

# Expectation Levels in Dictionary Consultation and Compilation\*

Inge Kosch, *Department of African Languages,  
University of South Africa, Pretoria, South Africa*  
([koschim@unisa.ac.za](mailto:koschim@unisa.ac.za))

---

**Abstract:** Dictionary consultation and compilation is a two-way engagement between two parties, namely a dictionary user and a lexicographer. How well users cope with looking up words in a Bantu language dictionary and to what extent their expectations are met, depends on their consultation skills, their knowledge of the structure of a Bantu language and the dictionary design. There are also demands that the lexicographer places on the user. These demands are of varying degrees, ranging from low-level expectations, where basic look-up skills according to the letters of the alphabet are assumed to have been mastered already (as needed when using a word-based dictionary), to medium-level expectations (such as being able to look up words in a stem-based dictionary), to high-level demands (where intuitive dictionary skills no longer suffice). A fine balance must be struck between the user's guidelines and the presumed dictionary consultation skill of the target user. This article endeavours to look into the matter of various levels of expectation of Bantu language dictionaries, both from the perspective of the user and the compiler.

**Keywords:** DICTIONARY CONSULTATION, DICTIONARY COMPILATION, EXPECTATION LEVELS, CONSULTATION SKILLS, WORD-BASED DICTIONARY, STEM-BASED DICTIONARY, DICTIONARY DESIGN, TARGET USER, LEXICOGRAPHER, LEMMATISATION

**Opsomming: Vlakke van verwagting in woordeboekgebruik en -samestelling.** Die raadpleging en samestelling van 'n woordeboek behels 'n wisselwerking tussen twee partye, naamlik 'n woordeboekgebruiker en 'n leksikograaf. Hoe goed gebruikers vaar wanneer hulle woorde in 'n Bantoetaalwoordeboek opsoek en die mate waarin daar aan hulle verwagtinge voldoen word, hang af van hulle eie naslaanvaardighede, hulle kennis van die Bantoetaalstruktuur en woordeboekstruktuur. Daar is ook sekere vereistes wat die leksikograaf aan die gebruiker stel. Hierdie vereistes is op verskillende vlakke, wat strek van laevlakverwagtinge waar aangeneem word dat basiese naslaanvaardighede volgens die letters van die alfabet reeds bemeester is (soos nodig wanneer 'n woord in 'n woord-gebaseerde woordeboek opgesoek moet word), tot mediumvlakverwagtinge (waar woorde volgens 'n stam-gebaseerde woordeboek nageslaan moet word), tot hoëvlakverwagtinge (waar intuïtiewe woordeboekvaardighede nie meer toereikend is nie). 'n Fyn balans moet gevind word tussen die gebruikersleiding en die veronderstelde naslaanvaardighede van die teikengebruiker. Hierdie artikel poog om die kwessie van verskillende vlakke van verwagting van Bantoetaalwoordeboeke te beskou, beide uit die perspektief van die gebruiker en die samesteller.

---

\* This article was presented as a paper at the Seventeenth Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by the Department of African Languages, University of Pretoria, Pretoria, South Africa, 2–5 July 2012.

**Sleutelwoorde:** WOORDEBOEKRAADPLEGING, WOORDEBOEKSAMESTELLING, VLAKKE VAN VERWAGTING, NASLAANVAARDIGHEDE, WOORD-GEBASEERDE WOORDEBOEK, STAM-GEBASEERDE WOORDEBOEK, WOORDEBOEKONTWERP, TEIKENGEBRUIKER, LEKSIKOGRAAF, LEMMATISERING

## 1. Introduction

Dictionary consultation and compilation is a two-way engagement between two parties, namely a dictionary user at the consumer end and a lexicographer at the production end. Both the consumer and the producer have mutual expectations when consulting or compiling a dictionary respectively. The nature of these expectations and the level at which they are pitched will be the focus of this article.

Most casual users of dictionaries may not have any other expectation besides establishing the meaning of a word which, according to Svensén (1993: 14), is the highest ranked type of information for which a dictionary is consulted. This is followed by the need to establish correct spelling, pronunciation, grammar, application in context, synonyms and etymology. After just a few consultation attempts, users will be able to tell whether a dictionary is user-friendly or not by the degree of satisfaction or frustration experienced during their look-up activity and the time it took to retrieve required information. How well they cope with looking up words in a Bantu language dictionary and to what extent their expectations are met, are largely dependent on the following three factors, namely

- their consultation skills, that is, their previous exposure to dictionary pedagogy,
- their knowledge of the structure of a Bantu language, and
- the dictionary design itself.

Not all users have been exposed to dictionaries to the same extent and hence they will experience different degrees of difficulty when consulting a dictionary. Abecassis (2007: 248) rightly states that "there is not a prototypical user, but a whole variety of users, with different cultural backgrounds which may affect their abilities".

Users of Bantu language dictionaries, especially inexperienced second language learners, are invariably guided by previous encounters with non-Bantu language dictionaries and will expect to find the required information in the first place they look, namely in the appropriate alphabetical slot according to the first letter of the word. However, they may be confronted with an unknown dictionary design owing to the agglutinative structure of the Bantu languages, which calls for a specialised approach to lemmatisation, different from the European languages, for example.

Lexicographers consciously and purposefully design their dictionaries with certain target users in mind. The demands they place on users' consultation abilities to successfully retrieve information may range from low-level, to medium-level, to high-level expectations.

## 2. Levels of expectation

### 2.1 Low-level expectations

At a low level of expectation the lexicographer can assume that the user, at the very least, has mastered basic look-up skills according to the letters of the alphabet. This requirement is so basic that it is taken for granted among literate people. In the context of the Bantu languages, this means that a user would be able to easily find items that have been lemmatised according to their first letter in a predictable alphabetical order — the so-called word-based approach. The advantage of this approach is that the consultation success is immediate and almost guaranteed without the user first having to resort to the user's guidelines. The word-based arrangement of items makes such dictionaries very user-friendly, particularly for beginners and pupils. At the same time, however, it obscures the connection between inflectionally or derivationally related lemmas that are removed from each other alphabetically as a result of phonological changes as in the case of *-leka* 'try' under the letter 'L' versus its deverbative *tekô* 'test' under the letter 'T' in Northern Sotho.

A word-based dictionary, though user-friendly and popular among beginners, has its limitations as a printed work. Owing to space and cost considerations it can impossibly include all the lemmas in a language. However, such challenges are no obstacle to modern technology, which has revolutionised dictionary compilation and consultation by electronic means as stated by Prinsloo (2005: 11): "The great capacity and speed characteristic of electronic products combined with enhanced query and data retrieval technology, indeed pave the way to a new generation of dictionaries unimagined in the paper-dictionary era."

### 2.2 Medium-level expectations

In agglutinating languages it is customary to compile dictionaries in terms of stems. The fact that expansions containing prefixes and or infixes must be searched for under the appropriate stem, according to Haas (1967: 49) "requires considerably more grammatical sophistication on the part of the user than is ever required, for example, on the part of the user of the ordinary English or French dictionary. Words found in actual texts would have to be analyzed before they could be looked up".

A stem approach assumes that the user is familiar with the morphological structure of words to the extent of being able to identify and then dismiss the

prefixes when looking up lemmas under the first letter of the stem, for example, to access the meaning of *seatla* 'hand' in Northern Sotho, the user is expected to truncate the prefix *se-* and look up the meaning of the word under the first letter of the stem, namely 'A'.

Most dictionaries using the stem approach actually follow a mixed approach of stem and word lemmatisation, as done in the *Comprehensive Northern Sotho Dictionary* of Ziervogel and Mokgokong (1975). Word lemmatisation is applied to nouns with irregular or non-overt class prefixes. These nouns are entered according to the first letter of the word, for example, *mmuši*, 'ruler', a noun in class 1, will be found under the letter 'M', while *pono* 'vision', a noun in class 9, will be lemmatised under the letter 'P'. The look-up success is not immediate, however, as the meaning is not given under *mmuši* or *pono*. Instead a cross-reference is given to the stems *-buša* 'rule' and *-bona* 'see' respectively from which these words have been derived. Though user-friendly, this method is time-consuming as it requires two look-up actions instead of one before the required information is obtained. The mixed design, whereby a word approach is used within a dictionary which is overridingly stem-based, constitutes a compromise to facilitate access by the not so advanced user. Lemmatisation according to stems is less user-friendly than lemmatisation according to words, but, in the opinion of Ziervogel and Mokgokong (1975: Preface, p. 87) it is the only scientific method. This scientific method has come at a price, though, namely the loss of accessibility, and the rise in the degree of difficulty as well as the time in finding desired information.

### 2.3 High-level expectations

High-level expectations can be regarded as those where users are required to apply skills that are language and dictionary specific and where intuitive dictionary skills no longer suffice. Referring once more to the example *mmuši*, it is interesting to note that Endemann (1911) took the scientific approach of stem lemmatisation to the extreme, as he does not list this word at all under its first letter, but instead only under the first letter of the stem. Without knowledge of the phonological changes in the Sotho languages or without a prior in-depth familiarisation with the front matter, the user is at a loss as to where to look up the meaning of this word.

The necessity to complete paradigms may also be experienced by a user as a medium- to high-level challenge. In a paper dictionary completion of paradigms is a practical requirement for space- and cost-saving reasons. Entries are often omitted from the central list, if the lexicographer is of the opinion that the user can complete the relevant paradigms, based on directives given in the user's guidelines or in the treatment of articles themselves. As a case in point, most dictionaries supply limited entries of reflexive forms, and only include those of which the meaning is not simply an accumulation of the meaning of the verb plus 'self', as in the case of *-ithuta* ('study' < teach oneself) and *-itshwara* ('behave' < carry oneself). To avoid duplication of verb stems Snyman, Shole

and Le Roux (1990), for example, have generally omitted verb stems with the reflexive prefix *i-* in the central list. However, for users to be able to find the meaning of reflexive verb stems, the compilers have provided the necessary support in the form of a table of sound changes in the front matter. Kriel (1983), on the other hand, lists a number of reflexive forms under the letter 'I', but provides no instructions on how to look up the meaning of non-listed reflexive forms. In the *English-Zulu Dictionary* by Doke, Malcolm and Sikakana (1971) reflexive forms of verbs are not listed under *zi-*. Instead, the function of this morpheme is briefly mentioned under the entry *zi-* with one example, *ukuzibona* 'to see oneself'. For the rest, the user is expected to find the meaning of reflexive forms by truncating the prefix *zi-*. This is incidentally not such a daunting task in Zulu as in the Sotho languages, where phonological changes accompany the prefixation of the reflexive morpheme.

As far as deverbatives are concerned, Endemann (1911: 36) explains that users can create nouns on their own according to a set pattern. In his front matter he gives one example of a regular formation and one where sound strengthening is involved in the initial sound of a verb stem as in the case of prefix-less classes. Expecting users to form their own deverbatives based on these limited instructions, can probably count as a high-level activity as no further support is provided at that point and no cross-reference is given to page 15 where sound strengthening is explained. Ziervogel and Mokgokong (1975: 117-123), on the other hand, offer a more comprehensive overview of the sound changes in their front matter. Items that users would expect to find in the central list should only be phased out of the central list, if their formation is sufficiently covered in the front matter.

A fine balance must be struck between the instructions given and the presumed dictionary consultation skills of the target user. The effort of the user to access the required information, including consultation of the outer text, should not be perceived as outweighing the informative value gained from the research result. A user-friendly dictionary would enable a user to derive 'missing' forms in a paradigm with the maximum amount of support and the minimum amount of effort. This once again underscores the fact that the front matter is pivotal for effective dictionary consultation, but it is unfortunate that lexicographers can generally not rely on the average dictionary users to consult the guidelines as they "usually move directly to the central list where they try to solve the problem that motivated their search. Too often the search does not lead them to the required destination because the user is not familiar with the system of the dictionary" (Gouws and Prinsloo 2005: 85).

One could also regard as medium- to high-level expectations those where users are required to cope with practices which they do not anticipate, such as the unconventional access alphabet in Ziervogel and Mokgokong (1975) where combinations of letters serve as article stretches. Without an understanding of this design, the user will try to look for a word starting with KG (for example *kgomo* 'head of cattle') before words starting with KO (for example *kolobe* 'pig') under the letter 'K'.

Compilers should make sure that scientific needs are not served to the detriment of the practical needs of users. The level of expectation should be pitched at the level of the target user. In this vein Gouws and Prinsloo (2005: 39) endorse that "[m]odern dictionaries are judged by the success with which the *user* is able to retrieve the sought information and not in the first instance by 'linguistic achievement'". Thus the ultimate test of a good dictionary is how well it can be used. This can be established from extensive empirical user research with or without preliminary drafts of (sections of) the dictionary. Often the compiler can only gauge whether the demands placed on the user were realistic once the dictionary is in active use. Lessons learnt can then be implemented in the revision of the dictionary.

### **3. Pre-empting users' queries and needs**

Dictionaries may seem to provide "too little" information in some respects and "too much" in others. The compiler should address such apparent inconsistencies in the front matter or within article treatments, thereby pre-empting queries most likely to come up in the minds of users. Lexical gaps and semantic elaboration are among the issues that would need to be brought to the user's attention.

#### **3.1 Lexical gaps**

Any perceived gaps in the central list should be accounted for in a lucid manner, such as the practice of paradigm completion as a space-saving device. The formation of reflexive forms of verb stems is a case in point as discussed in section 2.3.

Lexicographers can choose not to include certain lemmas for various reasons. In the *South African Multi-language Dictionary and Phrase Book* (Reader's Digest 1991: 6) the compilers, anticipating queries about the absence of some entries, offer the following as an explanation:

When using the word lists, it is important to remember that no language translates literally into another — and where no simple translation exists in a particular African language, our contributors have chosen not to translate the word in that language.

If a word does not appear in a dictionary, the reason may not necessarily be that it was overlooked. The selection of words may seem arbitrary, but it is governed by incidence of use and need in the fields of thought of the particular society who speaks the language.

#### **3.2 Semantic elaboration**

Dictionary compilation is steeped in culture and therefore the lexicographer

has to take cognisance of all possible meanings of a word if he/she wants to be culturally correct. To this end the user is in some cases provided with unexpected (but important) semantic elaboration.

Although Martin (1967: 156) states that "[i]n the interest of conciseness, we should aim at a single translational equivalent whenever possible", semantic elaboration is unavoidable in some cases. The foreign language learner might, for example, quickly want to look up a translation equivalent for the word 'brother', only to find that a simple equivalent does not exist. Instead, he/she is confronted with a list of possible words, wherein the meaning of 'brother' is shown to be determined by this person's relation to other people in the family, e.g. younger brother, elder brother, mother's brother, father's elder brother, father's younger brother, etc. (e.g. Snyman, Shole and Le Roux 1990: 226-227). The user is thus confronted with more than the expected information and this may slow down successful retrieval of the meaning of the desired word. However, as stated by Nkabinde (2003: 180): "Culture is indispensable in the description of a language in general and in the compilation of a dictionary in particular".

Obviously, in a monolingual dictionary, a lengthy treatment of a lemma like 'brother' would be the norm, as users of the dictionary consult it for this very purpose. Abecassis (2007: 251) points out that high proficiency learners (unlike low proficiency learners) often give preference to monolingual dictionaries "when bilingual dictionaries do not suffice for finding extra information on a word meaning". The relative article length assigned to words like 'brother' is a reflection of its importance in the society.

#### 4. Conclusion

This article considered various levels of expectation, both from the perspective of the user and the compiler. At a low level of expectation basic look-up skills according to the letters of the alphabet are assumed to have been mastered already (as needed when using a word-based dictionary). At a medium level of expectation users are expected to be able to look up words in a stem-based dictionary. High-level expectations are those where the user has to be able to look up words which have undergone phonological changes or to complete paradigms where lemmatisation is limited. The latter is acceptable as long as there are appropriate guidelines given in the front matter. The fine balance that was referred to between the assumed dictionary consultation skill of the user and the expectations of the compiler is aptly summarised as follows by Svensén (1993: 16): "It is not merely the information needs of different categories of user that are of importance. An equally important factor is their ability to find and make use of the information given. Here the lexicographer is forced into a difficult balancing act. On the one hand, the dictionary must contain as much information as possible without becoming unnecessarily bulky; on the other hand, this information must not be so compressed that the intended user will not be able to utilize it."

## References

- Abecassis, M.** 2007. Is Lexicography Making Progress? On Dictionary Use and Language Learners' Needs. *Lexikos* 17: 247-258.
- Doke, C.M., D.McK. Malcolm and J.M.A. Sikakana.** 1971. *English-Zulu Dictionary*. Johannesburg: Witwatersrand University Press.
- Endemann, K.** 1911. *Wörterbuch der Sothosprache*. Abhandlungen des Hamburgischen Kolonialinstituts. Band VII. (Reihe B. Völkerkunde, Kulturgeschichte und Sprachen. Band 4.) Hamburg: L. Friederichsen & Co.
- Gouws, R.H and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Haas, M.R.** 1967. What Belongs in a Bilingual Dictionary? Householder, F.W. and S. Saporta (Eds.). 1967. *Problems in Lexicography*: 45-50. Bloomington: Indiana University.
- Kriel, T.J.** 1983. *Pukuntšu Woordeboek: Noord-Sotho-Afrikaans, Afrikaans-Noord-Sotho*. Third edition. Goodwood: Nasionale Boekdrukkery.
- Martin, S.E.** 1967. Selection and Presentation of Ready Equivalents in a Translation Dictionary. Householder, F.W. and S. Saporta (Eds.). 1967. *Problems in Lexicography*: 153-159. Bloomington: Indiana University.
- Nkabinde, A.C.** 2003. The Implications of Culture for Dictionaries of the African Languages. *Lexikos* 13: 168-182.
- Prinsloo, D.J.** 2005. Electronic Dictionaries Viewed from South Africa. *Hermes, Journal of Linguistics* 34: 11-35.
- Reader's Digest.** 1991. *South African Multi-Language Dictionary and Phrase Book*. Cape Town: The Reader's Digest Association South Africa.
- Snyman, J.W., J.S. Shole and J.C. le Roux.** 1990. *Dikišinare ya Setswana-English-Afrikaans Dictionary/Woordeboek*. Pretoria: Via Afrika.
- Svensén, B.** 1993. *Practical Lexicography. Principles and Methods of Dictionary-Making*. Oxford/New York: Oxford University Press.
- Ziervogel, D. and P.C. Mokgokong.** 1975. *Pukuntšu ye kgolo ya Sesotho sa Leboa, Sesotho sa Leboa-Seburu/Seisimane/Groot Noord-Sotho-woordeboek, Noord-Sotho-Afrikaans/Engels/Comprehensive Northern Sotho Dictionary, Northern Sotho-Afrikaans/English*. Pretoria: J.L. van Schaik.