Discipline, Dilemmas, Decisions and Data Distribution in the Planning and Compilation of Monolingual Dictionaries*

Rufus H. Gouws, Department of Afrikaans and Dutch, University of Stellenbosch, Stellenbosch, Republic of South Africa (rhg@sun.ac.za)

Abstract: Bilingual dictionaries play an important role in the standardisation of a language and are often the first dictionary type to be compiled for a given speech community. However, this may never lead to an underestimation of the role and importance of monolingual descriptive dictionaries in the early lexicographic development of a language. In the planning of first descriptive dictionaries the choice of the proper subtype and a consistent application of theoretical principles should be regarded as of extreme importance. Even the compilation of a restricted descriptive dictionary should be done according to similar theoretical principles as those applying to comprehensive dictionaries. This contribution indicates a number of dilemmas confronting the lexicographer during the compilation of restricted monolingual descriptive dictionaries. Attention is given to the role of lexicographic functions and the choice and presentation of lexicographic data, with special reference to the presentation of certain types of polysemous senses which are subjected to frequency of use restrictions. Emphasis is placed on the value of a heterogeneous article structure and a micro-architecture in the articles of restricted dictionaries.

Keywords: ACCESS STRUCTURE, DATA DISTRIBUTION, FRAME STRUCTURE, FREQUENCY OF USE, HETEROGENEOUS ARTICLE STRUCTURE, LEXICOGRAPHIC FUNCTIONS, LEXICOGRAPHIC PROCESS, MICRO-ARCHITECTURE, MONOLINGUAL DICTIONARY, POLYSEMY, SEMANTIC DATA, TEXT BLOCK, USER-FRIENDLINESS, USER-PERSPECTIVE, VERTICAL ARCHITECTONIC EXTENSION

Opsomming: Displine, dilemmas, besluite en dataverspreiding in die beplanning en samestelling van eentalige woordeboeke. Tweetalige woordeboeke speel ’n belangrike rol in die standaardisering van taal en is dikwels die eerste woordeboektype wat vir ’n bepaalde taalgemeenskap saamgestel word. Dit mag egter nie tot ’n geringskatting lei van die rol en waarde van eentalige verklarende woordeboeke in die vroeë leksikografiële ontwikkeling van ’n taal nie. Van belang in die samestelling van eerste eentalige verklarende woordeboeke is die keuse van ’n korrekte subtip en ’n konsekwente toepassing van gesonde teoretiese beginse in die beplanning van die woordeboek. Selfs in die samestelling van ’n beperkte verklarende woordeboek moet die beplanning volgens soortgelyke teoretiese beginse gedoen word as wat die geval is in

* This article was presented as a paper at the Tenth International Conference of the African Association for Lexicography, organised by the Sesiu sa Sesotho Lexicography Unit, University of the Free State, Bloemfontein, Republic of South Africa, 13–15 July 2005.

Lexikos 16 (AFRILEX-reeks/series 16: 2006): 84-94
die samestelling van ‘n omvattende woordeboek. Hierdie bydrae wys op ‘n aantal dilemmas waarneem die leksikograaf gekonfronteer word in die samestelling van ‘n beperkte eentalige verklarende woordeboek. Aandag word gegee aan die rol van leksikografiese funksies, die keuse en aanbod van leksikografiese data met spesifieke verwysing na die aanbieding van sekere tipes betekenisonderskeidinge wat aan gebruiksfrekvensiebeperkinge onderhewig is. Die waarde van ‘n heterogene artikelstruktuur en ‘n mikro-argitektuur in die artikels van beperkte woordeboeke word beklemtoon.

Sleutelwoorde: DATAVERSPREIDING, EENTALIGE WOORDEBOEK, GEBRUIKERSSPREKTEF, GEBRUIKERSVRIENDELIJKHEID, GEBRUIKSFREKWNSE, HETEROGENE ARTIKELSTRUKTUIR, LEKSIKOGRAFISE FUNKSIES, LEKSIKOGRAFISE PROSES, MIKRO-ARGITEKTUUR, POLISEMIE, RAAMSTRUKTUIR, SEMANTIESE DATA, TEKSBLOK, TOEGANGSTRUKTUIR, VERTIKALE ARGITEKTONIESE UITBREIDING

1. Introduction

The planning and compilation of a dictionary may never be done randomly. At the centre of all the decisions regarding the compilation of dictionaries stands the person who has to open the dictionary and use it. This dictionary user is no longer the well-known stranger it used to be because lexicographers are fully aware of the compelling need to identify their potential target user before they even start with the planning of a dictionary. Modern-day lexicography demands a clear indication of the genuine purpose of each dictionary and the fulfilment of that genuine purpose is only possible if the intended target user can achieve an optimal retrieval of information from the dictionary.

Within a multilingual environment it is extremely difficult to decide whether a monolingual or a multilingual dictionary would be the better choice for a given speech community. This question should not be isolated from the bigger lexicographic picture and the real lexicographic needs of the relevant speech community. Looking at these needs, one should also distinguish between the needs of the users in terms of their day-to-day communication and their needs in terms of having a standardised language. Gallardo (1980: 61) maintains that, when a non-standardised language is involved, bilingual dictionaries would be the typical type available, whereas the existence of monolingual dictionaries implies that the cultural properties of the standard language are well established. According to Gallardo, the very existence of monolingual dictionaries testifies to a highly sophisticated native linguistic scholarship and a sizeable literate speech community. However, one should never underestimate the important role of bilingual dictionaries, especially in a multilingual society. Moreover, one should never underestimate the importance of moving on and working towards monolingual dictionaries. “Working towards” implies that it is not an either/or choice when it comes to bilingual versus monolingual dictionaries. The emergence of the bilingualised dictionary or the monolingual dictionary with a bilingual dimension, i.e. a hybrid form which contains features of both monolingual and bilingual dictionaries, has given
enough proof that such a dictionary type can often be regarded as the best lexicographic response to the diverse needs of a specific user group.

The realisation of the need, also in the South African national lexicography units, for bilingual dictionaries will direct the focus of the remainder of this article to various aspects regarding decisions in the planning and compilation of monolingual dictionaries.

2. Dilemmas and decisions in the planning and compilation of a monolingual dictionary

2.1 Lexicographic functions

The compilation of any dictionary needs to be preceded by a thorough planning phase which gives attention to the users, the type, the functions, the structures and the contents of the envisaged dictionary. This planning should be done in a structured way. Research in the field of metalexicography has provided enough guidelines not only to assist lexicographers in their endeavours to formulate the necessary dictionary conceptualisation plan, but also to ensure that the lexicographic process follows the necessary direction for the eventual success of the dictionary. For the lexicographic process of any dictionary there are a number of crucial route markers which cannot be ignored without lexicographers running into difficulties. The lexicographer should not only know at whom the dictionary is directed but also how the target user will use it and in what user situation it will be used. This implies that once the target user, the needs of the target user and the situation of use have been ascertained, the lexicographer can determine the dictionary type and the functions of the dictionary. Regarding the functions it has to be decided whether the dictionary should help to fulfil a knowledge-directed function or a communication-directed function, and, if the latter is the case, most importantly whether it should aim at helping with text production, text reception or both these functions. These questions regarding lexicographic functions (cf. Tarp 2000, Bergenholtz and Tarp 2002, Tarp and Gouws 2004) have to be dealt with at a very early stage of the lexicographic process. Once the functions have been determined the lexicographer can decide on the relevant lexicographic structures to provide a venue for the contents so that the functions can be fulfilled (cf. Gouws to appear).

2.2 Subtypological dilemmas

One of the decisions to be made at an early stage regards the typological nature of the dictionary. The typological category of monolingual dictionaries includes a number of subtypes. Even a broad subtypological category like “general monolingual dictionary” has a number of members. These dictionary subtypes do not represent distinctly isolated categories but rather a continuum of dic-
tionaries ranging from the comprehensive multivolume product to the smallest of dictionaries for tourists, or an initial source of reference like *My First Three Hundred Words for a young child acquiring his/her mother-tongue*.

Although the comprehensive multivolume product, especially if it is based on historical principles, represents the ultimate lexicographic aim, it should not, on account of its size or typological classification, be regarded as a better dictionary than any other monolingual dictionary of a much more modest extent. The compilation of a monolingual dictionary should be the result of a well-informed decision based on the needs of a specific user group. These needs should also determine the subtypological category of the dictionary to be compiled. In this regard it is important to recognise the fact that the lexicographic process of any dictionary has to be regarded as a sophisticated endeavour that should reflect the relevant theoretical approach. Although the compilation phase of the lexicographic process of a comprehensive dictionary takes much longer than that of a standard or a school dictionary, the planning of all dictionary types needs to be done along the same lines. To ensure that the compilation of a small monolingual dictionary results in the publication of a good product a planning exercise similar to that of a multivolume dictionary is demanded.

### 2.3 Selection and presentation of data and the lexicographic treatment

The nature and extent of the treatment on offer in different types of monolingual dictionaries differ vastly and this creates various dilemmas for the lexicographer. One of the first dilemmas regards the macrostructural selection. When confronted with all the interesting words in one’s corpus it is difficult to limit the list of lemma candidates to a small number. One should, however, resist the temptation to incorporate more lemmas than originally planned. Discipline to stick to one’s assignment is extremely necessary. A further dilemma is posed by the microstructural entries. Here it is of paramount importance that a microstructural programme should be devised in the early phases of the lexicographic process and that lexicographers should adhere to the constraints of such a programme. Yet another dilemma confronts the lexicographer when it comes to decisions regarding the data-distribution structure. Modern-day lexicography (cf. Kammerer and Wieand 1998, Gouws 2004) has emphasised the need for a frame structure with front- and back-matter texts complementing the central list and adding venues that can accommodate lexicographic data. The use of front- and back-matter texts should not be seen as applying exclusively to bigger dictionaries. Users in an environment where a dictionary culture thrives, should be made aware that the dictionary as a carrier of text types has much more on offer than the data included in the central list. What better way of helping to establish a dictionary culture than by introducing the use of a frame structure in the first dictionary products with which a member of a speech community is confronted?
For the average dictionary user the central list remains the primary target of the dictionary consultation procedures. Consequently the presentation and treatment in the central list should be of such a nature that the intended target user can find answers to the questions that prompted the consultation. This has implications for both the macrostructural and microstructural selection and presentation. As far as the macrostructure is concerned, a choice has to be made between a straight alphabetical ordering and a sinuous lemma file with nested and niched lemmata (cf. Gouws 2005, to appear). On a microstructural level, a variety of relevant topics demands a thorough planning during the early phases of the lexicographic process (cf. Gouws 2003). These comprise the type of microstructure, the data categories to be included and the organisation and presentation of the different entries in the dictionary article. When it comes to the data on offer, it is important that lexicographers should select data in accordance with the criteria applicable to the subtypological classification of the dictionary. Besides the data categories, the extent of the treatment allocated to each class and the presentation thereof are crucial factors. When the organisation and presentation of entries are planned, the notion of the micro-architecture of the article structure comes to the fore.

The main focus of the remainder of this article will be on the selection of some types of semantic data and the development of a micro-architecture for smaller monolingual dictionaries.

2.4 Aspects of the selection of semantic data

The most typical consultation procedure in monolingual dictionaries is directed at the retrieval of information regarding the meaning of the word represented by the lemma sign. Consequently, in spite of the importance of all the other microstructural entries, lexicographers may never neglect the treatment allocated to the meaning of a word. Attention should not only be given to an explanation of the meaning of the word but it should also be shown that the lexicon consists of a network of semantic relations. For the lexical items being treated relations like semantic opposition, e.g. antonymy and complementarity, and semantic inclusion, e.g. synonymy and hyponymy, should therefore be indicated. The paraphrase of meaning, often referred to as the lexicographic definition, remains the most prominent microstructural entry to convey semantic data.

The dynamics of language leads to constant changes on all levels of language and language use. This includes the lexicon, with new items being added and others being omitted, as well as the semantic load of single lexical items, with changes in the polysemous paradigm by the addition or omission of polysemous senses. Decisions regarding the number of polysemous senses to be presented in the treatment of a lexical item pose a real challenge to lexicographers. In this regard the lexicographers of a comprehensive multivolume dictionary have fewer problems. Their historical approach compels them to
give account of both old and new senses and even where the dictionary does not follow an historical approach they can rely on a balanced and representative corpus to identify all the senses that pass the minimum frequency of use test, and they have to include these senses. A standard dictionary has a lower density of data and usually fewer co-text entries but it still contains a reasonable number of senses based on frequency of use. In using this criterion one often finds that figurative and metaphoric senses are listed before the literal senses of a word, and that the original senses display a lesser frequency of use than new senses. Bigger monolingual dictionaries have the luxury to include the lesser used literal and older senses along with the figurative or newer senses, allowing users to make a link between figurative and literal and between old and new.

In learners’, desk and smaller dictionaries the strict adherence to usage frequency can become a problem. Fox (2005) indicates that a sole reliance on usage frequency often leaves the user stranded for lack of a link between a metaphoric sense of a word and its literal sense. This also applies to new literal senses. In the Oxford Basic English Dictionary the word tablet is treated as a monosemous item with only one paraphrase of meaning, i.e. "a small hard piece of medicine that you swallow". The user is left stranded regarding the well-known older sense "flat piece of clay, wood or stone with words inscribed on it". The South African Oxford School Dictionary assists its users by including three senses, i.e. "a pill", "a solid piece of soap" and "a flat piece of stone or wood etc. with words carved or written on it". In the compilation of a small dictionary the lexicographer needs clear guidelines to decide how the choice of polysemous senses should be made and whether it is sufficient only to rely on frequency of use as a criterion for inclusion. Consistency remains a cornerstone of the lexicographic practice and lexicographers should not follow two different approaches in their selection and ordering of senses. However, it may become necessary to complement the criterion of frequency of use with an additional procedure in order to give users a fuller picture of the meaning of the word. This is especially true in the treatment of words where the metaphoric sense is much more frequent than the literal sense.

Where the lexicographer opts for a selection and ordering of senses according to frequency of usage this criterion should be applied in all articles. The number of senses allowed in the article of a lemma representing a polysemous lexical item should not only be determined purely mathematically but also on usage frequency. Instead of allowing only two senses per article provision should be made that the treatment of a given word with three frequently used senses should have three subcomments on semantics to include these senses. However, depending on the size of the dictionary, the lexicographer could decide beforehand not to go further than, say, three or four senses in any given article. The inclusion of a given sense should satisfy a minimum frequency of usage. This may mean that the metaphoric sense qualifies for inclusion but not the literal sense. If the lexicographer feels it is important that users
should get information on a literal or an old sense of the word, the article structure should make provision for the inclusion of an additional text block in which the complementing data is presented. This should not be an additional subcomment on semantics, because all senses allocated to subcomments on semantics qualify on the grounds of their frequency of use. In modern-day Afrikaans the word *pitkos* is primarily used in a figurative sense, i.e. “substantial food for thought”. The literal sense of this word is “concentrated food like grain”. Very few people actually use, or even know, the word *pitkos* in this sense. In a comprehensive dictionary each one of these senses should be allocated a separate subcomment on semantics but in a dictionary of limited extent only the figurative sense will qualify if frequency of use is the criterion. For the user of such a dictionary it might be necessary to link the figurative sense to a literal one. However, the lexicographer should adhere to the principles governing the sense selection and should not include the literal sense if it does not satisfy these principles, e.g. the frequency of usage criterion.

One of the valuable insights of modern-day metalexicography is that dictionaries do not necessarily have to display a homogeneous article structure. A specific article structure needs to be devised to which the default article should adhere but provision should also be made for motivated deviations from this structure. The introduction of a heterogeneous article structure allows the lexicographer to include some additional or alternative article-specific entries, irrespective of whether they have a lemmatic or a non-lemmatic addressing.

Where the criteria for the selection of polysemous senses for inclusion in the subcomments on semantics exclude a specific sense which the lexicographer deems necessary for the user, the lexicographer may opt for a heterogeneous article structure to include an additional entry complementing the default article structure. This could, for instance, be done by means of an inserted inner text, or a comment block. In the case of *pitkos* the treatment of the figurative sense could be followed by a postcomment in which an indication is given of the original meaning of the word.

**pitkos** s.nw.
Kragtige voedsel vir die gees ...

Oorspr.: gesonde kos wat bestaan uit mieliepitte.

**Dictionary article 1**

If the lexicographer regards this as important data, it should be presented in a position of salience. This demands that the planning of the article structure should make provision for positioning data in such a way that the user can be made aware of it and have unimpeded access to it. In this regard the lexicographer of any dictionary, even a restricted monolingual dictionary, should embrace the notion of a micro-architecture.
2.5 The micro-architecture of dictionary articles

Bergenholtz, Tarp and Wiegand (1999) make a distinction between dictionary articles with and those without a micro-architecture. In an article with an architecture the article-internal items display a left-right as well as a top-bottom relation. The following article from The South African Pocket Oxford Dictionary of Current English does not display a micro-architecture:

ventilat\(e\) ... \(v\). (-ting) 1 cause air to circulate freely in (a room etc.). 2 air (a question, grievance, etc.). 3 \(Med\). \(a\) oxygenate (the blood). \(b\) admit or force air into (the lungs).

**Dictionary article 2**

The introduction of a micro-architecture could lead to the following presentation:

ventilat\(e\) ... \(v\). (-ting)
  1 cause air to circulate freely in (a room etc.).
  2 air (a question, grievance, etc.).
  3 \(Med\). \(a\) oxygenate (the blood). \(b\) admit or force air into (the lungs).

**Dictionary article 3**

From a user perspective dictionary article 3 is more user-friendly than dictionary article 2. This is primarily the result of article 3 displaying four text blocks compared to the single text block of article 2. Because of these text blocks article 3 has a micro-architecture (cf. Bergenholtz, Tarp and Wiegand 1999: 1770). These text blocks play an important role in the recognition of the entries accommodated in them. Bergenholtz, Tarp and Wiegand (1999: 1770) argue that they increase the access confidence of the user and diminish the access time.

In dictionary article 3 the first text block only contains the comment on form. It is separated from the comment on semantics, presented in the following three text blocks where a subcomment on semantics, containing a polysemous sense of the lexical item represented by the lemma sign, appears in each one of the text blocks, preceded by a polysemous marker. The comment on form precedes the comment on semantics in the top-bottom relation and within the comment on semantics the different subcomments on semantics are also in a top-bottom relation.

Following the arguments of Bergenholtz, Tarp and Wiegand (1999: 1771) the higher degree of user-friendliness of dictionary article 3 compared to dictionary article 2 is also evident from the rapid inner access structure flowing vertically in article 3 compared to the sinuous flowing in article 2. The rapid inner access structure of article 3 shows a vertical-architectonic extension, and vertical-architectonic access structures are regarded as straight access structures.
Compare dictionary articles 4 and 5 below. Dictionary article 4 from the electronic version of the Afrikaans monolingual dictionary HAT has no micro-architecture:

**bad**  s.nw. (-dens; -jie) 1 Groot bak waarin 'n mens jou liggaam kan was. 2 Die was van die liggaam in so 'n bak: 'n Bad neem. 3 Hoeveelheid water, vloeistof waarin die liggaam gewas word: jou bad is glad te warm. 4 Bak waarin bv. olie gehou word en waarin 'n deel van 'n masjien werk: Oliebak 5 (mv. baaiie) Bloatstelling van die liggaam aan water, sonstralé, stoom, modder, elektrisiteit, ens.: 'n Sonbad, modderbad. 6 (mv. baaiie) Badplek met geneeskragtige water: Die baie gebruik. Vgl. WARMBAD. UITDR.: Iemand 'n koue bad gee, hom ontgoël. 6 vv. (gebad) Die liggaam in 'n bad was: 'n Mens behoort elke oggend te bad.

**Dictionary article 4**

In dictionary article 5, the above HAT article has been adapted to display a micro-architecture:

**bad**  s.nw. (-dens; -jie) 1 Groot bak waarin 'n mens jou liggaam kan was. 2 Die was van die liggaam in so 'n bak: 'n Bad neem. 3 Hoeveelheid water, vloeistof waarin die liggaam gewas word: jou bad is glad te warm. 4 Bak waarin bv. olie gehou word en waarin 'n deel van 'n masjien werk: Oliebak 5 (mv. baaiie) Bloatstelling van die liggaam aan water, sonstralé, stoom, modder, elektrisiteit, ens.: 'n Sonbad, modderbad. 6 (mv. baaiie) Badplek met geneeskragtige water: Die baie gebruik. Vgl. WARMBAD. UITDR.: Iemand 'n koue bad gee, hom ontgoël. 6 vv. (gebad) Die liggaam in 'n bad was: 'n Mens behoort elke oggend te bad.

**Dictionary article 5**

In dictionary article 5 the comment on form is yet again ordered in a top-bottom relation to the comment on semantics. Within the comment on semantics a vertical-architectonic relation exists between the different subcomments on semantics. Separate text blocks for fixed expressions and the occurrence of the lemma in a different part of speech function follow the subcomments on semantics.

The user-friendly nature of a dictionary article with a micro-architecture is also seen in the following excerpt from the article of the lemma sign **because** in the Collins COBUILD on CD Rom:

**because**  1 because

You use because when stating the reason for something.

He is called Mitch, because his name is Mitchell.

Because it is an area of outstanding natural beauty, the number of boats available
for hire on the river is limited.
Women are doing the job well. This is partly because women are increasingly
moving into a man’s world.
‘Why didn’t you tell me, Archie?’ ‘Because you might have casually mentioned it
to somebody else.’
CONJ-SUBORD
2 because
You use because when stating the explanation for a statement you have just
made.
Maybe they just didn’t want to ask too many questions, because they rented us a
room without even asking to see our papers.
The President has played a shrewd diplomatic game because from the outset he
called for direct talks with the United States.
I had a sense of deja vu because I could recognise everything in London.
CONJ-SUBORD
3 because
If an event or situation occurs because of something, that thing is the rea-
on or cause.
Many families break up because of a lack of money.
Because of the law in Ireland, we had to work out a way of getting her over to Brit-
ain.
PHR-PREP
4 because
You use just because when you want to say that a particular situation
should not necessarily make you come to a particular conclusion. (SPO-
KEN, INFORMAL)
Just because it has a good tune does not mean it is great music.
Just because something has always been done a certain way does not make it right.
PHR-CONJ-SUBORD

Dictionary article 6

Using a micro-architecture does have space and therefore also cost implica-
tions. But it also has implications for the success of the dictionary consultation
procedures of the target users and it can enhance the rapid and unambiguous
retrieval of information and play an important role in fulfilling the lexicog-
graphic functions and the genuine purpose of the dictionary.

3. In conclusion

The compilation of monolingual dictionaries confronts lexicographers with
decisions and dilemmas, and demands a well-disciplined approach to the task.
It is important to regard all dictionaries as sophisticated lexicographic products
and not to argue that a dictionary of limited extent is inferior to its more com-
prehensive family members. Lexicographers need to conduct the planning and
compilation of smaller dictionaries in a similar theoretically-based fashion as
they would use in the planning and compilation of comprehensive dictionaries.
Care should be taken that the data distribution and presentation reflect the needs of the users and the typological criteria of the specific dictionary. A consistent application of these criteria is of extreme importance. Provision should be made for a deviation from the default article structure in order to present lexical item-specific data needed by the user. To ensure the best possible presentation of all data types the use of a micro-architecture will lead to macro benefits for the user.

References


