Defining Formats and Corpus-based Examples in the General Ndebele Dictionary, *Isichazamazwi SesiNdebele* [*]

Langa Khumalo, *Ndebele Lexicography Unit, African Languages Research Institute, University of Zimbabwe, Harare, Zimbabwe (langa@arts.uz.ac.zw)*

Abstract: In this article the writer evaluates the defining formats that were used in defining headwords in the first monolingual General Ndebele Dictionary, *Isichazamazwi SesiNdebele* (ISN). The emphasis in the ISN was on the concept of user-friendliness. The article establishes that defining formats in the ISN are a judicious mixture mainly of the defining formats of the Collins Birmingham University International Language Database (COBUILD) and of what has been referred to as *traditional* formats. The first part of this article is an analysis of the decisions taken by the ISN editors in formulating their defining formats. It assesses the COBUILD defining principle vis-à-vis its application in defining headwords in the ISN and the impact of this principle on the user-friendliness of the dictionary. It further discusses other formats, including the decision to retain *traditional* defining formats for defining headwords. One of the traditional defining styles agreed upon was that the editors were to give the hypernym in the case of semantic sets, and then to identify the concept being defined by specifying aspects that distinguish it from others of its type. The second part of the article evaluates the importance and use of the corpus in providing both definitions and examples for the ISN. However, it is further argued that since a corpus has to be "representative" in terms of size in order to be appropriately used as basis for such corpus-based dictionaries, the ISN editors whose corpus was relatively small, could not avoid relying on intuitive knowledge in constructing some examples.

Keywords: COBUILD, CORPUS, DEFINITION, DEFINING FORMATS, TRADITIONAL FORMATS, EXAMPLES, USER-FRIENDLINESS, COTEXT, CONCORDANCE.

Opsomming: Definieerpatrone en korpusgebaseerde voorbeelde in die Algemene Ndebelewoordeboek, *Isichazamazwi SeSiNdeble*. In hierdie artikel beoordeel die skrywer die definieerpatrone wat in die eentalige Algemene Ndebelewoordeboek, *Isichazamazwi SesiNdebele* (ISN) gebruik is. Die klem in die ISN was op die konsep gebruikersvriendelikheid. Die artikel toon aan dat definieerpatrone in die ISN 'n oordeelkundige mengsel is hoofsaaklik van die definieerpatrone van die Collins Birmingham University International Language Database

---

(COBUILD) en van waarna verwys word as *tradisionele* patrone. Die eerste deel van hierdie artikel is 'n ontleding van die besluite geneem deur die redakteurs van die ISN by die formulering van hul definieerpatrone. Dit beoordeel die definieerbeginsel van COBUILD ten opsigte van die aanwending daarvan by die definiëring van trefwoorde in die ISN en die uitwerking van hierdie beginsel op die gebruikersvriendelikheid van die woordeboek. Verder bespreek dit ander patrone, inbegrepe die besluit om *tradisionele* definieerpatrone te behou vir die omskrywing van trefwoorde. Een van die tradisionele definieerstyle waaroor ooreengekom is, was dat die redakteurs die hiperniem sou gee in die geval van semantiese stelle, en dan die konsep wat gedefineer word, te identifiseer deur die spesifisering van aspekte wat dit onderskei van ander van sy soort. Die tweede deel van die artikel beoordeel die belangrikheid en gebruik van die korpus in die verskaffing van sowel definisies as voorbeelde vir die ISN. Daar word egter verder geredeneer dat, aangesien 'n korpus "verteenwoordigend" moet wees in terme van grootte ten einde toepaslik gebruik te word as basis vir sulke korpusgebaseerde woordeboeke, die redakteurs van die ISN wie se korpus relatief klein was, nie kon vermy om op intuïtiewe kennis te steun by die maak van sommige voorbeelde nie.

Sleutelwoorde: COBUILD, KORPUS, DEFINISIE, DEFINIEERPATRONE, TRADISIONELE PATRONE, VOORBEELDE, GEBRUIKERSVRIENDELIKHEID, KOTEKS, KONKORDANSIE

1.   Introduction

A simple definition of a dictionary is a book or reference work that lists words found in a language in alphabetical order and gives their meanings or definitions, and these definitions may or may not include examples. It is, therefore, common or basic knowledge that a dictionary lists words and their meanings. The importance of a definition in a dictionary cannot, therefore, be overemphasized. It is an intrinsic component of a dictionary. In fact, defining can be considered as one of the most important aspects of the dictionary-making process. It is as a result of this rigorous process that users can interact with the definition and know what a particular word means as explained in it.

This makes defining formats — a format being "the way in which something is arranged and presented" (Sinclair 1990: 218) — very crucial in composing definitions. Defining formats are pertinent in so far as they help the compilers or editors to write clear, specific, systematic and culturally relevant definitions that make dictionaries user-friendly. These are formats that provide the editor with certain guidelines to follow in constructing a definition, hence they contribute to consistency in the treatment of related sets of headwords.

Examples can also be seen as an important part of the definition. This is so because they further help to illuminate the definition. With their aid the editor can write a clear, specific definition, which is culturally and contextually relevant. The article will, therefore, evaluate the importance and use of the corpus in providing both definitions and examples in the *Isichazamazwi SesiNdebele* (ISN). According to Sinclair (1991: 171), "a corpus is a collection of naturally-occurring language texts, chosen to characterize a state or variety of a language". The ISN is largely a corpus-aided dictionary. If instances of usage from

the corpus are studied, it is found that the surrounding words and phrases help considerably in determining the meaning. Each word, therefore, occurs in its own special environment, called its *cotext*, and by studying the cotext it can often be established which meaning is intended. Since concordances provide real text examples, hence have a status which made-up examples do not have, the ISN editors obtained most of their examples from the corpus. This, it will be argued, provided the ISN with "more natural" examples. However, since a corpus has to be "representative" in terms of size in order to be appropriately used as the basis for such corpus-based dictionaries, the ISN editors whose corpus was relatively small, could not avoid relying on intuitive knowledge in constructing some examples.

2.   Definitions and Defining Formats in the ISN

As has already been stated in the introduction, a definition is an intrinsic component of any dictionary. A definition is itself "a statement of the meaning of a word" (Waite 1998: 163) which is distinct in outline. It explains the meaning and use of a particular word in a particular speech community. It thus projects the "world-view" of a people. Two facts result from these observations: firstly, that a mother-tongue speaker is usually best equipped to formulate effective definitions of words in a particular language; and, secondly, that it is a rigorous process that follows set principles so as to provide a brief, distinct and conceptually appropriate statement of the meaning of a particular word.

Defining can be considered as one of the most important processes in dictionary making. It is the process through which users may become more aware of the meanings and explanations of words. It is important for users to know the meanings of words so that they can use the words in a *prescribed* manner in both spoken and written media. A dictionary is prescriptive in so far as it distinguishes the formal and colloquial uses of any particular word. A set of defining principles is therefore of paramount importance during the defining stage so that the editors can write clear, concise, consistent, culturally relevant, and specific definitions.

Defining formats are very crucial in that they provide the editor, when defining a particular word, with a scheme to follow. It is a framework from which an editor can construct a definition. Defining formats are also important in that they contribute to consistency in the treatment of related sets of headwords. This view is accentuated by Gellerstam's (1993: 47) observation that "the development of definition formats is advantageous in that once a particular format has been agreed upon, a number of words can be defined with it". The ultimate goal of any defining guideline is to enable an editor to write what is widely referred to as a *user-friendly* definition.

The defining formats designed for the dictionary based on the Collins Birmingham University International Language Database (COBUILD) are well-known. The following are specifications for COBUILD defining formats as out-

lined by Moon (1987) and Hanks (1987):

(a)   the headword should appear in the definition,

(b)   the headword should be shown in a typical context of common usage because "context disambiguates" (Hanks 1987: 125),

(c)   definitions should consist of full sentences,

(d)   explanations should consist of two parts, with the first part showing the word in use and the second part explaining the meaning, and

(e)   the format should suggest a preference rather than a restriction.

For COBUILD, a good definition should be short, user-friendly, culturally relevant, consistent with others of its type and should use superordinate terms (Hanks 1987: 118). When a definition has been made simple and familiar, it helps the dictionary as a whole to be user-friendly. The COBUILD system demonstrates a marked departure from traditional lexicography which, according to Hanks (1987: 117), "heavily relies on the use of parentheses whose function is not based on the standard usage of ordinary written English". COBUILD has done away with the use of parentheses. These are not user-friendly since considerable sophistication is required on the part of the dictionary user to determine what precisely is meant by each set of parentheses.

Evident characteristics of the COBUILD system of defining formats are that:

(a)   definitions should be as natural as possible,

(b)   definitions should project the typical usage of a headword,

(c)   definitions should be easy to understand, and

(d)   a "dictionary as prose" (Hanks 1987: 118) should be created, hence there should be done away with the use of parentheses.

These guidelines are instrumental in organizing, formulating and presenting definitions that are user-friendly. There will be focused on the formats used by the editors of the ISN for defining the noun, verb and the ideophone in Ndebele.

The following are some of the COBUILD defining formats used for explaining the meanings of nouns (cf. Ndlovu 1998: 21):

(i)   **A(n) + Noun Headword is ...**, e.g.
       A **pond** is a small area of water that is smaller than a lake, or
       An **ostrich** is a large African bird that cannot fly.

(ii)   **Your + Noun Headword is/are ...**, e.g.
       Your **father** is your male parent.

(iii)   **If you call someone a(n) + Noun Headword ...**, e.g.
If you call someone a **bitch**, you mean that she behaves in a very un-
pleasant way.
This format is used to indicate a rude and offensive sense of a word, in
this case *bitch*.

(iv)   **Noun Headword is ...**, e.g.
Love is a very strong feeling of affection or liking for someone or some-
thing.

The following is a selection of three of the COBUILD defining formats for verbs
(cf. Ndlovu 1998: 23):

(i)   **To + Verb Headword + Someone or Something means ...**, e.g.
To **destroy** something means to damage it so much that it is completely
ruined or ceases to exist.

(ii)   **If/When you + Verb Headword ...**, e.g.
If you **jump** something such as a fence, you jump over it or across it.

(iii)   **If Someone or Something + Verb Headword ...**, e.g.
If someone or something **activates** something he/she or it spurs it into
action.

In the style manual of the ISN, it is clearly stated that the COBUILD formats
should be used for styling definitions.
    Nouns in the ISN have been defined using the following five defining
formats:

(i)   **Noun Headword is ...**, e.g.
indwangu bz 9. *Indwangu* yinyamazana elesiphongo, ifana lenkawu
kodwa yona inkulu. FAN imfene. Isib. *Indwangu ziqede umumbu ema-
simini.*
(**baboon** A *baboon* is an animal that has a forehead, it is like a monkey
but is bigger. Example: *Baboons ate and finished all the maize in the maize
fields.*)

(ii)   **If you are + Noun Headword ...**, e.g.
iqhawe bz 5. Nxa ul*iqhawe* empini uyabe utshengise isibindi lokukha-
lipha.
(**hero** If you are a *hero* in the army you would have demonstrated brav-
ery and skill.)

(iii)   **You call someone a(n) + Noun Headword ...**, e.g.
isitha bz 7. Uthi umuntu y*isitha* sakho nxa lizondana.
(**enemy** If you call someone your (an) *enemy* you hate each other.)
As in the COBUILD, this format was mainly used to indicate a rude and

offensive sense.

(iv)   **In [specialised field, e.g. medicine, business, law, etc.] + Noun Head-word ...**, e.g.
umethironi bz 1a. Kwezokwelapha *umethironi* ngunesi omkhulu.
(**matron** In a hospital a *matron* is a very senior nurse.)
This format situates the headword in the context in which it is typically used.

(v)   **Definition + Noun Headword**, e.g.
amachaphaza bz 6. Izibungwana ezimhlophe ezingakabi zinyosi ng*ama-chaphaza*.
(**larva** Small white insects at the stage before they become adult bees are *larvae*.)
This format does not occur in COBUILD. Its use demonstrates, as has been stated in the introduction, that defining formats in the ISN are a judicious mixture of essentially COBUILD formats and some modifications of traditional formats.

Verbs in the ISN have a typical subject. Giving a typical subject is meant to contextualise the definition and makes it more lucid. The following formats have been used in defining verbs in the ISN:

(i)   **To + Verb Headword + Someone or Something means** ..., e.g.
-tshengisa sz mwa Uku*tshengisa* umuntu ulutho yikwenza ukuthi alubone.
(**show** To *show* someone an object is to make them see it.)
This format was used alternately with the following format which, in the ISN, is a slight modification of the COBUILD format:

(ii)   **To + Verb Headword + is to ...**, e.g.
-labhula sz mwa Uku*labhula* yikufuqa ulutho engathi uyaluphosa ngesandla ufuna ukuthi luye khatshana ngoba lukunenga.
(**shove** To *shove* is to push something as if throwing it contemptuously far away because it irritates you.)

(iii) **If/When you + Verb Headword ...**, e.g.
-thenga sz mwa Nxa u*thenga* impahla uyabe uzibhadalela.
(**buy** When you *buy* property you obtain it by paying money for it.)

(iv)   **If Someone or Something + Verb Headword ...**, e.g.
-sisela sz mwa Nxa umnumzana e*sisela* omunye inkomo uyabe emgcinise ezinye zezinkomo zakhe.
(**loan** If a rich person *loans* someone cattle he will be asking him to keep some of his cattle.)

The ISN formats are a judicious mixture of COBUILD and traditional defining styles, since the COBUILD dictionary was designed to meet the perceived needs of advanced users of English as second language, while the ISN targeted speakers of Ndebele as first language. Not every aspect of the COBUILD would therefore be relevant to the ISN, hence the need for modification. The ISN editors agreed that consistency in the treatment of items from the same word-class was to be maintained. For the definitions to be user-friendly, the defining style should have to reflect a Ndebele world-view. One of the traditional defining styles agreed upon was that the definers were to give the hypernym or class containing similar animals or plants, and then identify the concept being defined by specifying aspects that distinguish it from others of its type. An example of the treatment of such classes is as follows:

> **inyamazana** bz 9. *Inyamazana* yisidalwa esilenyawo ezine esilesikhumba kanye loboya.
> (**animal** An *animal* is a creature that has four legs and has hide and fur.)

Other animals would then be defined in the context of this definition, as is shown by the following definition of *baboon*:

> **indwangu** bz 9. *Indwangu* yinyamazana elesiphongo, ifana lenkawu kodwa yona inkulu. FAN imfene. Isib. *Indwangu ziqede umumbu emasimini.*
> (**baboon** A *baboon* is an animal that has a forehead; it is like a monkey but is bigger. Example: *Baboons ate and finished all the maize in the maize fields.*)

One other modification of the COBUILD for defining verbs in the ISN is **Infinitive + Copulative Form ...** which can be exemplified as follows:

> **-khala** sz mwa uku*khala* yikulila ikakhulu uphuma izinyembezi ngokuzwa ubuhlungu.
> (To *cry* is usually to have tears coming out of your eyes as a result of feeling pain.)

The use of the phrase **Copulative Form ...** does not appear in the COBUILD.

The ISN had to develop definition formats for ideophones since the COBUILD dictionary does not give any definitions for the English ideophones **grr** (sound made by a lion) or **cockadoodle-doo** (sound made by a cock). To create a format for defining ideophones, the ISN had to modify the COBUILD formats for nouns and verbs. The following are formats that were used in the definition of ideophones:

(i)   **This is an ideophone that means ...**, e.g.
        vumbu szk. Lesi yisenzukuthi esitsho ukuvela kolutho lungalindelwe.

Isib. *Sithe sisalibele vumbu umfana owayegalule*.
(**appear suddenly** This is an ideophone that means something appears suddenly and unexpectedly. Example: *When we were relaxed suddenly and unexpectedly there appeared a boy who had drowned*.)

(ii)   **If Something or Someone + Copulative Form + Ideophone Headword ...**, e.g.
phamu szk. Nxa ulutho lugcwele *phamu* luyabe lugcwele lwaze lwagamanxa.
(**full to the brim** If something is *full to the brim* it would have filled its container without leaving space.)

(iii)   **Ukuthi + Ideophone Headword + Verb Form ...**, e.g.
bani szk. Ukuthi *bani* yikuphazima kombani kumbe ukukhanyisa ngokuphangisa.
(**spark** To *spark* is to appear as a sudden flash of lightning or a short burst of light.)

These formats comply with the specification for defining headwords in contexts of their common usage. They therefore clearly follow the COBUILD practice.

Note that, as Ndebele ideophones usually do not have English equivalents, the above can only be rough translations.

The ISN editors formulated the following as their defining principles, which are essentially those of the COBUILD dictionary:

(a)   definitions should be brief,

(b)   separate entries should be made for each sense,

(c)   the primary sense will determine the sense listed first, and

(d)   definitions should be long enough to explain the concept, but over-particularization should be avoided, e.g. concrete nouns should be defined just far enough to suggest the reality of the object.

The definitions in the ISN are therefore brief, but long enough to capture the main sense.

Words with the same form, but unrelated meanings (homonyms) were entered separately. The following are examples of the treatment of such words:

umuzi bz 3. *Umuzi* yindawo kumbe ikhaya elihlala abantu.
(**home** A *home* is a place or house in which people live.)
umuzi bz 3. *Umuz*i lichatha elimnyama emzimbeni umuntu azalwa elalo.
(**birthmark** A *birthmark* is a black spot with which a person is born.)
umuzi bz 3. *Umuzi* butshani obude obukhula ikakhulu emaxhaphozini.
(**Umuzi** is a tall grass which grows in very wet places.)

Definitions of words with different, but related meanings were given after the particular headword, beginning with the primary definition. The following can serve as an example:

>   **impilo** bz 9. 1. *Impilo* yisikhathi esiphilwa ngumuntu kusukela ezalwa aze afe. *UMahlangu waba lempilo ende ngoba waze wabhubha eselekhulu leminyaka*. 2. *Impilo* yindlela abantu abaphila ngayo. *Labana baphila impilo enhle ngoba bafundile*. 3. *Impilo* yikuphila kuhle emzimbeni komuntu engelamkhuhlane omkhathazayo.
>   (**life** 1. *Life* is the time that a person lives from his birth until his death. Example: *Mahlangu had a long life because he died when he had turned a hundred years old*. 2. *Life* is the way people carry on with their daily chores. Example: *Those people live a good life because they are educated*. 3. *Life* is physical good health without any pain or illness of one's body.)

Concrete nouns were also defined in such a way that the user will understand the structure of the real object referred to by the noun in question. The following is an example:

>   **isihlahla** bz. 7. *Isihlahla* yisithombo esihlanyelwayo kumbe esizimilelayo esikhula sibe side, silengatsha lamahlamvu.
>   (**tree** A *tree* is a tall plant that grows on its own or is grown, it has a firm trunk, branches and leaves.)

It is evident from this that definitions in the ISN are long enough to explain the concept but at the same time avoiding over-particularization.

3.   Examples in the ISN

As source of data for linguistic analysis a corpus plays an important role in linguistic research. In lexicography particularly it plays an even greater role. Kennedy (1998: 88) observes that a linguistic corpus, in whatever form, is an important basis for more accurate and reliable descriptions of the structure and use of languages. Sinclair (1991: 171) has defined a corpus as "a collection of naturally-occurring language texts, chosen to characterize a state or variety of a language". It is a body of collected texts of the written or spoken word, which is stored and processed on computer for the purposes of linguistic research (Renouf 1987: 1).

Because of the importance of the corpus in any dictionary-making process, that is, for purposes of headword selection, sense selection and obtaining examples of natural language usage, the ISN team collected a corpus of about a million running words before compiling the dictionary. The Ndebele corpus at the moment comprises about 1,2 million words. The ISN editors benefited from

the Ndebele corpus in the following ways:

(a)   Each word occurs in its own special environment. This is referred to as
       the *cotext*. By studying the cotext the editors managed to distinguish
       various senses of each word in these different environments. This made
       a seemingly difficult task very simple by consulting the corpus for the
       various instances of the usage of a particular word. The senses of most
       headwords in the ISN were ascertained through this method.

(b)   Concordances provide real text examples. A text example demonstrates
       a particular word in a typical sentence in agreement with other naturally
       co-occurring words. According to Sinclair (1991), concordances have a
       status which made-up examples do not have. They are typical. The
       examples for most headwords in the ISN therefore came from the cor-
       pus. This provided the dictionary with more "natural" examples.

However, the ISN is not exactly a corpus-based dictionary, but rather a corpus-
aided dictionary. Because of the limited size of the corpus at the time of the
compilation of the dictionary, the corpus was not considered large enough to
be "representative". Since a corpus has to be "representative" in terms of size in
order to be appropriately used as the basis for corpus-based dictionaries, the
ISN editors whose corpus was then relatively small, could not avoid relying on
the intuitive knowledge of the language in constructing some examples. After
all, it has been argued that constructed examples are bound to be more capti-
vating and lucid. This observation is, however, open to debate. In the case of
the ISN, it provided a balance by including both corpus examples and con-
structed examples.

       Examples in the ISN were viewed as important components of the defini-
tions. The ISN editors resolved that:

(i)   since examples are part of the definition, they be used only when neces-
        sary to complement the definition in clarifying the headword, that is, to
        illustrate meanings or uses,

(ii)   a maximum of one example be given for reasons of saving space, and

(iii)   like definitions, they also be short.

The examples in the ISN were therefore used only as far as they served to illu-
minate the definitions. Their inclusion was to achieve the goal of producing a
user-friendly dictionary.

4.   Conclusion

The article has outlined the defining formats for nouns, verbs and ideophones

that were used in the ISN. It is clear from this outline that the formats are essentially those of the COBUILD dictionary although there were very minor modifications especially with regard to verbs. The defining formats for the ideophone were based on the defining styles of the noun and verb. It can be concluded that the ISN mainly followed the COBUILD defining formats as stated in its style manual. While many examples were taken from the corpus to illustrate the meaning of the headwords in instances where it was really necessary, the editors also used their intuitive knowledge to construct some examples. This had the advantage that the example, just like the definition, would be short, direct and lucid. On the other hand, the use of corpus examples provided fresh and natural real text examples. A balance between these two sources can be considered to form a sound basis for examples.

References

Gellerstam, M. 1993. The Use of the Corpus. Chimhundu, H. (Ed.). *Report on the African Languages Lexical (ALLEX) Project Planning and Training Workshop*. Harare: University of Zimbabwe.

Hadebe, S. et al. (Ed.). 2001. *Isichazamazwi SesiNdebele*. Harare: College Press.

Hanks, P. 1987. Definitions and Explanations. Sinclair, J.M. (Ed.). 1987: 115-136.

Kennedy, G. 1998. *An Introduction to Corpus Linguistics*. London/New York: Longman.

Moon, R. 1987. The Analysis of Meaning. Sinclair, J.M. (Ed.). 1987: 86-103.

Ndlovu, F. 1998. *Considering Factors that are Relevant to Adopting and Adapting COBUILD Defining Formats for the General Ndebele Dictionary*. Unpublished M.A. Thesis. Harare: University of Zimbabwe.

Renouf, A. 1987. Corpus Development. Sinclair, J.M. (Ed.). 1987: 1-40.

Sinclair, J. 1990. *Collins COBUILD Student's Dictionary*. London: Collins.

Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, J.M. (Ed.). 1987. *Looking Up: An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*. London/Glasgow: Collins ELT.

Waite, M. (Ed.). 1998. *The Little Oxford Dictionary*. Oxford: Clarendon Press.