

---

# Supplying Syntactic Information in a Quadrilingual Explanatory Dictionary of Chemistry (English, Afrikaans, isiZulu, Sepedi): A Preliminary Investigation\*

Elsabé Taljard (*etaljard@postino.up.ac.za*) and Rachéle Gauton (*rgauton@postino.up.ac.za*), Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa

---

**Abstract:** In response to a number of recent government publications on the promotion of science and technology training, the *Suid-Afrikaanse Akademie vir Wetenskap en Kuns* (South African Academy for Science and Art) initiated the compilation of a multilingual explanatory dictionary for chemistry. The need for such a dictionary is especially urgent in learning environments where learners receive tuition through a medium other than their mother tongue. This is particularly relevant for learners whose mother tongue is an African language. Taking into account the target user's fragmented knowledge of the subject field and low second language competence, the planners decided that certain syntactic information should be specified for the isiZulu and Sepedi entries; specifically the part of speech to which a lemma belongs. This poses certain problems for the lexicographer, since the issue of word categorisation is, especially in Sepedi, an unresolved one. A second problem that presents itself to the lexicographer is the lexicographic convention which is used to indicate the word class to which a specific lemma belongs. Abbreviations referring to the parts of speech are normally used for this purpose. No standardised abbreviations for the different parts of speech exist in isiZulu or Sepedi. Principles for the formation of abbreviations have also not been formulated for the official orthographies of these languages. The only solution to this problem is to study existing abbreviations in order to abstract the principles of abbreviation formation. Only then can abbreviations indicating the parts of speech be formulated for use in the dictionary.

**Keywords:** SPECIAL FIELD DICTIONARY, CHEMISTRY, ISIZULU, SEPEDI, WORD CATEGORISATION, LEXICOGRAPHICAL CONVENTION, ABBREVIATIONS, MULTILINGUAL DICTIONARY

**Opsomming: Die insluiting van sintaktiese inligting in 'n viertalige verkla- rende chemiewoordeboek (Engels, Afrikaans, isiZulu, Sepedi): 'n Voorlopige ondersoek.** In reaksie op 'n aantal onlangse regeringspublikasies rakende die bevordering van wetenskap- en tegnologieopleiding, het die Suid-Afrikaanse Akademie vir Wetenskap en Kuns die

---

\* This paper was presented at the Fifth International Conference of the African Association for Lexicography, held at the University of Stellenbosch, Stellenbosch, 3-5 July 2000.

inisiatief geneem om 'n viertalige verklarende chemiewoordeboek saam te stel. Die behoefte aan so 'n woordeboek is veral dringend in leeromgewings waar leerders onderrig in 'n ander taal as hul moedertaal ontvang. Dit is veral van toepassing op leerders wat 'n Afrikataal as moedertaal het. Met die teikengebruiker se gefragmenteerde kennis van die studieveld en swak tweedetaalvaardigheid in ag geneem, het die beplanners besluit dat bepaalde sintaktiese inligting ook ten opsigte van die isiZulu- en Sepedi-inskrywings gespesifiseer moet word; veral die woordklas waartoe 'n lemma behoort. Dit veroorsaak sekere probleme vir die leksikograaf, aangesien die aangeleentheid van woordkategorisering veral in Sepedi, 'n onopgeloste een is. 'n Tweede probleem waarmee die leksikograaf gekonfronteer word, is die leksikografiese konvensie wat gebruik word om die woordkategorie waartoe 'n spesifieke lemma behoort, aan te dui. Afkortings wat na die woordklasse verwys, word normaalweg vir hierdie doel gebruik. In nóg isiZulu, nóg Sepedi bestaan daar egter gestandaardiseerde afkortings vir die verskillende woordklasse. Geen beginsels vir die vorming van afkortings is ook nog in die amptelike ortografieë van hierdie tale geformuleer nie. Die enigste oplossing vir hierdie probleem is om bestaande afkortings te bestudeer ten einde die beginsels vir die vorming van afkortings daaruit te abstraher. Slegs dan kan afkortings wat die woordklasse aandui, geformuleer word vir gebruik in die woordeboek.

**Slutelwoorde:** SPESIALISWOORDEBOEK, CHEMIE, ISIZULU, SEPEDI, WOORDKATEGORISERING, LEKSIKOGRAFIESE KONVENSIE, AFKORTINGS, VEELTALIGE WOORDEBOEK

## 1. Introduction

The compilation of the *Quadrilingual Explanatory Dictionary of Chemistry* (QEDC) is a project which was initiated by the *Suid-Afrikaanse Akademie vir Wetenskap en Kuns* (South African Academy for Science and Art) in response to a number of recent Government policy publications in which strong emphasis is placed on education and training in the field of science, engineering and technology. Especially relevant in this regard are the *White Paper on Reconstruction and Development* of 1994, *South Africa's Green Paper on Science and Technology* (s.a) and the *White Paper on Science and Technology* of 1996. The key concept in these papers is the development of human potential through optimal access to information. Carstens (1997: 2) argues "that multilingual, explanatory special-field dictionaries can be implemented to realise these objectives by providing easy access to new or incompletely learnt concepts".

## 2. Role players

The initial planning of the dictionary project was to a large extent done by Prof. A. Carstens of the Department of Afrikaans and Prof. D.J. Prinsloo of the Department of African Languages, both from the University of Pretoria, in collaboration with Dr Van Zyl de Villiers of the Atomic Energy Corporation and Dr M. Alberts of the National Language Services. An article providing

detailed information on the planning of the project "Issues in the Planning of a Multilingual Explanatory Dictionary of Chemistry" was published by Carstens in 1997 in *Lexikos* 7.

In the aforementioned article, it is indicated that a so-called "multispectral" approach to data-collection for the QEDC should be followed, involving numerous role players from different fields. Since it is a chemistry dictionary, consultation with experts in the various fields of chemistry is a prerequisite. The multilingual aspect necessitates the involvement of mother-tongue speakers of Afrikaans, isiZulu and Sepedi; not only as lexicographers and terminologists, but also as teachers, lecturers and experts in the chemistry field. Carstens (1997: 11) states that the involvement of mother-tongue speakers in various capacities serves a number of purposes. Firstly, it can legitimise the dictionary amongst the speakers of the various language groups; secondly, it keeps the lexicographer/terminologist in touch with the linguistic reality, thus preventing him/her from coining terms where terms already exist; and thirdly, it ensures continuous monitoring of the relevance and usefulness of the dictionary with regard to the communicative functions it has to fulfil.

### 3. Target users

The heterogeneous nature of the target users of the QEDC impacts on all aspects of the dictionary and the planning thereof. The target user of this dictionary has been identified as either a learner in the senior secondary school phase (grades 10-12), or a pregraduate university or technikon student in the physical sciences. This user group is characterised by varying levels of encyclopaedic knowledge of chemistry and second language competence. One could safely assume that the average target user would have a fragmented knowledge of the conceptual system of the subject field and, generally speaking, an average to low second language competence. Learners who have one of the African languages as mother-tongue are further disadvantaged by the fact that they receive their tuition through the medium of English and not, as is mostly the case for Afrikaans-speaking learners, through the medium of their mother-tongue.

### 4. Function of the dictionary

Carstens (1997: 10) refers to three possible communicative functions which are to be fulfilled by a dictionary, viz. **production** (native language or foreign language), **reception** (native language or foreign language) or **translation** (into or from a foreign language). In the case of the QEDC, the reception function would be of primary importance, since the majority of target users are taught in English, which is their second or even third language. The appearance of both term and concept definition in Afrikaans, isiZulu and Sepedi would assist

in the cognitive function of conceptualisation. Production would only come at a later stage, with translation being a long-term possibility. Bergenholtz and Tarp (1995: 25) do however warn that where a special field dictionary is designed for users with different mother-tongues, an integration of the above functions presents problems.

## 5. Format of the dictionary

After an extensive needs assessment, it was decided that English should be the language of lemmatisation in the primary word list. All the encyclopaedic and linguistic information supplied for the English lemma would then also be supplied for the translational equivalents in Afrikaans, isiZulu and Sepedi. Compare example (1)<sup>1</sup> by way of illustration:

### (1) English

**acid** (*n*) A substance that produces an excess of H<sub>3</sub>O<sup>+</sup> (hydronium) ions when it dissolves in water, e.g. HCl (hydrochloric acid); HNO<sub>3</sub> (nitric acid); CH<sub>3</sub>COOH (acetic acid). *Opposite* **base**. ♣ **acidic** (*a*); **acidify** (*v*).

### Afrikaans

**suur** (*n*) 'n Stof wat 'n oormaat H<sub>3</sub>O<sup>+</sup>-ione (hidroniumione) vorm wanneer dit in water opgelos word, bv. HCl (soutsuur); HNO<sub>3</sub> (salpetersuur); CH<sub>3</sub>COOH (asynsuur). *Teenoorgestelde* **basis**. ♣ **suur** (*a*); **aansuur** (*w*).

### IsiZulu

**i-asidi** (*bz*) Isiqo noma into ekhiqiza ama-ayoni ehayidroniyamu (H<sub>3</sub>O<sup>+</sup>) ngokweqile uma sinyibilika emanzini, izib. HCl (i-asidi yehayidroklorikhi); HNO<sub>3</sub> (i-asidi initric); CH<sub>3</sub>COOH (i-asidi emuncu). *Umqondophika* **isiqu**. ♣ **-neasidi** (*isich.*); **-ba yiyasidi** (*sbnj.*) / **-muncisa** (*sz*)

### Sepedi

**esiti** (*l*) Selo se se tšweletšago diayone tša haedroniamo (H<sub>3</sub>O<sup>+</sup>) tše di feleletšego ge di tologa ka meetseng, mohl. HCl (esiti ya haedrokloriki); HNO<sub>3</sub> (esiti ya naetriki); CH<sub>3</sub>COOH (esiti ya asetiki). *Lelatodi* **peisi**. ♣ **-nago esiti** (*tlha*); **esitifatša** (*led*).

Reverse word lists for the other three languages will then be supplied after the primary word list. The QEDC thus shares some features with translation dictionaries. The reason for selecting English as the language of primary lemmatisation is that lexical gaps abound in especially isiZulu and Sepedi, and to a lesser extent in Afrikaans. Swanepoel (1989: 244) points out there tend to be lexical gaps in languages of developing communities and/or languages which have only recently become standard languages. He contends that in such cases it is advisable to compile a translation dictionary in which the language(s) containing the lexical gaps is (are) used as the target language(s). The language

from which such languages obtain their scientific and technical information, in this case English, then serves as the source language. In this manner the lexical gaps in the so-called "younger language(s)", or "languages of limited diffusion (LLDs)" as they are also called, can be identified more easily.

## 6. Two phases of dictionary compilation

The dictionary is compiled in two phases, the first of which has been completed. During this first phase  $\pm 500$  chemistry terms and their conceptual definitions were terminologically and lexicographically processed for English and Afrikaans and entered in an electronic database. The second phase consists of adding the isiZulu and Sepedi equivalents of both terms and definitions. It was decided during the planning stages that all encyclopaedic and linguistic information supplied for the English lemmas would also be supplied for the translation equivalents. With reference to the inclusion of linguistic information in a chemistry dictionary, Carstens (1997: 15) indicates that traditional monolingual science dictionaries provide very little linguistic information, the reason being that "many dictionaries of science and technology have been designed by experts in the subject-field without any linguistic background or interest in grammar". Experts in the natural sciences may consider grammar as less important and not crucial for communication in a scientific or technical domain, thus putting a higher premium on economy, conceptual precision and accessibility. Carstens (1997: 15), however, argues convincingly that a dictionary of the type in question is not adequate without linguistic information such as part of speech, morphological formation, contextual information and information on linguistic and stylistic usage. The decision to include this type of linguistic information was apparently motivated by the low second language proficiency of the target user.

Furthermore, as was stated earlier, the QEDC, being a multilingual term dictionary, shares certain features with translation dictionaries. With reference to translation dictionaries, Swanepoel (1989: 234) argues that an indication of the part of speech to which the main lemmas belong is of the utmost importance, especially where no examples providing information on the syntactic behaviour of the lemma are supplied. This would seem to be the case with the QEDC and other technical dictionaries where the minimum of syntactic information is provided. He further contends that there is a direct correlation between the amount of grammatical and syntactic information supplied, and the envisaged functions of the dictionary. If the aim of the dictionary is to assist the user in **decoding** texts, then less syntactic information is supplied than would be supplied if the aim of the dictionary is to enable the user to **generate** texts. Since the ultimate aim of the dictionary is to enable the target user to generate texts in the target language, the inclusion of syntactic information is an essential aspect of the QEDC.

Although the usefulness of such an approach cannot be questioned, providing linguistic information does pose certain practical problems regarding the entering of Sepedi and isiZulu equivalents. For the purpose of this discussion, it was decided to focus on one aspect of linguistic information to be provided in the dictionary, viz. specifying the part of speech to which every lemma belongs. It should be noted that the problems which have been identified are primarily of a **linguistic nature**, but they do have terminological and/or lexicographical implications not only for the compilation of the QEDC, but perhaps even to a larger extent for the compilation of future monolingual dictionaries.

## 7. Specification of the part of speech to which a lemma belongs

Two problems regarding the specification of the part of speech to which a lemma belongs have been identified. The first is one with which linguists studying the indigenous South African languages have long been grappling, viz. that of word categorisation. The second concerns the lexicographic convention used to indicate to which word class a specific lemma belongs.

### 7.1 Word categorisation in Sepedi and isiZulu

Firstly, specifying the part of speech to which a specific lemma and/or its derivatives belong, is a relatively simple exercise in Afrikaans and English. Unfortunately, this is not the case in the African languages, especially in Sepedi. Due to the different approaches to word identification and categorisation, two sets of word classes or parts of speech are currently recognised in grammars of Sepedi. The one, with which most Sepedi linguists are probably familiar, is that formulated by Van Wyk (1961: 69 et seq.), and adapted by Lombard (1985). This classification has dominated the Sepedi linguistic scene in South Africa over the past four decades, since it represents the first attempt at scientifically identifying the word classes of especially the Sotho languages. In explaining the theoretical principles on which his identification of the parts of speech are based, Van Wyk (1961: 69 et seq.) emphasises the fact that all linguistic features of words must be taken into account when words are to be classified into word categories. Consequently, he distinguishes four principles according to which the words of a language should be classified, namely syntactic, morphological, phonological and semantic principles. He regards the syntactic and morphological principles as dominant principles of classification, since they take the word into account as a complete linguistic sign, i.e. as a unit of sound and meaning. By making use of the dominant principles of classification, the linguist can identify the essential features of a specific word category. Phonological and semantic principles are termed subsidiary principles of classification, the use of which leads to the identification of the additional features of word categories. Van Wyk distinguishes the following eight word classes

based on these principles:

- (2) Van Wyk's word classes (Kosch 1994: 59, 60):
- I. Words with normal phonological structure
    - A. Substantives
      1. Nouns
      2. Pronouns
    - B. Predicatives
      3. Verbs
      4. Locative copulative demonstratives  
(Interjective demonstratives)
    - C. Morphologically heterogeneous words
      5. Adverbs
      6. Particles
  - II. Words with paranormal phonological structures
    7. Ideophones
    8. Interjections

A ninth class "conjunctions" was added at a later stage, presumably by Lombard (1985). This system of word categorisation is not without its problems, as has been pointed out by Louwrens (1994: 133, 134). The recognition of a word class "particles", which implies that these linguistic units enjoy autonomous word status, is especially problematic and has been questioned by a number of scholars. In 1994, in their publication *A Linguistic Analysis of Northern Sotho*, Poulos and Louwrens (1994: 9, 10) propose an alternative set of word classes. Where Van Wyk's classification is based mainly on the structural features of words, Poulos and Louwrens' classification is based on the function of words in sentences, thus showing a strong Dokeian influence:

Words perform various functions when they are used in sentences, and on the basis of these functions, they are classified into the different word categories, or to use the traditional term, "parts of speech" (Poulos and Louwrens 1994: 9).

The parts of speech they distinguish for Sepedi are the following:

- (3) Poulos and Louwrens' word classes (Poulos and Louwrens 1994: 10):
- |      |                   |  |
|------|-------------------|--|
| i)   | The noun          |  |
| ii)  | The pronoun       | Absolute<br>Quantitative                           |
| iii) | The demonstrative |  |
| iv)  | The qualificative | Adjective<br>Possessive<br>Relative<br>Enumerative |
| v)   | The verb          |  |
| vi)  | The copulative    | Identifying<br>Descriptive                         |

- vii) The adverb Associative  
Locational  
Numerous subtypes (formed in many different ways)
- viii) The ideophone
- ix) The interjection
- x) The conjunction
- xi) The interrogative

The problem of two systems of word-class categorisation in Sepedi is illustrated in (4) below, using the term "constant" as an example:

(4)

| English  | Part of speech | Afrikaans   | Part of speech | Sepedi                       | Part of speech  |
|----------|----------------|-------------|----------------|------------------------------|---|
| constant | Adjective      | konstant(e) | Adjective      | -sa fetogego<br>(NST&O: 134) | Relative verb<br>(Van Wyk) or<br>qualificative<br>(Poulos and Louwrens) |

Due to the different classificatory criteria being used, the contents of the word classes of Van Wyk differ from that of Poulos and Louwrens. An example such as *-sa fetogego* 'constant' is regarded as a verb within the Van Wyk framework, based on the fact that its structure consists of *inter alia* a subject concord and a verb stem *-fetog(a)*. Based on its qualificative function however, Poulos and Louwrens (1994: 90) classify examples such as these as qualificatives, since they serve to qualify or describe a noun. It is also clear from example (4) that word categories in Afrikaans and English do not correspond with word categories in Sepedi.

The problem of word-class categorisation does not occur to the same extent in isiZulu as in Sepedi, as Doke's (1927) word-class classification is generally accepted for isiZulu.<sup>3</sup> Compare (5) which shows his slightly revised classification:

(5) Doke's word classes (Doke 1945<sup>4</sup>: 34):

- I. Substantive: (a) Noun 1.  
(b) Pronoun 2.
- II. Qualificative: (a) Adjective 3.  
(b) Relative 4.  
(c) Enumerative 5.  
(d) Possessive 6.
- III. Predicative: (a) Verb 7.  
(b) Copulative 8.
- IV. Descriptive: (a) Adverb 9.  
(b) Ideophone 10.
- V. Conjunctive 11.
- VI. Interjective 12.

From the lexicographer's point of view, a more pressing problem regarding word-class categorisation is similar to that found in Sepedi, namely that the word categories of isiZulu do not correspond to those of Afrikaans and English. Compare for example:

(6)

| English  | Part of speech | Afrikaans   | Part of speech | IsiZulu                                  | Part of speech |
|----------|----------------|-------------|----------------|--|----------------|
| constant | Adjective      | konstant(e) | Adjective      | -ngaguqukiyo<br>(ZT&O <sup>+</sup> : 60) | Verb           |

As can be deduced from the Sepedi and isiZulu examples (4) and (6), this problem is especially relevant in the case of adjectives (and this also applies to adverbs), since there is very little correspondence between these parts of speech in English and/or Afrikaans and their equivalents in Sepedi and isiZulu.

The problem concerning the specification of word categories is further compounded by the extensive use of paraphrases in isiZulu and Sepedi to supply translational equivalents for English terms. Zgusta (1971: 319), as well as Al-Kasimi (1983: 60), terms such paraphrases "explanatory" or "descriptive" equivalents, as opposed to "translational" or "insertible" equivalents. Whereas a translational equivalent can be immediately inserted into a target-language (TL) sentence, the explanatory equivalent cannot always be directly inserted into such a TL sentence.

### 7.1.1 The use of paraphrases

As was stated earlier, the QEDC and translation dictionaries have certain features in common, which means that the compilers of the QEDC experience problems similar to those of the compilers of translation dictionaries. The underlying cause of these problems is the lack of equivalence or "anisomorphism"<sup>5</sup> between languages. Swanepoel (1989: 205) points out that a consequence of the basic anisomorphism between languages is that few multilingual dictionaries have been compiled. He argues that the problems relating to establishing equivalence between languages increase proportionally with the number of languages being incorporated into such multilingual dictionaries. It must be pointed out, however, that there are far fewer problems caused by the lack of equivalence between languages in the case of multilingual scientific term dictionaries such as the QEDC, than is the case with (multilingual) general language dictionaries. Pinchuck (1977: 165) points out that scientific language is specialised and tends to become more so, in contrast with the versatility of ordinary language. Scientific language seeks the most economic use of linguistic means in order to achieve standardisation of terms and usage. It also seeks to avoid ordinary language associations and endeavours to define its terms accurately. It is therefore far easier to fit a term directly into a target-language text than would be the case with a word from the general vocabulary.

Languages such as isiZulu and Sepedi — so-called languages of limited diffusion or LLDs — make extensive use of paraphrases to fill onomasiological gaps that occur because scientific and technological terminology in the source language (in this case English), do not exist in the target languages, i.e. isiZulu and Sepedi. Consider (7) and (8) in this regard:

- (7) **IsiZulu**
- (a) **Activated complex:** Ukuphithana okubukhuphekhuphe / okuxwayile (lit. complex (lit. entanglement) that is active)
  - (b) **Activation energy:** Umdlandla wobukhuphekhuphe (lit. enthusiasm/keenness of activity)
  - (c) **Acid anhydride:** I-anidrayidi ye-asidi (lit. anhydride of acid)
  - (d) **Chemical:** (adj): -kwesayensi yekhemistri (lit. of (pertaining to) the science of chemistry) (ZT&O: 51)
- (8) **Sepedi**
- (a) **Activation energy:** Mafofolo a maatla (lit. enthusiasm of power)
  - (b) **Chemical:** (adj): sa khemisi (lit. of (pertaining to) the science of chemistry) (NST&O: 134)
  - (c) **Constant:** (adj): -sa fetogego (lit. which does not change) (NST&O: 134)

The problem for the compilers of the QEDC regarding word-class categorisation is that these paraphrases consist of more than one (linguistic) word and can therefore not be said to belong to a single word category.

### 7.1.2 A possible solution for the problems regarding word-class categorisation

As was indicated earlier, two problems regarding word-class categorisation were identified. The first of these problems concerns the existence of two competing word-class categorisations for Sepedi. The lexicographer must take cognisance of this issue and its linguistic implications and will therefore have to make a choice as to the word categorisation to be used. The most logical option would be to select the system of word categorisation with which the target user is familiar. This will in all probability be the categorisation which is used in the teaching of Sepedi in secondary schools. It will therefore be the responsibility of the lexicographer to ascertain which of the two systems enjoy preference in the teaching of Sepedi at secondary school level in order to enable him/her to make an informed choice in this regard.

Secondly, the extensive use of paraphrases in Sepedi and isiZulu in order to supply translational equivalents for English terms, creates difficulties for the lexicographers of the QEDC in that a term consisting of more than one linguistic word does not easily lend itself to satisfactory categorisation. The compilers of the English and Afrikaans sections of the dictionary were clearly also con-

fronted with similar problems. Compare the examples taken from the QEDC given in (9) below:

(9) **English**

**homologous series** (*n*) A series of compounds containing common structural elements, but differing in the number of atoms making up the molecule, e.g. alkanes such as  $C_nH_{2n+2}$ .

**Afrikaans**

**homoloë reeks** (*n*) 'n Reeks verbindings wat dieselfde struktuurelemente het, maar verskil ten opsigte van die aantal atome waaruit die molekule bestaan, bv. alkane soos  $C_nH_{2n+2}$ .

In both the English and Afrikaans entries, the lemma consists of an adjective followed by a noun. However, the part of speech is simply indicated as (*n*), apparently since the lemma refers to a single concept.

A similar solution can be applied in the case of the isiZulu and Sepedi paraphrases as listed in examples (7) and (8) earlier. The isiZulu translational equivalents in examples (7)(a), (7)(b) and (7)(c) could all be labelled *noun*, whilst the example in (7)(d) could be labelled *qualificative*, even though the lemmas themselves consist of more than one linguistic word. This would be acceptable since the referent of the lemma in question is a single concept in each case. The same principle can be applied to the Sepedi examples in (8). Even though, linguistically speaking, the Sepedi term in (8)(a) consists of three linguistic words, it can be labelled *noun*, since it refers to a single concept. The examples in (8)(b) and (8)(c) can be labelled as *qualificatives*, since they serve to qualify whichever noun might precede them. This underscores the difference between words as found in general language and terms used in languages for special purposes. The lemma in general language dictionaries consists of a single linguistic word which lends itself easily to categorisation, whereas a term is often complex or may even constitute a phraseological unit as illustrated in (7) and (8) above. It is therefore clear that in terms of categorisation, words and terms cannot be handled in the same manner.

Earlier, we identified two problems regarding the specification of parts of speech in the QEDC. Whereas the first problem (as discussed in the previous sections) dealt with word categorisation in Sepedi and isiZulu, the second problem concerns the lexicographic conventions to be used in indicating the word-class affiliation of a particular lemma. This issue will be discussed in detail in the next section.

**8. Lexicographic conventions used to indicate to which word class a specific lemma belongs**

The second problem to be addressed in this article concerns the use of abbreviations to indicate the part of speech to which a specific lemma belongs. A standardised list of abbreviations for word categories does not exist in either

Sepedi or isiZulu. Furthermore, no rules on the formation of abbreviations are explicitly stated in dictionaries or in official orthographies of these languages. This makes the formation of new abbreviations very difficult.

Therefore, the only viable option available to us was to abstract principles for the formation of abbreviations from existing examples of abbreviation. The first step was to obtain as many examples of isiZulu and Sepedi abbreviations as possible. The official terminology and orthographies of these languages, as well as various dictionaries, were consulted.<sup>6</sup> It must be noted, however, that this was by no means an exhaustive study, and that it should be seen as constituting a preliminary investigation only. After these examples had been studied, a number of rules determining the formation of abbreviations could be formulated. These principles were then implemented in drawing up lists of proposed abbreviations for the various word categories of Sepedi and isiZulu. Although a number of abbreviations referring to the isiZulu parts of speech are used in the three monolingual explanatory dictionaries (Nkabinde 1982 and 1985, and Nyembezi 1992), it was found that there are word categories which are not referred to in abbreviated form in these dictionaries. By identifying the principles underlying abbreviation formation, it was possible to devise abbreviations for all the word categories of isiZulu.

The suggested principles for abbreviation formation appear as appendixes A (isiZulu) and B (Sepedi). The proposed abbreviations for the word classes/parts of speech for isiZulu and Sepedi are listed in (10) and (11) respectively.

In example (10), Doke's (1927: 34-35) so-called "fundamental or basic parts of speech" (identified according to the work that they do in the sentence) are given in upper case characters. The so-called "real parts of speech" (identified according to the form in which they appear) are written in lower case characters. Also note that in example (10), the parts of speech for which abbreviations already exist, appear in bold type. These existing abbreviations have been taken from the isiZulu monolingual dictionaries of Nkabinde (1982 and 1985) and Nyembezi (1992). Lastly, note that in the formation of the proposed abbreviations in (10), **any** of the identified principles of abbreviation formation (as set out in appendix A) could have been used. The suggested abbreviations have therefore been chosen relatively arbitrarily, but cognisance was taken of (a) the shape of those abbreviations already in use in Nkabinde's and Nyembezi's dictionaries, and (b) the frequency of occurrence of each principle of abbreviation formation as set out in appendix A. Consider:

(10) **Proposed Abbreviations for the Word Classes/Word Categories of isiZulu**

| ENGLISH          | ISIZULU           | ABBREVIATION | PRINCIPLE |
|------------------|-------------------|--------------|-----------|
| SUBSTANTIVE      | USOBIZO           | usbz.        | 1         |
| <b>noun</b>      | <b>ibizo</b>      | <b>bz.</b>   | 1         |
| <b>pronoun</b>   | <b>isabizwana</b> | <b>sbz.</b>  | 1         |
| QUALIFICATIVE    | ISICHASISO        | isich.       | 4         |
| <b>adjective</b> | <b>isiphawulo</b> | <b>ph.</b>   | 1         |

|                  |                    |             |   |
|------------------|--------------------|-------------|---|
| <b>relative</b>  | <b>isibaluli</b>   | <b>bl.</b>  | 1 |
| enumerative      | inani              | nan.        | 3 |
| possessive       | ubumnini           | ubumn.      | 4 |
| PREDICATIVE      | ISILANDISO         | slds.       | 1 |
| <b>verb</b>      | <b>isenzo</b>      | <b>sz.</b>  | 1 |
| copulative       | isibanjalo         | sbnj.       | 1 |
| DESCRIPTIVE      | ISIKHANYISO        | skhny.      | 1 |
| <b>adverb</b>    | <b>isandiso</b>    | <b>sd.</b>  | 1 |
| <b>ideophone</b> | <b>isenzukuthi</b> | <b>szk.</b> | 1 |
| CONJUNCTIVE      | ISIHLANGANISO      | isihl.      | 4 |
| INTERJECTIVE     | ISIBABAZO          | bbz.        | 1 |

Note that in the case of the suggested Sepedi abbreviations in (11), word categories not distinguished by Van Wyk (1961), but listed by Poulos and Louwrens (1994), are marked with an asterisk:

(11) *Proposed Abbreviations for the Word Classes/Word Categories of Sepedi*

| ENGLISH                  | SEPEDI     | ABBREVIATION      | PRINCIPLE |
|--------------------------|------------|-------------------|-----------|
| Noun                     | Leina      | Lna. / L.         | 3 / 1     |
| Pronoun                  | Lešala     | Leš.              | 2         |
| Verb                     | Lediri     | Led. / Ldr.       | 2 / 1     |
| Demonstrative-copulative | Lešupileba | Ll.               | 1         |
| Adverb                   | Lehlathi   | Lehl.             | 2         |
| Particle                 | Sekantšu   | Sek. / Sn. / Skn. | 2 / 1 / 1 |
| Conjunction              | Lekopanyi  | Lkp.              | 1         |
| Ideophone                | Leekiši    | Lki.              | 3         |
| Interjection             | Lelahlelwa | Lla.              | 3         |
| Demonstrative            | *Lešupi    | Lpi.              | 3         |
| Qualificative            | *Tlhaodi   | Tlha.             | 2         |
| Copulative               | *Leba      | Leb.              | 2         |
| Interrogative            | *Lebotšiši | Ltši.             | 3         |

Clearly, not all of the listed abbreviations will be relevant for use in the QEDC, but we are of the opinion that they could be of use in future, specifically when compiling monolingual dictionaries for isiZulu and Sepedi. This study will therefore contribute not only to lexicography in the African languages, but also to isiZulu and Sepedi linguistics. The principles that have been formulated could also be helpful to other grammarians and linguists for the formation of abbreviations in these languages.

## 9. Conclusion

It was found that the inclusion of syntactic information in the isiZulu and Sepedi entries of the QEDC, led to certain practical lexicographic problems. These problems are caused by the unique linguistic structure of the African

languages and cannot be solved without considering the grammatical structure of these languages as well as the unresolved linguistic issues present in these languages.

## Notes

1. The isiZulu and Sepedi equivalents of this entry are provisional.
2. *Northern Sotho Terminology and Orthography* (1988).
3. This does not mean, however, that other word-class categorisations have not been suggested for this language, *inter alia* those of Van Eeden (1956), Cope (1957), Ziervogel (1964), Van Wyk (1958, 1961 and 1967) and Nkabinde (1975), to name the most important classifications.
4. *Zulu Terminology and Orthography* (1980).
5. Anisomorphism < **an-** 'not' + **-iso-** 'equivalent' + **-morph** 'form' (Swanepoel 1989: 205).
6. These works are listed under (a) in the bibliography.

## Bibliography

### A. Dictionaries, terminologies and orthographies

- Dent, G.R. and C.L.S. Nyembezi. 1977. *Scholar's Zulu Dictionary*. Pietermaritzburg: Shuter and Shooter.
- Department of Education and Training. 1980. *Zulu Terminology and Orthography*. Pretoria: The Government Printer.
- Department of Education and Training. 1988. *Northern Sotho Terminology and Orthography*. Pretoria: The Government Printer.
- Doke, C.M., D.M. Malcolm, J.M.A. Sikakana and B.W. Vilakazi. 1990. *English-Zulu/Zulu-English Dictionary*. Johannesburg: Witwatersrand University Press.
- Nkabinde, A.C. 1982. *Isichazamazwi 1*. Pietermaritzburg: Shuter and Shooter.
- Nkabinde, A.C. 1985. *Isichazamazwi 2*. Cape Town: Oxford University Press.
- Nyembezi, S. 1992. *AZ Isichazamazwi sanamuhla nangomuso*. Pietermaritzburg: Reach Out Publishers.

### B. Other literature

- Al-Kasimi, A.M. 1983. *Linguistics and Bilingual Dictionaries*. Leiden: Brill.
- Bergenholtz, H and S. Tarp. 1995. *Manual of Specialised Lexicography*. Amsterdam: Benjamins.
- Carstens, A. 1997. Issues in the Planning of a Multilingual Explanatory Dictionary of Chemistry for South African Students. *Lexikos* 7: 1-24.
- Cope, A.T. 1957. The Grammatical Structure of Zulu. *African Studies* 16: 210-220.
- Doke, C.M. 1927. *Text Book of Zulu Grammar*. Johannesburg: The University of the Witwatersrand Press.
- Doke, C.M. 1945<sup>4</sup>. *Text-Book of Zulu Grammar*. London: Longmans, Green and Co. Ltd.
- Kosch, I. M. 1994. *A Historical Perspective on Northern Sotho Linguistics*. Pretoria: Via Afrika.
- Lombard, D.P. 1985. *Introduction to the Grammar of Northern Sotho*. Pretoria: J.L.van Schaik.
- Nkabinde, A.C. 1975. *A Revision of the Word Categories in Zulu*. Unpublished D.Litt. et Phil. thesis. Pretoria: University of South Africa.

- Pinchuck, I.** 1977. *Scientific and Technical Translation*. London: Deutsch.
- Poulos, G. and L.J. Louwrens.** 1994. *A Linguistic Analysis of Northern Sotho*. Pretoria: Via Afrika.
- South African Government.** s.a. *Green Paper on Science and Technology*. Pretoria: Government Printer.
- South African Government.** 1994. *White Paper on Reconstruction and Development*. Pretoria: Government Printer.
- South African Government.** 1996. *White Paper on Science and Technology*. Pretoria: Government Printer.
- Swanepoel, P.H.** 1989. *Only Study Guide for LEKPER-Q*. Pretoria: University of South Africa.
- Van Eeden, B.I.C.** 1956. *Zoeloe-grammatika*. Stellenbosch/Grahamstad: Die Universiteitsuitgewers en -Boekhandelaars.
- Van Wyk, E.B.** 1958. *Woordverdeling in Noord-Sotho en Zoeloe*. Unpublished D.Litt. thesis. Pretoria: University of Pretoria.
- Van Wyk, E.B.** 1961. Die woordklasse van Noord-Sotho. *Feesbundel vir Prof. Dr. Jan Antonie Engelbrecht*: 69-85. Johannesburg: Afrikaanse Pers Beperk.
- Van Wyk, E.B.** 1967. Word Classes: Northern Sotho. *Lingua* 17: 230-260.
- Zgusta, L.** 1971. *Manual of Lexicography*. The Hague: Mouton.
- Ziervogel, D.** 1964. Bantoetaalstudie in Suid-Afrika. Van der Merwe, H.J.J.M. (Ed.). 1964. *Studierigtings in die Taalkunde*: 208-224. Pretoria: Van Schaik.

## Appendix A The Formation of Abbreviations in isiZulu

Some 73 abbreviations were identified in the dictionaries and terminology and orthography listed under (a) in the bibliography, and the following principles of abbreviation formation were abstracted from these examples:

### PRINCIPLE 1: (Approx. 38% of all examples.)

The first letter of one or more syllables of the word is used. This principle also applies to abbreviations of compound nouns and where more than one word is abbreviated. Where two or more separate words are abbreviated, the first letter of every *stem* is used, e.g.:

| ENGLISH                          | ISIZULU                               | SOURCE   |
|----------------------------------|---------------------------------------|----------|
| C.O.D. (cash on delivery)        | K.F. (Khokha i-fika)                  | ZT&O: 55 |
| C.W.O. (cash with order)         | K.O. (u-kheshe ne-oda)                | ZT&O: 67 |
| G.C.M. (greatest common measure) | F.V.K. ((i)-fektha (i)m-vama-en-kulu) | ZT&O: 99 |

It is interesting to note that in the case of syllables containing nasal compounds, the nasal in the compound is never used in the abbreviation, but rather the consonant that follows the nasal. (However, this does not seem to be the case with the so-called "vocative forms", as is illustrated further on.) See the following examples in this regard, where the nasal compounds concerned have been highlighted:

| ENGLISH                        | ISIZULU   | SOURCE        |
|--------------------------------|---|---------------|
| cwt. (hundredweight)           | ct. (i- <b>ce-nta</b> -li)  | ZT&O: 67      |
| lb. (pound)                    | pd. (i- <b>pha-wu-ndi</b> )   | ZT&O: 120     |
| L.C.M. (least common multiple) | P.V.C. ((isi) <b>phindo-</b> (i)m <b>vama-nci</b> )<br>Take note of the nasal compound in the last syllable where the consonant (not the nasal) is used, i.e. n + c > ngc | ZT&O: 120     |
| aux. (auxiliary)               | sgsz. (i- <b>si-nga-se-nzo</b> )  | Nyembezi 1992 |

### PRINCIPLE 2: (Approx. 34% of all examples)

The abbreviation is borrowed from English, i.e. the same abbreviation as in English is used, e.g. a.m., p.m., i.e., cm., P.S.

### PRINCIPLE 3: (Approx. 5% of all examples)

The first syllable of the stem is combined with the first letter of the next syllable, e.g.:

| ENGLISH       | ISIZULU                        | SOURCE    |
|---------------|--------------------------------|-----------|
| doz. (dozen)  | daz. (i- <b>da-zi</b> -ni)     | ZT&O: 78  |
| h (height)    | phak. (uku- <b>pha-ka</b> -ma) | ZT&O: 103 |
| sec. (second) | sek. (i- <b>se-ke</b> -ndi)    | ZT&O: 178 |

**PRINCIPLE 4:** (Approx. 5% of all examples)

The (full) class prefix is followed by the first letter of the stem. These abbreviations all seem to be of words of a general type commonly used in textbooks, i.e. "for example", "answer", "exercise" and "figure". See:

| ENGLISH        | ISIZULU                     | SOURCE                                  |
|----------------|-----------------------------|---|
| e.g.           | izib. (izi-bonelo)          | Dent and Nyembezi 1977: ix;<br>ZT&O: 82 |
| ans. (answer)  | umph. ( <b>um-phumela</b> ) | ZT&O: 27                                |
| ex. (exercise) | ums. ( <b>um-sebenzi</b> )  | ZT&O: 87                                |
| fig. (figure)  | umd. ( <b>um-dwebo</b> )    | ZT&O: 91                                |

**Vocative Nouns / Terms of Address:**

IsiZulu vocative nouns / terms of address seem to form a class of their own. They only conform to the general principles in the broadest of terms. Rather, they seem to follow their own set of rules as determined by their morphological structure, e.g.:

| ENGLISH         | ISIZULU   | SOURCE  |
|-----------------|---|---|
| Mr.             | Mnu. ( <b>Mnumzane</b> )                            | Dent and Nyembezi 1977: ix                            |
| Rev. (Reverend) | Mfu. ( <b>Mfundisi</b> )<br>Mf. ( <b>Mfundisi</b> ) | ZT&O: 171<br>Dent and Nyembezi 1977: ix;<br>ZT&O: 133 |
| Miss            | Nks. ( <b>Nko-sa-za-ne</b> )                        | Dent and Nyembezi 1977: ix                            |
| Mrs.            | Nkz. ( <b>Nko-si-ka-zi</b> )                        | Dent and Nyembezi 1977: ix                            |

No further general principles could be abstracted from the remainder of the examples.

**Punctuation:**

It would seem that in isiZulu a full-stop is generally used after every abbreviated word. This is also the case with (conjunctively written) compound nouns that consist of more than one word, e.g.:

| ENGLISH                        | ISIZULU                                    | SOURCE    |
|--------------------------------|--|-----------|
| H.C.F. (highest common factor) | F.V.K. (i-fektha + (i)-mvama<br>eli-khulu) | ZT&O: 106 |
| cub. yd. (cubic yard)          | yd. kb. (i-ya-di- + (i)-kyu-bhu)           | ZT&O: 66  |
| sq. ml. (square mile)          | ml. sk. (i-ma-ye-la + (i)si-<br>kwele)     | ZT&O: 189 |

Where the Zulu abbreviation has been borrowed from English, however, the same punctuation as in English is used, e.g. a.m., i.e., cm, g, ha, P.S.

## Appendix B The Formation of Abbreviations in Sepedi

The only Sepedi source in which abbreviations are listed, is *Northern Sotho Terminology and Orthography* (1988). It must be pointed out that in this specific source there is no consistency regarding the use of full-stops after abbreviations. The original punctuation as it appears in the said work has thus been retained, pending further research into this matter.

**PRINCIPLE 1:** (Accounts for approx. 57% of all examples)

The first letter(s) of all of/some of the syllables of the word is (are) used:

| ENGLISH       |              | SEPEDI    |              |
|---------------|--------------|-----------|--------------|
| Full form     | Abbreviation | Full form | Abbreviation |
| Wednesday     | Wed.         | Laboraro  | Lbr.         |
| post meridiem | p.m.         | thapama   | tp.          |
| height        | ht.          | bogodimo  | bg.          |

In the case of compounds, the first letter(s) of the words forming the compound should be used. This is also the case where two or more separate words are abbreviated:

| ENGLISH          |              | SEPEDI             |              |
|------------------|--------------|--------------------|--------------|
| Full form        | Abbreviation | Full form          | Abbreviation |
| etcetera         | etc.         | bjalobjalo         | bj.bj        |
| cash on delivery | C.O.D.       | kešhe kamogedišong | k.k.         |

**PRINCIPLE 2:** (Accounts for approx. 33% of all examples)

The first syllable only, or the first syllable and the first letter(s) of the following syllable are used:

| ENGLISH   |              | SEPEDI     |              |
|-----------|--------------|------------|--------------|
| Full form | Abbreviation | Full form  | Abbreviation |
| Monday    | Mon.         | Mošupologo | Moš          |
| dozen     | doz.         | tosene     | tos.         |
| Reverend  | Rev.         | moruti     | mor.         |
| figure    | fig.         | seswantšho | ses.         |

**PRINCIPLE 3:** (Accounts for approx. 10% of all examples)

First and last letters of the word and in some cases, the first letter(s) of other syllables should be used:

| ENGLISH   |              | SEPEDI     |              |
|-----------|--------------|------------|--------------|
| Full form | Abbreviation | Full form  | Abbreviation |
| Mister    | Mr.          | Morena     | Mna          |
| Mistress  | Mrs.         | Mohumagadi | Mdi          |

**Note:** When a syllable starts with hl, the use of h in the abbreviation should be avoided. A better option for the abbreviation of Labohlano would therefore be Lno or possibly Lbo, which would be in accordance with principle 3.