The Use of Examples in Polyfunctional Dictionaries

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

Abstract: Dictionary compilation for the African languages in South Africa stands on the eve of the African Renaissance and prospective compilers of dictionaries for African languages are in need of a sound perspective and guidelines in respect of the compilation and choice of examples. The aim of this article is to analyse and evaluate some current approaches towards the handling of examples of usage as a data category in modern dictionaries and to suggest ways in which this information category can be improved by compiling, selecting and shaping examples to render optimal transfer of information and to enhance information retrieval. The emphasis will be on example phrases and sentences given in the dictionary article.

Keywords: LEXICOGRAPHY, DICTIONARY, METALEXICOGRAPHY, AFRICAN LANGUAGES, EXAMPLES, AUTHENTIC EXAMPLES, CONSTRUCTED EXAMPLES

Opsomming: Die gebruik van voorbeelde in polifunksionele woordeboeke. Die samestelling van woordeboeke vir die Afrikatale in Suid-Afrika staan op die vooraand van die Afrika Renaissance en voornemende samestellers van woordeboeke vir die Afrikatale benodig 'n gesonde perspektief en riglyne ten opsigte van die samestelling en die keuse van voorbeelde. Die doel van hierdie artikel is om sommige van die jongste benaderings ten opsigte van die hantering van gebruiksvoorbeelde as 'n datakategorie in moderne woordeboeke te analiseer en te evalueer en om metodes aan die hand te doen waarvolgens hierdie inligtingskategorie verbeter kan word deur die samestelling, keuse en afronding van voorbeelde om optimale inligtingsoordrag te bewerkstellig en om inligtingsontsluiting te verbeter. Die klem sal val op voorbeeldfrases en -sinne wat in die woordeboekartikel aangebied word.

Sleutelwoorde: LEKSIKOGRAFIE, WOORDEBOEK, METALEKSIKOGRAFIE, AFRIKATALE, VOORBEELDE, OUTENTIEKE VOORBEELDE, GEMAAKTE VOORBEELDE

Introduction

One of the developments in modern-day lexicography, as indicated in numerous publications, e.g. Hartmann (1989), Gouws (1989), Wiegand (1998), Van der Merwe-Fouché (1999), is the emphasis on an approach according to which the compilation of dictionaries is guided by the user perspective. A user-orientated dictionary should lead to enhanced information retrieval procedures.

Improving the quality and appropriateness of examples is one of the ways of enhancing the process of information retrieval. The inclusion of examples should benefit all target users of a dictionary but it is of special importance to the encoding user who needs maximum guidance within the physical limitations of a dictionary article. Illustrative examples play a vital role in dictionaries, and the dictionary conceptualisation plan of any new lexicographic project should make provision for a systematic presentation of this data type in the data distribution structure. This also applies to the planning of dictionary projects to be compiled by the National Lexicography Units (NLUs) of all the official languages of South Africa. These dictionaries should primarily be directed at the communicative needs of the members of the relevant speech communities and should endeavour to equip these target users with the necessary communicative skills to function successfully in a multilingual and multicultural environment. Consequently, the functional inclusion of examples, illustrating actual everyday language usage, is of prime importance. It adds to the quality and user-friendliness of a dictionary.

Examples can have different functions, including an explanatory function, in a dictionary article (cf. Rademeyer 1992), which may never be underestimated but has to be maximized and utilised to the utmost. This is of particular interest to the dictionaries compiled for the African languages. It has been stated repeatedly in publications such as Mbogho (1985) and Gouws (1990) that to date, dictionaries for African languages are the products of limited efforts. When planning the dictionaries to be compiled by the NLUs the lexicographers have to be aware of the problems of the past and they have to identify and avoid these pitfalls. In the case of examples, attention should focus on e.g. information regarding the purpose and characteristics of good examples, the question of authentic versus constructed examples, the decoding and encoding functions of examples and the needs of the target user in especially learner's dictionaries. One of the problems confronting the lexicographer is the fact that, also with regard to the presentation and treatment of examples, theoreticians offer different views and suggestions. They are biased towards one or other ideological method, often reflecting a one-sided view of the compilation or choice of examples. In this article an attempt will be made to create a more balanced perspective on the selection and presentation of examples and their distribution in the article structure. Guidelines for the use of examples will be formulated, especially for prospective compilers of dictionaries for African languages.

The dictionary plan, which has to be implemented when the dictionaries of the NLUs are compiled, should display a sound theoretical basis (cf. Gouws 2000b). The lexicographic process may include no haphazard procedures or an arbitrary and random choice and presentation of data categories. This also applies to the selection and presentation of illustrative material. The role of the dictionary compiler as "the deliberate controller and manipulator of illustrative

material" (Cowie 1989: 63) has to be reassessed to ensure the consistent application of a well-designed lexicographic system.

Examples and dictionary typology

No single dictionary can be everything to everyone. It is therefore of paramount importance that the planning of every dictionary has to be preceded by an in-depth analysis of the needs and reference skills of the intended target users in order to determine the typological criteria to be adhered to (cf. Hartmann 1989). Dictionaries represent a reliable form of communicative empowerment. The dictionaries compiled for a specific speech community should be a response to the specific communicative needs of that speech community. A speech community existing in a monolingual environment has different needs compared to a speech community which is part of a multilingual society.

Within multilingual South Africa, dictionaries compiled for the African languages should aim to meet the needs of the members of the relevant speech communities. Although it has to be seen as the ultimate ideal to compile a comprehensive monolingual dictionary for each one of the South African languages, such a dictionary may not be regarded as a short-term target of the NLUs. Given the communicative needs of the members of the different South African speech communities, the compilation of more restricted dictionaries, e.g. learner-orientated bilingual or descriptive dictionaries or school dictionaries should be regarded as one of the primary objectives of the NLUs. In the past dictionaries of these typological categories have often neglected illustrative examples as a functional data category.

The comprehensive lexicographic process in South Africa (cf. Gouws 2000) has to include the formulation of a dictionary plan to be applied within each one of the NLUs. The application of the dictionary plan should result in the publication of a first as well as subsequent dictionaries. More specifically, the dictionary plan should not only aim at the publication of the first dictionary but also at the development of a range of dictionaries for each of the NLUs. Such an approach will have definite implications for the typological choice and all the structural components of the target dictionaries. This includes the macro- and microstructures as well as the access and mediostructure and, very important, the data distribution structure. The development of the South African lexicographic process has to be regarded as an evolutionary process. This has direct implications for the structure of the dictionaries to be compiled during the first phase of the process because these dictionaries should constitute the basis for the dictionaries of the next phase.

The presentation and treatment of examples should be seen within the context of the more comprehensive lexicographic process where both the dictionary typology and the choice of the type of microstructure play a decisive role. A decision by a specific National Lexicography Unit to start with a mono-

lingual or a bilingual dictionary of a restricted nature, e.g. a school dictionary or a desk dictionary, does not imply the omission of illustrative examples. Instead, it implies a much stricter set of selection criteria for the inclusion of entries representing this data category. Contrary to the system adhered to in many older dictionaries, it should be emphasised that the treatment of entries in dictionaries of a restricted nature may not leave the intended target user with the feeling that the dictionary has not assisted him/her in achieving the purpose of his/her dictionary consultation effort. In the past, school dictionaries have too often been the result of an editorial cut and paste process. A more comprehensive dictionary had been used as a point of departure and the school dictionary was often compiled by deleting macrostructural and microstructural entries in an arbitrary way. Illustrative examples represent one of the data categories which often did not survive to feature in a school dictionary. Such an approach totally eschews the needs of the target users of the dictionary.

This paper presupposes that illustrative examples represent a functional data category in any general translation or descriptive dictionary and that dictionary typology, within this broad category of general descriptive and translation dictionaries, should not detract from the value of examples. In such dictionaries, the comment on semantics should contain a search zone allocated to this data type. Any new dictionary project plan should make provision for an evolutionary lexicographic process which allows lexicographers to use the first dictionary as a basis for the following ones. This implies that each data category presented in the first dictionary, should be dealt with in such a way that it could be expanded on in a second, more comprehensive dictionary. The responsibility is on the lexicographer to ensure that each data category in a restricted dictionary is presented in a scientific and theoretically sound and adequate way. The typological nature of a dictionary may never be an excuse for the presentation and treatment of any given data category in a manner not motivated by sound theoretical principles.

It should thus be emphasized that dictionaries should display a structure aimed at the successful transfer of those data categories needed by the target users. This is of particular relevance to the dictionaries compiled for the African languages in the newly established NLUs. The specific nature of the microstructure and the article structure is of extreme importance for the inclusion, positioning and treatment of examples. Having decided on a specific microstructure and an article structure, the lexicographer should apply and adhere to these decisions in a consistent way.

Moving from an unintegrated to an integrated and eventually to a semi-integrated microstructure

Dictionaries can display different types of microstructures and the compilation of any dictionary should be in accordance with the criteria of the relevant type. Wiegand (1996b) makes provision for, among others, three major types of

microstructures: integrated, unintegrated and semi-integrated microstructures. For both monolingual descriptive and translation dictionaries, a semi-integrated microstructure could be seen as an ideal. However, this is a type of microstructure best suited for a more comprehensive type of dictionary and not for desk and school dictionaries. Furthermore, it presupposes quite sophisticated dictionary-using skills. In the remainder of this paper the emphasis will be on the other two types of microstructures. The discussion of a semi-integrated microstructure could come to the fore again in a model for a more comprehensive dictionary.

The use of an unintegrated microstructure implies that the translation equivalents and the cotext entries do not appear in the same text block. Positioning the translation equivalents and the cotext entries in different text blocks does not necessarily imply a lack of a coordination between these entries or the absence of an addressing procedure between cotext entry and translation equivalent. In a dictionary with short articles a well-devised unintegrated microstructure could serve the user well. However, the presentation of two separate search zones should not lead to an arbitrary ordering of entries on an intra-search zone level (cf. Gouws 2000a). This is unfortunately the case in many of the existing South African bilingual dictionaries due to an unpredictable and inconsistent selection and presentation of examples and other cotext entries.

The use of an unintegrated microstructure in many South African dictionaries often tends towards uncoordinated text blocks. The situation is often aggravated by the lack of a microarchitecture and consequently the absence of a clear indication of the boundaries of the different search zones. An unintegrated microstructure compels the lexicographer to use structural markers to ensure the identification of the different search zones as well as the necessary inter-search zone coordination of translation equivalents and cotext entries. The absence of such structural markers can lead to a special form of internal textual condensation in an article which requires more sophisticated dictionary using skills than the lexicographer should expect from the target user of the dictionary. Although the examples, given as cotext entries, can be coordinated with the relevant member of the translation equivalent paradigm if a relation of addressing equivalence prevails (cf. Gouws 2000b), this nonlemmatic addressing procedure illustrates the occurrence of nonspecific distant addressing. The nonspecific and nondirect link between the cotext entry and its address makes it more difficult to achieve successful dictionary consultation. Compare the article of the lemma sign **stryd** in *Groot Woordeboek/Major Dictionary*:

> **stryd**, fight, strife, struggle, contest, conflict, combat, action, war; *die ~ AANBIND met*, join issue with; *die ~ BESLIS*, win the battle; *DIE STRYD (pol., hist)*, the Struggle; *die ~ om (die) BESTAAN*, the struggle for existence, *die ~ GEWONNE gee*, admit defeat; *IN ~ met*, in

conflict with, contrary to; ... *die goeie ~ STRY*, fight the good fight; ... *ten ~e TREK teen*, go to war against; ...

Contrary to this example, the functional application of an unintegrated microstructure presents the illustrative examples in a text block separated from either the text block containing the translation equivalent paradigm or the different subcomments on semantics given for the various polysemous senses of the lemma sign. The success of such a microstructural approach is the clear and unambiguous coordination between these text blocks so that each illustrative example entry has a definite and clearly indicated address in the preceding text block.

An integrated microstructure contains both the translation equivalents and the cotext entries in the same text block with a cotext entry following each translation equivalent to assist the encoding user in his/her attempt to master the correct usage. The microarchitecture of an article displaying an integrated microstructure can confuse users if the lemma sign represents a polysemous lexical item and translation equivalents have to be given for each one of the polysemous senses of the lexical item represented by the lemma sign. Users then may have to struggle through a lengthy article to reach the needed translation equivalent. One of the major advantages of an integrated microstructure is the coordination between translation equivalent and cotext entry, as a result of specific and direct, i.e. nondistant, nonlemmatic addressing. Compare the article of the lemma sign **maak** in *Tweetalige Aanleerderswoordeboek/Bilingual Learner's Dictionary*:

maak **1.** make [a] *My mother can **make** clothes.* My ma kan klere **maak** ... **2.** go *Ducks go "quack".* Eende **maak** "kwaak". ... **6.** put *She tried to put the baby to sleep.* Sy het die baba aan die slaap probeer **maak**.

If applied well and consistently, both integrated and unintegrated microstructures assist the user to access the needed data categories presented in the comment on semantics of a bilingual dictionary. However, both these types of microstructures can also impede the information retrieval process if applied inconsistently or in an article not suited for the specific type of microstructure. Wiegand (1996b) has developed an additional microstructure, which includes the advantages and excludes the disadvantages of both the other types. This type of microstructure is called a semi-integrated microstructure. Articles with a semi-integrated microstructure display a characteristic microarchitecture with differentiated text blocks in the comment on semantics, to allow for both an integrated and an unintegrated lexicographic treatment. For reasons mentioned earlier in this section, the semi-integrated microstructure will not be discussed in more detail in this paper.

The illustrative examples in a bilingual dictionary can be regarded as the data category of which the presentation is most influenced by the specific

microstructural type. The plan of any new dictionary project should therefore also pay thorough attention to the specific type of microstructure most suitable to respond to the real needs and reference skills of real target users.

The purpose and properties of good examples

Just as any dictionary needs to have a genuine purpose (cf. Wiegand 1998), each and every data category included in a dictionary article should adhere to the realisation of the purpose of the dictionary. This can only be done if the lexicographer has a definite and clear view of the function of each data category. The inclusion, presentation, nature and extent of each data category should be governed by a set of criteria which places that data category within the broader frame of the purpose of the dictionary. Some data categories may have a monofunctional occurrence whereas others may be multifunctional. Although illustrative examples constitute a data category with a whole range of functions, one of the basic aims of the lexicographer by including and presenting examples should be to guide the user on a variety of characteristic features of the lexical item represented by the lemma sign which functions as guiding element of the specific article. Examples play an important role in guiding the user to *know* the word. Laufer (1992: 71) formulates this as follows:

> Knowing a word would ideally imply familiarity with all its properties ... When a person "knows" a word, he/she knows the following: the word's pronunciation, its spelling, its morphological components, if any, the words that are morphologically related to it, the word's syntactic behaviour in a sentence, the full range of the word's meaning, the appropriate situations for using the word, its collocational restrictions, its distribution and the relation between the word and other words within a lexical set ... The foreign language learner knows a much smaller number of words ... In many cases word knowledge is only partial, i.e. the learner may have mastered some of the word's properties but not the others.

Research over a broad spectrum (cf. Cowie 1989, Gouws 1989, 2000, Rademeyer 1992, Rundell 1998, and Atkins, Rundell and Gouws 1998) reveals some of the main characteristics of examples. According to this research, examples should help to

— disambiguate senses,
— distinguish one meaning from another,
— clarify an abstract definition,
— supplement the information in a definition,
— show or indicate the selectional range,

- place the word in context,
- place the word in cotext,
- specify the semantic range,
- indicate the collocational behaviour, including typical collocations,
- illustrate the grammatical patterns,
- specify the word order,
- give pragmatic uses,
- note stylistic features,
- indicate appropriate registers,
- reflect the word history,
- be accurate, especially those quoting measurements, technical data, etc., and
- stimulate the user to capture the features or characteristics of the word in question and use the examples as a model to create examples of his/her own.

Characteristics of bad examples

There are numerous pitfalls in the way of constructing good examples, such as register mismatch, examples that are confusing, distracting, contain irrelevant detail, etc. Atkins, Rundell and Weiner (1997: slide 9c25) list the following typical problems:

- natural, typical ... but completely pointless (e.g. Sicilian: *a quarrel between two Sicilians*)
- includes distracting or irrelevant detail
- seriously atypical (e.g. by proxy: *You can create an international incident by proxy*)
- highly context-dependent (e.g. gravitate: *He gravitated, naturally, to Newmarket*)
- register mismatch (e.g. latter: *We have to decorate the kitchen and the hall — I'd rather do the latter (room) first*)
- confusing (e.g. black vb: *They blacked all coal from mines that had continued working during the strike)*
- or just missing when you need it ...

The incorrect handling or unnecessary introduction of extralinguistic factors such as race, sex, politics, culture, etc. also render bad examples, e.g. reference to female subjects, companies or the morally unaccepted. Compare Atkins, Rundell and Weiner (1997: slide 9c26) in this regard:

> **nag** His wife nags (at him) all day. | She kept nagging (her husband) for a new car. | She nagged her son into leaving the house ... [LDOCE1]

crap [Company name]'s is crap — let's go to Burger King! [Longman Language Activator — early draft]

just the job Thanks for that screw — it was just the job! [LDOCE1]

It is also not advisable to refer to living persons or current events. Such examples enhance the relevance of the dictionary and are interesting to the user but easily become outdated, consequently the user might judge the whole dictionary as dated.

Corpus examples versus constructed examples: Introduction

The past few decades saw the rise of huge computerized corpora which brought a new dimension to dictionary compilation. The traditional method according to which examples were "made up" became increasingly challenged by the so-called "authentic" examples (unmodified utterances). The debate intensified since 1987 when Cobuild started using authentic examples in dictionaries. It unfortunately became an ideological struggle resulting in loss of perspective where supporters of a strategy only emphasized the virtues of their strategy and highlighting the shortcomings of the alternative method rather than taking the best of both alternatives. A more balanced perspective will be attempted in the following paragraphs.

In favour of corpus (authentic) examples and biased against constructed examples

Arguments in favour of authentic examples reflecting negatively on constructed examples are summarized in terms of Cowie (1978: 129), Fox (1987), Sinclair et al (1987, as quoted by Cowie 1989: 58), Laufer (1992), and Potter (1998: 357-362).

Authentic examples

— are almost always superior to constructed examples,
— are grammatically correct,
— are situationally appropriate,
— support the explanations,
— give accurate collocations,
— illustrate usage,
— provide a reliable guide for speaking and writing in the English of today,
— have actually occurred in the language,
— represent the language as it is actually spoken and written,

— show all the features of normal discourse, such as the use of pronouns rather than nouns, of linking words such as *and* and *therefore*, etc.,
— provide valid and accurate information on which learners can rely,
— include one or more useful collocations and, where appropriate, a range of grammatical patterns, and
— give the learner the guarantee that a piece of language does occur.

Constructed examples

— are often isolated, self-contained sentences because lexicographers tend to produce sentences with too much information in them,
— are sometimes odd and not very likely to occur in a communicative act,
— are artificial and may not reveal the most typical usage of a word,
— are really part of the explanations,
— have no independent authority or reason for their existence, and in many cases are used to clarify the explanations,
— give no reliable guide to composition in English and would be very misleading if applied to that task,
— do not say "This is how the word is used" but rather "This will help you to understand the sense", and
— could be hazardous without consulting a corpus.

Finally it is stated that

— usage cannot be invented, it can only be recorded,
— much of the information given in constructed or corpus-aided examples is reliable neither about the contexts in which a word or phrase is typically used, nor about the words that typically occur with it, and
— in some cases it can actually be misleading.

Potter (1998: 359-362) gives a detailed discussion on how constructed examples such as the one for the entry *bloom*, "I was just admiring the blooms in your garden", could be hazardous without consulting a corpus.

> It would therefore have been ridiculous to have studied real language in order to find out the facts of the language, and then to have abandoned this and concocted fake examples for the dictionary. (Fox 1987: 138)

In favour of constructed examples and against corpus examples

Hornby (as quoted by Cowie 1978: 58), Laufer (1992), Minaeva (1992) and others argue that learners are better served by examples that are to a greater or lesser extent invented by lexicographers rather than by examples selected from a corpus. According to them constructed examples

— are as useful or more useful to learners than those taken directly from a corpus with little or no modification,
— demonstrate the linguistic points the lexicographer wishes to convey, without any distraction or added difficulty which may be introduced by examples taken directly from real texts,
— are more helpful in the comprehension of new words than authentic examples,
— can be made to include detail, whether syntactic or lexical, which throws light on the meaning or use of the entry word,
— can be judiciously shaped in the interest of the learner,
— can include the significant syntactic detail, and
— are more useful because they seem to be less dependent on the learner's general lexical knowledge than authentic examples.

Authentic examples

— do not, even when used extensively, guarantee that the various needs which dictionary examples have to meet, will automatically be fulfilled,
— place on the compiler the additional burden of selecting (or even editing) examples so that meaning, syntax and style are effectively illustrated, and
— are overloaded with irrelevant extralinguistic information.

Perspective on corpus examples versus authentic examples

From the arguments in favour of or against authentic examples and constructed examples presented above, it is clear that both sides claim superiority. It is also unfortunate that most lexicographers defend only one alternative by focusing only on its virtues whilst ignoring its deficiencies, rather than trying to capitalize on the virtues of both methods. Such an ideological struggle does not serve the interests of lexicography. The two extremes are manifested or summarized in the following quotations:

> There is rarely any need to consider thinking up examples. There is sufficient evidence of how the word has been used: we can see what its collocates are; we can check whether the subject or the object of a verb is more typically a person or a thing ... (Fox 1987: 147)

versus:

> Lexicographer's examples are more helpful in comprehension of new words than the authentic ones. In production of the new word, lexicographer's examples are also more helpful, ... the usefulness of constructed

> examples seems to be less dependent on the learner's general lexical knowledge than the usefulness of the authentic examples. (Laufer 1992: 75)

One can also assume that in many cases lexicographers are not "free" to choose between authentic and constructed examples but have to follow "company policy". Furthermore, both sides claim that their viewpoints are supported by user feedback and experiments conducted on groups of users.

As a step towards a clearer perspective, general claims such as the superiority of one method over the other, authority, natural versus unnatural, more versus less thought-provoking, typical versus not typical, situationally appropriate versus inappropriate, user preference of one over the other, etc. should be avoided. It is risky to claim that the positive objectives listed here can only be obtained by using authentic examples. It will be more sensible to adopt a policy of taking the best from both options and avoiding the weak points of both at the same time.

To deal with the issue of natural versus unnatural: It is risky to generalize in stating that corpus examples are natural and constructed one's are not.

> Lexicographers who are educated native speakers of the language are bound to have correct intuitions about their mother tongue, about the grammaticality of the word, its typical use and its typical environment. These intuitions are not necessarily less correct than the intuitions of those language users who are represented in the corpus and are therefore not less reliable. (Laufer 1992: 72)

> Editors are constantly constructing examples to meet the learner's productive needs. Such examples ... often achieve their aim precisely because of some "unauthentic", pedagogically contrived feature of wording or typography ... Whatever the merits of authentic specimens of usage, it still has to be convincingly shown that they can take the place of invented examples for several productive purposes. (Cowie 1989: 58)

Thus it has to be concluded that a debate on this level has no gain. If some constructed examples are less "natural" than authentic ones, the problem does not lie in the inability of the lexicographer as a mother-tongue speaker to produce sentences as well-formed or natural as those uttered or written by other speakers of the language as taken from the corpus, but unnatural because too much information is stuffed into a single sentence, rendering it unlikely to occur in a natural conversation.

Fox (1987: 141-144) formulates this as follows:

> Isolated examples carry too much meaning content for one sentence. ... *"The teacher used to cane me when I behaved badly"* ... Sentences occurring in

> a natural context are not neat little isolated wholes. They carry a lot of loose ends — they follow on from what has been said and they lead into what will be said ... Language is not a series of isolated sentences ...

And Cowie (1989: 59) says the following:

> Compilers who invent their examples may, of course, fall into the opposite trap of sacrificing linguistic naturalness to explanatory fullness.

On the other hand, Minaeva (1992: 78-79) points out that corpus examples, although "very natural", can carry an abundance of irrelevant sociolinguistic information such as names of people and places:

> To use these dictionaries to the best advantage, one should possess a considerable amount of background knowledge because illustrative word-combinations and sentences abound in sociolinguistic information. The user should be an expert in science, literature and music ("Einstein was a mathematical *genius*." OALD; "I *myself* feel that Muriel Spark is very underrated." COBUILD; ...), know geography and history ("They've got a small farm in Devon." ... COBUILD ...); ... They are overloaded with irrelevant extralinguistic information ... irrelevant sociolinguistic information ...

Compare also the comments of Rundell (1998: 335) in this regard:

> The risk here, illustrated rather too often in COBUILD1 but only very occasionally in COBUILD2, is that wholly authentic examples can sometimes show mystifyingly irretrievable contexts (for example in COBUILD1's example at *gravitate*: *He gravitated, naturally, to Newmarket*); atypical uses ... or too much irrelevant and — to the learner — distracting material ... So there is still a place for the more "pedagogical" example, typically now produced by modifying an actually occurring sentence, which allows the lexicographer to focus on specific linguistic points without baffling the user.

A second pitfall related to naturalness is to formulate constructed examples in one tense, mood or actuality, such as in the present tense positive, whilst the specific word more frequently occurs in, say, the past tense negative. The same holds true for constructing active examples, whilst such words are more frequently used passively.

There are certain clear limitations to the lexicographer's intuition. Without the use of a corpus, the lexicographer has no chance of finding the hundreds of senses and subsenses of a word like *run* on intuition. Furthermore, Fox (1987: 146) states that an experiment on intuition in respect of collocates revealed that

on intuition the lexicographer is inclined to think in terms of "semantic sets" rather than of "words which are actually likely to occur in the near vicinity of each other".

> When we sit and intuit how words are used, we are likely to get it wrong. We also know that as soon as we start playing around with examples, making them more "accessible" or more "regular", we are liable to take the life out of them, or worse, mislead the user of the dictionary. (Fox 1987: 148)

The first step towards a clearer perspective on the issue of authentic versus constructed examples will be to view them as a *continuum* with end points *authentic (taken from the corpus)* as the one extreme and *constructed examples* as the other extreme with certain discreet *categories* in between, such as *edited corpus examples, corpus-based examples,* and *constructed examples taking corpora into consideration.* Compare the following oversimplified table illustrating this continuum:

Extreme	**Intermediate categories**			**Extreme**
Authentic (corpus examples) taken directly from a corpus without editorial modification	Slightly *edited/modified* corpus examples	Heavily *edited/modified* corpus examples	Partially invented, based on a corpus	Constructed examples

Humble (1998: 593) rightfully concludes:

> The bone of contention came to be considered almost as an *ideological problem* and not as a practical one for which the solution *may lie somewhere in the middle*: each kind of example was considered intrinsically better or worse (emphasis added). (Humble 1998: 593).

Although lexicographers in principle opt for one of the extremes, there are certain factors which necessitate the selection of examples from the intermediate stages. Humble (1998: 593) considers factors such as (a) encoding versus decoding, (b) the target user's skills e.g. "beginners", "intermediate" or "advanced" learners, and (c) frequency of use of words.

> Despite a certain amount of research into the issue ... the jury is still out on the relative merits of corpus-based and lexicographer-produced examples. But it is really no longer relevant to characterize the argument as concerning a simple choice between the authentic and the invented. All reputable monolingual learners' dictionaries now base every aspect

of their text on corpus data, so the differences now lie in the degree to which corpus material is "processed" on its way into the examples. (Rundell 1998: 334)

Encoding versus decoding and the learner's profile

According to some traditional points of view (cf. Kromann, Riiber and Rosbach 1984a, 1984b), the ideal situation would necessitate the compilation of at least four and possibly eight bilingual dictionaries for any given language pair. Certain applications of the active-passive principle imply that one has to distinguish four different functions for each member of any given language pair and that a separate dictionary should be compiled to comply with each one of these functions. This implies the compilation of dictionaries for text reception and text production for mother-tongue and non-mother-tongue speakers of both languages involved (cf. Hausmann 1977, 1986). The specific function of a dictionary should then determine the nature and extent of the illustrative examples to be used in that dictionary. However, even for the most sophisticated languages the idea of eight or even four dictionaries per language pair has to be regarded as totally impractical. From a user-perspective, such an idea is also not attainable. The current approach in dictionary research, which favours a user-driven lexicography, is also in opposition to such an impractical point of view.

There is quite a difference between dictionaries with a passive and those with an active function. Wiegand (1996a: XV) emphasises the fact that the formulation of a theory of bilingual lexicography may not lead to a situation where theory is isolated from lexicographic practice. Consequently Wiegand (1996b: 2) indicates quite clearly that the active-passive principle should not be regarded as a principle determining the typological nature of a dictionary. Within a theory of bilingual lexicographic texts and with a distinct focus on the needs of the target users of any given dictionary, it should rather be used to differentiate between various functions prevailing in one bilingual dictionary and even in one article of a given dictionary. Wiegand argues convincingly against the need to compile four (or even eight) bilingual dictionaries for a given language pair. According to him, only one general translation dictionary is sufficient for a given language pair if the data distribution structure and presentation in such a dictionary make provision for the different needs of different users by adopting a polyfunctional approach.

The dictionary conceptualisation plan of any bilingual dictionary should clearly indicate whether it is a monofunctional or a polyfunctional product and the lexicographer should know well in advance what the profile of the intended target user is and whether the search zone in which the illustrative examples are accommodated should focus on an encoding or a decoding function or on both these functions. The importance of examples for both the

encoding and the decoding function may never be underestimated. Fox (1987: 138-139) says:

> It is important that we realize that learners' needs for encoding are at least as urgent as their needs for decoding. More so, perhaps, in these days when such strong emphasis is placed on communication ... We must ensure that the information we give students will genuinely help them to produce language of their own which is as near as possible to that produced by native speakers. In the past, dictionaries have been seen more as aids to decoding than aids of encoding. There is no reason why they should not succeed in doing both ...
>
> Dictionaries have traditionally concentrated more on the grammatical, partly perhaps because they have been seen as tools for interpreting language rather than for creating it.

The use of examples for an encoding and/or decoding function is not only relevant to learner's dictionaries or bilingual dictionaries. Each and every lexicographer who decides to include illustrative examples in a dictionary should be aware that the entries given in the presentation of this data category, have to be functional text segments of the specific dictionary in which they appear. In a decoding function (cf. Cowie 1989: 57), examples can help to clarify individual meanings and can help the user to distinguish between related meanings. On an encoding level they help the user to select the correct grammatical pattern(s) for a given word or sense, to form acceptable collocations and to produce language according to stylistic norms typical of native speakers of the given language.

Although extremely important, the mono- versus polyfunctional nature of a dictionary may not be regarded as the only determining factor with regard to the choice of examples. Besides the function of the dictionary, the user's level of proficiency as well as the lemma in question should not only determine the choice of examples, but it should also influence the lexicographer's choice between constructed versus authentic examples. The dictionary conceptualisation plan should therefore provide for a data distribution structure which allows a differentiated approach with regard to the use of illustrative examples.

The treatment of words referring to very specific concrete objects, e.g. *radio* and *television*, do not benefit greatly from exemplification. Fox (1987: 137) states that examples in articles headed by lemma signs representing this kind of words teach the user nothing more about the actual object, but instead give information on how the word is likely to be used, e.g. *turn on the television* and *watch television*.

The lexicographer has to realise that the functional use of examples is a choice which will lead to entries accessible to the user. They should not be too difficult to understand, although, according to Humble (1998: 594), learners

profit most from an input which is slightly beyond their capacities. The examples must however be comprehensible and reproducible. The appropriate cotext, the typical grammatical structure as well as the pragmatic context in which a word occurs should form part of the information transfer achieved by the correct choice of examples. Whether the user is a learner or a mother-tongue speaker, whether the dictionary has a monofunctional or a polyfunctional approach, the choice of each example has to be done in such a way that it responds to the needs of at least one section of the target users of the dictionary.

If the dictionary is aimed at learners, Humble (1998: 596) stresses the special qualities of examples:

> They should be given examples resembling the sentences they aim to produce. Words can combine to make sentences according to the *open-choice principle* or to the *idiom principle*. Frequent words have a tendency to form idioms often, but can be used on their own as well, in an *open-choice* setting, even if this modality is not always the vernacular one and examples of these congruent uses are hard to find in a corpus. Constructed examples, however, if congruent, use the words in their core meaning, and are especially useful for beginning and intermediate learners.

The data distribution structure of a dictionary (cf. Bergenholtz, Tarp and Wiegand 1999) should also determine the extent of the use of examples and the consequent balancing of data categories in the dictionary article. The extended use of examples at the cost of the inclusion or treatment of other data categories needs to be discussed at an early stage of the formulation of the dictionary conceptualisation plan. Although their functional value may never be underestimated, illustrative examples should always be regarded as one of the data categories which assist the user towards a better understanding and use of the treated language.

Decisions regarding the use of examples may not be taken in an arbitrary way. The dictionary conceptualisation plan should give a meticulous exposition of the function, choice, nature and extent of examples. Lexicographers should then adhere to these criteria.

References

Atkins, B.T. Sue, Michael Rundell and Edmund Weiner. 1997. *Salex'97. A Training Course in the Compilation of Monolingual Dictionaries.* Unpublished course material of a tutorial held at the Dictionary Unit for South African English, Rhodes University, Grahamstown, 15-26 September 1997.

Atkins, B.T. Sue, Michael Rundell and Rufus H. Gouws. 1998. *Afrilex-Salex'98. A Training Course in the Compilation of Bilingual Dictionaries.* Unpublished course material of a tutorial held at the University of Pretoria, Pretoria, 7-18 September 1998.

Bergenholtz, H., S. Tarp and H.E. Wiegand. 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft. Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research*: 1762-1832. Berlin: De Gruyter.

Cowie, A.P. 1978. The Place of Illustrative Material and Collocations in the Design of a Learner's Dictionary. Strevens, P. (Ed.). 1978. *In Honour of A.S. Hornby*: 127-139. Oxford: Oxford University Press.

Cowie, A.P. 1989. The Language of Examples in English Learner's Dictionaries. R.R.K. Hartmann (Ed.). *Lexicographers and their Works*. Exeter: University of Exeter Press.

Du Plessis, M. 1993. *Tweetalige Aanleerderswoordeboek/Bilingual Learner's Dictionary*. Cape Town: Tafelberg.

Eksteen, L.C. et al. (Eds.). 1997[14]. *Groot Woordeboek/Major Dictionary*. Cape Town: Pharos.

Fontenelle, Thierry et al. (Eds.). 1998. *Euralex '98 Proceedings*. Louvain: University of Louvain.

Fox, Gwyneth. 1987. The Case for Examples. Sinclair, John M. (Ed.). 1987: 137-149.

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

Gouws, R.H. 1990. Information Categories in Dictionaries with Special Reference to Southern Africa. Hartmann, R.R.K. (Ed.). *Lexicography in Africa*: 52-65. Exeter: University of Exeter Press.

Gouws, R.H. 2000. Toward the Formulation of a Metalexicographic Founded Model For National Lexicography Units in South Africa. Wiegand, H.E. (Ed.). 2000. *Wörterbücher in der Diskussion IV*: 109-133. Tübingen: Max Niemeyer.

Gouws, R.H. 2000a. Doeltaalgerigtheid teenoor lemmagerigtheid in vertalende woordeboeke. *Tydskrif vir Geesteswetenskappe*. 40(1): 39-47.

Gouws, R.H. 2000b. Strategies in Equivalent Discrimination. Zettersten, A. and V.H. Pedersen. (Eds.). 2000. *Symposium on Lexicography IX*. Tübingen: Max Niemeyer.

Hartmann, R.R.K. 1989. Sociology of the Dictionary User: Hypothesis and Empirical Studies. Hausmann, F.J. et al. (Eds.) 1989-1991: 102-111.

Hausmann, F.J. 1977. *Einführung in die Benutzung des neufranzösischen Wörterbücher*. Tübingen: Max Niemeyer.

Hausmann, F.J. 1986. Grundprobleme des zweisprachigen Wörterbuchs. *Symposium on Lexicography III*: 137-154. Tübingen: Max Niemeyer.

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.

Humble, Philippe. 1998. The Use of Authentic, Constructed and "Controlled" Examples in Foreign Language Dictionaries. Fontenelle, Thierry et al. (Eds.). 1998: 593-599.

Kromann, H.-P., T. Riiber, P. Rosbach. 1984a. "Active" and "Passive" Bilingual Dictionaries. The Ščerba Concept Reconsidered. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83*: 207-215. Tübingen: Max Niemeyer.

Kromann, H.-P., T. Riiber, P. Rosbach. 1984b. Überlegungen zu Grundfragen der zweisprachigen Lexikographie. Wiegand, H.E. (Ed.). Studien zur neuhochdeutschen Lexikographie. *Germanistische Linguistik* 3-6/81: 159-238. Hildesheim: Georg Olms.

Laufer, Batia. 1992. Corpus-based versus Lexicographer Examples in Comprehension and Production of New Words. Tommola, Harnu et al. (Eds.). 1992: 71-76.

Mbogho, Kalumbo. 1985. Observations on Bilingual Lexicography involving Bantu and Indo-European Languages. *Babel* 31(3): 152-162.

Minaeva, Ludmila. 1992. Dictionary Examples: Friends or Foes. Tommola, Harnu et al. (Eds.). 1992: 77-80.

Otto, A.N. 1989. *Kriteria vir 'n Afrikaanse aanleerderwoordeboek*. Unpublished D.Litt. thesis. Stellenbosch: University of Stellenbosch.

Potter, Liz. 1998. Setting a Good Example. What Kind of Examples Best Serve the Users of Learners' Dictionaries? Fontenelle, Thierry et al. (Eds.). 1998: 357-362.

Rademeyer, L. 1992. *Die funksies van verbale voorbeeldmateriaal in eentalige woordeboeke*. Unpublished M.A. thesis. Stellenbosch: University of Stellenbosch.

Rundell, M. 1998. Recent Trends in English Pedagogical Lexicography. *International Journal of Lexicography* 11(4): 315-342.

Sinclair, John 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: Collins ELT.

Tommola, Harnu et al. (Eds.). 1992. *Euralex '92 Proceedings*. Tampere: University of Tampere.

Van der Merwe-Fouché, M. 1999. *Gebruikersperspektief in die Afrikaanse leksikografie*. Unpublished D.Litt. thesis. Stellenbosch: University of Stellenbosch.

Wiegand, H.E. (Ed.) 1996. *Wörterbücher in der Diskussion II*. Tübingen: Max Niemeyer.

Wiegand, H.E. 1996a. Zur Einführung. Wiegand, H.E. (Ed.). 1996: VII-XV.

Wiegand, H.E. 1996b. Das Konzept der semiintegrierten Mikrostrukturen. Ein Beitrag zur Theorie zweisprachiger Printwörterbücher. Wiegand, H.E. (Ed.). 1996: 1-82.

Wiegand, H.E. 1998. *Wörterbuchforschung*. Berlin: De Gruyter.