
Challenges of Predictability and Consistency in the First Comprehensive Sotho Dictionary

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

Abstract: Predictability and consistency are requirements that should run like a golden thread through the macro-, medio- as well as the microstructure of dictionary articles. Adherence to these requirements is one of the marks of a user-friendly reference work that will allow for easy access and trouble-free retrieval of required information. This article aims to investigate some features of Endemann's (1911) *Wörterbuch der Sotho-Sprache* (Dictionary of the Sotho language) with the focus on challenges of predictability and consistency in the lemmatization approach, the access alphabet, cross references and article treatments. The dictionary has hitherto remained at the outskirts of scholarly investigation, one of the reasons being the fact that the target language is German and that, as such, it is not readily accessible to every scholar of the Bantu languages. A further reason is that it is aimed at subject specialists rather than the general public and places a high demand on the lexicographic skill of the user. Some interesting insights can be gained from this example of an early attempt at dictionary compilation and its shortcomings can serve as a springboard for continual improvement of access routes and user-friendliness of dictionaries in the Sotho languages.

Keywords: PREDICTABILITY, CONSISTENCY, COMPREHENSIVE DICTIONARY, SOTHO LANGUAGES, DICTIONARY USER, LEMMATIZATION STRATEGY, CROSS REFERENCES, ALPHABETIZATION, WORD LEMMATIZATION, STEM LEMMATIZATION, DATA CATEGORIES

Opsomming: Uitdagings van Voorspelbaarheid en Konsekwentheid in die Eerste Omvattende Sothowoordeboek. Voorspelbaarheid en konsekwentheid is vereistes wat soos 'n goue draad deur die makro-, medio-, sowel as die mikrostruktuur van woordeboekartikels behoort te loop. Nakoming van hierdie vereistes is een van die kenmerke van 'n gebruikersvriendelike naslaanbron wat maklike en probleemvrye toegang tot verlangde inligting sal verleen. Die doelwit van hierdie artikel is om enkele eienskappe van Endemann (1911) se *Wörterbuch der Sotho-Sprache* (Woordeboek van die Sothotaal) te ondersoek met die klem op uitdagings betreffende voorspelbaarheid en konsekwentheid in die lemmatiseringsbenadering, die toegangsalfabet, kruisverwysings en beskrywing van artikels. Die woordeboek lê buite die kollig van wetenskaplike navorsing, een van die redes synde die feit dat die teikentaal Duits is en dat dit as sulks nie gereedelik toeganklik is vir elke Bantoetaalkundige nie. 'n Verdere rede is dat dit gemik is op vakkundiges eerder as die gewone publiek en hoë vereistes stel aan die

naslaanvaardighede van die gebruiker. Hierdie vroeë poging tot woordeboeksamestelling bied 'n aantal interessante insigte en die tekortkominge kan op hul beurt weer dien as vertrekpunt vir voortdurende verbetering van toegangroetes en gebruikersvriendelikheid van woordeboeke in die Sothotale.

Slutelwoorde: VOORSPELBAARHEID, KONSEKWENTHEID, OMVATTENDE WOORDEBOEK, SOTHOTALE, WOORDEBOEKGEBRUIKER, LEMMATISERINGSTRATEGIE, KRUISVERWYSINGS, ALFABETISERING, WOORDLEMMATISERING, STAMLEMMATISERING, DATAKATEGORIEË

1. Introduction

The aim of this article is to focus on some problems of predictability and consistency with reference to the *Wörterbuch der Sothosprache*, compiled by the Berlin missionary Karl Endemann (1911). The work can be regarded as the first attempt at a comprehensive dictionary, because the term 'Sotho' in the title does not refer to one language only, but to three, namely Northern Sotho (also known as Sesotho sa Leboa or Sepedi), Tswana (Setswana) and Southern Sotho (Sesotho), including their dialects where such are distinguished (that is, in Northern Sotho and Tswana). These languages constitute the Sotho group of languages and are spoken in South Africa and parts of southern Africa. The dictionary is mono-directional with Sotho (as defined above) as the source language and German as the target language. The examples chosen for discussion have been extracted randomly from the dictionary, but with the express purpose of raising some interesting points that can lead to a continual improvement of an optimal retrieval experience for the user.

Defining a dictionary simply as a book containing the words of a language arranged in alphabetical order and describing their meanings, underestimates the complexities associated with dictionary compilation in the Bantu languages. In the first instance, the so-called words are not all words in the sense of autonomous words with independent meanings — in fact a large number of the dictionary entries (or lemmata) are items which are smaller than words and only acquire a 'meaning' or rather 'function' by virtue of their combination with other linguistic elements. Secondly the access alphabet may differ from the traditional alphabet and either display gaps or a proliferation of more article stretches than the 26 letters provided for in the traditional alphabet. Where clusters of two or three letters serve as article stretches, a user may not find a particular word in its expected alphabetic slot. In their *Comprehensive Northern Sotho Dictionary*, Ziervogel and Mokgokong (1975) make extensive use of such clusters, but to the dictionary user, as rightly pointed out by Prinsloo and De Schryver (1999: 261), "it is nothing more than sheer frustration to eventually find for example a word commencing on *bj* alphabetically after *bu* in the dictionary". The reason for this apparent anomaly lies in the fact that the cluster "BJ" occurs as an alphabetical stretch after the letter "B". These are just some of

the unexpected features a first-time user of a Bantu language dictionary will have to contend with.

The dictionary under investigation, though rich in information, did not impact lexicographic practices in South Africa to any noteworthy degree. This situation can in part be ascribed to the fact that it was accessible only to a select number of target users, namely native speakers of German. It also had limited success as a learner's dictionary as it propagated a new, and, in the author's opinion, more scientific orthography and was intended rather for the subject specialist than for the beginner.

Predictability relates to the user being able to find the information where he or she expects to find it. **Consistency** implies that a predetermined pattern is followed in the way the data is presented. These two elements are essential parts of the dictionary conceptualisation plan or style manual. As pointed out by Gouws and Prinsloo (2005b: 16), the result of the lack of a properly planned microstructural programme leads to "the situation where the lexicographers decide in a haphazard way to include a certain data category in a specific article and omit it from the next". They also state that "one of the most annoying experiences in the process of dictionary consultation is to be confronted with an inconsistent presentation of data" (Gouws and Prinsloo 2005b: 9). A random approach to the microstructure with an arbitrary presentation of data negatively impacts the quality and status of the dictionary as an authoritative and accessible source of information (Gouws 2003: 109).

2. Lemmatization strategies

The agglutinative nature of the Bantu languages has led lexicographers to adopt either a *stem* approach or a *word* approach to the lemmatization of words. In *stem lemmatization* words are alphabetized on the first letter of the stem, which means the prefix, if there is an overt prefix, is disregarded. In *word lemmatization* words are alphabetized on the first letter of the word, which would be the first letter of the prefix, if the word has a prefix. Some dictionaries exclusively make use of stem lemmatization for all word categories (e.g. Endemann 1911, Ziervogel and Mokgokong 1975), whereas others make use of both approaches, depending on the word category, e.g. the stem approach for verbs and the word approach for nouns (e.g. Kriel 1983, de Schryver 2007). For a detailed discussion of the difference between stem lemmatization and word lemmatization and the implications they have on user-friendliness, the reader is referred to Prinsloo (2009).

Endemann followed a stem approach like Ziervogel and Mokgokong (1975), but with two major differences (Kosch 2011: 115). Firstly, in the case of nouns, the prefix is not indicated to the right of the stem, but to the left, separated from it by means of a hyphen, for example, under the letter "A" we find the following derivations from *aga* 'build' (Endemann 1911: 55):

- (1) *mo-àxi*, Pl. *va-*,¹ "Erbauer, Ansiedler, Bewohner".
 ↓
se-àxi, Pl. *li-*, "Maurer, der viel und gut baut".
 ↓
axiǝa, Pass. v. *axa*.
 ↓
axiša, Kaus. v. *axa*, "bauen helfen, siedeln helfen, benachbart sein".

The entries are arranged in a way that is appealing to the eye, because the first letter of the stems ("A" in this case) is aligned vertically (as indicated by arrows for the purpose of this discussion). This type of presentation has been referred to as the left-expanded article structure. The benefits of this approach are expressed as follows by Gouws and Prinsloo (2005a: 38):

Giving the complete orthographic words, in spite of a procedure of stem lemmatisation, by including the prefixes in the article-initial position is user-friendly and reassuring, especially to the inexperienced user [knowing] that (s)he is dealing with the right word.

Secondly, the majority of entries are presented as main lemmata. Lemmata which belong together semantically and structurally are thus not presented as sub-lemmata under one lemma sign. For example, *agela* 'build for', *agelela* 'fence in, build up' and *agelelana* 'fence in for each other', are all accorded their own lemma status, just like *aga* 'build', from which they are derived.

The left-expanded article arrangement may have its advantages in that it allows the user to recognize full orthographic words at one glance, but it also presents a challenge. It presupposes knowledge of the prefix-truncation rule, whereby the user needs to be able to identify the prefix of a word (e.g. *mo-* in *moagi* 'builder'), then to remove it successfully and to recognise that the letter under which *moagi* will have to be looked up is "A". Besides this, the user is required to know the morphophonological rules of the language. For example, should he/she want to look up the word *mmelegi* (Endemann: *mmelexi*) 'child bearer/burden bearer' (from *-belega* 'give birth, bear, carry on back'), he/she would have to know that the prefix *mo-* followed by a verb stem commencing on *b*, changes to *mm* (*mo-b* > *mm*), hence *mmelegi* (< *mo-belegi*, Endemann: *mo-velexi*). The word would hence have to be looked up under the first letter of the stem, in this case "B" for *-belegi* (or "V" for *-velexi* according to Endemann's access alphabet as will be explained in section 4). It is clear that lemmatization according to the first letter of the stem is not user-friendly in all cases and requires a higher level of access skill, since "such stems hardly come naturally into the minds of the users, when they look up the words in a dictionary" (Kiang 2005: 269).

3. Cross references

In order to indicate the relationship between lemmata, Endemann used various mechanisms to link the entries either to previous or following entries. Space does not allow for an in-depth discussion of these, but example (2) should serve as an illustration: The relationship of *agela* to *aga* is indicated by means of the description 'Direkt. v. *aχa*' (Directive/Applicative of *aga*). In other cases the derivational relationship of one entry to the next is indicated by means of the marker 'Davon...' (i.e. from whence ...), e.g. (Endemann 1911: 55):

- (2) *aχa*, V. tr., "bauen; Gebäude, Gehege aufführen; aufmauern; sich ansiedeln; horsten". ...

aχela, **Direkt. v. *aχa***, "bauen für ..., bebauen". **Davon**

aχelela, Direkt., "einfriedigen (Garten, Hof usw.), umhegen" ... **Davon**

aχelelana, Rezipr., "im freundlichen Verkehr mit einander sein" (wörtl. "einander umbauen, umhegen").

Space should always be used judiciously in the layout of a dictionary. There is no denying the fact that the many cross references in the dictionary under review take up quite some space. However, given the lemmatization strategy that is adopted here, cross referencing cannot be done away with and is in the interest of users as it alerts them to grammatical and semantic relationships between entries. Endemann made extensive use of cross references of the type illustrated in example (2) above, especially in the case of verbs. In the case of deverbatives, however, the relationship with the basic lemma is not always indicated, e.g. for *moagi* 'builder', *seagi* 'expert builder' and *moagiši* 'assistant builder'² there are no cross references to *aga*. One could argue that Endemann may have regarded cross references in cases like *moagi*, *seagi* and *moagiši* to *aga* as obsolete, because these lemmata appear in close proximity and in the same article stretch as *aga*. Nevertheless, it would have been good pedagogical practice to be consistent with the cross references, especially in cases where the related words are somewhat removed from each other by unrelated lemmata and occupy another slot in the article stretch, where they belong alphabetically, but not semantically.

Where a related entry is accommodated in a different article stretch altogether as a result of morphophonological changes, the user could expect a cross-reference to the word it derives from, but this is not always the case. In *kago* 'building', for example, which appears under the letter "K" and *ikagela* 'build for oneself' which appears under "I", cross references have been given to *aga* and *agela* respectively. However, for *tefo* 'payment' under "T", there is no cross reference to *lefa* 'pay'. Gouws and Prinsloo (2010) rightly observe that alphabetical ordering for the African languages has serious detrimental conse-

quences for grammatical relations. Without cross references, users will be the poorer for their look-up activity.

4. Access alphabet

Regrettably, users very seldom read (still less memorize) the instructions for use (Svensén 1993: 230), when consulting a dictionary. Instead, they go straight to the central list, where they anticipate finding items in strict alphabetical order. It is only when the search proves unsuccessful — either because there are inexplicable gaps in the alphabet or because the words are not found where the users expected to find them — that they might reluctantly resort to the front matter of the dictionary for specific conventions and an explanation regarding the ordering of lemmata. Users of Endemann's dictionary will only grasp the format of the access alphabet once they understand that the orthography was designed to be as close to a phonetic orthography as possible. The reason why the letters 'B', 'D' and 'G', for example, are absent from the access alphabet, is because Endemann did not regard the sounds these orthographic symbols are used for in a phonetic alphabet, as part of the Sotho sound inventory, namely the bilabial plosive stop [b], the apico-alveolar plosive stop [d] and the dorso-velar plosive stop [g] respectively. He ascribed the incorrect usage of these symbols to ignorance of phonetics, e.g. with regard to the letter B, he declares: "Die bisher übliche Schreibung *b* statt *v* im Sotho und Tžöana beruht auf Unkenntnis bezüglich der Phonetik" ("The customary spelling of *b* instead of *v* in Sotho and Tswana used up to now is based on ignorance regarding the phonetics") (Endemann 1911: 3). He proposed other symbols to reflect the correct pronunciation of these three sounds respectively, namely the Greek letter ν (V), an l with a dot as a diacritic under the symbol ! (L) and the Greek letter χ (X).

Standard alphabet:	Endemann:
B, b [b]: bilabial plosive stop	V, ν [β]: bilabial fricative continuant
D, d [d]: apico-alveolar plosive stop	L, ! [l]: palato-alveolar flap (retroflex)
G, g [g]: dorso-velar plosive stop	X, χ [γ , x]: dorso-velar fricative continuant ⁸

From a scientific point of view Endemann's propagation of new symbols may have been a laudable attempt to enforce correct pronunciations, but it also diminished the appeal of the dictionary as a useful learning tool to its potential users. Especially the beginner, oblivious to finer nuances of pronunciation, would probably not have searched for an entry like *-belega* 'give birth' under the letter 'V' but rather under 'B', only to find this letter missing completely from the access alphabet.

Table 1 presents Endemann's access alphabet (lower case letters have been added, because they differ in type in some cases from the capital letters), com-

pared to the traditional alphabet. The rightmost representations in phonetic script indicate all the initial sounds of lemmata accommodated under each letter of the access alphabet (in broad phonetic script, i.e. without indication of ejection, vowel raising, etc.).

Table 1: Access alphabet in Endemann (1911)

Traditional alphabet	Endemann's access alphabet	Initial sounds of lemmata accommodated under letter of access alphabet
A	A/a	[a]
B	–	
C	–	
D	–	
E	E/e	[e, ε, j ⁴]
F	F/f	[ϕ, ϕs, ϕj]
G	–	
H	H/h ⁵	[ɦ, h]
I	I/i	[i]
J	J/j	[dʒ]
K	K/k	[k, kh, kxh]
L	L/l Ḷy/ly	[l, ɹ] [ʒ]
M	M/m	[m]
N	N/n Ṇ/ṇ Ñ/ñ	[n, ɲ] [ŋ ^{ll}] [ŋ]
O	O/o	[o, ɔ, w ⁶]
P	P/p	[p, ph, ps, psh, pʃ, pʃh]
Q	–	
R	R/r	[r]
S	S/s Š/š Ş/ş	[s] [ʃ] [ʂ]
T	T, Th/t, th Ṭ, Th/ṭ, th	[t, th, ts, tsh, tʃ, tʃh] [ṭl, ṭlh]
U	U/u	[u, w ⁷]

V	V/v	[β, β3]
W	–	
X	X/χ X̄/χ̄	[x, ʎ] [h]
Y	–	
Z ^s	Z/z Ž/ž	(dial.: only in one example) (dial.: only in one example)

Endemann only made limited use of clusters of sounds as article stretches, unlike Ziervogel and Mokgokong (1975) who used many digraphs (combinations of two symbols) and trigraphs (combinations of three symbols) in their dictionary. To amplify the symbols Endemann added diacritics like a tilde or a dot. These peculiar combinations as well as the absence of certain letters render the access alphabet unpredictable. Endemann compiled the dictionary according to what he thought would be the most suitable for his intended users, but as shown in Table 1, the access alphabet is not necessarily logical to the user. In compiling dictionaries, lexicographers should therefore be wary of serving technical and scientific considerations above the users' interests.

In two cases Endemann gives two symbols as an article stretch, namely 'T', 'Th' and 'Ṫ', 'Ṭh'. This is a confusing notation, leaving the user to anticipate one of two possibilities: either that the lemmata with the unaspirated initial consonants will be treated exhaustively first, followed by the lemmata beginning with the aspirated consonants, or that the stems with aspirated plosives as initial sounds are integrated alphabetically with the rest of the stems. The latter is indeed the case, which means that an entry such as *taelq* 'command' occurs before *thatanq* 'mutual love', which in turn will precede *timelq* 'loss'. One wonders why two symbols for an article stretch are only given in the case of 'T', 'TH' and 'Ṫ', 'ṬH' and not in some of the other article stretches where the same principle would apply.

5. Data categories

In each language the data dictate the microstructure, but for each article there should at least be an obligatory microstructure and the ordering of the data types should be fixed (e.g. morphological data, part of speech, translation, usage in context, etc.). Endemann did not seem to adhere to a consistent pattern as far as the microstructural components are concerned. The article treatments rather suggest that he improvised as he compiled the dictionary, dictated by the nature of the material that presented itself to him. He was led by intuition in his treatment of each article as to what information to include and what to omit — at least one can accept that he did not record non-attested forms under the urge of completing paradigms. No phonetic transcription is given — the very purpose of the orthography that Endemann proposed, was that the pronunciation would be correctly derivable from the

practical orthography, making a phonetic transcription obsolete.

The grid in Table 2 below captures the data categories encountered in the dictionary for the treatment of the lemmata. Not all slots are consistently filled for every entry: some entries of which the meaning is transparent, have limited treatment and some articles only supply the minimum information such as a translation equivalent, for example:

- (3) p. 53 *vo-qtá* "Torheit" (i.e. negligence, stupidity)

In many cases the relevant sense of a lemma sign is not explained by means of its usage in a cotext; in other cases lengthy explanations, personal experiences, opinions or a word of caution regarding correct pronunciation are recorded as first item under an article stretch, e.g. under the letter "R":

Endemann (1911: 418):

Das *r* des Sotho darf nur alveolar gesprochen werden. Der Europäer der sein *r* velar spricht, muss sich hier Mühe geben, die richtige Aussprache zu lernen.

The *r* of Sotho may only be pronounced as an alveolar. The European who pronounces his *r* in a velar manner, must make an effort to acquire the correct pronunciation. (own translation)

The numbering in the following table corresponds with the paragraphs below the table where each data category is discussed in more detail.

Table 2: Basic indicators of articles

5.1 Origin/source of words
5.2 Part of speech and grammatical data (may include notes on spelling or pronunciation)
5.3 Semantic data (Translation equivalent, paraphrase on meaning, meaning relations); etymological data
5.4 Cotext (Application in cotext/proverbs/idioms); data on cultural value; pragmatic data; usage notes
5.5 Elaboration of articles (i.a. reference to other languages, personal note, reprimand or word of caution)
5.6 Sketches

5.1 Origin/source of words

The first entry after the lemma sign is invariably an indication of where the

item was attested, either in a language, dialect or particular region. This category is marked by bracketed abbreviations, which are explained in the front matter (Endemann 1911: 36, 37). For example, in (4) the word was attested among Tswana-speakers (Tž.). Sometimes more than one origin for an item is included between brackets, e.g. (Tž, Kχ for 'Tswana, Kgatla') or an "etc." is added which leaves the user somewhat perplexed, as he/she would not know what the "etc." refers to, compare example (5).

- (4) p. 232 *mo-kχōkχōma* (Tž.) "Lufttröhre"
(*mokgokgoma*, attested in Tswana: "air pipe")
- (5) p. 49 *aōā!* (P. etc.), Interj., "nein!", *aōa se!* "nein doch!" (erfreut, auch spöttisch). (*aowa*, attested in Pedi, etc. Interjection "no!", *aowa hle* "no, please!" pleased, also mocking)

The author explains that where such bracketed labels occur, this does not necessarily mean that the items are not in use elsewhere. Where the lemma is deemed to occur across the board in the Sotho languages and dialects under investigation, this data category is absent as in example (6).

- (6) p. 233 *vo-kχōlē*, "Ferne, Entfernung" (*bokgole* i.e. distance)

The label marking the origin is thus not an essential part of every article treatment.

5.2 Part of speech and grammatical data

The indication of the part of speech is not altogether transparent or consistent. The part of speech is omitted in the case of nouns with overt class prefixes. This is a space-saving device and probably adopted because the user is expected to know that the presence of a class prefix (cf. *vo-* and *mo-* in examples (7) and (8)) automatically implies a noun. If the noun is countable, the plural class prefix will follow the lemma, compare *va-* in example (8):

- (7) p. 55 *vo-aχišō* "Nachbarschaft". (*boagišo* 'assistance'⁹)
- (8) p. 55 *mo-aχiši*, Pl. *va-*, "Nachbar". (*moagiši* 'assistant builder'¹⁰)

In contrast to these examples, there seems to be no pattern when the class prefix is not overt, as in the case of nouns in class 9. In some cases the nominal category of the noun is marked by a capital letter N, compare example (9), but in other examples the noun is left unidentified for part of speech, compare example (10):

- (9) p. 561 *thōta*, N., Pl. *li-*, "Bug, Erdhügel, Erdwall, kleiner Hügel, steigender Boden, Hochland". Siehe *marota*, *morota*. (i.e. rising land, hunch or hump)

- (10) p. 561 *thope*, Pl. */i-*, "kurzer, platter, breiter Speer" (i.e. short, flat, broad spear)

In irregular nouns, the plural form is written in full, e.g. *vānana* (*banana* 'little girls') in (11) and *nyōákø* (*nywako* 'houses') in (12). Examples like these constitute a challenge to users, because they would not think of looking up *ngwanana* 'little girl' or *ngwako* 'house' under the letter "A", but under the letter "N" for *ng* (Endemann *n*). In (12) the user is given cases of similar sound changes as examples, e.g. *ngwaga* from **moaga* 'year'. This was meant as an educational exercise, but it sidetracks the user and increases his/her look-up time.

- (11) p. 48 *mo-anana*, gespr. *nōanana*, Pl. *vānana*, "Mädchen, weibliches Kind".
 (12) p. 41 *mo-ákø*, gespr. *nōákø*, Pl. *nyōákø*, "Haus". Der Pl. *ny.* ist = *myako* aus *me-ako*; vergl. *moaxa* = *nōaxa*, *anyōa* f. *amyā*, *enyōa* f. *emyā*.

5.3 Semantic data

A translation or explanation of the meaning is present in all article treatments or alternatively the user is referred to another form with a different spelling under which the lemma can be looked up. The meaning explanation follows the word category indicator (where present) and is indicated between double quotation marks, but not where the lemma needs to be circumscribed, e.g.

- (13) p. 253 *molālq̄* (N.). Pl. *me-*, Name einer Baumart, deren Bast zu Stricken geflochten wird und von deren Saft Bier gemacht wird. (Name of a type of tree, of which the bark is plaited into cords and of which beer is made from its sap.)

There are a fair number of entries about which uncertainty exists regarding the precise meaning equivalent. This negatively affects the user's confidence because his or her consultation effort does not reach a satisfactory conclusion. The semantic gaps are probably due to the fact that Endemann prepared the dictionary for publication after his return to Germany, where he was not able to obtain missing information in the field among the Sotho speakers themselves, but had to rely only on his notes and information from some informants who were in Germany at the time, cf. example (14):

- (14) p. 163 *mo-kikiana* (Tž), Pl. *me-*, Name einer hohen Baumart (näheres bietet die Quelle nicht). (Name of a tall type of tree — the source does not provide more detail).

It is evident that the final version of the dictionary should have been completed in the field for increased accuracy and completeness of information.

5.4 Cotext

Information that may be encountered in the treatment of an article is its application in a cotext (such as in idiomatic expressions or proverbs), e.g. (15).

- (15) p. 54 *ma-ávola* "Anteile". *Leǰú ke maávola, le ávotžǒǒ* "Tod ist Anteil (d.h. das bestimmte, zuerteilte Los), er ist zuerteilt". (Redensart beim Tode eines Kindes).

("Shares". "Death is a share, i.e. the designated, allocated destiny), it is dished out." (Expression at the death of a child).

Such examples are regrettably in the minority, although the meaning equivalents of most lemmata are explained as they were applied in the traditional milieu. Applications in cotexts would have been a great learning tool.

5.5 Elaboration of article

Even though Endemann provides users' guidelines in the front matter, he relegated many observations to the individual entries themselves, rendering some dictionary articles disproportionately long. The dictionary articles thus often provide more than the user bargains for, and this interferes with the user's information retrieval process. As pointed out by Gouws and Prinsloo (2005b: 170): "... the lexicographer should realise that users very seldom want to read through an article. The typical dictionary consultation procedure is aimed at finding one specific item or data type". Descriptions should be short and compact. Endemann particularly falls prey to digression. In the case of the verbal particle *ke* (p. 150) "it is", for example, he takes pains to point out its wrong application based on a wrong translation in German, taking up more than one column of space.

The treatments of some lemma-signs are used as a forum to give lengthy explanations of phonological rules, rules which were evidently felt not to have been addressed adequately in the front matter. Furthermore, contrary to good lexicographic practice, the author expresses personal preferences and judgments regarding wrong perceptions and pronunciations, often as the first component of an article.

In a number of articles the compiler demonstrates his wide knowledge of other languages outside the Sotho group and one wonders how much the user gains through comparisons with other languages, compare the reference to the languages Nyamwezi, Yao and Zulu in example (16):

- (16) p. 167 *kǒàla* (S.-S.), V. tr. = *tzǒàla*, "verschliessen". Vergl. Nyamw.: *lugala*, Yao: *ugala*, Zu.: *vala*.

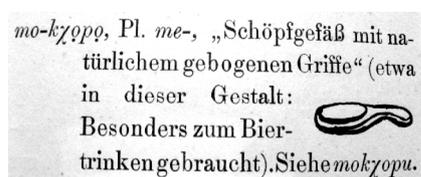
Over-completeness where too much extra-linguistic information is given, can

be counterproductive.

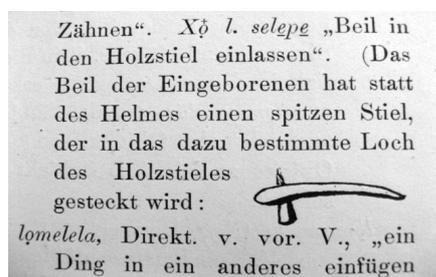
5.6 Sketches

If one decides to support certain data with illustrations, this approach should be followed consistently throughout a dictionary and not in an ad hoc way. In Endemann's dictionary, only two instances were found where a sketch formed part of the article. The sketches depict cultural objects or tools.

Sketch 1: *mo-kχopo* (Endemann 1911: 241)



Sketch 2: *selepe* (Endemann 1911: 281)



Obviously space is a restricting factor, but then again one wonders why the sketches were limited to these particular two entries. Were they just entered as an afterthought or because the description was felt to be inadequate? What about the many other lemmata in the dictionary referring to cultural objects, unknown in the culture of the target user, which could equally have benefited from a pictorial illustration for extra-linguistic information? Landau (1989: 112) mentions that "Zgusta sees the primary purpose of illustrations as that of depicting unusual or unfamiliar things".

6. Space saving strategies/Visual presentation

Limitations of space and cost restrictions will always remain a major con-

sideration while lexicographers are bound to the paper dictionary. The lemmata in Endemann's dictionary are displayed in alphabetical order in two columns on a page, the columns being separated by a vertical line. In some instances the treatment of a particular lemma exceeds its allotted space within the given column and spills over into the line of the previous or next lemma. In such a case it is separated from the treatment of the other lemma by a square bracket, as can be seen in Sketch 3a, where the phrase "Maurer, der viel und gut baut" (Builder that builds much and well) spills over onto the next line. In Sketch 3b, the translation of *mo-köalava* "steiniger Rücken" (stony ridge) is completed in the preceding line. This may be a convenient space saving strategy on the side of the publishers, but to the user it may convey the message of poor dictionary planning and slow down his or her look-up time.

Sketch 3: Overlapping of lines (Endemann 1911: 55)

- (a)
-
- mo-ázi*, Pl. *va-*, „Erbauer, Ansiedler, Bewohner“.
se-ázi, Pl. *li-*, „Maurer, der viel und
áziá, Pass. v. *áza*. [gut baut“.
áziá, Kaus. v. *áza*, „bauen helfen,
siedeln helfen, benachbart sein“.
Davon
ázišana, Rezipr., „einander bauen
helfen, einander siedeln helfen,“

- (b)
-
- le-köalava* (S.-S.), Pl. *ma-*, „Berg“.
Vergl. *mokóló*. [Rücken“.
mo-köalava (S.-S.), Pl. *me-*, „steiniger
köqlela (P.), Direkt. v. *köqala*, „wehe-
tun (intr.)“. (Wörtl. „empfindlich
werden für . . .“).
köqlela (N.), Direkt. v. *köqala*, = *tzöa-*
lala, „verschließen“.
köqletza (P.), Faktit. v. *köqlela*, „wehe-
tun (tr.)“.

7. Conclusion

Predictability and consistency play a crucial role in optimising the user's effective and trouble-free retrieval of information. The word approach is the most user-friendly, because it functions in a predictable way as far as the access alphabet is concerned. However, it is also the most space-consuming in a paper dictionary. At face value, Endemann's approach looks like a word approach and does not create unnatural forms, because entries can be recognised as full orthographic words with prefixes preceding their stems. However, all the entries are still lemmatized on their stems. It may be true that the left-expanded article structure boosts the confidence of users as they are able to recognise full

words at one glance, but this approach requires more than just basic reference skills. It expects the user to have an understanding of the word structure and phonological changes operating in the particular language. For any scholar, especially someone with no prior knowledge of the structure of a Bantu language, the look-up time will be slowed down significantly. Before being able to look up the meaning of words, the user would first have to consult the explanatory introduction which should form part of any good dictionary (Lombard, 1970: 12).

Using a dictionary as a forum to introduce a new orthography based on pronunciation, proved to be a deterrent to users as it complicated access to the central list. Already the first encounter with the access alphabet diminishes a user's appreciation of the work. Landau (1989: 76) aptly observes that "before a dictionary can be written for a language, the language must have developed more or less standard spellings". Differences in the data categories accommodated in the treatment of articles were observed, with some data categories absent, or discussed at length, leading to imbalanced entries.

The solution to the problem of user-friendly dictionaries for the Bantu languages, and one which dispels the greatest obstacles of the lexicographer, obviously lies in the electronic dictionary, where lemmata can be listed according to the first letter of the word in the case of nouns without space saving concerns. Illustrations, especially of cultural objects, of which sadly only two appear in Endemann's dictionary, can be of great historical value for posterity and with today's technology, these could be facilitated by pop-up functionalities from within an entry, without affecting the layout of the article in any way.

Gouws and Prinsloo (2005a: 34) mention examples of Bantu language dictionaries that employed a lemmatisation strategy resembling left-expanded article structures and that went unnoticed in South African Bantu language lexicography (e.g. the *Concise SiSwati Dictionary* by Rycroft, 1982 and the *Southern Sotho-English Dictionary* by Mabile and Dieterlin, 1988). Endemann's work could just as well have been included here, as it followed the same approach and "went unnoticed". However, it deserves recognition, despite its shortcomings as far as predictability and consistency in certain respects are concerned, because it was the very first dictionary compiled for the Sotho languages that used the left-expanded approach for lemmatisation. The shortcomings in the dictionary that were pointed out during the discussion can hopefully serve as a springboard for continual improvement of access routes and user-friendliness of dictionaries in the Sotho languages.

Endnotes

1. Examples extracted from Endemann are given in the author's orthography.
2. Endemann's meaning equivalent of *moagiši* in German is *Nachbar* "neighbour". The correct word for neighbour is *moagišani*, but the latter is not given by Endemann.

3. Regarding the voiced and voiceless variants of χ ([Y] and [x] respectively) Endemann did not see the need to distinguish them in writing in the practical orthography (1911:8).
4. For the palatal glide [j] Endemann uses the symbol $\dot{\epsilon}$.
5. Endemann compromises here for space saving reasons and uses the letter "H" for stems which are also encountered elsewhere in the Sotho languages with an initial "F" ([f]) or "X" ([x] or [Y]).
6. For the labio-velar glide [w] Endemann uses the symbol δ .
7. Only one example where [w] is represented as \ddot{u} .
8. The sounds represented by "Z" result from the omission of a preceding υ or \imath according to Endemann. The precise phonetic representation cannot be established from the information given.
9. The German translation equivalent given for *boagišo* is 'neighbourhood' instead of 'assistance'.
10. The translation equivalent in German is given as *Nachbar* which means 'neighbour'. See also endnote 2.

References

- De Schryver, G.-M. (Editor-in-chief).** 2007. *Oxford Bilingual School Dictionary: Northern Sotho and English/ Pukuntšu ya Polelopedi ya Sekolo. Sesotho sa Leboa le Seisimane. E gatišitšwe ke Oxford.* Cape Town: Oxford University Press Southern Africa.
- Endemann, K.** 1911. *Wörterbuch der Sothosprache.* Abhandlungen des Hamburgischen Kolonialinstituts, Band VII. (Reihe B. Völkerkunde, Kulturgeschichte und Sprachen Band 4). Hamburg: L. Friedrichsen & Co.
- Gouws, R.H.** 2003. Aspekte van mikrostrukturele verskeidenheid en inkonsekwentheid in woordeboeke. *Lexikos* 13: 92-110.
- Gouws, R.H. and D.J. Prinsloo.** 2005a. Left-Expanded Article Structures in Bantu with Special Reference to IsiZulu and Sepedi. *International Journal of Lexicography* 18(1): 25-46.
- Gouws, R.H. and D.J. Prinsloo.** 2005b. *Principles and Practice of South African Lexicography.* Stellenbosch: SUN PReSS.
- Gouws, R.H. and D.J. Prinsloo.** 2010. Lexicography Workshop Notes: Dictionaries — the Theory and Practice of the Lexicographic Process. Workshop held at the University of Pretoria, 10–11 May 2010.
- Kiango, J.G.** 2005. Problems of Citation Forms in Dictionaries of Bantu Languages. *Nordic Journal of African Studies* 14(3): 255-273.
- Kosch, I.** 2011. Innovation and Compromise in K. Endemann's Dictionary of the Sotho Language (1911). *South African Journal of African Languages* 31(1): 110–120.
- Kriel, T.J.** 1983. *Pukuntšu Woordeboek (Noord-Sotho–Afrikaans, Afrikaans–Noord-Sotho).* Pretoria: J.L. van Schaik.
- Landau, S.I.** 1989. *Dictionaries: The Art and Craft of Lexicography.* Cambridge: Cambridge University Press.
- Lombard, D.P.** 1970. Bantoetaalstudie: Dictionaries in Northern Sotho. *Bantu Education Journal* 16(3): 12-13.

- Prinsloo, D.J.** 2009. Current Lexicography Practice in Bantu with Specific Reference to the Oxford Northern Sotho School Dictionary. *International Journal of Lexicography* 22(2): 151-178.
- Prinsloo, D.J. and G.-M. De Schryver.** 1999. The Lemmatization of Nouns in African Languages with Special Reference to Sepedi and Cilubà. *South African Journal of African Languages* 19(4): 258-275.
- Svensén, B.** 1993. *Practical Lexicography: Principles and Methods of Dictionary-Making*. Oxford/New York: Oxford University Press.
- Ziervogel, D. and P.C. Mokgokong.** 1975. *Groot Noord-Sotho Woordeboek/Comprehensive Northern Sotho Dictionary*. Pretoria: J.L. van Schaik/Unisa.