of the cultural superiority of inherited western European traditions, than as a crude or doctrinaire racism. However, it undoubtedly contributed to the inadequacies that we have noted above. For example, the shortage of documentation may have been partly due to a reading programme which devoted less effort (proportionately) to sources concerned with Africa than to those concerned with Britain, Europe, or North America. The brevity of definitions and absence of finer sense discriminations may have been due to a feeling that the vocabulary of the outposts of the Empire was not as important as that of the well-established English-speaking areas. And the skimpiness of the etymologies could have been due to an unconscious prejudice in favour of Indo-European and other languages with a long scholarly tradition.

Such attitudes may be detected in an entry like that for **Eboe** (the ethnonym, now usually spelt **Ibo**) which reads as follows:

> A name applied in the W. Indies and U.S. to the Negroes from Benin. Hence *attrib.*, **eboe-tree** (*Dipteryx eboensis*), a tree of Central America, yielding **eboe oil**, a name also given to the oil of a different tree; see quot.

Normal entry structure is suspended in the presentation of this information. The cursoriness of the wording seems to convey a distinctly dismissive attitude towards a whole nexus of interesting questions relating to the cultures of both the West Indies and West Africa. (The 1972 volume of the OED *Supplement*, incidentally, merely added "The same term as IBO in West Africa, but used with wider application in the W. Indies".)

Another example is the entry for **cam-wood**:

> [According to some, ad. native African name *kambi*.] The hard red wood of *Baphia nitida* (N.O. Leguminosae), imported from West Africa.

The etymology is hardly a model of exact scholarship. Notice also the implicit Britocentrism of the last four words.

2.   African words in *OED*[3]

The third edition of the *OED*, as is now well known, is in progress and due for completion in 2005. Whereas the second edition represented the integration of *OED*[1] and its four-volume *Supplement*, for the third edition all aspects of the dictionary are being revised. The new words which are at present being published separately will all be included in this edition. In working on the third edition, we should, in theory, bestow the same degree of attention on words of all kinds, whatever variety of English they may belong to or have arisen in. This has implications for data collecting, defining, and etymological work.

2.1   Implications for data collecting

The collection of data, from the reading of primary sources, potentially includes all varieties of World English. We have a British and a North American Reading Programme, which take in not only British and American English, but also Caribbean, South Asian, West and East African, and East Asian varieties. But in fact the analysis of lexical data from a variety of English outside Britain is better done by local scholars than by British lexicographers. Once a dictionary covering one such variety is complete, it is relatively easy for us to scale down its very detailed treatment to the dimensions of the *OED*, ensuring that the overall perspective remains intact. Hence, the *Dictionary of South African English on Historical Principles* is not only a landmark in the study of World Englishes but will also make an enormous difference to *OED*[3]'s effectiveness in covering English of African origin.

2.2   Implications for defining

OED[3]'s description and definition of vocabulary will be undertaken from a more neutral standpoint. It is, of course, impossible to write a dictionary without making a number of cultural assumptions. However, it is possible to try to accord every variety equal treatment, within the limitations of the materials available, and it ought to be possible to avoid the grosser kinds of cultural superiority complex.

2.3   Implications for etymological work

Etymological research and documentation are to be as full as is practicable. We are sending the majority of etymologies that have any significant non-English input to be checked by people who are proficient in the relevant languages. In the case of very frequently cited languages, we obtain guidance on overall pol-

icy and style, and on exceptionally difficult words, from experts, but have the bulk of the forms checked by graduates who are proficient in those languages. We shall therefore send most of the items of African origin to linguistic specialists for correction and amplification. (In this regard, again, it is very helpful to have the *DSAE* available, since most of the work on items of South African origin (at least 12%) has already been done by those editors and checked by local experts.) Furthermore, we confidently expect that this checking will bring to light items not recognized hitherto as of African origin. There are a number of borrowings from Brazilian Portuguese which, when traced through Portuguese etymological sources, turn out to be of African origin.

For example, the word **macumba** (the Brazilian religious system), whose etymology in *OED*[2] reads simply "Portuguese", is almost certainly derived from a language of the Zaïre region. Recent scholarship makes it very likely that the word **mahogany** is from the Yoruba *m'oganwo*, which in Nigeria is applied to a different but not dissimilar tree.

A further stage of checking is in store for etymologies which have been extensively rewritten in-house and newly provided with foreign-language etyma. These are all flagged in the database and will be sent to experts after the main phase of revision. This process will probably bring yet more African borrowings to light.

3.   Results so far

The revision of the *OED* began with the letter M, because it was from this point that our extensive quotation files had been systematically sorted into the categories needed for efficient revision. Revision begins with the documentation, definitions, and sense-structure of the scientific and general vocabulary. Etymological work on a given range of words is done after the work of the science revisers and general revisers has been keyboarded, so that all changes thrown up by new forms, new dating, and new ordering can be taken into account. Already in the part of the letter M that has been etymologically revised we have seen: (a) a gain in the number of African etymologies, and (b) an improvement in the accuracy of information given for such words. Examples of (a) are **machila** (hammock-like conveyance), formerly ascribed (via Portuguese) to Tamil, but now believed to originate in a Moçambican language; **macumba**, mentioned above as probably Zaïrean; and **many-root**, a West Indian plant name previously considered to be a purely English compound, but now believed to come from a Twi word influenced by folk-etymology. An example of (b) is **marimba** (musical instrument) previously said to be from "Congo", but now ascribed to a specific Kimbundu word.

There are also other words which look as if an African origin, not previously mentioned, is likely; for example **malambo bark**, which entered English from Spanish or Portuguese, looks quite likely to turn out to have an African etymon. It is not, of course, that we are giving preferential treatment to

words of African or any other origin outside the historic English-speaking areas. Loanwords and local words of all origins will be given as much attention as our resources allow. But it may be the case that there are more deficiencies to be made good, and more cultural slants to be adjusted, in this sector of the lexicon than in many others.

References

Burchfield, R.W. (Ed.). 1972-1986. *A Supplement to the Oxford English Dictionary.* Four volumes. Oxford: Oxford University Press.

DSAE = Silva, P.M. (Ed.). 1996.

Murray, J.A.H. (Ed.). 1884. *A New English Dictionary on Historical Principles.* Volume 1. Oxford: Clarendon Press.

Murray, J.A.H. et al. (Eds.). 1933. *The Oxford English Dictionary.* Oxford: Clarendon Press.

OED[1] = Murray, J.A.H. et al. (Eds.). 1933.

OED[2] = Simpson, J.A. and E.S.C. Weiner (Eds.). 1989.

Silva, P.M. (Ed.). 1996. *A Dictionary of South African English on Historical Principles.* Oxford: Oxford University Press.

Simpson, J.A. and E.S.C. Weiner (Eds.). 1989. *The Oxford English Dictionary.* Second edition. Oxford: Oxford University Press.

Simpson, J.A. and E.S.C. Weiner (Eds.). 1993. *OED Additions Series.* Volumes 1 and 2. Oxford: Oxford University Press.

Probleme der Textauswahl für einen elektronischen Thesaurus: Tagungsbericht

Michael Schlaefer, *Deutsches Wörterbuch,*
Akademie der Wissenschaften zu Göttingen, Deutschland

Abstract: Problems Concerning the Choice of Text Materials for an Electronic Thesaurus: Conference Report. At the invitation of the Akademie der Wissenschaften in Göttingen a colloquy was held on 1 and 2 November 1996. Experts in lexicographical studies and single word research participated in a discussion of problems concerning the choice of text materials for an electronic thesaurus.

Keywords: DEUTSCHES WÖRTERBUCH, CORPORA, TEXT SELECTION, GERMAN LANGUAGE

Zusammenfassung: Auf Einladung der Göttinger Akademie der Wissenschaften wurde am 1. und 2. November 1996 ein Kolloquium abgehalten. Lexikographen und Wortforscher diskutierten Probleme der Textauswahl für einen elektronischen Thesaurus.

Stichwörter: DEUTSCHES WÖRTERBUCH, CORPORA, TEXTSELEKTION, DEUTSCHE SPRACHE

Das im Jahr 1838 von den Brüdern Jacob und Wilhelm Grimm begonnene *Deutsche Wörterbuch* wird seit 1960 in der Trägerschaft der Akademien in Göttingen und Berlin im Buchstabenbereich A - F neu bearbeitet.

Dem Göttinger Neubearbeitungsteil des *Grimmschen Wörterbuchs* liegt eine Quellenauswahl zugrunde, die etwa 6000 Texte des 8. - 20. Jahrhunderts umfaßt. Dieses Quellenmaterial ist etwa zwischen 1960 und 1970 über Mikroverfilmung erfaßt und in ein Zettelarchiv für die Buchstaben D - F umgesetzt worden. Dieses Zettelarchiv umfaßt einen Bestand von etwa 2,5 Millionen Belegen. Durch technische Veränderungen bedingt, ist seit etwa 1980 eine finanziell vertretbare künftige Weiternutzung der Quellenfilme fraglich geworden. Aus diesem Sachverhalt ergeben sich weitreichende Konsequenzen für Überlegungen zur Fortführung der Neubearbeitung des *Grimmschen Wörterbuchs* über F hinaus.

Die Göttinger Kommission für das *Deutsche Wörterbuch* hat daher seit Mitte der achziger Jahre nach Wegen gesucht, den Zugang zum Göttinger Quellenmaterial zu erhalten, um für die historische Lexikographie und Wortforschung auch über den engeren Planungsbereich des laufenden Neubear-

beitungsvorhabens hinaus eine wissenschaftliche Arbeit in den Standards des *Grimmschen Wörterbuchs* zu ermöglichen. Ab dem Jahr 1986 beauftragte die Kommission die Göttinger Arbeitsstelle des Grimmschen Wörterbuchs, Quellen des Göttinger Neubearbeitungsteils in eine maschinenlesbare Form umzusetzen und daraus Belegmaterial für den historischen deutschen Wortschatz zu gewinnen. Das Arbeitsvorhaben erhielt die Bezeichnung "Sicherungsmaßnahme". Für eine erste Phase, die vor allem der Entwicklung und Erprobung der theoretischen Grundlagen, der praktischen Vorgehensweise und der Auswertbarkeit gewonnener Materialien dienen sollte, wurde das in der Sicherungsmaßnahme zu erfassende Textvolumen auf einen Bestand von 50 Texten begrenzt. Als erreichbares Belegvolumen wurde etwa eine Million elektronisch verfügbarer Stichwortnachweise angenommen.

Dieses erste Arbeitsphase der Sicherungsmaßnahme kam im Jahr 1994 weitgehend zum Abschluß, so daß es wünschenswert erschien, die gewonnenen Erfahrungen und Ergebnisse im Rahmen eines Arbeitsgesprächs mit Fachleuten für deutschsprachige Textcorpora zu diskutieren und Perspektiven für die Weiterarbeit zu entwickeln. Am 1. und 2. November 1995 fand daher auf Einladung der Kommission für das Deutsche Wörterbuch (Göttingen) in der Trägerschaft der Göttinger Akademie der Wissenschaften ein Kolloquium zu "Problemen der Textauswahl für einen elektronischen Thesaurus" statt.

Dieses Kolloquium wurde zugleich als Einleitung zu einer Reihe von wissenschaftlichen Veranstaltungen verstanden, die sich mit den Problemen der historischen deutschen Wortforschung als einem Teil der Rahmenbedingungen für die "Sicherungsmaßnahme" befassen sollten.

Außer den im weiteren genannten Referenten nahmen der Präsident und der Sekretär der Göttinger Akademie, Mitglieder der Göttinger und der Berliner Wörterbuchkommissionen, das Kollegium der Göttinger Arbeitsstelle sowie einzelne Gäste an der Veranstaltung teil.

Das Programm des Arbeitsgesprächs für den 1. November 1996 umfaßte drei Themenbereiche. Unter dem Schwerpunkt "Zielsetzungen und Vorarbeiten" war aus dem Projekt der "Sicherungsmaßnahme" zu berichten. Daran schlossen sich die Themenblöcke "Corpuserfahrungen in sprachhistorischen und gegenwartssprachlichen Projekten" und "Sprachgeographische Corpusaspekte" an.

Im ersten Themenschwerpunkt stellte der Vorsitzende der Göttinger Kommission für das Deutsche Wörterbuch, R. Bergmann (Bamberg), Geschichte und Arbeitsstand des Göttinger Digitalisierungsvorhabens dar. Er erläuterte die unternehmensstrategischen Grundzüge und ging in diesem Zusammenhang auf die in den letzten Jahren stark veränderten Rahmenbedingungen für die Wortforschung ein. Eine zukunftsorientierte Gestaltung der Grundlagen für historische deutsche Wortforschung könne nicht mehr in isolierten Einzelvorhaben, sondern nur in Kooperation vieler Wissenschaftler und Institutionen des gesamten deutschsprachigen Raumes erreicht werden. Der Leiter der Arbeitsstelle, M. Schlaefer (Göttingen), referierte über den aktuellen Stand der

Arbeiten an der "Sicherungsmaßnahme". Er zeigte Nutzungsmöglichkeiten der verfügbaren Materialien auf und stellte schwerpunktmäßig Grundlagen der Textauswahl bzw. der Lemmatisierung und der Belegauswahl für die Sicherungsmaßnahme vor.

Die Diskussion zu den Einführungsvorträgen ließ erkennen, daß für einen Teil der Anwesenden die "Sicherungsmaßnahme" einen Schritt in Richtung auf einen sprachnationalen Thesaurus darstellte. Die unter einer solchen Annahme zu behandelnden wissenschafts- und fachpolitischen Perspektiven traten daher gegenüber den konzeptionellen arbeitspraktischen Fragen des Göttinger Projekts in den Vordergrund. Angesichts der knapp bemessenen Diskussionszeit wurde vereinbart, am Ende des Kolloquiums das Thema "Gesamtthesaurus" im Rahmen einer Grundsatzdiskussion zu behandeln.

Im Themenschwerpunkt "Corpuserfahrungen" gingen zunächst die Referenten K.P. Wegera (Bochum) und H.-J. Solms (Halle) unter verschiedenen Gesichtspunkten auf Erfahrungen mit dem Bonner Corpus frühneuhochdeutscher Quellen ein, das seinerzeit als eines der ersten maschinenlesbaren Corpora für sprachhistorische Arbeiten erstellt worden war. Dabei standen für Herrn Wegera Fragen der Textauswahl und der Textumfänge im Vordergrund, für Herrn Solms Fragen des Zusammenhangs von sich wandelnden Erkenntnisinteressen und Corpusstrukturen sowie Aspekte der polyfunktionalen Nutzbarkeit elektronischer Corpora.

Mit unterschiedlichen Problemen der Strukturierung historischer Quellencorpora befaßten sich auch die Referate von U. Goetz (Bamberg), H. Kämper-Jensen (Mannheim) und U. Haß-Zumkehr (Mannheim). U. Goetz stellte am Beispiel des seit 1990 laufenden und kurz vor dem Abschluß stehenden Bamberg-Rostocker Gemeinschaftsprojekt "Die Entwicklung der Großschreibung im Deutschen von 1500 bis 1700" Möglichkeiten für eine straffe raum-zeitliche und quantitative Corpusgliederung dar. H. Kämper-Jensen referierte über das "IdS-Fremdwörtercorpus", das aufgrund der langen Entstehungszeit sowie der unterschiedlichen Schwerpunktsetzung in seiner Struktur heterogene Züge aufweist. Den Gesichtspunkt der Polyfunktionalität betonte U. Haß-Zumkehr in ihrem Bericht über das im Aufbau befindliche sogenannte Historische Corpus des Instituts für deutsche Sprache in Mannheim. Der Beitrag von M. Wermke (Mannheim) über Vorüberlegungen zum Aufbau elektronischer Textcorpora in der Dudenredaktion machte in sehr anschaulicher Weise den Unterschied zwischen Textcorpusorientierung und Belegcorpusorientierung deutlich. In diesem Zusammenhang hob er die Bedeutung des Modells einer lemmatisierten exemplarischen Belegsammlung gegenüber einem unbegrenzten Wortformenreservoir für die Wörterbucharbeit hervor.

Im zweiten Themenschwerpunkt gingen W. Bauer (Wien), R. Ris (Zürich) und P. Ott (Zürich) auf unterschiedliche Zusammenhänge der sprachgeographischen Schichtungen des Deutschen mit vorhandenen oder künftigen Corpusbildungen ein. Zuerst berichtete W. Bauer über "Historische Quellen des Wörterbuches der bairischen Mundarten in Österreich zu den bairisch-öster-

reichischen Sprachvarietäten des 14. bis 19. Jahrhunderts". Über "Möglichkeiten eines schweizerischen Corpusteils" sprachen R. Ris und P. Ott. R. Ris erörterte die zentrale Frage nach der Tragfähigkeit der dem [1]DWB zugrundeliegenden Quellenbasis für eine geschichtlich zuverlässige Abbildung des Schweizerischen. Er wies eine starke Einseitigkeit der DWB-Quellenbasis im Bereich protestantischer Autoren aus Zürich nach und entwickelte eine alternative Corpusstruktur für die jüngere schweizerische Varietät des Deutschen. P. Ott stellte anhand detaillierter sprachgeschichtlicher, sprachsoziologischer und sprachgeographische Merkmale einzelner Quellen und Quellengruppen ein exemplarisches Corpus für das schweizerische Deutsch des 16. Jahrhunderts vor.

Für den zweiten Tag des Kolloquiums waren Beiträge zu exemplarischen fachsprachlichen Aspekten eines historischen Corpus sowie zur Frage von Wörterbüchern als Quellen vorgesehen. K. Jacob (Dresden) befaßte sich in diesem Zusammenhang mit "Techniksprachlichen Quellen des 17.-19. Jahrhunderts". Er machte deutlich, daß gerade unter lexikalischen Gesichtspunkten eine Ausblendung der jüngeren Techniksprache, wie sie auch noch das *Grimmsche Wörterbuch* zeigt, kaum vertretbar ist. G. Wagenitz (Göttingen) stellte in seinem Vortrag "Wortbildung und Textarten in der Biologie" Übergänge von fachsprachlich-botanischen Bildungen in die Gemeinsprache dar und zeigte eine Reihe von Textsorten auf, die als Vermittler zwischen den beiden sprachsoziologischen Bereichen wirkten.

Am Beispiel der "Wörterbücher des 16. Jahrhunderts als Thesaurusquellen" erläuterte P.O. Müller (Erlangen) die besondere Problematik, die sich in diesem metasprachlichen Quellensektor durch immanente philologische Bedingungen und Traditionen ergeben.

H. Henne (Braunschweig) ging unter dem Thema "Ein Deutscher Thesaurus und die deutschen Wörterbücher des 17. und 18. Jahrhunderts" über den engeren thematischen Rahmen ausgreifend auf Möglichkeiten und Erwartungen ein, die er mit elektronischen Textcorpora und Wörterbüchern verbunden sah.

Für die Weiterarbeit in der Sicherungsmaßnahme haben die Beiträge eine Reihe wichtiger Anregungen gegeben. Auf konzeptioneller Ebene wird die Abwägung zwischen Text- und Belegarchiv ebenso kritisch vorzunehmen sein, wie die ausschließliche Stützung auf das Göttinger Quellencorpus zu hinterfragen ist. Eine noch größere Aufmerksamkeit dürfte künftig der Auswahl und der Umfangsbegrenzung des digitalisierten Quellenmaterials zukommen. Für den arbeitspraktischen Bereich stellt sich die Aufgabe der Bündelung der Ressourcen. Angesichts der Vielfältigkeit der Corpusinitiativen liegen hier große Aufgaben für die Koordination, aber auch Chancen dafür, kooperativ eine tragfähige Plattform für die historische Wortforschung zu schaffen. Angesichts der Bedeutung, die den Referaten für die Sicherungsmaßnahme, aber auch für die Erstellung historischer Corpora allgemein zukommt, strebt der Vorsitzende der Göttinger Kommission für das Deutsche Wörterbuch die Veröffentlichung

264   Michael Schlaefer

eines Sammelbandes für 1997/98 an.

Die zum Abschluß des Kolloquiums angesetzte Grundsatzdiskussion über einen Gesamtthesaurus für die deutsche Sprache führte gedanklich weit über den Rahmen hinaus, der mit der Göttinger "Sicherungsmaßnahme" und dem Thema des Kolloquiums gegeben war. In der Diskussion und einzelnen Statements wurde die Frage nach der Aktualität und Dringlichkeit eines Thesaurusaufbaus ebenso angesprochen, wie die Stellung eines künftigen Gesamtthesaurus gegenüber den in Bearbeitung befindlichen Wörterbüchern. Dabei spielte das Problem der Bearbeitungszeit gerade des *Grimmschen Wörterbuchs* eine besondere Rolle. Ein weiterer Diskussionsschwerpunkt betraf Überlegungen zum Zusammenwirken der deutschsprachigen Länder und deren Forschungseinrichtungen. Schließlich wurden mehrfach Entwürfe über verschiedene Thesauruskonzeptionen mit Text- bzw. Belegarchiven, aber auch mit verschiedenen Wörterbuchprojekten gedanklich angerissen. Der Wunsch, den Plan für einen solchen Thesaurus der deutschen Sprache in weiteren Kolloquien zu diskutieren, wurde allgemein unterstützt. Als eine fokusartige Bündelung der mit der Grundsatzdiskussion eröffneten fachpolitischen und konzeptionellen Dimensionen läßt sich die mehrfach benutzte Formulierung betrachten, es gehe um die deutsche Lexikographie des 21. Jahrhunderts.

A Dictionary of South African English on Historical Principles:[*]
A Case of Lexical Invasion or Corpus Enhancement?

Liesel Hibbert, Department of English,
University of the Western Cape, Bellville, South Africa

Abstract: The *DSAE* (*A Dictionary of South African English on Historical Principles*) embraces linguistic diversity by including many Afrikaans lexemes and lexemes from African languages. In the advertising pamphlet the dictionary proposes to, among other things, improve communication, give access to education, change perceptions of SAfE (South African English) locally and internationally, improve historical and political perspectives and create a new South African identity. These statements are discussed in relation to popular local debates around "standards", language variation and policy. An overview is given of the current status of SAfE in the context of Southern Africa and Africa. Finally I argue in favour of the dictionary as documentary evidence of a living spoken language at a given point in history.

Keywords: DICTIONARY, LEXICOGRAPHY, SOUTH AFRICAN ENGLISH, VARIETIES OF ENGLISH, INTERCULTURAL COMMUNICATION, MULTILINGUALISM, SOUTH AFRICAN IDENTITY, LANGUAGE PLANNING, DISCOURSE COMMUNITIES, ORGANIZATION OF AFRICAN UNITY, COLLOQUIALISMS, REGIONALISMS, LANGUAGE SHIFT, LANGUAGE CHANGE, DIALECT, LOAN-WORDS, BORROWINGS, AFRICANISMS, AFRIKAANSISMS, BLACK SOUTH AFRICAN ENGLISH, LINGUISTIC CHARTER FOR AFRICA, STANDARDIZATION, CORPUS ENHANCEMENT, LEXICAL INVASION, SOCIOLINGUISTICS

Opsomming: *A Dictionary of South African English*: **'n geval van leksikale indringing of korpusuitbreiding?** Die *DSAE* (*A Dictionary of South African English on Historical Principles*) omvat taalverskeidenheid deur baie Afrikaanse lekseme en lekseme uit Afrikatale in te sluit. Die woordeboek onderneem in die advertensiepamflet om onder andere kommunikasie te verbeter, toegang tot onderrig te verleen, persepsies oor SAfE (Suid-Afrikaanse Engels) binnelands en internasionaal te verander, historiese en politieke perspektiewe te verbeter en 'n nuwe Suid-Afrikaanse identiteit te skep. Hierdie stellings word met verwysing na populêre plaaslike debatte oor "standaarde", taalvariasie en beleid bespreek. 'n Oorsig word gegee van die huidige status van SAfE binne die konteks van Suider-Afrika en Afrika. Ten slotte lewer ek 'n betoog ten gunste van die woordeboek as dokumentêre bewys van 'n lewende gesproke taal op 'n bepaalde oomblik in die geskiedenis.

Sleutelwoorde: WOORDEBOEK, LEKSIKOGRAFIE, SUID-AFRIKAANSE ENGELS, VARIËTEITE VAN ENGELS, INTERKULTURELE KOMMUNIKASIE, VEELTALIGHEID, SUID-AFRIKAANSE IDENTITEIT, TAALBEPLANNING, SPRAAKGEMEENSKAPPE, ORGANISASIE VIR AFRIKA-EENHEID, SPREEKTAALUITDRUKKINGS, STREEKTAALUITDRUKKINGS, TAALVERSKUIWING, TAALVERANDERING, DIALEK, LEENWOORDE, ONTLENINGS, AFRIKANISMES, AFRIKAANSISMES, SWART SUID-AFRIKAANSE ENGELS, TAALMANIFES VIR AFRIKA, STANDAARDISERING, KORPUSUITBREIDING, LEKSIKALE INDRINGING, SOSIOLINGUISTIEK

Introduction

The *DSAE* (*Dictionary of South African English on Historical Principles*) is an open invitation to all South Africans to come out of their linguistic laagers. The dictionary cuts through common discourse systems and presents us with what has so often been termed "levelling the playing fields". Voluntary and involuntary discourse systems are laid bare, stripped of stigma or elevated status, and are listed alongside each other.

But is this desirable? Do we regard our dialects as acceptable? Is local really *lekker*? Not everyone would agree. The *DSAE* which appeared in September 1996 challenges many notions of what may be regarded as standard English in South Africa today. Surveying the responses that have already appeared, we find that some are outraged and others charmed and delighted by the appearance of the dictionary. A selection of titles of the reviews reveal a wide range of attitudes, e.g. "Lapping up our lekker lexicon" (Bundy 1996: 1-2) "Dictionary has cultural bias" (Landau 1996: 16) and "Dictionary deserves a good word" (Hughes 1996: 23). What most readers seem to react to in their different ways is the vast extent to which loan-words, borrowings, dialectical lexemes, regionalisms and colloquialisms are recorded and painstakingly glossed. Issues which the dictionary raises for linguists are questions regarding lexical invasion of core vocabularies vs corpus extension and enhancement, condoning diversity vs maintaining "standards" and the maintenance vs blurring of distinctions between first- and second-language Englishes, a distinction which has been heavily entrenched in the South African education systems.

A very general question which would suitably frame this discussion would be: Does the *DSAE* do what it is designed to? In the glossy four-page advertising pamphlet distributed by Oxford University Press, it is stated that the dictionary "extends far beyond the field of South African language studies" and that it contributes towards improved communication, education, historical understanding, political perspective, the development of the arts, a South African identity, changing the local perception of South African English and international lexicography. Each of these is qualified in a brief paragraph. Most of my discussion will be centered around those assertions.

The advertisement claims that: "With English becoming an increasingly important lingua franca in South Africa it is essential that people have tools like

the *DSAE* to help them use and understand the language." It is also stated that the dictionary is "certain to be found and frequently consulted in educational and public libraries". The assertion that the dictionary will be widely consulted by secondary and tertiary students and by the public is based on the assumption of consensus about the centrality of English among the population at large. There is also the assumption that everyone wants to overcome communication barriers and embrace linguistic diversity and to see the *DSAE* as the answer to a common national goal. In response to this one may ask whether, as more and more African language speakers move into high status educational-, political- and business-related contexts, the lingua franca in these contexts will remain English. Numerous linguists and sociologists have developed strong arguments to the contrary, i.e. Heugh (1996: 17) and Prah (1995: 15). All these writers dispute the notion that English is the "natural" African or Pan-African lingua franca. They also reject a number of current assumptions about intercultural communication, i.e. that English is dominant internationally, that English is the language of international trade and future African trade and that intercultural competence is synonymous with fluency in English. But if we accept for the sake of argument that English has or will have the dominant status claimed for it, the appearance of the *DSAE* has foregrounded the question of which variety of English is to perform this national and international function. To answer this question, a closer look at the current status of different Englishes in Southern Africa is needed.

The Current Status of Englishes in Southern Africa — An Overview

Southern African Englishes are currently categorized according to the sociopolitical order constructed under apartheid rule. In order to recognize diversity in the reconstruction of society, the labels and hierarchies put in place to elevate some dialects and varieties above others, I believe, need to be re-evaluated. Over and above that, colonialism has entrenched the notion of English as the chief source of knowledge which was already a commonsensical notion at the turn of the century, almost 100 years ago:

> Probably everyone would agree that an Englishman would be right in considering his way of looking at the world and at life better than that of the Maori or Hottentot, and no-one will object in the abstract to England doing her best to impose her better and higher view on these savages ... Can there be any doubt that the white man must, and will, impose his superior civilization on the coloured races? (Earl Grey 1899, quoted in Hodson 1902: 158 and cited in Skutnabb-Kangas and Phillipson 1994: 336-337)

Has anything changed since then? Where do South African linguists in the 1990s stand in relation to this view? Titlestad (1996: 169) maintains:

> A notably different SAfE (South African English) would merely mean
> that two varieties of English would have to be taught and learned. The
> damage, once done, would be difficult to undo, and South Africa would
> have lost one of its vital resources, its chief source of knowledge.

Titlestad seems to subscribe to the notion of two clear varieties of English, the
first- vs second-language variety. What is also clear is that he is in favour of the
L1 variety for educational purposes. In contrast to this, the *DSAE* unquestion-
ably blurs the distinction between these two varieties.

According to Lanham and MacDonald (1979) there are four common cur-
rent labels used to refer to SAfE (South African English). All of them are prob-
lematic in today's situation.

Significant sociolinguistic shifts since 1992 have been brought about by the
removal of the Group Areas Act, the removal of the homeland policy, constitu-
tionally enshrined multilingualism (11 official languages), mobility of the pre-
viously disenfranchised sector of the population into Model C schools (previ-
ously "White schools"), increased access to tertiary education and the lifting of
the international cultural and economic boycotts bringing about, in addition to
links with other continents, new links with the rest of sub-Saharan Africa and
Africa as a whole.

The first of these labels is "conservative SAfE". The implication of this label
is that it links accent and political affiliation thereby oversimplifying sociolin-
guistic realities, resulting in inaccuracies in grouping of speakers. Secondly,
"acceptable" SAfE implies that the accent is too broad to be regarded as
"standard English". Thirdly, "extreme" SAfE implies that it is not "proper"
English at all. Fourthly, Black South African English (BSAfE) as it is referred to
by Gough (1996) and Buthelezi (1995) could imply that all black South Africans
(which would constitute the vast majority of the population) all speak the same
English.

The sociolinguistic shifts and discourses on language planning in Africa in
the current literature on Pan-African language policy, suggest two necessary
shifts in point of view on English in South Africa. Firstly, a shift from refer-
ences to ethnically defined speech communities to Pan-Southern African
speech communities with many overlaps, taking into account social shifts. Sec-
ondly, a shift from references to colour as the only variable to consideration of
multiple variables or combinations of variables such as, for instance, region-
ality, socioeconomic status of educational institutions and levels of income. The
necessary paradigm shift would constitute not viewing SAfEs in terms of
British English (measured against current British standards) to viewing it in
relation to other African Englishes in sub-Saharan Africa, and defining it as a
modern African language. Another good reason for looking at English in Africa
differently would be to fall in line with the Linguistic Charter for Africa. In
terms of the Charter, according to Dalby 1985: 29 cited in Skuttnabb-Kangas
and Phillipson (1994: 344-345), it is proposed:

That the *equal linguistic rights of every individual* be recognised, together with the need to provide *access to literacy in every living African language.*

That as many languages as possible in each African state, depending on the number of speakers, be given the status of *national languages,* with an established place in the national education system and in the media.

That at least one African language in each state be given the status of *official language,* to replace or be used alongside any existing "foreign" official language.

What parallels can be drawn between the manifesto above and the emergence of a high-status variety of African English in South Africa as reflected in the *DSAE?* According to some (e.g. Gough 1996 and Buthelezi 1995) "Black English" as it is commonly referred to in the current literature on South African sociolinguistics published post-1992, has already emerged, according to others it will still emerge. This raises questions about the spread and ownership of this variety of English as well as the label. The label "Black English" raises two fundamental issues: the labelling of speech communities on ethnic lines, and the notion that South Africa is on its way to becoming a melting-pot. Apartheid has caused speech communities to be conceived of as separate and distinct, but this can no longer hold. As is pointed out by De Klerk (1996: 9), the label "Black English" has two connotations: one, of subscribing to the American melting-pot philosophy, which I think is too simplistic a view given our very multilingual population and two, of separating speech communities along ethnic lines, which she too points out, is no longer regarded as "politically correct".

Due to the high degree of language contact between English, Afrikaans and African language speakers over a great number of decades, it is often difficult to distinguish between English, Afrikaans and African language origins, as in the case of the following two entries, *jol* and *kwela. Jol* is a word so commonly used in South Africa that it hardly needs an explanation, yet the origins of the word remain reasonably obscure. It is said that the pronunciation of the word is either derived from Afrikaans or is an English pronunciation of the Afrikaans word. Here one might speculate whether the word was perhaps derived from the Afrikaans word *jolyt* or from the English word *jolly.* This demonstrates how intermingled the two languages are historically.

Kwela is said to have its origins in Xhosa and Zulu from the word *khwela* which means "climb on or in, to mount". The word has now taken on a variety of different connotations. Three different but related meanings are listed. *Kwela* is said to be a popular music style characterized by repetitive penny-whistle sounds, or *kwela* is the penny-whistle itself, or it is a dance-style which emerged around the late 1950s and early 1960s in the black ghettos around large African cities such as Johannesburg. *Kwela-kwela* is also listed and refers firstly to police vans and secondly to taxis. Again, there probably aren't that many South Africans who haven't heard the word before.

Facilitating Better Communications?

Is this dictionary really going to add to the empowerment of local users as it proposes to? If it is, then one would be able to say that the editors are making a positive contribution to corpus enhancement. The phrase "lexical enhancement" would be inappropriate, as the word *invasion* implies that unwanted neologisms which muddy the so-called "purity" of the language are imposed, even forced on users, and which have negative effects on its users. The stance against "lexical invasion" has clearly been a very powerful one in the past and it seems, is being deliberately opposed by the editors.

The advertising pamphlet reminds us that: "The language includes both formal and colloquial words, which are indispensable to the proper understanding of fields as diverse as politics, the arts, flora and fauna ... Terms such as influx control, Africanism, indaba ... enable the environment to be fluently and accurately described." My point here is that focusing on fluency and accuracy glosses over social class in relation to variety spoken in different contexts. It would be true to say that the dictionary facilitates enhanced communication and access to education if it lends any guidance at all on desirability of specific lexical items in specific discourse contexts. This is, however, not the case. The dictionary tries to be as unprescriptive as possible. Therefore, one assumes, the listings for each lexeme do not include field labels (indicating certain disciplines or areas of knowledge in which the words might appear) status labels or usage labels (indications of where and when the words might be appropriate), or geographic labels and temporal labels.

The *DSAE* attempts to destigmatize South African varieties of English which, even in the international sphere, have elicited much-documented ridicule. D. Chrysal (1995: 357) caricatures South African English as spoken by white South Africans. He ridicules the pronunciation as a patois-type, uneducated, crude version of "standard" English, in an attempt, it seems, to appear politically correct. To enhance his point of view his text is accompanied by a visual of a naively laughing fresh-faced, wide-eyed male and supposedly signifying Dutch origin accompanied by the text "*AH BIG YAWS?*" (which is supposed to mean "pardon" or "I beg your pardon"). His treatment of South African English is superficial, generalized and therefore insensitive to the local user spectrum, especially as it disregards entirely the English spoken by the majority of South Africans who do not fall into the category he chooses to pick on.

The series *Varieties of English around the World* edited by Manfred Gorlach, a collection of papers incorporating sociolinguistic descriptions of Englishes around the globe, represents the way in which language dictionaries, lexicons and ways of thinking about standard written forms is seen to have taken on a descriptive rather than prescriptive function. Parakrama (1995) on the dehegemonizing of English also seems to subscribe to the trend towards descriptivism in sociolinguistics.

In the advertising pamphlet, the importance of the *DSAE* to international lexicography is said to be the following:

> South African English is considered an important variety and has given the English-speaking world such words as *apartheid* and *trek*. While there are systematic and comprehensive dictionaries of the English used in, for example, the USA, Canada and Australia, there is no such work on South African English. The *DSAE* will fill this gap in the lexicography of English as a world language, and will be recognized and used as an authoritative source.

This assertion raises the issue of authority of sources. Cameron (1995: 50) explains how Oxford University Press publications are regarded as Coke (the real thing) while any others may be regarded as Pepsi (not quite the real thing). While Oxford University Press is certainly regarded in South Africa too as an authoritative source of legitimate knowledge, it needs to be acknowledged that much of the information in for instance the *DSAE* has been drawn from previous reliable and substantial sources. Two of these were also published by Oxford. They are D.R. Beeten and H. Dorner (1975) *A Dictionary of English Usage in Southern Africa* and J. and B. Branford (1991) *A Dictionary of South African English*. The *DSAE*, though, has made the previously collated material more readily accessible while furthering the project of recording language in use, which in addition assists people in studying the evolution of a living spoken language. This once again reflects the author's feeling that the language is being enhanced and extended in a positive way rather than "invaded".

Contributing to Historical Understandings and Political Perspectives and a New South African Identity?

It is stated in the advertising pamphlet that by systematically recording the changing use of words from the late 1500s to the present and the changing ideas which they represent, the dictionary will make a significant contribution to the understanding of South African history and of political terms and concepts. Varying views have been expressed about the usefulness of recording historical origins and understandings of local lexemes in this dictionary.

Peter Wilhelm (1996: 22) is quoted as having written off the 25 years of dictionary compilation conducted by the Rhodes University Dictionary Unit as a "pointless exercise". His major objection is that the dictionary "puts all our linguistic scars on display". By this one assumed he is alluding to the sociopolitical atrocities of our national history which have brought into being a language which now provides blatant irreversible evidence of an embarrassing heritage. It may be said that Wilhelm suffers from an "ostrich mentality" (the tendency of denying evidence, as listed in the *OED*[19] as *ostrich behaviour*: "the

ostrich habit of burying their head in the face of everything they don't like"). Wilhelm is not going to be willingly politically repositioned in order to face his own heritage in a local language.

A related confusion is expressed by Julia Landau (1996: 16) who has pointed out that the dictionary suffers from a cultural bias and that the lexemes have been recorded from a colonial perspective which becomes apparent in definitions such as "bushman—a small nation of savage people ... low in stature, tawny colour'd with crisped hair", recorded in 1968. The point which this reviewer misses, is that the *DSAE* is not pretending to be anything other than a reflection and record of our colonial history, if one believes that what language does is tell one's history.

A different stance is taken by Geoffrey Hughes (1996: 23). He insists that the origins of phrases and words "are important pieces of evidence in our complex, cruel and rich society history". This would be in line with the Foucauldian view that in democratizing institutionalized forms of voluntary and involuntary discourse, we need to focus on the historical conditions of the context and on the rules inherent in the formulation of such discourses.

The review of Colin Bundy (1996: 1-2) strikes a similar, more politically responsible note. He states that: "The complex and often brutal social history of the countryside is also represented by its specialist vocabulary." In his review, conducted from the perspective of a social historian with specific interests in land, its ownership, use and inhabitants, he does a word search through these themes and discovers much interesting background information, some missing lexemes, identifies some omissions and discrepancies, but is generally approving of the dictionary which he maintains "will delight and intrigue anyone interested in the language they speak and read and hear. It rewards serendipity and systematic inquiry with even-handed largesse."

This suggests that the *DSAE* may be able to act as a touchstone of perspectives on current cultural life in South Africa, or provide a new context to frame fractured societies and disruptive existences by simply legitimatising diversity. But rather than fostering a South African identity, should the dictionary not rather foster a Southern African identity? Does the relaxation of political boundaries between states in South Africa and in the Southern Africa region not render the label South African English as too limiting? Should one not rather speak of *A Dictionary of Southern African English*? Evidence has shown that African speakers of English have more commonalities than differences in the way in which Africanisms are incorporated into the English language. The present distribution of Southern Bantu languages, according to Herbert (1990), shows Nguni (which is Zulu and Xhosa) and the Sotho languages (Northern and Southern Sotho and Tswana) to be dominant in the Southern African region. This means that these two language groups are likely to have the heaviest impact on the English emerging in the Southern African region as a whole. South Africa's neighbours have, however, had more standard dialect in their school systems over a number of decades and therefore do not suffer from the

interlanguage fossilization problem as much as South Africans do. In compiling a dictionary of Southern African English such as the *DSAE*, it might have been useful to conduct a comparative study of a specific list of lexical items and their origins and current meanings in Zimbabwe, Botswana and perhaps Namibia.

Can the *DSAE* really contribute to the development of local art-forms and literatures, as it is said to be able to? Performing arts and literature may be directly influenced. Writers of new South African multicultural texts which have become so popular, will probably have to subscribe to the definitions of the local works as prescribed. This raises a very important issue in regard to dictionary compilation. If lexical items for inclusion in this dictionary were last recorded in 1991, then what about the meanings of those words as they stand now in 1997? Significant political changes have resulted in substantial policy changes in, for instance, the media. The SABC is producing more locally flavoured advertisements and more local soap operas in an attempt to affirm and cater for African viewers and appealing to local music fans. Accent and pronunciation seem to have become a non-issue in the media. The scale seems to have been tipped from a high degree of correlation with British standard English to the "anything goes" zone in the last few years.

Let us look, for instance, at the well-known phrase *yebo gogo*. There has been much debate around the meaning of *yebo gogo* since it appeared in the Vodacom advertisement on national television three years ago. Everyone has a different definition for it. Now, to my surprise, I see that *yebo* and *gogo* are listed in the *DSAE* as English words.

In order to get a sense of what the current popular definition of the expression is, I conducted a survey with a group of 150 first-year English students at the University of the Western Cape. To the question, what does *yebo* mean and where does it come from, the most common responses were that it means "to agree with something" or "to admit something", that the meaning is "yes" or that it is a greeting. The most interesting response I got, was from one student who wrote: "It means 'hallo' or 'yes'. It has its origin in the Anglo-African fusion of *yes* and *bona*. *S* and *na* was dumped to give *yebo*." Most students said that *gogo* meant "grandmother". Some said it came from Zulu, some it came from Xhosa.

To the question, what does *yebo gogo* mean and where does it come from, most students wrote that it means "yes grandmother" and it comes from Zulu. Some said that it means that you are agreeing with your grandmother.

The *DSAE* explains *yebo* as being a term of approval, agreement or consent, i.e. in English "We say yes to ..." and is said to be pronounced as "yeh baw". According to the *DSAE*, *gogo* is a term of respect for an elderly person and is also used as title. An alternative meaning of *gogo* is "prophet" or "seer".

Is this corpus enhancement or lexical invasion? If *yebo gogo* is known and commonly used throughout South Africa, this must be an incident of corpus enhancement. The expression crosses all age, racial, ethnic and status barriers.

However in this edition *yebo* and *gogo* are listed separately, while in the next edition they will have to be listed as one, because that has become the popular use of the expression since the appearance of the television advertisement for Vodacom through which it bacame popularized. The social and cultural tapestry in South Africa is rapidly being transformed by media policy and this in turn impacts on language change.

Conclusion

The *DSAE* subscribes to what may be termed "the variation ideology" (Neustupny 1993 quoted in Cameron 1995: 28) which "valorizes linguistic (and ethnic) diversity as a social good in itself" (Cameron 1995: 28), and condones the accommodation of diversity because it is said to enable minority participation in public discourse. It favours innovation over conservatism. In referring to the *DSAE*, one therefore cannot speak of "lexical invasion" as such, because many of the loanwords, borrowings, neologisms and innovations have existed in a very wide user system in the past, but have never before been acknowledged, recognized and accepted in any significant official way or by means of any official language-related policies in South Africa.

In the light of this, the *DSAE* does not make the task of educators any easier or the responsibility of language education any lighter. On the contrary, it amplifies the conflict between two fundamental concepts of linguistic democracy, one being that linguistic varieties should be encouraged, and the other that every individual in society has a right to access fluent Standard English and that it is the duty of the schools not to withhold this right. If linguists, educationalists and policy-makers resolve to work within the Organization of African Unity Language Plan for Action for Africa, this might entail subscribing to a resolution such as the one taken at the PRAESA (Project for Alternative Education in South Africa) Conference in July 1996 at the University of Cape Town. At this international seminar on language and education on the African continent, the resolution taken reads as follows:

> Our promotion of L1 education goes hand-in-hand with our understanding that the people of Africa should be multilingual citizens in multilingual societies. For this reason, we believe that learners should be encouraged to acquire other relevant languages including languages of international communication and vehicular languages of cross-border, intra-African communication.

This is meant to imply that education and access to institutions, general language ability would be assessed, rather than proficiency in L1, L2, L3 or proficiency in medium of instruction. Would intercultural or cross-cultural competence be a factor? And, what is meant by intercultural competence? Would it mean understanding dialects, metaphors etc. of other varieties of English, or

subscribing to a list of nuclear lexemes which are applicable internationally? The dictionary's existence puts the notion of a "nuclear vocabulary" into question. It seems to acknowledge that the boundaries between languages and language varieties are always blurred and that SAfE is a "moving target" as Bundy calls it in his review.

This dictionary, though painstakingly researched and compiled by a team of highly specialized linguists, serves to reduce the "user vs observer" paradox (Stubbs 1986: 71, from Labov 1972: "We would like ideally to observe how languages is used, when no one is observing it.") While this is not possible, the dictionary does blur traditional notions of distinctions between description and prescription and does not pretend to be a style-guide for appropriate usage. This might seem like an "anything goes" philosophy, but while a society is engaged in large-scale socio-political transformation, a context-appropriate style guide would probably be a bit premature.

This dictionary is potentially a first in line of a multitude of different South African English dictionaries still to come (as planned by the Dictionary Unit, Rhodes University). There will eventually be regional dictionaries, school dictionaries, academic dictionaries, historical dictionaries, geographical, cultural, musical, literary dictionaries and dictionaries such as the edition of a dictionary of political terminology published by the *Mail and Guardian*. Another useful publication would be a cross-cultural, cross-lingual, cross-regional reference thesaurus. This would be used in contexts of intercultural communication such as interpretation procedures in court and for worker vs management negotiations where it is crucial to accurately interpret meanings from non-standard dialects.

By producing the *DSAE*, Oxford University Press has finally and officially severed the cross-continental umbilical chord with the "mother language" and has put SAfE decisively on the international linguistic map. So, is the *DSAE* a dictionary of slang? Aikona! Is it a must for every shelf? Yebo gogo! It's lank cool.

References

Beeton, D.R. and H. Dorner. 1975. *A Dictionary of English Usage in Southern Africa.* Cape Town: Oxford University Press.

Branford, J. and B. Branford. 1991. *A Dictionary of South African English.* Cape Town: Oxford University Press.

Bundy, Colin. 1996. Lapping Up our Lekker Lexicon. *Mail and Guardian Review of Books*, Johannesburg, October 1996: 1-2.

Buthelezi, Q. 1995. South African Black English: Lexical and Syntactic Characteristics. Mesthrie, R. (Ed.). 1995: 242-250.

Cameron, D. 1995. *Verbal Hygiene.* London / New York: Routledge.

Chrystal, D. 1995. *The Cambridge Encyclopedia of the English Language.* Cambridge: Cambridge University Press.

Dalby, D. 1985. The Life and Vitality of African Languages: A Charter for the Future. Mateene, Kahombo, John Kalema and Bernard Chomba (Eds.). 1985: 29-34.

De Klerk, Vivian. (Ed.). 1996. *Focus on South Africa.* Amsterdam: John Benjamins.

Gough, D. 1996. Black English in South Africa. De Klerk, Vivian (Ed.). 1996: 53-78.

Herbert, Robert K. 1990. The Sociohistory of Clicks in Southern Bantu. *Anthropological Linguistics* 32(3-4): 295-315, Fall and Winter.

Heugh, K. 1996. English Fiction. *Bua!* 10(2): 17-19, 33, August 1996.

Hughes, Geoffrey. 1996. Dictionary Deserves a Good Word. *Sunday Times*, 15 September 1996: 23.

Hodson, J.A. 1902. *Imperialism, a Study.* London: Allen and Unwin.

Labov, W. 1972. Some Principles of Linguistic Methodology. *Language in Society* 1: 97-120.

Landau, Julia. 1996. Dictionary has Cultural Bias. *Cape Times*, 13 September 1996: 16.

Lanham, L.W. and C.A. MacDonald. 1979. *The Standard in South African English and its Social History.* Heidelberg: Julius Gross.

Mateene, Kahombo, John Kalema and Bernard Chomba (Eds.). 1985. *Linguistic Liberation and Unity of Africa.* OAU/BIL Publication 6. Kampala: OAU Inter-African Bureau of Languages.

Mesthrie, R. (Ed.). 1995. *Language and Social History: Studies in South African Sociolinguistics.* Cape Town: David Philip.

Neustupny, J.V. 1993. History of Language Planning: Retrospect and Future Prospects. Keynote address to the tenth AILA World Congress, Amsterdam, 13 August.

Oxford University Press. Advertising pamphlet. *A Dictionary of South African English on Historical Principles.*

Parakrama, A. 1995. *De-hegemonizing Language Standards: Learning from Post-Colonial Englishes about English.* London: Macmillan.

Prah, Kwesi. 1995. Foreign Languages "Hold Back Africa's Genius". Reported by Mxolisi Mgxashe. *Saturday Weekend Argus*, 15/16 April 1995: 15.

Silva, Penny (Managing Ed.). 1996. *A Dictionary of South African English on Historical Principles.* Cape Town: Oxford University Press.

Skutnabb-Kangas, Tove and Robert Phillipson (Eds.). 1994. *Linguistic Human Rights: Overcoming Linguistic Discrimination.* Berlin / New York: Mouton de Gruyter.

Stubbs, M. 1986. *Educational Linguistics.* Oxford: Basil Blackwell.

Titlestad, Peter. 1996. English, the Constitution and South Africa's Language Future. De Klerk, Vivian (Ed.). 1996: 163-173.

Wilhelm, Peter. 1996. Gatvol with this Pointless Exercise. Letter. *Sunday Times*, 8 September 1996: 22.

Klein Woordeboek / Little Dictionary

Phillip Adriaan Louw, *Departement Afrikaans en Nederlands,*
Universiteit van Stellenbosch, Stellenbosch

Abstract: *Klein Woordeboek / Little Dictionary.* Bilingual translation dictionaries
play an important part in modern user orientated lexicography in South Africa. An affordable bidi-
rectional pocket translation dictionary, such as *Klein Woordeboek / Little Dictionary*, with English and
Afrikaans as language pair, is growing in value as a carrier of necessary everyday linguistic infor-
mation. At present there are no criteria by which these dictionaries can be judged or that can be of
help in their compilation. In this review article an attempt will be made to find typological answers
with *Klein Woordeboek / Little Dictionary* as a guide. This search encompasses traditional and less
traditional typological considerations. It focuses on the needs of target users and specifically the
role which contextual guidance plays in the furthering of communicative equivalence.

Keywords: ACCESS STRUCTURE, ADDRESSING STRUCTURE, ARTICLE, BIDIREC-
TIONAL TRANSLATION DICTIONARY, COMMUNICATIVE EQUIVALENCE, CONTEXTUAL
GUIDANCE, DUAL USER ADDRESSING, EQUIVALENT DISCRIMINATION, EXPLANATORY
INTRODUCTION, INFORMATION DENSITY, LEMMA, LEXICOGRAPHY, MEANING DIS-
CRIMINATION, MEDIOSTRUCTURE, MICROSTRUCTURE, POCKET TRANSLATION DIC-
TIONARY, TARGET LANGUAGE SYNONYM PARADIGM, TARGET USER GROUP, THEORY
OF ORGANISATION, TRANSLATION EQUIVALENT, TYPOLOGY, SEMANTIC RELATIONS,
SOURCE AND TARGET LANGUAGE POLYSEMY, USER ADDRESSING, USER PERSPECTIVE,
USER PROFILE

Opsomming: Tweetalige vertalende woordeboeke speel 'n belangrike rol in die moderne
gebruikersgerigte leksikografie in Suid-Afrika. 'n Bekostigbare tweerigting vertalende sakwoorde-
boek soos *Klein Woordeboek / Little Dictionary* met Engels en Afrikaans as behandelde taalpaar het
groeiende waarde as draer van noodsaaklike alledaagse taalkundige inligting. Tog is daar nog
geen maatstawwe gestel waarvolgens hierdie woordeboeke beoordeel kan word of wat as hulp kan
dien by die opstel daarvan nie. In hierdie resensieartikel word daar met *Klein Woordeboek / Little
Dictionary* as gids gepoog om tipologiese antwoorde te vind. Die soektog betrek tradisionele en
minder tradisionele tipologiese oorwegings. Dit fokus op die behoeftes van die teikengebruiker en
veral die rol van konteksleiding in die bevordering van kommunikatiewe ekwivalensie.

Sleutelwoorde: ADRESSERINGSTRUKTUUR, ARTIKEL, BETEKENISBETREKKINGE,
BETEKENISDISKRIMINASIE, BRON- EN DOELTAALPOLISEMIE, DOELTAALSINONIEM-
PARADIGMA, DUBBELE GEBRUIKERSADRESSERING, EKWIVALENTDISKRIMINASIE,
GEBRUIKERSADRESSERING, GEBRUIKERSPERSPEKTIEF, GEBRUIKERSPROFIEL, INLIG-
TINGSDIGTHEID, KOMMUNIKATIEWE EKWIVALENSIE, KONTEKSLEIDING, LEKSIKOGRA-

FIE, LEMMA, MEDIOSTRUKTUUR, MIKROSTRUKTUUR, ORGANISASIETEORIE, TEIKEN-
GEBRUIKERSGROEP, TIPOLOGIE, TOEGANGSTRUKTUUR, TOELIGTING, TWEERIGTING
VERTALENDE WOORDEBOEK, VERTAALEKWIVALENT, VERTALENDE SAKWOORDEBOEK

Die Suid-Afrikaanse leksikografie met sy multilinguale konteks bied 'n eiesoor-
tige uitdaging aan die vertalende leksikograaf. Die opdrag om suksesvol kom-
munikatiewe ekwivalensie te bevorder, behoort die leksikograaf te noodsaak
om indringend aandag te skenk aan die gebruikersperspektief wat dan moet lei
tot die opstel van 'n fyn gedefinieerde tipologiese profiel vir die betrokke woor-
deboek. Veral dié woordeboeke met Afrikaans en Engels as behandelde taal-
paar skep 'n unieke geleentheid vir diegene wat werk aan 'n standaard verta-
lende woordeboek, om tradisionele tipologiese grense veral op die terrein van
"gebruikersadressering" (Louw en Gouws 1996: 98) oor te steek. Ongelukkig is
tipologiestudies met betrekking tot die vertalende leksikografie in Suid-Afrika
'n metaleksikografiese veld wat nog braak lê. Dit het dan ook 'n nadelige
invloed op die sukses van die praktiese leksikografie.

      Die gebrek aan navorsing is nie net beperk tot standaardwoordeboeke nie.
In Suid-Afrika is daar 'n uiters produktiewe kategorie, naamlik die vertalende
sakwoordeboek, gewoonlik met Engels en 'n ander landstaal as bewerkte tale.
Die taalpaar Afrikaans en Engels beskik oor vier verteenwoordigers in hierdie
kategorie, naamlik *Klein Woordeboek / Little Dictionary* (voortaan afgekort as
KW) en *Mini Woordeboek*, voorheen uitgegee deur J.L. van Schaik, *Tweetalige
Sakwoordeboek*, voorheen deur Nasou uitgegee en *Juta se Sakwoordeboek / Juta's
Pocket Dictionary*. Ten spyte van dié oorvloed, is daar nog geen gestandaardi-
seerde kriteria waarvolgens hierdie onderskatte tipe woordeboek beoordeel
kan word nie. In hierdie resensieartikel sal daar gepoog word om, met KW as
voorbeeld, sekere kriteria te stel en derhalwe 'n voorlopige tipologie te bena-
der.

Tipologiese vereistes

Seker die mins tradisionele, maar mees belangrike tipologiese beperkings van
'n vertalende sakwoordeboek, is dat dit die teikengebruiker se sak moet pas:
letterlik en figuurlik. Eerstens moet dit 'n gemaklike grootte boekie wees om in
'n sak of 'n tas te dra. KW met sy klein formaat is hier 'n sterk mededinger wat
byvoorbeeld gemakliker is as die groter *Tweetalige Sakwoordeboek*. Mary Haas se
primêre vereiste aan 'n ideale vertalende woordeboek, naamlik "compactness"
(Haas 1967: 47) is dus hier van kardinale belang. Dit het dan ook verreikende
implikasies vir die gebruikersadressering en die onderskeie strukture van so 'n
woordeboek.

      Tweedens is dit belangrik dat sodanige woordeboek bekostigbaar moet
wees. Teen ongeveer R30 vervul KW 'n waardevolle rol in die Suid-Afrikaanse
mark. Dit is goedkoop genoeg om (ideaal gesproke) aangekoop te word as 'n

tweede vertalende woordeboek vir diegene wat 'n hulpmiddel nodig het vir situasies wat nie linguisties te komplekse vereistes stel nie. Die realiteit is egter dat dit ook soms dien as enigste woordeboek vir diegene wat nie 'n duurder standaard vertalende woordeboek kan bekostig of wil aankoop nie.

Hierby is dit dus nodig dat die opstellers van 'n vertalende sakwoordeboek hulle teikengebruiker noukeurig sal bepaal, maar ook voorsiening sal maak vir mense wat die woordeboek buite sy tipologiese grense wil / moet gebruik. Die implikasie is (vir byvoorbeeld KW) dat dié woordeboekie primêr gerig sal wees op die gebruiker met 'n redelike kennis van beide tale (dus nie aanleerders nie) wat vinnig antwoorde op relevante, alledaagse vertalingsvrae en dus vernaamlik vertaalekwivalente, soek.

Buiten die woordeboekvaardighede van die woordeboek se teikengebruikers moet die leksikograaf ook bepaal aan watter taalgroep hulle behoort. Een van die positiewe punte om te noem ten opsigte van KW is dat dit 'n tweerigting vertalende woordeboek is. Hierteenoor is een van die sterkste punte van kritiek dat dit klaarblyklik albei taalgroepe probeer dien in albei afdelings van die sentrale teks. Haas (1967: 47) se algemene stelling is van toepassing ook op die opstellers van KW: "Thinking they are preparing a dictionary for speakers of both languages, they ... end up producing a dictionary which is not as useful as it should be to speakers of either language".

Louw en Gouws (1996: 92-100) wys daarop dat 'n standaard vertalende woordeboek met Afrikaans en Engels as behandelde taalpaar, deur op 'n vernuwende wyse die mikrostruktuur en adresseringstruktuur te benader, wel 'n dubbele gebruikersadressering kan implementeer in elke deel van die woordeboek. Indien die konvensionele benadering van die Afrikaanse leksikografie gevolg word waar slegs karige konteksleiding gegee word, sal die sisteem misluk in selfs 'n groter standaardwoordeboek. Ook in KW word die gebrekkige sisteem aangetref wat die suksesvolle betekenisoordrag en dubbele gebruikersadressering van sy groter standaardstalmaat, *Groot Woordeboek / Major Dictionary* (voortaan GW), kelder.

Verder tel KW se belangrikste bate, sy grootte, ook teen 'n suksesvolle bewerking volgens beginsels wat by 'n dubbele gebruikersadressering in die woordeboek geld. Daar is bloot nie genoeg ruimte om twee sprekersgroepe binne een tweerigting vertalende sakwoordeboek van genoeg konteksleiding te voorsien om 'n toereikende funksionele ekwivalensie te verseker nie. In hierdie resensieartikel sal daar geagiteer word vir twee aparte tweerigting vertalende sakwoordeboeke: een vir Afrikaanssprekendes en 'n ander vir Engelssprekendes, beide van die grootte van KW.

Die bewyse wat KW lewer vir hierdie aanvraag lê veral op die vlak van die mikrostruktuur. 'n Vollediger bewerking van die lemmas is 'n noodsaaklike verandering. Hoewel 'n sakwoordeboek uiteraard 'n laer inligtingsdigtheid as 'n standaardwoordeboek het, geld die kriterium dat 'n vertalende woordeboek genoeg inligting aan die gebruiker moet bied om hom / haar in staat te stel om die regte vertaalekwivalent te kies, steeds. KW se gebrekkige en inkonsekwente

verskaffing van noodsaaklike betekenisinligting weerspreek deurgaans hierdie kriterium. Die miskenning van die gebruiker se behoeftes lê een van die grootste gebreke van KW bloot. Die tipologiese keuses en die gevolglike mikrostrukturele bewerkings moet gemaak word aan die hand van 'n realistiese "gebruikersprofiel" (Gouws 1989: 49) wat vir die betrokke woordeboek opgestel is. Die tekort aan 'n goed gedefinieerde en verantwoorde teikengebruikersgroep is dus die beginpunt van KW 'se mikrostrukturele dilemma. Ter versagting moet dit egter genoem word dat die foute in KW deurgaans simptomaties is van kwale in die breë vertalende leksikografie in Suid-Afrika. Weens hulle noue historiese verbintenis is die meeste swakhede in KW dan ook na te speur in GW.

Mikrostrukturele bewerking

Soos by KW se stalmaat bemoeilik 'n gebrek aan 'n konsekwente sisteem van semantiese konteksleiding en gebruiksleiding die gebruiker se kommunikatiewe sukses. Twee vlakke van betekenisinligting word veral afgeskeep, naamlik betekenisdiskriminasie (beide van die lemma se polisemiese waardes en tussen die onderlinge vertaalekwivalente) en ook wat Gouws (1990: 57) noem "secondary semantic relations", met spesifiek die aandui van betekenisbetrekkinge as oogmerk. By eersgenoemde sal gefokus word op veral bron- en doeltaalpolisemie en diskriminering tussen doeltaalsinonieme.

Betekenisdiskriminasie

Die korrekte hantering van brontaalpolisemie voorveronderstel dat die vertalende leksikograaf sekerheid verkry oor die verband wat tussen die polisemiese waardes van die lemma bestaan voordat vertaalekwivalente gegee word. Hierdie verband moet ook vir die gebruiker duidelik gemaak word (in die brontaal) indien die spesifieke deel van die tweerigting vertalende sakwoordeboek 'n enkoderende funksie wil verrig. Slegs dan sal die moedertaal- en / of toegangstaalspreker van die brontaal die woordeboek kan gebruik om doeltaaltekste te produseer. By die gebruik van 'n vertalende sakwoordeboek kan die opsteller nie van die gebruiker verwag om hierdie informasie uit 'n verklarende woordeboek te kry nie. 'n Sakwoordeboekgebruiker wil juis dié woordeboek as eerste en enigste bron van inligting gebruik. 'n Sak het nie ook nog plek vir 'n tweede woordeboek nie.

Die bietjie toeligting wat wel voor die Afrikaans-Engelse deel van KW gegee word, is in Afrikaans. Dit impliseer dat hierdie deel primêr gerig sou wees op brontaalsprekers. Tog word daar geen betekenisleiding vir die Afrikaanssprekende gebied nie. Die inkonsekwent toegepaste sisteem maak slegs voorsiening vir vertaalkomplemente wat doeltaalinskrywings is. Die meeste polisemiese lemmas word egter sonder verdere leiding van vertaalekwivalente vir die betekenisonderskeidings voorsien.

'n Voorbeeld van dié gebruikersonvriendelike hanteringswyse is die bewerking van die lemma **premie** se optrede as selfstandige naamwoord. Die lemma word voorsien van drie vertaalekwivalente vir die onderskeie polisemiese waardes naamlik **premium; bounty; bonus**. Geen verdere konteksleiding word vir die Afrikaans- of Engelssprekende gegee nie. Die aard en graad van die onderlinge verwantskappe word nie geëkspliseer nie. Indien die leksikograaf hier 'n besluit neem om die taal waarin konteksleiding gegee word te laat korreleer met die taal waarin die grammatikale inligting en toeligting gebied word, sal 'n sisteem van "sense discriminations" (Iannucci 1985: 58) ingespan kan word. Hierdie brontaalinskrywing sal die artikelgleuf voor elke vertaalekwivalent of sinoniemparadigma vul. Gekoppel aan 'n sisteem wat polisemiese onderskeidinge numeries orden, sal so 'n konsekwente toepassing van hierdie diskriminators die kommunikatiewe waarde van die woordeboek vir Afrikaanssprekendes verhoog. Die diskriminator sal egter verkieslik net 'n enkele woord of frase moet wees om by die sakwoordeboek se kenmerkende vinnige soekroete te kan inpas. Dieselfde sisteem sal dan gevolg kan word in die Engels-Afrikaanse deel van 'n vertalende sakwoordeboek gerig op Engelssprekendes.

'n Ideale tweerigting vertalende sakwoordeboek behoort egter ook in die ander deel van die woordeboek vir sy teikengebruikersgroep se teenstellende behoeftes voorsiening te maak. In die hipotetiese voorbeeld sal dit beteken dat die Engels-Afrikaanse deel (wat nou op die doeltaalspreker gerig sal wees) sal optree as werktuig vir die dekodering van brontaaltekste asook die re-enkodering daarvan. Dit behels die oorgang vanaf die verkryging van 'n vertaling of vertaalekwivalent tot die formulering of enkodering daarvan in die doeltaal (dus Afrikaans). Ook hierby geld Gouws (1989: 151) se pleidooi: "Die vertaalekwivalent van elke polisemiese onderskeiding moet 'n konteksaanduiding kry."

Daar is reeds genoem dat KW wel doeltaalinskrywings gee in die artikelgleuf na die betrokke vertaalekwivalent. Hierdie inskrywings dien as vertaalkomplemente, dus "as doeltaalinskrywings in die mikrostruktuur wat daarop gerig is om die vertaalekwivalent uit te brei" (Louw en Gouws 1996: 93). Weens die beperkte spasie verkies KW hier die diskriminators wat Al-Kasimi (1977: 71) "context words or phrases" noem. Hierdie inskrywings word ook in die standaard vertalende woordeboeke soos GW gebruik.

In die dekoderende / re-enkoderende deel is hierdie inskrywings van onskatbare waarde. Dit bied nie net aan die gebruiker hulp ter afbakening van die lemma se polisemiese waardes nie, maar bied ook waardevolle leiding indien die vertaalekwivalent self 'n polisemiese leksikale item is. In KW se Engels-Afrikaanse deel kry die lemma **cup** drie vertaalekwivalente wat deur kommapunte geskei word: **koppie** en **beker** en **kelk**. Die primêre vertaalekwivalent **koppie** word onverklaard gelaat ('n praktyk wat nie onkrities aanvaar behoort te word nie). Die ander twee word egter elk van 'n kontekswoord bedien naamlik **(sport)** by **beker** en **(blom)** by **kelk**. Op dié wyse word die

potensiële verwarring wat bron- en doeltaalpolisemie kan skep, vir die Afrikaanssprekende uit die weg geruim.

Hoewel KW oor 'n werkbare inskrywingstipe beskik, verloor hierdie konteksleiding baie waarde weens 'n lukrake toepassing. Teenoor **cup** staan die swak hantering van die lemma **maiden** as selfstandige naamwoord wat bloot die volgende vertaalekwivalente vir sy polisemiese waardes kry: **meisie; nooi; maagd; leë boulbeurt. Leë boulbeurt** is 'n krieketterm wat 'n vae verwantskap met die ander waardes se vertaalekwivalente toon. Tog kry dit nie die verwagte kontekswoord **(krieket)** daarby nie. Dit word ongemerk aangebied. Tussen **meisie** en **nooi** is daar hierteenoor so 'n noue verwantskap dat die doeltaalvorm maklik as sinonieme in sekere kontekste gebruik kan word. Deur nie brontaalpolisemie aan te dui of die spesifieke betrokke polisemiese waardes van die vertaalekwivalent te ekspliseer nie, tree hierdie artikel eerder verduisterend as verhelderend op. Die woordeboek moet die gebruiker se vrae beantwoord en nie meer vrae of verwarring skep soos by die hantering van **maiden** nie.

Die leksikograaf moet as deel van die "organisasieteorie" (Wiegand 1984: 15) van sy / haar woordeboek 'n sisteem opstel vir die hantering van polisemie. Slegs dan kan 'n konsekwente toepassing van 'n gebruikersvriendelike sisteem volg. Hierdie kriterium geld dan ook vir die diskriminering tussen lede van 'n doeltaalsinoniemparadigma.

Ekwivalentdiskriminasie binne sinoniemparadigmas

In beide dele van 'n tweerigting vertalende sakwoordeboek met 'n enkele gebruikersadressering behoort daar ekwivalentdiskriminasie binne doeltaalsinoniemparadigmas toegepas te word. Dit gebeur dikwels dat verskillende sinonieme gegee word as vertaalekwivalente omdat die woordeboek daardeur vir fyner nuanses van die lemma voorsiening maak. Dit kan óf etiketteerbare kontekstuele verskille (byvoorbeeld styl- en registerverskille) óf verskillende gebruike van die lemma wees. In elke geval moet die doeltaalvorm binne 'n sekere konteks vir die lemma 'n vertaalekwivalent wees en hierdie konteks behoort geëkspliseer te word, hetsy in die brontaal of die doeltaal. Kromann e.a. (1989: 2724) beweer tereg: "It is one of the ancient and deadly sins of translation lexicography in bidirectional dictionaries to provide lists of equivalents and accompanying quasi synonyms without meaning discriminating comments". 'n Gebrek aan sodanige leiding is een van die grootste tekortkominge in KW.

Die lemma **hou**[1] kry onder meer die doeltaalsinonieme **blow** en **stroke**. Daar word egter geen leiding vir óf die bron- óf die doeltaalgebruiker gegee in verband met die spesifieke konteks waarin elk van hierdie woorde ekwivalent is aan **hou**[1] nie. Volgens HAT se hantering van **hou** verteenwoordig **blow** en **stroke** verskillende gebruike van die naamwoord **hou**. KW behoort hierby vir

konteksverklarende aantekeninge voorsiening te maak hetsy deur middel van 'n brontaalinskrywing voor elke sinoniem (vir Afrikaanssprekendes) of 'n doel-taalinskrywing na elke sinoniem (vir Engelssprekendes) afhangende van die deel se gebruikersadressering. Verder behoort KW ook sy keuse tussen vertaal-ekwivalente vir 'n lemma se gebruike en dié vir 'n lemma se polisemiese waar-des duideliker te maak en te motiveer. Die aanbied van byvoorbeeld **sick** en **dreary** as doeltaalsinonieme vir **naar**, kan nie sonder meer as reg aanvaar word nie.

Ook die hantering van die lemma **slym** is problematies met **slime, phlegm** en **mucus** wat onverklaard gegee word. Hier is egter duidelike stylverskille tussen **slime** en **phlegm**) en registerverskille (tussen **slime** en die biologieterm **mucus**). Etikette sou enige verwarring uit die weg kon ruim.

'n Toekomstige sakwoordeboek sal nie langer die beginsels van gebrui-kersvriendelikheid op dié wyse kan misken nie. Deeglike konteksleiding (semantiese en gebruiksleiding) sal op 'n sistematiese wyse ook by doeltaal-sinonieme gegee moet word. Die leksikograaf sal in die aard, adressering en die keuse tussen bron- en doeltaalinskrywings haar / hom moet laat lei deur die behoeftes en vaardighede van die teikengebruiker.

Betekenisbetrekkinge

'n Vernuwende sisteem van konteksleiding alleen sal nie die probleme wat in KW bestaan, kan oplos nie. Dit sal in kombinasie moet optree met 'n andersoor-tige benadering tot die markering van betekenisbetrekkinge. Die konvensionele sisteem van kommas tussen doeltaalsinonieme en kommapunte tussen vertaal-ekwivalente en sinoniemparadigmas wat 'n polisemiese waarde kan vervang, is verouderd en steeds verwarrend vir die minder bedrewe gebruiker. Hierdie probleem word vererger deur die gebrekkige toeligtingsteks wat nie die gebruik van hierdie kenmerkers verklaar nie. 'n Sisteem geskoei op die verkla-rende woordeboeke se numeriese en lettermerkers is 'n moontlikheid wat oorweeg behoort te word. Die toegangstruktuur, en veral die interne toegang-struktuur, van KW behoort dringend aandag te geniet.

Verder is ook die ordeningsisteme waarvolgens die posisies van die poli-semiese waardes en / of sinonieme in 'n artikel bepaal word, telkens lukraak en ongemotiveerd (vergelyk **skakel** as werkwoord en **ignore**). Dit is weer eens een van die sterkste punte van kritiek ook teen GW. 'n Doelbewuste en seman-ties motiveerbare ordeningsisteem is nog 'n kardinale komponent waaraan 'n vertalende leksikograaf aandag moet skenk in die vasstelling van die organisa-sieteorie van die woordeboek.

Die veranderinge wat tot dusver bepleit is, kan egter slegs effektief wees indien dit saamhang met 'n algehele vernuwing van die bestaande artikelstruk-tuur van KW. Op die oomblik mors die losstaande hantering van voorbeeld-materiaal kosbare spasie deurdat kollokasies en selfs voorbeeldsinne hulle dis-

kriminerende waarde verloor. Vergelyk hier die hantering van die lemma **aan**. Slegs die derde kollokasie, naamlik **aan tafel** bevat enige diskriminerende waarde omdat die vertaling hiervan een van die gegewe vertaalekwivalente vir **aan**, naamlik **at** bevat. Tog moet die gebruiker in dié geval verwyderde lemmatiese en nielemmatiese adresseringsprosedures benut (aangesien die verwante inskrywings nie bymekaar geplaas is nie) om die nodige inligtingskategorieë te laat korreleer. Die leksikograaf koester hier 'n onregverdige verwagting van die meeste gebruikers, wat die beste gedien sal word deur onmiddellike toegang tot konteksleiding. Voorbeeldmateriaal behoort in 'n vertalende sakwoordeboek in 'n artikelgleuf tussen die betrokke vertaalekwivalent en die volgende vertaalekwivalent gegee te word om sodoende vinnige toegang en maksimum diskriminerende waarde te verseker. Binne 'n nuwe aanduidingsisteem kan daar dan ook 'n aparte artikelposisie gevind word vir kollokasies waarby die vertalings nie verwant is aan die lemma se vertaalekwivalente nie. Enige veranderings vanaf die norm moet egter deeglik in 'n toeligtingsteks verwoord word.

Die enigste antwoord blyk dus 'n beginselverandering te wees. In plaas van leksikografiese apartheid, behoort dit wat semanties bymekaar hoort binne die mikrostruktuur bymekaar geplaas te word. Hierdie beginsel kan ook help om van die verwarring wat KW se nislemmatisering skep, op te klaar. By die lemma pig is nislemmatisering gekies, met die gevolg dat die lemma pigeon en later **pigment** tussen die lemma **pig** en verwante sublemmas geplaas word. Weens 'n tekort aan konteksleiding blyk dit verder ook nie altyd duidelik uit al die vertaalekwivalente van die nislemmas wat binne 'n artikel met pig: as lemmateken voorkom, wat die presiese verwantskap met die oorspronklike lemma **pig** is nie. Indien die leksikograaf die stelsel wil behou, sal 'n omvattende sisteem van kruisverwysings en konteksleiding by nislemmas 'n voorvereiste wees.

Daarteenoor kan nislemmas vervang word met neslemmas, wat die betrokke morfologiese derivate by die betrokke polisemiese onderskeidings plaas. Die vraag wat die leksikograaf sal moet stel, is of 'n neslemmatisering nie die soekroetes van die gebruiker meer sal versteur nie. Neslemmas is die semantiese ideaal. Binne die heersende Suid-Afrikaanse woordeboekkultuur moet daar egter gehou word by 'n streng alfabetiese ordening. Dit geld vir die makrostruktuur, maar ook veral vir die aanbied van sublemmas as nislemmas. Die semantiese ideaal volgens die anti-apartheidsbeginsel moet dus noodwendig in 'n sakwoordeboek ondergeskik gestel word aan nislemmas met deeglike kruisverwysings as deel van die mediostruktuur (dit is die struktuur wat alle leksikografiese kruisverwysings omvat) wat strek tot by die spesifieke polisemiese waarde of gebruik waaraan dit verwant is. Hoewel KW se hantering steeds gebrekkig is, is dit reeds 'n verbetering op die onlogiese inkonsekwenthede te vinde by GW (vergelyk dié woordeboek se hantering van die lemma **lock**). KW neig dus al na 'n verbeterde gebruikersvriendelike sisteem maar die mediostruktuur vereis veel meer noukeurigheid.

Toeligting

KW ignoreer egter nog 'n baie elementêre beginsel van gebruikersvriendelikheid. Gold (1993: 304) maak die volgende relevante stelling in sy resensie van *Bilingual Phrase Dictionary*: "Responsible dictionary compilers are expected not only to give careful thought to those questions [oor byvoorbeeld gebruikersadressering], take explicit decisions, and give effect to their decisions, but also to tell potential users of their works precisely what they can expect from them". Die toeligtingsteks in 'n vertalende sakwoordeboek behoort ook nie slegs 'n speelveld te bied vir die leksikograaf om sy / haar organisasieteorie te motiveer nie. Dit behoort primêr daarop gerig te wees om as 'n bemagtigingsmiddel vir die gebruiker te dien. In dié funksie mag die toeligting nie gesien word as 'n opsionele en plekvermorsende deel nie, maar as 'n noodsaaklike, verpligte deel van die voorwerk van 'n vertalende sakwoordeboek. Al die voorstelle wat dus gemaak is ter verbetering van KW kan slegs 'n daadwerklike impak hê indien die opstellers dit deeglik uiteensit en verantwoord in 'n toeligtingsteks.

Toegangstruktuur

Die bestaan van sodanige toeligtingsteks kan die leksikograaf ook in staat stel om meer innoverend met die woordeboek se toegangstruktuur te werk te gaan. Syfer- en lettermerkers, simboolmerkers, 'n ander kleur of twee, bykomende lettertipes, ens. kan ingespan word mits dit in die toeligting verantwoord word. KW se huidige aanduidingsisteme is te gebruikersonvriendelik om die gebruiker daadwerklik in veral sy artikelinterne soekroete by te staan. Verandering sal moet kom, maar die leksikograaf sal hom deur 'n paar praktiese kriteria moet laat lei. Eerstens sal die koste verbonde aan byvoorbeeld ander kleure drukwerk in ag geneem moet word. Tweedens sal daar rekening gehou moet word met die tekortkominge van die beskikbare rekenaarprogrammatuur en die voorkeure en afkeure van uitgewers. 'n Estetiese aanbod wat nie oogsteurende vernuwings invoer nie is 'n verdere voorvereiste. Verreweg die belangrikste kriterium is egter gebruikersvriendelikheid. Onder geen omstandighede mag die vernuwende toegangstruktuurkomponente die teksdigtheid so verhoog dat die inligting vir die teikengebruiker ontoeganklik word nie. Ook hierdie kriteria behoort deeglik in ag geneem te word by die opstel van 'n werkbare organisasieteorie. In die vertalende sakwoordeboek behoort die algehele aanslag daarop gerig te wees om die gebruiker so vinnig moontlik te bring by die inligting waarna hy soek. Indien die vernuwings hierteen indruis, misluk die leksikograaf in sy doel.

Slot

Uit hierdie resensieartikel behoort dit duidelik te blyk dat daar van enige leksikograaf verwag word om 'n goeie sisteem te hê, dit deeglik in die toeligtings-

teks te verwoord, en dit konsekwent toe te pas. In bykans al die belangrikste inligtingstipes en strukture misluk KW met betrekking tot al drie hierdie kriteria. Verder misluk 'n goeie sisteem van byvoorbeeld grammatikale inligting en een van konteksleiding bloot weens 'n inkonsekwente toepassing en 'n onvermoë om die noodsaaklikheid van 'n enkele gebruikersadressering in te sien. Feitlik al hierdie tekortkominge kan teruggevoer word tot 'n gebrekkige organisasieteorie.

Tog hoef die toekoms nie duister te lyk nie. Met die amalgamering van Nasionale Boekhandel se woordeboekuitgewers in een sentrale buro, Pharos, kan daar daadwerklik gekyk word na 'n nuwe benadering ten opsigte van hulle drie bestaande vertalende sakwoordeboeke. KW het reeds van die mikrostrukturele swakhede en oortollighede van GW uitgeskakel en sodoende die kommunikatiewe sukses verhoog en die makrostruktuur proporsioneel vergroot. Indien 'n gees van verdere vernuwing in die woordeboekburo posvat, kan die vertalende sakwoordeboeke 'n innoverende werkswyse toon wat vir die Suid-Afrikaanse vertalende leksikografie van onskatbare waarde sal wees en hopelik ook na hulle groter stalmaats sal versprei.

KW beset 'n deel van die mark wat deur talle van die tradisionele grense in die teikengebruikersverdeling sny. Indien leksikografiese sukses op hierdie vlak verkry kan word, kan dit 'n belangrike faktor word in die skep van 'n gunstige taalgebruiksklimaat waarin die wisselwerking tussen gebruikersvriendelikheid en woordeboekkultuur net kan lei tot die verdere verbetering van die leksikografiese standaard.

Bronnelys

A.   Woordeboeke

Barnes, A.S.V. e.a. 1993[3]. *Tweetalige Sakwoordeboek*. Kaapstad: Nasou.

Juta se Sakwoordeboek / Juta's Pocket Dictionary. 1983[6]. Kenwyn: Juta.

Kritzinger, M.S.B., P.C. Schoonees, U.J. Cronjé en L.C. Eksteen. 1986[13]. *Groot Woordeboek / Major Dictionary*. Pretoria: J.L. van Schaik.

Kritzinger, M.S.B. en J. Kromhout. 1992[11]. *Klein Woordeboek / Little Dictionary*. Pretoria: J.L. van Schaik.

Kromhout, J. 1993. *Mini Woordeboek*. Pretoria: J.L. van Schaik.

Odendal, F.F. (Red.). 1994[3]. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Doornfontein: Perskor.

B.   Ander Bronne

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

Gold, David L. 1993. Resensie van *Bilingual Phrase Dictionary / Tweetalige Frasewoordeboek (E-A)*. *Lexikos* 3: 303-313.

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

Gouws, R.H. 1990. Information Categories in Dictionaries, with Special Reference to Southern Africa. Hartmann, R.R.K. (Red.). 1990: 52-65.

Haas, M.R. 1967. What Belongs in a Bilingual Dictionary? Householder, F.W. en Sol Saporta (Reds.). 1967[2]: 45-50.

Hartmann, R.R.K. (Red.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983.* Lexicographica, Series Maior 1. Tübingen: Max Niemeyer.

Hartmann, R.R.K. (Red.). 1990. *Lexicography in Africa.* Exeter Linguistic Studies 15. Exeter: University of Exeter Press.

Hausmann, F.J., Q. Reichmann, H.E. Wiegand en L. Zgusta (Reds.). 1989-1991. *Wörterbücher. Dictionaries. Dictionnaires. An International Encyclopedia of Lexicography.* Berlyn: Walter de Gruyter.

Householder, Fred W. en Sol Saporta (Reds.). 1967[2]. *Problems in Lexicography.* Bloomington: Indiana University.

Iannucci, J.E. 1985. Sense Discriminations and Translation Complements in Bilingual Dictionaries. *Dictionaries* 7: 57-65.

Kromann, Hans-Peder, Theis Riiber en Poul Rosbach. 1989. Principles of Bilingual Lexicography. Hausmann, F.J. e.a. (Reds.). 1989-1991: 2711-2728.

Louw, P.A. en R.H. Gouws. 1996. Lemmatiese en nielemmatiese adressering in Afrikaanse vertalende woordeboeke. *Suid-Afrikaanse Tydskrif vir Taalkunde* 14 (3): 92-100.

Wiegand, H.E. 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.R.K. (Red.). 1984: 13-30.

R.R.K. Hartmann (Ed.). *Solving Language Problems: From General to Applied Linguistics,* 1st edition 1996, vi + 298 pp. ISBN 0-85989-484-3. Exeter: University of Exeter Press. Price £13,95.

Questions about the nature of linguistics and its contribution to human intellectual pursuit often appear in academic discussions and in casual conversations. While any language practitioner with a linguistic background can give an elaborate answer to these questions, the general perception, even amongst educated people, seems to be that linguistics is a somewhat esoteric discipline with little to offer towards "practical" issues such as language learning.

Hartmann's book, *Solving Language Problems,* offers an accessible survey of key areas in linguistics and their relevance in the field of applied linguistics. This edited volume, containing ten chapters, seeks to demonstrate that the relation between general and applied linguistics cannot be reducible to theory versus practice.

In their introductory chapter, Cahill and Hartmann defend the intellectual autonomy of linguistics and discuss the vast realm of applied linguistics. General linguistics has fragmented into a number of sub-areas and interdisciplinary fields and the more one moves from theory to practical problem solving, the more contact there is with other subject fields. This interdisciplinary nature lies at the heart of modern linguistics and at the same time is a defining feature of applied linguistics. The authors concede that the relation between the two appears to be mutually reinforcing rather then dichotomous. Interpretation also depends on the scope of meaning of the applied linguistics, which is viewed by some rather narrowly as a teaching-oriented field, and by others as a broadly understood study of the uses that a person makes of language and problems he or she encounters in the process.

The introduction also specifies the intended audience for the volume and recommends it as an "introduction to linguistics for students about to embark on a postgraduate course in the subject" (p. 11). The comment on the back cover of the book highlights its usefulness to undergraduates in the field of linguistics and applied linguistics. Ultimately, both these recommendations and the content of the whole volume leave one rather unclear about the target readers and how much knowledge of linguistics the authors assume a priori from their audience.

The topics selected for the volume fall into three groups: language theory, language description, and language use in context. The first includes a chapter on Transformational-Generative Grammar (Mersedeh Proctor) and a chapter devoted to Systemic Grammar (Steven Dodd). Language description comprises Phonetics and Phonology (Steven Dodd and Jon Mills) and Aspects of Lexical Semantics (Rufus Gouws) which — by the author's own admission — is not located in any particular theoretical framework. Language use in context is, predictably, most widely represented in the present volume and includes a contribution on Discourse (Mersedeh Proctor), chapters on Psycholinguistics

(Aidan Cahill) and Sociolinguistics (Mike O'Rourke), and three papers with a strong focus on applied linguistics as a teaching-oriented discipline: Orientations in Language Teaching and Testing (Diane Davies), Lexicography as an Applied Linguistic Discipline (Reinhard Hartmann) and Computers in Applied Linguistics (Jon Mills).

Given the number of contributions and the vast area of coverage, it is only to be expected that they will differ in depth of presentation; some offer a necessarily cursory view of a discipline (Phonetics and Phonology), others combine a fairly exhaustive theoretical exposition with a thorough analysis of data (Discourse). Stylistic variation is also palpable; ranging from a rigorous academic discourse of chapter 3 (Systemic Grammar), to a more casual mode of presentation in chapter 9 (Lexicography).

Although in general all contributions can be assessed as useful and instructive, a few contain certain problematic aspects.

The first contribution (by Steven Dodd and Jon Mills) provides an introduction to phonetics and phonology. While the need for the inclusion of a chapter on speech sounds and sound systems is indisputable, the description and presentation are inconsistent. The reader is acquainted with the IPA symbols, place and manner of articulation for such "exotic" sounds as the German voiceless palatal fricative, the uvular stop found in Arabic, or the uvular trill of French and German, yet the "commonplace" velar nasal found in English is referred to as the "sound written <ng>" (pp. 16, 17), with no accompanying symbol. The pulmonic airstream mechanism is described in detail but the glottalic airflow only receives a mention in the form of "implosive or explosive glottalic sounds" (p. 15), which can hardly satisfy the curiosity of a beginning student in linguistics for whom the book is intended.

The application of grammatical descriptions presented in chapter 2 (Transformational-Generative Grammar (TGG)) and in chapter 3 (Systemic Grammar), is, even according to the editor, not so clear-cut, although their relevance to linguistic theory in general is unquestionable. In fact, Proctor, after presenting her account of TGG gives three reasons why TGG cannot be used in classrooms, which is a laudable admission. She sees the lack of application of TGG in the classroom situation as resulting from the highly abstract nature of the theory, its disregard of the influence which the non-linguistic context exerts on language and its focus on a scientific description of language as an end in itself.

Proctor's treatment of thematic roles in the examples provided by her is somewhat unorthodox and I find myself in disagreement with her assignment of thematic roles. Her analysis of the sentence *Eliot drove her to hospital* is as follows (p. 55):

[Eliot] drove	[her]	to [hospital]
AGENT	BENEFACTIVE	SOURCE

While it has to be admitted that syntactic and semantic analyses are occasionally tinged with subjectivity (grammaticality judgement is perhaps the most salient example), the following assignment of thematic roles appears less controversial:

[Eliot] drove	[her]	to [hospital]
AGENT	THEME/PATIENT	GOAL

Steven Dodd's introductory account of Systemic Grammar presented in chapter 3 is packed with information and for that reason is not easily digestible for a beginning student of linguistics. It also contains a questionable exposition. On page 86, the author contradicts himself in the description and exemplification of the imperative mood, which, he says, is further divisible into **inclusive**, as in the example *Open the door!*, or **exclusive**, as in *Let's open the door*. The summing up of realisation rules at the bottom of the page presents us with an entirely different subdivision, namely: mood imperative inclusive contains the word *let*; mood imperative exclusive appears without it.

Rufus Gouws' chapter on aspects of lexical semantics offers a very informative introduction to the study of lexicon and sense relations holding between lexical items. However, bearing in mind the intended audience, one cannot shake off the impression that the wealth of factual information and new terminology can be overwhelming for a linguistic novice. For example, in his discussion of dictionaries, Gouws remarks that "[...] a dictionary has to convey a variety of information types, including information on the lexical meaning of the lemmas" (p. 100). While it has to be acknowledged that the term **lemma** has been in use for some time (e.g. Zgusta 1971, Levelt 1989), it has not been universally accepted by lexicographers and semanticists. Standard introductory textbooks in linguistics, like O'Grady *et al.* (1991) and Fromkin and Rodman (1988) also avoid the term. Therefore, it would be beneficial for the prospective student to have contentious elements of linguistic jargon explained before embarking on discussions of a more advanced level.

The chapter under discussion contains another potentially controversial point. In the section devoted to relations between lexical items and extralinguistic properties, Gouws offers an explanation of **reference** as the "relationship between an expression and its referents in the extralinguistic world". Further in the text, **denotation** is defined as the "relationship between a lexical item and the places, things, properties, persons, processes and activities outside the language system [...]" (p. 103). Given that a lexical item can also be an expression, these definitions do not seem to distinguish the two terms. Indeed, denotation is naturally connected with reference and some authorities do not draw any distinction between them. However, having these two terms discussed under separate headings might require some further elaboration. Perhaps it will be helpful to add that denotation belongs to a lexical item independently of its use on particular occasions of utterance, whereas the reference

of a referring expression is usually context-dependent.

The nature of this volume allows for a certain degree of proliferation of specialised linguistic terms but lack of adequate explanation can make the text impenetrable and unduly hard to read.   Ultimately, one has to admit that it is not the link between linguistics and applied linguistics that leaves the reader with unanswered questions, but the uneven level of presentation and difference in the assumed level of knowledge.

The bottom line, however, is how much the book meets what is seeks to do at the outset, i.e. explain the nature of linguistics, its relevance and contribution to applied linguistics understood as "dealing with the practicalities of language".   In this respect, *Solving Language Problems* is to be commended because of its contributors' dedicated and largely successful attempts to make the discipline more approachable and relevant. The book contains a wealth of information spanning a broad range of issues. The material presented is — on the whole — up to date, and despite stylistic and methodological heterogeneity, students with diverse academic interests and a common interest in linguistics are certain to find this volume a useful and valuable contribution.

References

Fromkin, V. and R. Rodman. 1988⁴. *An Introduction to Language*. New York: Holt.

Levelt, W.J.M. 1989. *Speaking: From Intention to Articulation*. Cambridge, Mass.: MIT Press.

O'Grady, W., M. Dobrovolsky and M. Aronoff. 1991². *Contemporary Linguistics: An Introduction*. New York: St. Martin's Press.

Zgusta, L. 1971. *Manual of Lexicography*. The Hague: Mouton.

Ewa Clayton
Department of Linguistics
University of the Witwatersrand
Wits
South Africa

Hausmann, F.J., O. Reichmann, H.E. Wiegand en L. Zgusta (Reds.).
Wörterbücher: Ein internationales Handbuch zur Lexikographie / Diction-
aries: An International Encyclopedia of Lexicography / Dictionnaires: Encyclo-
pédie internationale de lexicographie. Deel 1-3. 1989-1991, LII + 3355 pp.
(Handbücher zur Sprach- und Kommunikationswissenschaft, Band 5.1-
5.3) ISBN 3-11-009585-8. Berlyn / New York: Walter de Gruyter. Prys
Deel 1, DM740,00; Deel 2, DM880,00; Deel 3, DM780,00.

1.   Inleiding

Die saamstel van woordeboeke is 'n praktyk wat meer as 4 000 jaar oud is, en
vir sover woordeboeksamestellers hulle in hierdie praktyk deur enige soort
oorweging (van watter aard ook al) laat lei het, het die metaleksikografie seer
sekerlik 'n ewe lang geskiedenis.

Wat die moderne metaleksikografie betref, was die verskyning van Ladis-
lav Zgusta se nou al bekende *Manual of Lexicography* in 1971 'n hoogtepunt
omdat dit die eerste handboek van formaat oor die metaleksikografie was wat
die lig gesien het. As sodanig het dit internasionale aansien gekry en die basis
gevorm van praktiese opleidingskursusse vir leksikograwe en akademiese kur-
susse oor die leksikografie.

Sowel die praktiese leksikografie as die metaleksikografie het sedertdien
'n ontploffing beleef. Eersgenoemde is duidelik aan die ongelooflike groot aan-
tal woordeboeke wat die lig gesien het (en daagliks aangroei) en die groot mate
van variasie in woordeboektipes en -formate. Dis 'n bedryf wat daagliks inter-
nasionaal op die gevarieerde talige inligtingsbehoeftes van taalgebruikers rea-
geer, en, in baie gevalle, sulke behoeftes antisipeer (en ook skep).

Insgelyks het die metaleksikografie, en dan veral daardie vertakking daar-
van wat ondersoek instel na die teoretiese en metodologiese uitgangspunte wat
die saamstel van woordeboeke van alle tipes onderlê, 'n ongekende opbloei
beleef. Dis 'n navorsingsaktiwiteit wat gerig is op die probleme wat uit die lek-
sikografiepraktyk na vore kom en ten doel het om beter woordeboeke daar te
stel vir die uiteenlopende gebruike waarvoor verskillende gebruikersgroepe
verskillende soorte woordeboeke inspan.

Met die publikasie van hierdie driedelige en drietalige werk *Wörterbücher /*
Dictionaries / Dictionnaires wou die redakteurs 'n vakensiklopedie beskikbaar
maak wat die stand van die metaleksikografie (met die verskillende onder-
soeksterreine daarvan) op oorsigtelike wyse aan bod stel, nie net vir die stu-
dent van die leksikografie nie, maar ook vir die toenemende aantal navorsers
en praktisyns uit verskillende vakgebiede wat by die uiteindelike samestelling
van woordeboeke en hulle gebruik betrokke is.

Die verskyning van die eerste deel van die ensiklopedie in 1989 is vooraf-
gegaan deur 'n lang beplanningsfase. Die ontwerp en inhoud vir so 'n ensiklo-
pedie is reeds in 1983 deur die redakteurs bekendgestel en verskeie leksiko-

grawe en metaleksikograwe het in die tydperk daarna aan die finale konsep help formuleer. Die resultaat is 'n omvattende, driedelige werk bestaande uit 349 artikels deur 248 outeurs.

Soos die redakteurs dit in die voorwoord stel, weerspieël die geheel noodwendig ook die voorkeure van die redakteurs ten opsigte van die temas wat behandel word, die strukturering van die artikels, die keuse van outeurs en die tale waarin die artikels geskryf is. Die meer spesifiek Wes-Europese georiënteerdheid van die geheel blyk byvoorbeeld duidelik uit die keuse van tale waarin die ensiklopedie aangebied word: met die uitsondering van die voorwoord en die uitgebreide sake- / naamregister wat in Engels, Frans en Duits is, is die afsonderlike artikels in enigeen van hierdie drie tale. Soos die redakteurs dit in die voorwoord stel, is die keuse van hierdie drie tale ook daardeur gemotiveer dat 'n groot deel van die resultate van metaleksikografiese navorsing in hierdie drie tale gepubliseer word.

Uiteraard is dit onmoontlik om in 'n resensie van hierdie aard aandag te gee aan elke faset van dié omvattende werk. In die onderstaande word daar ter oriëntering van die leser dus 'n beskrywing gegee van die opbou van die ensiklopedie as geheel en daarbinne word in meer besonderhede ingegaan op die inhoud van enkele artikels oor die hooftemas wat behandel word. Ten slotte volg 'n aantal kritiese opmerkings.

2.   Algemene opbou

Die driedelige ensiklopedie bestaan uit 'n kort voorwoord (in elkeen van die drie tale), wat gevolg word deur die 349 artikels wat in 38 hoofstukke georganiseer is:

Deel 1: hoofstuk I-VII (artikel 1-100)
Deel 2: hoofstuk VII-XX (artikel 101-219)
Deel 3: hoofstuk XXI-XXXVI (artikel 220-334)

Deel 3 bevat ook 'n bibliografie van bibliografieë oor woordeboeke, 'n uitgebreide saakregister (in al drie die tale) en 'n naamregister.

Elke artikel bevat 'n kort inhoudsopgawe en word afgesluit met 'n geselekteerde bibliografie. Die artikels self wissel onderling in lengte, maar die inhoudsopgawe wat die nommering en opskrifte van die paragrawe weergee, vergemaklik die naslaan van spesifieke onderdele van enige artikel.

3.   Spesifieke temas

3.1   Aspekte van die eentalige leksikografie

In deel 1 kom in hoofsaak, maar nie uitsluitlik nie, aspekte van die eentalige

leksikografie aan bod. In hoofstukke I en II word daar in die artikels gefokus op aspekte buite om die interne struktuur van woordeboeke.

3.1.1   Woordeboeke en die publiek

Die 12 artikels in hoofstuk I belig vanuit die oorkoepelende tema "Dictionaries and their Public" 'n uiteenlopende aantal aspekte van die verhouding tussen woordeboeke en die gebruikerspubliek (bv. die normatiewe en ideologiserende funksie, die rol wat woordeboeke speel in die totstandkoming en ontwikkeling van 'n standaardtaal en gebruikers se siening van die funksie van 'n woordeboek). Die hoofstuk as geheel is tematies egter iets van 'n hutspot omdat dit byvoorbeeld ook artikels bevat oor die opleiding van die leksikograaf en verskeie oor die funksie van woordeboeke — 'n tema wat weer aangesny word in hoofstuk II.

In hoofstuk II (artikel 12-25) word, soos die hoofstuktitel "Dictionaries and their Users" aandui, gefokus op die komplekse sosiologiese profiel van die verskillende soorte gebruikers van woordeboeke, maar in hoofsaak op die verskillende funksies waarvoor woordeboeke ingespan word. Daarmee kom gevolglik twee van die belangrikste veranderlikes wat die makro- en mikrostrukturele aspekte van enige woordeboek behoort te bepaal, aan bod: die beoogde teikengroep en die funksies waarvoor die woordeboek ingespan (moet kan) word.

Die hoofstuk word ingelei met artikel 12 van R.R.K. Hartmann "Sociology of the Dictionary User: Hypotheses and Empirical Studies". In artikel 16-24 word spesifieke gebruike van woordeboeke belig, byvoorbeeld hulle gebruik in linguistiese en filologiese navorsing en in vertaling en tweede- / vreemdetaalonderrig. 'n Interessante artikel (artikel 13) is dié van Peter Kühn waarin 'n woordeboektipologie — die hooftema van hoofstuk VI e.v. — gegrond op die funksies van woordeboeke aan bod gestel word. Hierdie artikel dien dus ook as 'n inleiding en aanvulling tot die daaropvolgende artikels waarin die verskillende funksies elk afsonderlik bespreek word. Die hoofstuk word afgesluit met artikel 25 van Henri Béjoint oor die onderrig van woordeboekgebruik — 'n artikel wat as hoeksteen kan dien vir enige taalgebruikskursus op skool of op tersiêre vlak.

Hoofstuk III (artikel 26-35) dra die titel "History and Theory of Lexicography: General Aspects". Wat inhoud betref, is hierdie hoofstuk ook ietwat van 'n hutspot vir sover daaronder 'n aantal tematies uiteenlopende artikels gegroepeer is. Slegs die eerste drie van hierdie artikels het 'n historiese perspektief, nl. dié van F.J. Hausmann oor 'n aanloop tot die beskrywing van die geskiedenis van die metaleksikografie, dié van N.E. Osselton oor die geskiedenis van akademiese kritiek op woordeboeke en dié van Oskar Reichmann oor die geskiedenis van leksikografiese projekte in Duitsland. Gesien die feit dat daar later uitgebreid aandag gegee word aan die eentalige leksikografie van afsonderlike tale, is dié artikel van Reichmann hier ietwat misplaas.

In die res van die artikels word 'n uiteenlopende aantal metaleksikogra-
fiese kwessies behandel: die huidige status van die leksikografie en die verhou-
ding waarin dit tot ander dissiplines staan, die onderskeid tussen die linguis-
tiese woordeboek, ensiklopedie en die ensiklopediese woordeboek, die begin-
sels van die eentalige leksikografie, die metataal / terminologie van die leksi-
kografie en die verskillende vorme wat woordeboeke vandag aanneem (ge-
drukte boek, mikrofiche en die elektroniese woordeboek).

3.1.2   Die makro- en mikrostrukturele kenmerke van eentalige woordeboeke

Hoofstuk IV (artikel 36-67), "Components and Structures of Dictionaries",
beslaan 'n groot deel van die eerste deel, en ten regte omdat daarin aandag
gegee word aan die makro- en mikrostrukturele aspekte van die eentalig-ver-
klarende woordeboek.

Die hoofstuk word ingelei met 'n oorsigartikel (artikel 36) deur F.J. Haus-
mann en H.E. Wiegand oor die verskillende struktuurkomponente van eenta-
lige woordeboeke, en dan gevolg deur drie ander (artikel 38, 38a en 39) waarin
aspekte van die begrippe "makrostruktuur" en "mikrostruktuur" ontleed word.
Hierdie twee strukturele begrippe is nou al goed gevestig in die metaleksiko-
grafie en hulle vorm saam met veranderlikes soos die aantal tale, teikengebrui-
kers en funksies van woordeboeke die vernaamste parameters in terme waar-
van woordeboeke individueel ontleed en tipologies van mekaar onderskei
word.

In artikel 37 (ietwat misplaas tussen die eerste vyf) bring Werner Wolski
die lemma en verskillende lemmatipes aan bod, en vanaf artikel 40 word daar
dan sistematies aandag gegee aan alle aspekte rakende die mikrostrukturele
inligting wat in eentalige woordeboeke aangegee word: ortografie, uitspraak,
(fleksie)morfologie, betekenis, sintaktiese inligting, die hantering van vaste uit-
drukkings, sinonieme en antonieme, homonieme en paronieme, e.s.m. Die
reeks leksikografiese etikette wat in eentalige woordeboeke gebruik word,
word vanaf 'n oorsigartikel (artikel 53) deur Hausmann in 'n reeks artikels be-
spreek. Aandag word gegee aan die gebruik van nieverbale inligting (byvoor-
beeld illustrasies) en bibliografiese verwysings in artikels.

Artikel 63 tot 67a behandel aspekte van die voor- en agterwerk van woor-
deboeke, soos die insluiting van 'n grammatikadeel, gebruikershandleiding,
ander linguistiese en ensiklopediese inligting wat nie in woordeboekartikels
van die hoofdeel opgeneem is nie en die register. Wat die struktuurelemente
van die eentalig-verklarende woordeboek betref, bied die ensiklopedie dus 'n
omvattende behandeling. In al die artikels word daar breedvoerig gebruik
gemaak van illustrasiemateriaal uit 'n verskeidenheid woordeboektipes en in
verskeie van hulle word ook aandag gegee aan die funksies van elke inligtings-
tipe, enkoderingsformate, plasing in die woordeboek, integrasie met ander
inligting / inligtingstipes en spesifieke beskrywingsprobleme wat elkeen ople-
wer.

Die artikels van hoofstuk V "Selected Problems of Description in the General Monolingual Dictionary" (artikel 69-90a) benader die makro- en mikrostrukturele aspekte van die eentalige woordeboek vanuit 'n ander hoek, nl. dié van die beskrywingsprobleme wat ondervind word by die saamstel van hierdie tipe woordeboek. Daar word onder andere aandag gegee aan probleme met die keuse van lemmas vir opname, die beskrywingsprobleme wat homonimie en polisemie bied, die beskrywing en hantering van spesifieke lemmatipes (bv. modale partikels), lemmas van spesifieke registers (bv. omgangswoorde) of spesifieke semantiese gebiede (bv. die beskrywing van emosiewoorde). Sake wat ook aandag kry, is die benadering van leksikografiese definisies vanuit die raamteorie, die ordening van definisies en vorme van teksverdigting.

In sekere opsigte is daar 'n mate van (m.i. onvermydelike) inhoudelike oorvleueling tussen die artikels van hoofstuk IV en V, maar in die meeste gevalle word daar in (die beste van) die artikels van hoofstuk V op geselekteerde beskrywingsprobleme en die gekose oplossings gekonsentreer. Op hierdie wyse kry die gebruiker dus ook insig in die teoretiese en praktiese oorwegings wat die spesifieke aanbieding van die makro- en mikrostrukturele inligting in woordeboeke onderlê.

By die ondersoek na enige aspek van óf die makro- óf die mikrostruktuur van die eentalige woordeboek sal die navorser die grootste voordeel uit die inligtingsaanbod in die ensiklopedie trek deur na te gaan of 'n spesifieke tema wat in hoofstuk IV aan bod kom, nie ook vanuit die oog van die beskrywingsproblematiek in hoofstuk V behandel word nie.

Die omgekeerde is ook nodig omdat sommige artikels in hoofstuk IV juis aandag gee aan die beskrywingsprobleme ten opsigte van elemente van die makro- of die mikrostruktuur van eentalige woordeboeke. 'n Voorbeeld hiervan is artikel 46 waarin veral op die beskrywingsprobleme in verband met die behandeling van fraseologismes (vaste uitdrukkings) gekonsentreer word.

Betekenisbeskrywing/-verklaring, byvoorbeeld, is een van die kernaangawes in eentalig-verklarende woordeboeke en 'n aspek waarmee leksikograwe heelwat probleme ondervind. In die ensiklopedie kan die gebruiker 'n hele aantal artikels hieroor raadpleeg: artikel 30 van L.F. Lara oor die onderskeid tussen die linguistiese woordeboek, ensiklopedie en die ensiklopediese woordeboek, artikel 44 van H.E. Wiegand oor die leksikografiese definisie ('n hoogs tegniese en verteoretiseerde behandeling), artikel 69 van Ekkehard Zöfgen oor die hantering van polisemie, 'n hele aantal artikels wat (onder meer) handel oor die definiëring van leksikale items wat tot bepaalde funksionele of semantiese klasse behoort (vgl. die artikels in hoofstuk V) en artikel 84 van Immo Wegner waarin leksikografiese definisies vanuit die raamteorie benader word.

Al die artikels in hierdie twee hoofstukke is egter nie van gelyke gehalte nie. 'n Illustrasie van die soort "mindere" artikel wat die leser sal aantref, is artikel 71 van Günther Drosdowski oor die hantering van metaforiese uitdrukkings in die eentalige woordeboek. In die artikel word daar kortliks 'n onder-

skeid getref tussen enersyds metaforiese uitdrukkings en die probleem van hulle seleksie vir opname, en andersyds metaforiese prosesse en die neerslag daarvan in die polisemiese onderskeidings van leksikale items. Enkele voorbeelde van die wyse waarop hierdie twee aspekte in Duitse woordeboeke gehanteer word, word ontleed en die artikel word dan afgesluit met 'n lys geselekteerde bronne. Nóg in die artikel, nóg in die bronnelys vind 'n mens egter iets van die leksikografiese navorsing oor hierdie twee aspekte van metaforiek terug wat deur die baie bekende werk van G. Lakoff en M. Johnson *Metaphors we Live by* (1980) gestimuleer is. Tewens, met die uitsondering van Zgusta se *Manual of Lexicography*, word daar in die geselekteerde bronnelys slegs Duitse bronne verskaf.

3.1.3 Woordeboektipologie

Vanaf hoofstuk VI in deel 1 word 'n tema behandel wat oorloop na 'n groot deel van deel 2 toe, nl. die tipologisering van die magdom uiteenlopende woordeboeke wat op die mark beskikbaar is. In hoofstuk VI word eerstens in 'n paar artikels 'n aantal algemene aspekte van woordeboektipologieë bespreek. Die belangrikste artikel hier is artikel 91 van Hausmann waarin aandag gegee word aan die begrip "tipologie", adekwaatheidskriteria vir tipologieë, die verskillende onderskeidingskriteria waarop woordeboektipologieë berus en uiteindelik ook 'n uiteensetting van die beginsels wat in die ensiklopedie self met die tipologisering van die onderskeie soorte woordeboeke gevolg is.

In hoofstuk VII word verskillende soorte "sintagmatiese" woordeboeke behandel (bv. woordeboeke van sintaktiese patrone, kollokasies en idiomatiese uitdrukkings). Die tipologisering van woordeboeke word dan uitgebreid voortgesit in deel 2 van die ensiklopedie:

—   hoofstuk VIII (artikel 101-114): paradigmatiese woordeboeke (bv. onomasiologiese woordeboeke, woordeboeke van sinonieme en antonieme);
—   hoofstuk IX (artikel 115-125): woordeboeke van lemmatipes wat in standaardwoordeboeke in een of ander opsig geëtiketteer word (bv. woordeboeke van argaïsmes, neologismes en leenwoorde);
—   hoofstuk X (artikel 126-135): woordeboeke van ander lemmatipes as die voorafgaande (bv. woordeboeke van doeblette, morfeme en afkortings);
—   hoofstuk XI (artikel 136-139): onomastiese woordeboeke;
—   hoofstuk XII (artikel 140-147): woordeboeke wat spesifieke soorte inligting verskaf (bv. oor ortografie en frekwensie);
—   hoofstuk XIII (artikel 148-152): aanleerderswoordeboeke;
—   hoofstuk XIV (artikel 153-162): woordeboeke wat inligting oor spesifieke variëteite van 'n taal bevat; en
—   hoofstuk XV (artikel 163-166): woordeboeke wat op spesifieke tekste gerig is (bv. die konkordans).

Soos uit die voorgaande duidelik is, word feitlik elkeen van die bekende woordeboeksoorte aan die orde gestel. In die afsonderlike artikels word daar dan aandag gegee aan aspekte soos die makro- en mikrostrukturele kenmerke van elke tipe, hulle tipiese gebruike en teikengebruikers, en hierdie kenmerke word met voorbeelde uit 'n verskeidenheid woordeboeke van 'n bepaalde tipe geïllustreer.

Die aanleerderswoordeboek is byvoorbeeld 'n woordeboektipe waaroor daar die afgelope aantal jare heelwat navorsing gedoen is en waarvan verskeie die lig gesien het. Alhoewel artikel 151, "Dictionaries for Foreign Language Teaching: English", van Thomas Herbst gerig is op woordeboeke vir die aanleer van Engels, bied die artikel 'n goeie oorsig van die makro- en mikrostrukturele kenmerke van hierdie woordeboektipe. In die artikel word daar byvoorbeeld aandag gegee aan die algemene kenmerke van die verskillende aanleerderswoordeboeke wat in die handel beskikbaar is, en dan word in meer besonderhede ingegaan op die volgende mikrostrukturele kenmerke: definisies, voorbeelde, sintaktiese inligting, illustrasies, gebruiksaantekeninge en uitspraak.

In samehang met die ander artikels in hierdie hoofstuk wat aanleerderswoordeboeke in spesifieke tale behandel, sal die gebruiker 'n goeie beeld kry van die soort variasie wat hierdie woordeboektipe kenmerk ten opsigte van formaat, makro- en mikrostruktuur, inligtingstipes wat in die voor- en agterwerk ingesluit word, gebruik in samehang met grammatikas en die integrasie van die woordeboek by ander onderrigmateriaal.

In die voorwoord tot die ensiklopedie stel die outeurs dit duidelik dat daar nie veel aandag aan die vakleksikografie gegee is nie, hoofsaaklik vanweë die huidige (ontoereikende) stand van navorsing op hierdie gebied. Sedert die verskyning van die ensiklopedie het navorsing op hierdie gebied egter al toegeneem. Die ensiklopedie self bied vir die gebruiker wel byvoorbeeld iets meer oor die probleme met die opname en beskrywing van terminologie in die eentalig-verklarende woordeboek (vgl. bv. artikel 76 en 77 deur Walther Dieckmann en Günther Kempcke oor die politieke en vaktegniese taal onderskeidelik) en van verskillende soorte vakwoordeboeke vir die kenner en die leek. Artikel 159 deur Kurt Opitz behandel byvoorbeeld die tegniese woordeboek vir die kenner en artikel 160 deur Hartwig Kalverkämper die vakwoordeboek vir die leek. In artikel 170 skets Opitz die probleme wat in terminografiewerk met dataversameling ondervind word. In hierdie opsig bied die ensiklopedie vir belangstellendes 'n goeie oriëntering.

In hoofstuk XVI van deel 2 word daar in agt artikels onder die tema "Procedures in Lexicographical Work" in hoofsaak aandag gegee aan die rekenarisering van alle prosedures en prosesse in die saamstel en uiteindelike produksie van 'n aantal woordeboeke. Hierdie hoofstuk kom ietwat misplaas voor, aangesien dit tematies aansluit by sommige van die aspekte wat in hoofstuk V van deel 1 oor beskrywingsprobleme handel.

3.1.4   Die eentalige leksikografie van die verskillende tale

Die res van deel 2 en 'n groot deel van deel 3 van die ensiklopedie (hoofstuk XVII tot XXXII) word gewy aan 'n bespreking van die eentalige leksikografie van die afsonderlike tale van die wêreld (bv. van die klassieke tale, feitlik al wat 'n moderne taal is en van die tale met klein gebruikersgetalle). Soos 'n mens sou kon verwag, het die seleksie van die tale (veral die uitsluiting van sekere tale) en besluite oor die presiese omvang van die beskrywing van die leksikografiewerk oor elke taal die redakteurs heelwat hoofbrekens besorg (vgl. in hierdie verband die voorwoord tot die ensiklopedie). In die meeste gevalle is die omvang van die artikels egter bepaal deur die omvang van die leksikografiewerk in 'n spesifieke taal, wat die uitgebreide aandag aan die leksikografieë van byvoorbeeld Duits, Engels en Frans sou verklaar.

Die Afrikaanse leksikografie kry in hierdie omvangryke deel van die ensiklopedie dekking in ongeveer 'n enkele bladsy, waar dit onder die groep Germaanse tale saam met die Nederlandse leksikografie aangebied word. Vanselfsprekend word daar in hierdie enkele bladsy nie veel meer as 'n kort historiese oorsig gebied en van die vernaamste woordeboeke genoem nie. Al woordeboeke wat meer uitgebreide behandeling kry, is die *Woordeboek van die Afrikaanse Taal (WAT)* en *Handwoordeboek van die Afrikaanse Taal (HAT)*. Vir verdere inligting oor die woordeboeke wat in meertalige lande vir elkeen van die tale bestaan, sal die gebruiker sy weg moet vind met behulp van die saakregister.

3.2   Twee- / meertalige leksikografie

Die twee-/meertalige leksikografie is 'n navorsingsterrein wat die afgelope aantal jare 'n sentrale plek in leksikografiese navorsing ingeneem het. In die ensiklopedie begin hierdie tema met hoofstuk XXXIII, waarin aandag gegee word aan die beginsels en komponente van die twee-/meertalige leksikografie. Soos in die geval van die ander temas, word ook hierdie een ingelei met 'n oorsigartikel waarin algemene aspekte van die teikengroepe en funksies van hierdie soort woordeboeke geskets word (vgl. artikel 285 "Principles of Bilingual Lexicography" deur Hans-Peder Kromann, Theis Riiber en Poul Rosbach). In ooreenstemming met die behandeling van die eentalig-verklarende woordeboek, word dit dan opgevolg deur 'n aantal artikels waarin spesifieke aspekte van die makro- en die mikrostruktuur van hierdie woordeboektipe in meer besonderhede bespreek word (vgl. in hierdie verband veral artikel 286 van Hausmann, wat van 'n uitgebreide bronnelys voorsien is en besonder bruikbaar is, nie net as oriëntering nie, maar ook vir verdere navorsing).

Artikels waarin in meer besonderhede op die mikrostrukturele elemente van hierdie woordeboektipe ingegaan word, is artikel 287 tot 292. Aspekte wat aandag kry, is onder meer grammatiese konstruksies, kollokasies, die gebruik van voorbeelde, die aangee van paradigmatiese semantiese relasies en die gebruik van etikette.

Van die vernaamste beskrywingsprobleme wat met die saamstel van twee-/meertalige woordeboeke ervaar word, is die onderwerp van hoofstuk XXXIV. Die probleme wat van artikel 293 tot 302 behandel word, is onder meer die volgende: dataversameling en strukturering, lemmatisering, ekwivalensie-verhoudinge, grammatiese en leksikale kategorisering, die mikrostrukturele organisasie van 'n woordeboekartikel, vorme van teksverdigting, die kontrastiewe linguistiek as uitgangspunt vir tweetalige leksikografie, die metataal wat gebruik word, die probleme wat kulturele verskille tussen tale oplewer en die beskrywing van pidgin- en kreoolse tale.

In ooreenstemming met die opbou van die hoofstukke en artikels oor die eentalige leksikografie, word hierdie beskrywingsprobleme gevolg deur 'n bespreking van die verskillende soorte twee-/meertalige woordeboeke. Hausmann gee in artikel 303 kortliks 'n tipologie, en dié word gevolg deur 'n bespreking van spesifieke woordeboeksoorte: die "dictionary of false friends", tweetalige aanleerderwoordeboeke en die reiswoordeboek.

Artikel 306 oor meertalige woordeboeke en die probleme wat hulle oplewer, lei as 't ware 'n nuwe subtema in. In hierdie omvangryke artikel deur Günther Haensch word daar 'n historiese skets van die ontwikkeling van hierdie woordeboektipe verskaf, en dan word kortliks ingegaan op die funksies en makro- en mikrostrukturele kenmerke daarvan. Die probleme wat die gebrek aan (direkte) ekwivalensie tussen meerdere tale vir die leksikograaf lewer, word ook kortliks aangespreek. Die artikel is van 'n uitgebreide bronnelys voorsien wat verdere navorsing oor hierdie woordeboektipe sal vergemaklik.

Dieselfde soort behandeling (vgl. artikel 307 ook deur Haensch) word van die tweetalige vakwoordeboek verskaf, alhoewel daar nie in soveel besonderhede as in die geval van die meertalige woordeboek ingegaan word nie.

Die tweetalige leksikografie van spesifieke tale is die onderwerp van die artikels van hoofstuk XXXVI (artikel 308 tot 328). Meestal word daar in hierdie artikels 'n kort historiese skets van die ontwikkeling van die tweetalige leksikografie van die betrokke taal verskaf en spesifieke woordeboeke na hulle funksies, vertaalrigting en makro- en mikrostrukturele kenmerke bespreek. Saam met die artikels van hoofstuk XVII tot XXXII oor die eentalige leksikografie van die afsonderlike tale, sal die gebruiker van die ensiklopedie 'n omvattende beeld van die stand van die leksikografie in enige van die bespreekte tale kan rekonstrueer.

Die artikelreeks van die ensiklopedie word afgesluit met hoofstuk XXXVI bestaande uit 'n aantal artikels oor woordeboeke van hulptale (bv. vir die dowes) en ander kommunikasiestelsels.

4.   Slotsom

Uit die voorgaande behoort dit duidelik te blyk dat die ensiklopedie, met die uitsondering van die navorsing oor woordeboekkritiek, vir die leser 'n baie

goeie oorsig oor die belangrikste aspekte van metaleksikografiese navorsing bied. Soos 'n mens van 'n vakensiklopedie kan verwag, is die meeste artikels daarop gemik om die leser (leek, student, navorser) ten opsigte van die vernaamste navorsing te oriënteer, probleme te identifiseer en leemtes aan te dui. Met die uitgebreide bronnelyste aan die einde van elke artikel word die navorser se behoeftes ook aangespreek, vir sover dit hom / haar leiding vir verdere navorsing verskaf.

Soos reeds vermeld, is al die artikels nie van gelyke gehalte nie — iets wat verstaanbaar is as die hoeveelheid skrywers in ag geneem word wat by die projek betrokke was. Sommige van die artikels is niks meer as 'n oppervlakkige verkenning van 'n tema nie, in welke geval die inligtingsoeker noodwendig die genoemde werke in die bronnelys by elke artikel en die bronnelys aan die einde van die ensiklopedie sal moet raadpleeg.

'n Opvallende eienskap van die ensiklopedie is die bydrae van twee van die redakteurs, nl. Hausmann en Wiegand. Nie net het hulle self 'n hele aantal van die artikels geskryf nie, maar ook in talle van die artikels, in die besonder dié in Duits, vorm hulle metaleksikografiese beskouings die sentrale uitgangspunt by die bespreking van die makro- en mikrostrukturele aspekte van die eentalige en die twee-/meertalige woordeboeke.

Die drietalige benadering wat gevolg word, sal vir talle Suid-Afrikaanse gebruikers van die ensiklopedie probleme oplewer omdat Duits en Frans vir die meeste nie 'n tweede of derde taal is nie. Dit is egter wel so dat van die inleidende artikels tot die verskillende temas in Engels is en dat die gebruiker via die verwysings na Engelse bronne in die artikels wat in Frans en Duits is, verdere opleeswerk sou kon doen.

'n Besonder hinderlike aspek van die organisasie van die ensiklopedie is die feit dat twee van die hooftemas in een deel begin en dan in 'n ander voortgesit word. Dit dwing die voornemende koper om al drie die dele aan te skaf ('n stewige uitgawe). Dit sou miskien sinvoller gewees het om alle metaleksikografiese aspekte rakende die eentalige leksikografie (insluitende die bespreking van spesifieke beskrywingsprobleme) in een deel af te handel en om 'n tweede te wy aan die twee-/meertalige leksikografie (met heelwat verdere uitdieping van die tema as wat tans die geval is). Die derde deel sou dan gewy kon gewees het aan 'n oorsig van die een- en twee-/meertalige leksikografie van die afsonderlike tale. So 'n organisasie sou dit moontlik maak om die eerste twee dele as handboeke vir studente te kon gebruik (veral ook as al die artikels in Engels is). In die huidige formaat, en veral vanweë die prys, sal min leksikograwe, studente en navorsers hierdie werk self kan bekostig.

Sowel die praktiese as die teoretiese leksikografie is vandag' vinnig ontwikkelende gebiede, wat onvermydelik inhou dat die ensiklopedie voortdurend bygewerk sal moet word. Dit geld nie net die bronnelyste nie, maar ook die navorsingsgebiede wat die afgelope aantal jare besondere aandag geniet het. Voorbeelde hiervan is die massa literatuur oor navorsing oor en gebruik van die rekenaar in die leksikografiepraktyk, die ontwerp en saamstel van

leksika in elektroniese formaat, die integrasie van leksika met teksmateriaal (veral onderrigmateriaal), die gebruik van hiperteks en die multimediategnologie en veral die ontginning van groot korpusse en die insigte van die korpuslinguistiek. In Europa is daar byvoorbeeld ook groot navorsingsprojekte die afgelope aantal jare geloods waaroor nie in die ensiklopedie gerapporteer word nie.

Hierdie punte van kritiek neem egter nie weg dat die ensiklopedie as 'n vakensiklopedie vir die grootste deel in sy doel slaag nie, nl. om die leek, student en navorser op die gebied van die leksikografie te oriënteer en om hom / haar, veral ook via die uitgebreide bronnelyste, die rigting vir 'n diepere verkenning te wys. Vir enige studie van die metaleksikografie is hierdie 'n onmisbare bron van inligting.

Piet Swanepoel
Departement Afrikaans
Unisa
Suid-Afrika

G.H. Kocks (Samesteller), J.P. Vording, A. Beugels, H. Bloemhof, e.a. (Medewerkers). *Woordenboek van de Drentse Dialecten, A-L*, 1996, LXIX + 706 pp. ISBN 90-232-3176-7. Assen: Van Gorcum. Prys ƒ87,50. Intekenprys ƒ69,50.

Hierdie woordeboek van die dialekte in die Nederlandse provinsie Drente is gerig op die behoeftes van die breë publiek en nie net op taalkundiges en dialektoloë nie. Die uitgangspunt was dat dit ook toeganklik moes wees vir wetenskaplikes uit volkskundige, historiese en kultuurhistoriese gebiede, asook vir niewetenskaplikes uit die streek van Drente en ander belangstellendes. Die betekenisomskrywings is derhalwe in die standaardtaal aangegee en is nie teen die agtergrond van die een of ander semantiese teorie gedoen nie.

Die woordeboek is 'n opgawe van Drentse dialekwoorde met 'n aanduiding van die lokaliteit en die semantiese waarde daarvan. Talle voorbeeldsinne is ook opgeneem om die gebruik van die woorde te illustreer. Geen aanduiding is gegee in verband met uitspraak, etimologiese aspekte en saak- en volkskundige aspekte nie.

In die voorwoord is 'n duidelike uiteensetting van hoe te werk gegaan is om die woordeboek saam te stel, wat die aard van die woordeboek is en watter dialekte gedek word met enkele kaarte ter illustrasie. Daar is ook goeie riglyne vir die gebruiker.

Die periode wat hierdie woordeboek dek, strek vanaf 1800 tot 1950. 'n Deel van die inligting is verkry uit vroeë versamelings van woordelyste en amateurwoordeboeke waarvan die oudste dateer uit 1839. Die woordeskat is verder uit boeke en tydskrifte verkry en sedert 1969 is materiaal ook mondeling ingesamel deur veldwerkers.

Die werk aan die woordeboek is in September 1969 begin toe verskillende medewerkersgroepe gevorm is, nadat die ervaring geleer het dat groepe medewerkers beter funksioneer as individuele werkers. Die ongeveer 600 entoesiastiese medewerkers was versprei oor die sewe dialekgebiede wat in Drente onderskei is. Hulle het hul dialeknavorsing vanuit twee kerndorpe in elke gebied geloods.

Die metode was om met vrae op kaartjies te werk. Die vrae is op verskillende maniere aan die dialeksprekers gestel met die doel om sinne as antwoorde te kry. Dit het die leksikograwe in staat gestel om voorbeeldsinne by die meeste betekenisse te gee.

Die dialekte

Die Drentse dialekte is Nedersaksiese dialekte, maar tegelyk ook Nederlandse dialekte, omdat hulle binne die Nederlandse taalgebied lê. Hierdie dialekte is bloot geografies afgebaken. Wat die aard van die dialekverskynsels betref, hoort baie verskynsels eerder tuis by dialekte buite die gebied van Drente. Sommige dialekgrense kon maklik bepaal word. Daar is byvoorbeeld 'n duide-

like skeidslyn tussen die suidweste en die suidooste, en in Oos-Drente kon maklik onderskei word tussen een gebied met veengrond en 'n ander met sandgrond. Ook ander veengebiede kon maklik geïdentifiseer word op grond van klank-, morfologiese en leksikale verskille in die woordeskat. Daar is egter nie oral duidelike dialekgrense nie en daarom was dit nie maklik om op 'n werklik verantwoorde dialekindeling te besluit nie.

Verder is dit moeilik om tussen dialekvorme en Standaardnederlandse vorme te onderskei, omdat 'n definitiewe skeidslyn tussen die Drentse dialekte en Standaardnederlands nie moontlik is nie. Omdat die dialeksprekers steeds meer in aanraking met Standaardnederlands kom, is daar 'n sterk wisselwerking tussen dialek en standaardtaal. Hierdeur ontstaan daar 'n derde vorm, nl. 'n tussenvorm met elemente uit beide dialek en standaardtaal. Die tussentaal kan ook as dialek beskou word, omdat dit afwyk van die standaardtaal. Dit was egter moeilik om te besluit wat in die woordeboek opgeneem moet word.

Die dialekwoordeskat

Dialekwoordeskat is in 'n hoë mate gebaseer op dit waarmee die sprekers daagliks te doen het. Daarom hang hierdie provinsie se dialekwoorde ook nou saam met die verskillende omstandighede wat in elke gebied heers. Die materiaal van hierdie woordeboek het betrekking op algemene en alledaagse sake. Dit is opvallend dat sekere verskynsels in die provinsie se dialekte nou verband hou met die verskillende grondtipes van Drente. Groot dele is met veengrond en ander met sandgrond bedek. Dit word weerspieël, nie alleen op leksikale gebied nie, maar ook op fonologiese gebied. In die veengebiede kom rekking van die a en van kort vokale in geslote lettergrepe meer voor as in die sandgebiede.

Die meeste woorde kom uit die hoofwoordsoorte, nl. selfstandige naamwoorde, werkwoorde, byvoeglike naamwoorde en bywoorde. Omdat die beroep van die mans oor hierdie tydperk meestal op die gebied van die landbou, soos byvoorbeeld die veenbedryf of veeboerdery was, of op die gebied van die een of ander ambag het die woordeboek baie woorde uit hierdie terreine. Daar is byvoorbeeld baie terme in die woordeboek opgeneem van die hoefsmid, die kleremaker, die klompemaker, die skoenmaker, die kuiper, die loodgieter, die messelaar, die meulenaar, die slagter en die timmerman. Die vroue was in hierdie tydperk weer meestal huisvroue en naaldwerksters. Die woordeskat lê verder op die gebied van maatskaplike verhoudinge, segswyses, spreekwoorde, liedjies en gebruike.

Die woordeboekartikel

Die volledigste woordeboekartikels van hierdie woordeboek onderskei drie hoofrubrieke, nl. (a) die trefwoord met grammatikale informasie en die dialek-

streek waarin dit aangetref word, (b) die variante en die streke waarin dit voor-
kom en (c) die betekenisdeel en voorbeeldsinne met die plek waar dit opgete-
ken is.

Die artikel **hantam** I sien byvoorbeeld soos volg daar uit: I de <-men>
(ZOZ, MD, KVD, ZWZ). Ook **antam** (MD), **haantam** (ZW), **haantamme** (ZWN)
= 1. kind dat overal met de vingers aanzit en ongezeglijk is *Een kind dat overal
bijzit is een hantam* (And), *Ik snap niet dat ze dat jongie van heur metnimt, dat is jao
zo 'n antam* (Hijk), *Dat is zo 'n hantam van een jong* (Sle)

Indien meer as een betekenis voorkom, word elke betekenis met voor-
beelde toegelig. Daar is soms soveel as 47 voorbeeldsinne om die semantiese
onderskeidinge by 'n enkele betekenis in verskillende dorpsgebiede te illu-
streer. Ook die idiomatiese gebruik word in die voorbeeldsinne geïllustreer,
sonder dat dit as idiomaties aangedui word.

As trefwoord is telkens die dialekvorm van die dorp Sleen gekies. Daar
was verskeie redes hiervoor, onder andere dat die woordeboek in Sleen begin
is, die hoofredakteur in daardie omgewing gebore is en sy proefskrif oor die
dialekte van Suidoos-Drente geskryf het. Verder was daar uit vorige dialekna-
vorsing reeds baie materiaal uit Sleen beskikbaar. Waar 'n woord in die
Sleense dialek ontbreek, is een uit die Suidoos-Drentse sandgebied geneem en
as daar ook nie een is nie, is die vorm met die hoogste gebruiksfrekwensie
gekies.

Opvallend is verder die talle oorspronklike metaforiese en woordekono-
miese komposita wat kenmerkend is vir dialeksprekers. Voorbeelde hiervan is:

fladderspul "erg geil gras"
dunderdagskrant "nieuwsblad van het dorp, gezegd van een persoon. De
    eigenlijke krant kwam nl. maandag, woensdag en vrijdag."
kattelikkie "zeer weinig"

In sommige gevalle is reekse komposita met dieselfde eerste komponent almal
as aparte lemmas gedokumenteer, byvoorbeeld **baggelbak**, **baggelbult**, **bag-
gelgeut**, **baggelhaak**, **baggeljongs**, ens. uit die veenbedryf.

Baie woorde is alternatiewe vorme van Standaardnederlandse woorde,
soos **botter** (boter); **gutig** (guitig); **hiesig** (hees); **jaor / joor** (jaar), **liek** (lijk) en
behèurlijk (behoorlijk). Opvallend is ook die talle umlautvorme **körf** (korf);
dörst (dorst); **gres** (gras) en **hendig** (handig).

In die woordeboek is nie slegs dialektiese woorde opgeneem nie, maar
ook standaardwoorde met standaardbetekenisse, bv. **ambacht** waarby die
gelykvormige standaardekwivalent aangegee is, sonder betekenisomskrywing
met enkele voorbeelde waarin die woord in dieselfde betekenis as in die stan-
daardtaal gebruik word.

Enkele terreine waarop probleme ondervind is

(a)   Aanduiding van gebruiksomvang en -frekwensie

Indien 'n woord slegs in enkele dialekgebiede voorkom, is die gebiede genoem. Sommige woorde kom egter slegs in 'n klein gebied voor, maar oorskry verskillende dialekgrense. Die sisteem van herkomsaanduiding waarop besluit is, kan nou die wanindruk laat dat die woord oor 'n groot gebied gepraat word, omdat dit in verskillende dialekgebiede voorkom. Daar is verder woorde wat oor 'n wye gebied aangetref word, maar slegs sporadies. 'n Geografiese aanduiding toon slegs aan dat 'n bepaalde woord in 'n sekere gebied voorkom, maar toon nie die digtheid van die voorkoms aan nie. Die gebruiksfrekwensie was ook nie 'n bepalende faktor by die opname van woorde nie. In die voorwoord word vermeld dat familiewoorde of woorde wat beperk is tot 'n enkele dorp ook in aanmerking gekom het, ook enkele woorde waarvan die leksikograwe bewus was dat hulle by 'n individu ontstaan het en toe deur die radio of publikasies uitgedra is en byval gevind het by enkelinge.

(b)   Die spelling

Die spelling is hoofsaaklik gedoen in ooreenstemming met die riglyne in 'n Drentse spellingboekie wat onlangs verskyn het, nl. *Drentse spelling: Een handleiding voor de schrifwijze van de streektaal*, Assen 1987. Die uitgangspunt met spelling is dat dit so konsekwent moontlik moet wees, goed leesbaar, en sover moontlik aansluitend by die spelling van Standaardnederlands. Daar is dus nie gepoog om die spelling erg foneties aan te gee nie, want dit sou moeilik leesbaar gewees het en die leser verwar het. Dit lyk egter of hierdie beleid hier en daar gelei het tot 'n taamlik onjuiste weergawe van die uitspraak. Sch aan die begin van 'n woord word byvoorbeeld op sommige plekke as sk uitgespreek, maar daar is besluit om die spelling sch in alle gevalle te handhaaf. Dergelike uitspraakvariante word dus nie in die woordeboek aangedui nie, weens die beleid om so na as moontlik by Standaardnederlands se spelling te hou.

(c)   Datering van leksikale items

Die probleem is verder dat die woordeboek so 'n lang periode dek. Daar was in hierdie tyd groot veranderinge in die woordeskat, veral in dié opsig dat sommige woorde verdwyn het en ander bygekom het. Die woordeboek maak nie voorsiening vir die datering van leksikale items nie. Die leser weet dus nie of 'n woord nog in 1950 in die woordeskat bestaan het en of dit miskien in 'n vorige eeu al uitgesterf het nie. Chronologiese beskrywings soos dat vorige

geslagte uit **tellers** geëet het, latere geslagte uit **schöttels** en die huidige geslag uit **borde** kan dus nie in hierdie woordeboek opgeneem word nie, weens die beperkte aard daarvan.

(d)   Voorbeeldsinne

Die talle voorbeeldsinne met aanduiding uit watter dialek die betrokke woord kom, maak die woordeboek interessant. Waar die betekenis van die voorbeeldsin moeilik is om te snap, bv. in gevalle waar idiomatiese verbindings voorkom, word dit meestal verklaar, bv. *die koe is der goed an* met verklaring "dik bevlees".

Ongelukkig is daar nie altyd duidelike voorbeeldsinne nie. By **daardedrachts** ("voor de derde keer drachtig") byvoorbeeld, bevat die voorbeeldsin 'n onverstaanbare woord vir die oningewyde, nl. **mottie** *Hie verkochte mij een daardedrachts mottie.*

Die indruk wat geskep word, is dat die voorbeeldsinne nie in die eerste plek aangegee word om die betekenis van die woord verder toe te lig nie, maar eerder net as 'n bewys dien dat 'n bepaalde toepassing van die woord iewers opgeteken is. Voorbeelde hiervan kom by die volgende woorde voor: **femelkonte** ("iemand die te pas en te onpas vrome praatjes ten beste geeft") *Wat is het een femelkonte* en **fiebelefratsen** ("onzinnige praatjes, uitvluchten, gekke grappen") *Wij bint aal hiel wat fiebelefratsen tegenkommen.* Veral in die gevalle van woorde sonder verklarings, soos **bakkeran**, maar slegs met 'n streeksaanduiding en verder twee voorbeeldsinne met plekaanduiding, is dit vir 'n buitestaander onmoontlik om agter te kom wat die woord beteken uit 'n voorbeeldsin soos: *Aj een leeg nummer trökken mit 't lotten veur de dienst, dan waaj bakkeran.*

Interessante gegewens

Opvallend is die gevalle waar daar 'n Standaardnederlandse vorm en daarnaas 'n dialektiese vorm is, bv. **leeuw** versus die dialektiese **leeuwerd** in die uitdrukking **lettert of leeuwerd** (kruis of munt). Verder het baie woorde, net soos in Afrikaans, 'n Standaardnederlandse betekenis en daarnaas ook 'n dialektiese gebruik, byvoorbeeld **amper** wat eerstens ooreenstem met die standaardtaal in die betekenis van "nouliks, nog maar net", maar tweedens 'n dialektiese waarde het wat ooreenkoms toon met die Afrikaanse betekenis, nl. "byna, nog nie heeltemal nie". Nog 'n voorbeeld waar 'n woord beide 'n standaard- en 'n dialekbetekenis het, is **complot** met standaardbetekenis "sameswering" en streeksbetekenis "groep" soos in die voorbeeldsin: *Der lop 'n hiel complot van die belhamels.*

Talle interessante woorde is opgeneem, soos die klanknabootsende **klabatsen** ("vallen") *Hij leut alles op de grond klabatsen* (Hy laat alles op die grond val), klankskilderende woorde soos **fiebelefansie** ("onzinnige praatjes, uit-

vluchten, gekke grappen") en volksetimologiese vervormings soos **èendeke-lèesten** ("eindelijk en ten laatsten").

Baie persoonsname maak ook die belangstelling gaande, byvoorbeeld: **klapzeiker** vir 'n "wandelende nieuwsblad, praatjesverkoper"; **klauterboks** as benaming vir "iemand die overal op klautert"; **daorel** vir 'n "suffig, vergeetachtig persoon" en **drietkeutel** vir 'n "bangerd".

Daar word voorsien dat hierdie woordeboek van die Drentse dialekte in sy totaliteit, ten spyte van 'n tekortkoming hier en daar, 'n uitstekende naslaanbron sal wees om die Drentse dialekte te leer ken.

G.S. van der Merwe
Departement Afrikaans en Nederlands
Universiteit van Stellenbosch
Suid-Afrika

P.A. Mbenzi and O.N. Iithete. *English-Oshindonga Dictionary for Primary Schools*, 1st edition 1996, iv + 97 pp. ISBN 99916-37-33-8. Windhoek: Out of Africa. Price N$39,90.

1.   Introduction

I cannot over-emphasize the critical importance of this book.

Although the above statement by Tshali Iithete in the preface of this dictionary might be considered as somewhat extravagant, it is without a doubt true that the dictionary does make a valuable contribution to the language learning aids available to primary school learners and teachers in Namibia. Since English is the only medium of instruction in government schools from the junior secondary phase onwards, it is vital that learners should have a reasonable command of the language by the time they enter grade 8. This dictionary works towards this aim, as is stated in the preface:

> The dictionary provides a solid first start in building up a strong English vocabulary and eventually grammar.

With this purpose in mind, the dictionary does have certain shortcomings, the removal of which would greatly extend the usefulness of this work, even into the junior secondary phase. Some of the shortcomings, together with the dictionary's strong points, will be highlighted briefly in this review.

2.   Front matter

The front matter of the dictionary consists of a preface and three paragraphs on how to use the dictionary, both in Oshindonga and English.

The dictionary is described as "a useful tool in acquiring basic skills in English". The usefulness of the dictionary as the said tool may however be questioned, since no explanation of the English (or Oshindonga) grammar, however simple, is found in the front matter or anywhere else in the dictionary. It should be kept in mind that Oshindonga as African language and English as Germanic language differ significantly on syntactic and morphological levels. This fact makes it imperative that any bilingual learner's dictionary that contains these two languages as language pair should also contain some description of the grammars of the two languages, since the acquiring of language skills requires a great deal more than the mere acquisition of a collection of lexical items in the target language.

3.   Macrostructure

According to the compilers the lemmas are chosen from "the most commonly used words in everyday English". In some cases the method of selection of lemmas is not clear, especially when one bears in mind that the dictionary "should never be used alone, but always together with and in nurturing an understanding of other subjects such as History, Science, Geography, etc." An example is the incomplete representation of a highly relevant semantic field for the primary school learner, namely that of arithmetic: while the lexical items *add* and *divide* enjoy lemma status, the lexical items *equal, minus, multiply, plus* and *subtract* are not taken up in the macrostructure. In the same way the lexical items *circle* and *rectangle* are taken up as lemmas and given Oshindonga equivalents, while the lexical items *square* and *triangle* are not to be found as lemmas in the dictionary. Another example that illustrates the incomplete representation of a semantic field, is the presence of lemmas such as **fifteen, fifty** and **nine**, while lexical items such as *eight, seven, ten, twenty, twenty-five*, etc. are absent. In some cases simple opposites are not included as lemmas: *wife* is present as lemma, while *husband* is not; *hot* and *warm* are presented as lemmas, while *cold* is not. These discrepancies place a question mark over the criteria used to select lemmas for this dictionary. The recognition of semantic fields and lexical relations, however elementary they might be, plays an important role in acquiring language skills. Where these aspects are not realised in a dictionary aimed at providing such skills, the learner is deprived of valuable aids in attempting to reach his/her goals. This issue will be taken further in the review of the microstructure.

An aspect of the dictionary which certainly places a great handicap on its use to acquire language skills in English, is the fact that it is unidirectional, i.e. only English-Oshindonga; it does not have an Oshindonga-English section. This most probably makes the dictionary only useful for decoding purposes, i.e. for use by Oshindonga-speaking learners to understand English texts. The dictionary has little or no encoding value. It will be impossible for a learner to find an English equivalent for a certain Oshindonga lexical item in the dictionary, which makes it impossible for a learner to use the dictionary effectively, if at all, to construct English texts in any way. Taking the stated aim of the dictionary into consideration, this is probably the most serious deficiency.

4.   Microstructure

Each article consists of a lemma in bold print, one Oshindonga translation equivalent in italic print, and an example sentence in English containing the lemma, followed by its Oshindonga translation containing the Oshindonga translation equivalent. Some articles also contain graphic illustrations after the example sentences. In cases where the lemma has more than one polysemic value, they are clearly distinguished by the numbering of the respective trans-

lation equivalents and the providing of example sentences at each polysemic value.

4.1   Polysemy and hyponymy

Translation equivalents of polysemic values of lemmas are generally indicated successfully by clear numbering to distinguish between different values. This practice is also explained in the front matter in the section on "How to use this dictionary", and it will make the young learner aware of the fact that certain words may have more than only one meaning.

The value of the dictionary as language reference work could however have been increased greatly if word classes were indicated for all polysemic values. In this dictionary the user is fully dependent on the example sentences to find out if the equivalent presented is the one he/she is looking for. In this way the lemma **answer** is given the following translation equivalents:

1. *yamukula*
2. *eyamukulo*

The first equivalent is the equivalent for *answer* as a verb, while the second is the equivalent for *answer* as a noun. This is however not indicated. The user must work through the example sentences to find out which equivalent is applicable. If the user e.g. knows that he/she is looking for the Oshindonga equivalent of the noun *answer* in the dictionary, he/she would simply have to look for the polysemic value which is indicated as a noun, provided that word classes are indeed indicated. In this case the user would then immediately realise that he/she does not have to study the first polysemic value any further, because it is indicated as a verb; he/she can proceed to the next equivalent without spending any more time on eventually irrelevant information. Primary school learners should have some knowledge of word class by the time they are in the senior primary phase, therefore the argument that word-class indication would complicate dictionary use for the learner does not hold water. In fact, the learner might realise that knowledge of word class makes his/her dictionary enquiries easier, while it also refines his/her skills in using reference works in general.

The negation by the dictionary of homonyms could seriously affect its value as language reference work. An example which illustrates this is the handling of the lemma **row**. According to the editorial system of the dictionary, three polysemic values are ascribed to the lemma. However, the three identified values are not polysemic values of the same form, but rather three homonymic forms. The form of the value presented as second polysemic value (*row* meaning "disagreement or noise") is in fact pronounced differently in English from the (identical) forms of the other two values: [rau] vs. [rou]. This information is completely lost to the learner, and if the teacher does not draw

the learner's attention to this difference, the learner might be confused when he/she is later confronted with different pronunciations for what is to him/her the same word. It is clear that the dictionary has failed to provide adequate language skills to the learner in this respect. It is therefore important that the dictionary should also make the appropriate homonymic distinctions where applicable. The way in which this is done (especially in a primary school dictionary) may still have to be considered, but the idea is that the learner should understand that homonymic forms are not related in terms of meaning (and origin) and sometimes pronunciation.

4.2   Example sentences

The compilers of this dictionary have laid down a very sound principle, i.e. to provide example sentences for all translation equivalents presented in the microstructure. However, example sentences should have definite functions — they should not only be there for the sake of example sentences. Most importantly, especially in a dictionary which aspires to provide a basis for language-learning, example sentences should be grammatically absolutely correct. Unfortunately this is not the case in this dictionary. The following example sentences, extracted from the microstructure, contain grammatical errors:

> "My grandmother told us a story about a huge giant." (lemma **giant**; tautology)
> "Many men have a beard." (lemma **man**; subject-object correspondence)
> "He looked at his watch to see how late it was." (lemma **watch**; instead of "... to see *what time it was*")

Apart from the above errors, example sentences can, in the absence of a grammar description, assist the user in providing clues as to the existing differences between the two languages involved, thereby helping in the acquisition of language skills. But the guidance of the teacher in this is essential. This makes this dictionary not only a reference work, but in fact a language work-book. The full potential of the dictionary can only be realised if it is actively used in (language) teaching, and not only used as a book a learner is referred to if he/she does not understand an English word.

With polysemic lemmas example sentences can be used effectively to guide the user (learner) in choosing the correct polysemic value and therefore the correct translation equivalent for the given discourse situation. The following extract from the dictionary illustrates this point:

> **open**
> 1. *egulula*
>   Please **open** the door!
>   *Egulula omweelo alikana!*

2. *eguluka*
   This door is **open**.
   Omweelo ogwe eguluka.

Although the indication of word class at each Oshindonga equivalent would certainly have made the choosing of the correct translation equivalent easier, the example sentences here clearly demonstrate the use of the first equivalent as a verb, and the second as an adjective.

4.3   Illustrations

Like example sentences, illustrations should be used in dictionaries for specific purposes, and not merely for decoration. In the case of a translation dictionary for primary schools it may be argued that illustrations are essential for language-learning, but then those illustrations used in the dictionary, too, must be chosen with great care. Together with the great potential of illustrations in such dictionaries comes the risk of actually damaging the language-learning process with the careless handling and placing of illustrations. In the *English-Oshindonga Dictionary* this risk might unfortunately turn into actual loss in some cases. The use of some illustrations might create confusion in the mind of the young learner of English. In more than one instance the example sentence refers to a person of one gender, while the following illustration shows a person of the opposite gender. Some examples are the following:

> "My grandfather walks with a cane", while the accompanying illustration shows a grandmother (lemma **cane**).
> "He uses chalk to write on the blackboard", while the accompanying illustration shows a female teacher (lemma **chalk**).

It should be expected that the learner will associate an example sentence in the dictionary with the accompanying illustration. The use of illustrations in dictionaries like this one should therefore as far as possible complement the example sentences in order to assist in the language-learning process. Illustrations should not only be seen as additions to information presented in the dictionary article, but also as applications.

In at least one instance the print quality of an illustration makes it virtually useless, namely at the lemma **plough**: the actual instrument cannot be distinguished against the dark background.

A great deal of information is lost to the language-learner because illustrations have no annotations. At the lemma **plant**, for example, reference is made to *trees*, *bushes* and *flowers* in the example sentences, and an illustration of each is added. However, the learner may become confused as to which English word refers to which kind of plant. Annotations with translation equivalents would practically rule out this possibility.

Some semantic fields are only partially represented by illustrations. Again, the semantic field of arithmetic serves as an example. At the lemma **add** the illustration material consists of a sum ("4 + 4 = 8"), which is a fine contribution to conveying the meaning. However, at the lemma **divide**, no illustration is provided. (In fact, the meaning of **divide** is not represented in its arithmetic sense, which further deprives the dictionary of supplying means to effectively acquire language skills.) Similarly, an illustration of a rectangle is found at the lemma **rectangle**, while there is no illustration of a circle at the lemma **circle**.

Illustrations which make up a lexical field (e.g. illustrations of rectangles, triangles, circles, etc.) could have been placed together at one encapsulating lemma or superordinate (e.g. **figure**) with annotations. Reference could then be made to this lemma at the lemmas constituting the particular subordinates. In this way semantic relations could effectively be used in providing means for language-learning.

When paging through the dictionary, one realises that all illustrations are computer-generated. This has seemingly placed a limit on the collection of illustrations the compilers could choose from for use in the dictionary. One almost gets the impression that fitting illustrations were not searched for for given lemmas, but rather that place in the dictionary was searched for for available illustrations.

5.   Perspective

One of the compilers rightly states that "a dictionary does not have to be over-loaded with linguistic information and, should this occur, it should be made as simple as possible so as to ease the understanding of the target users" (Mbenzi 1996: 318). One should, however, distinguish between *overloading* and *effective use to reach the aim of a dictionary* of linguistic information. Equally valid is certainly the view that the total negation of linguistic information in a dictionary can cause the dictionary not to realise its aim.

The *English-Oshindonga Dictionary* is a valuable contribution to language learning materials in Namibia, and certain improvements will definitely bring it closer to its aim.

Reference

Mbenzi, P.A. 1996. Review of *English-Kwanyama Dictionary*. *Lexikos* 6: 318-319.

Herman L. Beyer
Languages Department
Swakopmund Secondary School
Swakopmund
Namibia

National Terminology Services of the Department of Arts, Culture, Science and Technology. *Dictionary of Geology / Geologiewoordeboek,* English-Afrikaans / Afrikaans-Engels, 1996. cv + 795 pp. ISBN 1-875061-23-1. Pretoria: Council for Geoscience of South Africa. Price R114,00.

In reviewing this phenomenal dictionary one can easily fall into the trap of losing oneself in the rich variety of terms and their creative equivalents. It would, however, be a disservice to all involved to view this product merely as a bilingual dictionary of specialist geological terminology. In the first instance, this is part of a new generation of dictionaries where the final paper product is based on a rich and much larger underlying lexical (terminological) database. In the second instance, this dictionary is the product of a process that can only be described as a model in both terminographical and in terminological practice. In this review, I will briefly discuss the terminographical practice used in the making of this dictionary, the terminological contribution that this dictionary and the underlying lexical database makes in South Africa, and the way in which this database can, in future, be utilised as a national resource.

Terminographical practice

It is currently accepted that a proper **needs analysis** is carried out to establish the needs of real language users before any large dictionary project is embarked on. However, in 1974 when this dictionary was officially initiated, the concept of sophisticated needs analyses in dictionary making was still far in the future. That this dictionary was nevertheless the result of a real need is, however, reflected in the involvement of all the role players in geology in South Africa, and the scope and the duration of the project. Each of these aspects will be discussed briefly.

It is a remarkable feat that all the role players who were involved in this dictionary right from its inception stayed with it to the last. The dictionary was initially started between 1963 and 1974 by the Geological Survey, but it was completed in consultation and with the cooperation of the Geological Society of South Africa and all the Geology departments at South African universities. The commitment of these role players is also reflected in the fact that the Geological Survey as an institution undertook the responsibility of this task. True commitment, however, is reflected on the level of the individual subject specialists who prepared for and attended meetings, and spent long hours editing term lists. The extensive list of collaborators and contributors testifies to the genuine involvement of the geological community in South Africa in this project.

The scope of this dictionary is very broad and goes beyond a mere bilingual glossary. It would have been a relatively simple matter to revise and up-

date the existing bilingual term list for geology. It would also have been a simple matter to translate the existing *Glossary of Geology* of the American Institute for Geology which was published in 1972 and therefore available at the time of the first official meeting of this project in 1979. It was, however, decided that a full-scale dictionary project which included all the terminographical phases of extraction, compilation, editing, translation and evaluation, be initiated.

This project was formally started on a full-time basis in 1974 and, with a break of 19 months in 1982/83, was completed in 1996; a period of just over 20 years. Had there not been a real need, there was ample time either to abandon the project, or at least to curtail its scope. As it was, it was carried to completion in spite of the time span and the continuous input required from the subject specialists.

The duration of this project points to a truism in all lexicographical work: *Good dictionaries take time*, and the reason this dictionary took twenty years was that the **terminographical method** was adhered to in all aspects and with a meticulousness that has come to be the hallmark of the lexicographical profession in general. The following brief overview of the method that was followed in the making of this dictionary will show that the duration of this project is mainly a reflection of the acute involvement of subject specialists at all times, which has, however, contributed enormously to the quality of the dictionary and the underlying database.

The extraction of geological terms was carried out in the following way: Although the existing term list was used as a basis, together with the *Glossary of Geology*, numerous new terms were extracted from existing, but also from newly published journals on an ongoing basis. Since the Library of the Geological Survey is the largest collection of geological publications in Africa and has maintained that collection over the years, the terms contained in this dictionary are not only up to date, but cover the whole range of geological publications.

The newly extracted terms (and other relevant information such as the etymology of the term, or known German and Dutch equivalents) were documented, first on a card system, and eventually directly in electronic form. Term lists for 37 subdomains in geology were compiled and sent to the collaborators.

These term lists went through several phases of editing and discussion by subject specialists both in term meetings and individually. Translation equivalents were suggested, discussed and evaluated by native speaker subject specialists. The term lists were also updated on a continuous basis. In fact, one of the main features of this dictionary is the sustained quality input from the subject specialists.

Through all the twenty plus years extensive documentation was maintained on all the relevant research by individuals and all the decisions made by the term committees. This information is now available for reference purposes

on the National Term Bank, and constitutes a national and an international resource.

The bilingual dictionary that was the main purpose of this project and can be said to be its final product, represents a finished and well-edited reference work that any geologist, terminologist or translator should have on his / her bookshelf. A few minor editorial inconsistencies were, however, noted. These inconsistencies refer mainly to the treatment of polysemous items and the indication of collocations and/or domains of occurrence. An example of these inconsistencies will be discussed briefly.

The term *level* has no less than six nominal entries with six different Afrikaans equivalents. Only three of them are clarified by means of a collocation (*vlak (bv in myn)*), a superordinate (*waterpas (instrument)*) and a synonym of the English term (*levelling instrument*). It would be quite unclear to anybody except a fully bilingual geologist which one of the three unclarified Afrikaans equivalents to use for *level* **in any given context**. Since both *waterpas* and *nivelleerder* are identified as instruments, it would also be quite difficult for an untrained person such as a translator to decide which instrument might be meant **in any given context**. Other examples abound, and include *drift, fracture, limit, inlet,* etc. These examples also serve to illustrate the importance of indicating **the linguistic context** in which a term is likely to occur (by means of collocations) or **the referential context** in which a concept is likely to occur (by means of a domain specification). Since this type of information is more than likely available on the electronic database it is not clear why it has not been included in the dictionary.

Terminological contribution

In terms of its contribution to terminological practice in South Africa, this dictionary and the underlying electronic database can hardly be overemphasized. As far as the translation of the specific geological terms from English to Afrikaans is concerned, many of the translation equivalents reflect the international status of geology and transliterations (such as *heksagonaal, longitudinaal* and *erupsie*) are therefore used in abundance. Other translation equivalents are, however, both creative and clarifying. Some interesting examples of translation equivalents include: **adolescent river** as *halfwasrivier,* **bottoming** as *afwaartse uitknyping,* **box fold** as *kofferplooi,* **daughter element** as *volgelement,* **erratic** adj. as *swerf-* (as in *swerfhorison* or *swerfvlak*), **grid method** as *ruitnetmetode,* **sponge-work** as *rysmierdoolhoof,* **stray sand** as *dwaalsand,* **scabland** as *skurweveld,* **squeaking sand** as *fluitsand,* **virgin flow** as *ongehinderde vloei,* **water economy** as *waterhuishouding,* and many more.

The true value of this dictionary for terminological practice in South Africa, however, also lies on two other levels. A major contribution lies in the extensive and rich introductory notes to this dictionary which can be applied in

the training of all terminologists in South Africa. Apart from a preface by the Chief Director of the Geological Survey, and the standard lists of collaborators, reference works and editorial abbreviations, the notes contain an historical overview of the project (which is interesting from a language planning and development point of view) as well as extensive explanatory notes on the linguistic and lexicographical terminology and method, and the editorial layout of the dictionary. The extensive discussion of the geological terms themselves contains a rich mine of information. For example, chronostratigraphic, lithostratigraphic, rock and mineral names have a particular status in geology and are discussed in terms which elucidate their significance for an educated layperson or translator without falling into the trap of becoming too specialised. After all, it is not the geologists themselves that are going to find their reading on these issues here. Interesting equivalents, loanwords, transliterations (especially of some common geological suffixes), and the relationship of the dictionary to the 1991 edition of the *Afrikaanse Woordelys en Spelreëls* are all discussed in detail.

Unfortunately, it is incomprehensible that not all this material is available in the English translation of these notes. Principles of term creation are relevant to all developing languages, and the significance and usefulness of this information does not only apply to Afrikaans, even though the actual terms being discussed are Afrikaans. The omission of this rich data is either an editorial oversight or a narrow view of the role and the general significance of term creation principles in developing all the languages in South Africa. It is also not clear on what principles the information in the *Introduction* and the *Explanatory Notes* was separated. It would have been easier, and more continuous, to keep all the general sociolinguistic, linguistic, terminological and terminographic information together. Nevertheless, the National Terminology Service, and especially Ms Judy de Beer, should be congratulated on this specific record of terminological practice, which is a first in South Africa.

Another major contribution of this dictionary lies in the fact that the bilingual dictionary that was initially to have been the final product, reflects an extensive electronic lexical database that can now serve as a base from which new terminological work and research can be launched. One of the main reasons for the original launch and sustained financial support of this large project, and of the National Term Bank, was the fact that the then government of the day had an active policy of developing Afrikaans. The investment in the electronic National Term Bank, and in the large dictionary projects of the past era can only be justified as a **national resource** if collections such as the geological terms on the Term Bank are transformed into truly national, multilingual collections in which all the languages of South Africa are developed. It is the responsibility of both the Council for Geoscience and the National Terminology Service to contribute to this development. This development can be carried out along the following lines:

(i)   Terminographers should extract the basic terminology in the environ-
        mental and earth sciences (which will be used in future primary and sec-
        ondary school syllabi) for translation into the other official languages in
        South Africa.
(ii)   Native speaker terminologists should translate these terms into all the
        official languages in South Africa.
(iii)   Terminologists together with subject specialists should write definitions
        and explanations for these terms.
(iv)   Terminologists should research the conceptual structure of specialised
        domains such as the earth sciences, so as to facilitate education and
        training in these domains at primary, secondary and tertiary level.
(vi)   Terminologists should investigate the general (morphological and other)
        processes in term creation so as to facilitate the translation of these terms
        into all languages in South Africa.
(vii)   The Council for Geoscience should actively market this resource (which
        is unique in Africa and abroad) for further funding and development.

Britta Zawada
Department of Linguistics
University of South Africa
Pretoria
South Africa

Publikasieaankondigings / Publication Announcements

H. Chimhundu (Mupepeti Mukuru). *Duramazwi Rechishona*, 1996, xxxiv + 504 pp. ISBN 1-77900-256-4. Harare: College Press. Price Z$110,99 (US$11,00)

J.J.N. Cloete. *Public Adminstration Glossary*, ii + 104 pp. ISBN 0-627-02089-5. Pretoria: Van Schaik. Price R35,75.

Michael Clyne (Editor). *Undoing and Redoing Corpus Planning*, 1997, viii + 520 pp. ISBN 3-11-015509-5. Berlin: Mouton de Gruyter. Price DM198,00.

P.J. Coertze en R.D. Coertze. *Verklarende Vakwoordeboek vir Antropologie en Argeologie*, 1996, ix + 424 pp. ISBN 0-620-19929-6. Pretoria: P.J. Coertze en R.D. Coertze. Prys R80,00.

R.R.K. Hartmann (Editor). *Solving Language Problems: From General to Applied Linguistics*, 1st edition 1996, vi + 298 pp. ISBN 0-85989-484-3. Exeter: University of Exeter Press. Price £13,95. (Review in this issue)

Joyce M. Hawkins. *The South African Oxford School Dictionary*, 1996, xxi + 551 pp. ISBN 0-19-571414-8. Cape Town: Oxford University Press. Price R39,95.

G.H. Kocks (Samensteller), J.P. Vording, A. Beugels, H. Bloemhof, e.a. (Medewerkers). *Woordenboek van de Drentse Dialecten A-L*, 1996, LXIX + 706 pp. ISBN 90-232-3176-7. Assen: Van Gorcum. Prijs ƒ87,50. Intekenprijs ƒ69,50. (Resensie in hierdie uitgawe)

T.J. Kriel, D.J. Prinsloo and B.P Sathekge. *Popular Northern Sotho Dictionary*, 4th extensively revised edition, 1997, ii + 335 pp. ISBN 1-86890-001-0. Cape Town: Pharos. Price R29,95.

Joep [T.J.W.M.] Kruijsen. *Geografiese patronen in taalcontact*, 1995, xv + 221 pp. ISBN 90-70389-46-0. Amsterdam: P.J. Meertens-Instituut. Prijs ƒ39,95.

Peter Rolf Lutzeier *et al.* (Editors). *Lexicology: An International Journal on the Structure of the Vocabulary*, Vol. 1 (2 issues per year), 1995, x + 218 pp. ISSN 0946-9400. Berlin: De Gruyter. Price DM184,00 (2 issues).

P.A. Mbenzi and O.N. Iithete. *English-Oshindonga Dictionary for Primary Schools*, 1st edition 1996, iv + 97 pp. ISBN 99916-37-33-8. Windhoek: Out of Africa. Price: N$39,90. (Review in this issue)

Nasionale Terminologiediens van die Departement van Kuns, Kultuur, Wetenskap en Tegnologie / National Terminology Services of the Department of Arts, Culture, Science and Technology. *Geologiewoordeboek / Dictionary of Geology*, English-Afrikaans / Afrikaans-Engels, 1996, cv + 795 pp. ISBN 1-875061-23-1. Pretoria: Raad vir Geowetenskap van Suid-Afrika / Council for Geoscience of South Africa. Prys / Price R114,00. (Review in this issue)

Rod Nesbitt. *Chambers-Macmillan South African School Dictionary for Junior Primary*, 1996, xx + 187 pp. ISBN 0-7978-0545-1. Manzini: Macmillan Boleswa. Price R 27,15.

Van Schalkwyk, D.J. (Hoofredakteur). *Woordeboek van die Afrikaanse Taal*, Deel X, 1996, xx + 660 pp. ISBN 0-9584120-1-4 (plastiekband); ISBN 0-9584120-2-2 (leerband). Stellenbosch: Buro van die WAT. Prys plastiekband R79,80; leerband R148,20.

A.A. Weijnen. *Etymologisch Dialectwoordenboek*, 1996, xxvii + 269 pp. ISBN 90-232-2967-3. Assen: Van Gorcum. Prijs ƒ65,00.