DEFINING FORMATS AND CORPUS-BASED EXAMPLES IN THE GENERAL NDEBELE DICTIONARY, *ISICHAZAMAZWI SESINDEBELE*: A CRITICAL ANALYSIS.*

Langa Khumalo: African Languages Research Institute (ALRI), University of Zimbabwe, Harare, Zimbabwe (langa@arts.uz.ac.zw)

Abstract: In this paper the writer critically evaluates the defining formats that were used in defining headwords in the first-ever-monolingual General Ndebele Dictionary, *Isichazamazwi SeSiNdebele*(ISN). The emphasis in the ISN was on the concept of user-friendliness. The paper will establish that defining formats in the ISN are a judicious mixture of mainly the Collins Birmingham University International Language Database (COBUILD) defining formats and what has been referred to as the traditional formats. The first part of this paper will be an analysis of the decisions taken by the ISN editors in formulating their defining formats. It will assess 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 will further discuss other formats, including the decision to retain what could be referred to as traditional defining formats, which were used in defining headwords. One of the traditional defining styles they agreed upon was that the definers were to give the hypernym in the case of semantic sets, and then identify the concept being defined by specifying aspects that distinguish it from others of its type. The second part of the paper evaluates the importance and use of the corpus in providing both definitions and examples in the ISN. However, we further argue that since a corpus has to be “representative” in terms of size, in order to be appropriately used as the basis of generalizations concerning such corpus-based dictionaries, the ISN editors whose corpus is relatively small, could not avoid relying on the intuitive knowledge in formulating some examples.

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

1. INTRODUCTION

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

This makes defining formats, a format being “...the way in which something is arranged and presented” Sinclair (1990:218), therefore, very crucial in composing definitions. Defining formats

* This paper was originally presented at the 6th International Conference of the African Association for Lexicography, at the University of the North in South Africa. The paper was later published in LEXIKOS 12.
are pertinent insofar as they help the compilers or editors to come up with clear, specific, systematic and culturally relevant definitions that make dictionaries user-friendly. These are formats that are geared towards providing the editor with certain guidelines to follow in constructing a definition hence contribute to consistency in the treatment of related sets of headwords.

It is our view also that examples are an important part of the definition. This is so because they help to illuminate further the definition. They help the editor in coming up with a clear, specific, definition, which is culturally and contextually relevant. The second part of the paper will, therefore, evaluate the importance and use of the corpus in providing both definitions and examples in the ISN. According to Sinclair (1991:171) “a corpus is a collection of naturally-occurring language text, chosen to characterize a state or variety of a language.” The ISN is largely a corpus-aided dictionary. From the corpus, if we study instances of usage, we find that the surrounding words and phrases help a lot in determining the meaning. Each word, therefore, occurs in its own special environment, called its COTEXT, and by studying the cotext we can often work out which meaning is intended. Since concordances provide us with real text examples hence have a status which made-up examples do not have, the ISN editors lifted most of their examples from the corpus. This, it would 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 of generalizations concerning such corpus-based dictionaries, the ISN editors whose corpus is relatively small, could not avoid relying on the intuitive knowledge in formulating some examples.

2. DEFINITIONS AND DEFINING FORMATS IN THE ISN.

It is a well-known fact that a dictionary lists words and their meanings and as already been stated in the introduction, a definition is an intrinsic component of any dictionary. A definition is itself “a statement of meaning of word”(Waite: 1998) 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 an effective definition can only come from a mother tongue speaker of that particular language. Secondly, that it is a process, a rigorous process that follows set principles so as to provide a brief, distinct and conceptually appropriate statement of meaning of a particular word.
Defining is, in our view, one of the most important processes in dictionary making. This is so because it is the process through which users may become more aware about the meanings and explanations of words. It is important for users to know meanings of words so that they can use the word in a *prescribed* manner in both spoken and written media. A dictionary is prescriptive insofar 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 come up with clear, concise, consistent, culturally relevant, and specific definitions.

Defining formats are very crucial in that they provide the editor, who is 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 consistence 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 come up with what is widely referred to as a *user-friendly* definition.

The following are COBUILD defining formats as outlined 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 one explaining the meaning.
(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 it should use super-ordinate terms (Hanks, 1987:118). When a definition has been made simple and familiar it thus helps the dictionary as a whole become 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 because they are not user-friendly since considerable sophistication is
required on the part of the dictionary user for him/her to work out 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 typical usage of a headword.
(c) definitions should be easy to understand.
(d) to create a “dictionary as prose...” (Hanks: 1987:118) hence doing away with the use of parentheses.

These guidelines would then be an instrument for the editors to organize, craft and present definitions that are user-friendly. We will focus on the defining formats for the Noun, Verb and (for Ndebele) the ideophone.

The following are some of the COBUILD defining formats used for explaining the meanings of nouns:

(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 can not fly.

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

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

(iv) Noun Headword is ...
   e.g Love is a very strong feeling of affection or liking for someone or something.

I have also selected some of the COBUILD defining formats for verbs. Here are three of them:

(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 their style manual the ISN editors state clearly that they used COBUILD format for their definition style. The following are examples of nouns in the ISN which follow the above formats:

Nouns in the ISN have been defined using the following five defining formats.

(i) The first one is **Noun Headword is** ... Here is an example below:

**indwangu** bz 9. *Indwangu* yinyamazana elesiphongo, ifana lenkawu kodwa yona inkulu. FAN imfene Isib. *Indwangu* ziqede umumbu emasimini.

(ii) The second definition format is **If you are** + **Noun Headword** ... Here is an example below:

**iqhawe** bz 5. Nxa uliqhawe empini uyabe utshengise isibindi lokukhalipha.

(iii) The third definition format is **You call someone a(n) + Noun Headword** ... Here is an example below:

**isitha** bz 7. Uthi umuntu yisitha sakho nxa lizondana
Note that as in the COBUILD this format was mainly used to indicate a rude and offensive sense.

(iv) The fourth sense situates the headword in the context in which it is typically used.

**In** [specialised field, e.g. medicine, business, law, etc] **+ Noun Headword** ... Here is an example:

**umethironi** bz 1a. Kwezokwelapha umethironi ngunesi omkhulu.

    (In medicine/hospital a matron is a very senior nurse.)

(v) The final format used for nouns in the ISN is the one in which a definition ends with the Headword. For example:

**amachaphaza** bz 6. Izibungwana ezimhlophe ezingakabi zinyosi ngamachaphaza.
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 were 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.

This format was used alternately with the following format, which is a slight modification of the COBUILD format, in the ISN:

(ii) **To + Verb Headword + is to ...**

* e.g. -labhula sz mwa Ukulabhula yikufuqa ulutho engathi uyaluphos ngesandla ukuna ukuti luye khatshana ngoba lukunenga.

(iii) **If/When you + Verb Headword ...**

* e.g. -thenga sz mwa Nxa uthenga impahla uyabe uzibhadalela.

(iv) **If Someone or Something + Verb Headword ...**

* e.g. -sisela sz mwa Nxa umnumzana esisela omunye inkomo uyabe emgcinise ezinye zezinkomo zake.

We stated elsewhere above that 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 English as a Second Language, while the ISN was targeted at first language Ndebele speakers. Not every aspect of the COBUILD would 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, it was agreed that the defining style reflects 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** bż 9. *Inyamazana* yisdalwa 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 the definition of a baboon below:

**indwangu** bz 9. *Indwangu* yinyamazana elesiphongo, ifana lenkawu kodwa yona inku. 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.

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

-**khala** sz mwa ukukhala yikulila ikakhulu uphuma izinyembezi ngokuzwa ubuhlunlu.

(To cry is to usually 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 the ideophones since the COBUILD dictionary does not have any definition for the English ideophones **grr** (sound made by a lion) or **cockadoodle-doo** (sound made by a cock). The ISN had to modify the COBUILD formats for nouns and verbs to come up with a format for defining ideophones. The following were formats that were used in the definition of ideophones:

(i) **This is an ideophone that means** .... Here is an example below:

 **vumbu** szk. Lesi yisenzukuthi esitsho ukuvela kolutho lungalindelwe. Isib. *Sithe sisalibele vumbu umfana owayegalule.*

(ii) **If Something or Someone + Copulative Form + Ideophone Headword** ... Here is an example below:

 **phamu** szk. Nxa ulutho lugcwele *phamu* luyabe lugcwele lwaze lwagamanxa.

(iii) **Ukuthi + Ideophone Headword + Verb Form** ... Here is an example below:

 **bani** szk. Ukuthi *bani* yikuphazima kombani kembe ukukhanyisa ngokupha-ngisa.

Format (c) share the element of defining headwords in contexts of their common usage. This therefore clearly follows the COBUILD practice.
The ISN editors then formulated the following as their defining principles, which are essentially COBUILD format:

(a) be brief
(b) separate entries are to be made for each sense
(c) primary sense will determine which sense is listed first
(d) aim for definitions that are long enough to cover the concept, but avoid over-particularization.

For example, with concrete nouns, define just far enough to suggest the reality of the object.

The definitions in the ISN are therefore brief enough to capture the main sense. There are those headwords that have different senses, and these were entered separately. Here is an example for such words with different senses:

**umuzi** bz 3. *Umuzi* yindawo kumbe ikhaya elihlala abantu

**umuzi** bz 3. *Umuzi* lichatha eliminyama emzimbeni umuntu azalwa elalo

**umuzi** bz 3. *Umuzi* butshani obude obukhula ikakhulu emaxhaphozini

There are some headwords that have different definitions that are, however, related. Such definitions were listed after that particular head word beginning with the primary definition. Here is an example below:

**impilo** bz 9. 1. *Impilo* yisikhathi esiphilwa ngumuntu kusukela ezalwa aze afe. UMahlangu waba lempilo ende ngoba waze wabhubha eselekhu leminyaka. 2. *Impilo* yindlela abantu abaphila ngayo. Labana baphila impilo enhle ngoba bafundile. 3. *Impilo* yikuphila kuhle emzimbeni komuntu engelamkhuhlane omkhathazayo.

Concrete nouns were also defined in such a way that the user will understand the structure of the noun in question. Here is an example:

**isihlahla** bz. 7. *Isihlahla* yisithombo esihlanyelwayo kumbe esizimilelayo esikhula sibe side, silengatsha lamahlamvu.

It is evident that definitions in the ISN are long enough to cover the concept but at the same time avoiding over-particularization.
3. EXAMPLES IN THE ISN.

Corpus work has played a very important role in linguistic research as a source of data for linguistic analysis. It has particularly played an even greater role in lexicography. Kennedy (1998:88) observes that a linguistic corpus, in whatever form, is important as a basis for more accurate and reliable descriptions of how languages are structured and used. Sinclair (1991:171) has defined corpus as “a collection of naturally-occurring language text, 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 exercise, that is, for purposes of headword selection, sense selection and lifting of examples of natural language usage, the GND team collected a corpus of about a million running words before compiling the dictionary. The Ndebele corpus now stands at around 1.2 million words. The ISN editors benefited the following from the Ndebele corpus in their defining tasks:

(a) Each word occurs in its own special environment. This is referred to as the COTEXT. By studying the cotext the editors managed to work out various senses of each word in these different environments. This made a seemingly difficult task very simple by asking the corpus to bring out various instances of usage of a particular word. Senses such as listed below were figured out through this method:

\textbf{umuzi} bz 3. \textit{Umuzi} yindawo kumbe ikhaya elihlala abantu
\textbf{umuzi} bz 3. \textit{Umuzi} lichatha elimnyama emzimbeni umuntu azalwa elalo
\textbf{umuzi} bz 3. \textit{Umuzi} butshani obude obukhula ikakhulu emaxhaphozini

(a) Concordances provide us with real text examples. It demonstrates a particular word in a typical sentence in agreement with other naturally co-occurring words. According to Sinclair, concordances have a status which made-up examples do not have. They are typical. The examples for most headwords in the ISN were therefore from the corpus. The ISN editors took most of their examples from the corpus. This provided the dictionary with more “natural” examples.
However, it has to be noted that the ISN is not a corpus-based, but a corpus-aided dictionary. This is to say that because of the limited size of the corpus at the time of the compilation of the dictionary, the corpus was not 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 of generalizations concerning such corpus-based dictionaries, the ISN editors whose corpus was then relatively small, could not avoid relying on the intuitive knowledge of the language in formulating some examples. After all, it has been argued that formulated examples are bound to be more captivating and lucid. This observation is, however, open to debate. In the case of the ISN, it provided a balance by including both the corpus examples and formulated examples.

Examples in the ISN were viewed as a very important component of the definition. The ISN editors resolved that:

(i) since examples are a part of the definition, they be used only when it serves to complement the definition in clarifying the headword. That is, to illustrate meanings or uses only when it is necessary.
(ii) that a maximum of one example be given for obvious reasons of saving space.
(iii) that they be short too (like definitions).

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

4. CONCLUSION
The paper has outlined the defining formats for nouns, verbs and ideophones that were used in the ISN. It is clear from the outline that the formats are essentially COBUILD although there were very minor modifications especially with regards to verbs. The ideophone defining formats also followed the noun and verb defining style. It can be concluded that the ISN followed mainly the COBUILD defining format as stated in their style manual. While many examples were lifted from the corpus to illuminate the meaning of the headwords in instances were it was really necessary to do so, the editors also used their intuitive knowledge to construct some definitions. This had an advantage that the example would be short, direct and lucid just as the definition. While on the other hand the use of corpus examples provided fresh and natural real text example. In our view the balance of the two strands is very health.
REFERENCES


