An Advanced Dictionary?  
Similarities and Differences between *Duramazwi ReChiShona* and *Duramazwi Guru ReChiShona*

Emmanuel Chabata, African Languages Research Institute, University of Zimbabwe, Harare, Zimbabwe (echabata@arts.uz.ac.zw)

Abstract: In this article a comparative analysis of *Duramazwi ReChiShona* (DRC) and *Duramazwi Guru ReChiShona* (DGC) is made. Both DRC and DGC are monolingual Shona dictionaries compiled by a team of researchers under the African Languages Lexical (ALLEX) Project, now the African Languages Research Institute (ALRI). During the compilation process, DRC was known as the General Shona Dictionary and DGC as the Advanced Shona Dictionary. A simple analysis of these titles shows that the dictionaries are similar in some ways and also different in others. The writer tries to show the ways in which DGC is regarded as a more advanced dictionary when compared to DRC. Although the argument of the article is mainly built on those differences which make DGC the more advanced, attention is also paid to the similarities between the dictionaries.

Keywords: ALLEX PROJECT, ALRI, DURAMAZWI RECHISHONA, DURAMAZWI GURU RECHISHONA, DICTIONARY, SHONA, HEADWORD, SENSE, MONOLINGUAL DICTIONARY, CORPUS, IDIOM, PROVERB, PITHY SAYING

*This article is a slightly revised version of a paper presented at the Sixth International Conference of the African Association for Lexicography, organised by the Sepedi National Lexicography Unit, University of the North, Pietersburg, 2–4 July 2001.*

http://lexikos.journals.ac.za
1. Introduction

The history of Shona lexicography dates back to the last quarter of the 19th century. Missionaries from different mission stations, strategically located in different parts of the country, were the initiators. Early publications from such stations include small dictionaries and vocabulary lists for dialects of the Shona language spoken in communities surrounding the respective mission stations. The dictionaries published were largely bilingual because they were mainly meant for second-language speakers of Shona. Examples of such dictionaries include English–Mashona Dictionary (Hartman 1894), English–Chiswina Dictionary (Biehler 1906) and ChiNdau–English and English–ChiNdau Vocabulary (Wilder 1915). Although many of these dictionaries were not comprehensive, they were at least adequate for the users targeted. However, they were not very useful to native speakers of Shona because they only provided English lexical equivalents to Shona headwords. The editors did not give explanations that enhance the development of the language they described.

Later, more comprehensive bilingual Shona–English dictionaries, which covered vocabulary from all dialects of Shona were compiled. This was after Doke’s (1931) recommendation for the unification of all Shona dialects. Dictionaries such as the Standard Shona Dictionary (Hannan 1959) and Duramazwi (Dale 1981) were published. These dictionaries are bigger, both in terms of the number of pages covered and the number of headwords defined. Unlike in earlier dictionaries where dialects of Shona were treated as different “languages”, headwords and senses for these two dictionaries came from all five dialects of the Shona language, namely Zezuru, Karanga, Manyika, Ndau and Korekore. Today these dictionaries are still being used as reference works for Shona.

Work on monolingual Shona lexicography only started with the launch of the ALLEX Project in September 1992. Through the Project, now transformed into ALRI, the researchers published Duramazwi ReChiShona (DRC) in 1996 and Duramazwi Guru ReChiShona (DGC) in 2001. As suggested in its title, the second dictionary is a more advanced lexicon compared to the first one. This article discusses the ways in which DGC is considered a more advanced dictionary in comparison with the former. The differences are discussed under the following subheadings: headword selection, sense selection, defining formats and the structure of the dictionaries. The article also describes some of the similarities between the two dictionaries.

2. Differences

2.1 Headword selection

One of the most important stages in the compilation of dictionaries is that of
selecting the words or phrases to be entered as headwords. The decisions to be taken on the type and number of headwords to be included or excluded are influenced by a number of factors, such as the purpose of the dictionary, its intended size as well as its target audience. For example, the issue of the target users is of paramount importance, for it is obvious that a dictionary compiled for primary-school children cannot be expected to be similar to the one intended for students at tertiary institutions. The complexity or simplicity of headwords and other lexicographic information to be included in a dictionary depend on the intended users of the dictionary.

The target groups for DRC and DGC are different. DRC was mainly meant to fulfil the needs of lower secondary school learners. Because of this, it was supposed to be a reference work containing basic Shona vocabulary. The dictionary was also an experiment in monolingual Shona lexicography and was therefore supposed to be relatively small and manageable. On the other hand, DGC was "intended to be a comprehensive reference work, which will serve as a resource for more advanced users, especially those at higher secondary and tertiary education levels" (Chitauro-Mawema 2000: 209). What this entailed was that in addition to all the headwords in DRC, many more had to be selected for DGC. For example, terms from specialised fields, which were not part of DRC, were to be entered in DGC.

Most of the additional headwords for DGC came from the Shona corpus, of which a list of all the items in it was made, and in which the researchers searched for new headwords. Other sources of new headwords included written materials in specialised subject areas. In fact, the selection exercise was comprehensive for it included words from all word categories and also from all spheres of life. Notable additions include phrasal headwords, that is, idioms, proverbs and pithy sayings. Svensen (1993: 108) defines an idiom as "a fixed group of words with a special meaning which is different from the meanings of the individual words". Like idioms, proverbs and pithy sayings are generalising statements whose purpose is to convey certain assertions about life (Svensen 1993: 110). Structurally, they are fixed word combinations that must be shown in full in a dictionary since they mean more than what is implied in their constructions. Because these phrases can only be understood as wholes, they should be presented as such in texts. As a result, they occupy considerable space. This is one of the reasons why they were excluded from DRC. Another reason is that they express meaning in a metaphorical way, and such meanings are captured in long explanations which often need illustrative examples. Zgusta (1971: 153) notes that such long headwords, which also need long explanations, are usually included in comprehensive dictionaries. It is because of the nature of their form and meaning that the title of the dictionary had to be followed by a subtitle: *rine zvirungamutauro* (with idiomatic expressions).

Another category of headwords added in DGC is technical terms. Svensen (1993: 49) notes that technical language arises as a consequence of constant development and specialisation in the fields of science, technology and sociol-
ogy. New concepts are constantly being defined, and in order to exchange information about them, new linguistic expressions have to be found for them. There are many specialised terms used in technical subjects in education, economics, sport, law, medicine and others. However, despite their origin in the terminology of various technical fields, many of these terms make their way into general language and become known to lay-people. Svensen (1993: 49) notes that general and broad terms within a certain area tend to move more readily across from technical to general language than terms representing specific concepts. Not all technical terms were eligible for selection as headwords for DGC. The eligibility of specialised and technical terms was not based on their importance in their respective subject areas, but on their use in general language. Only those that the editors felt were generally or commonly used in the Shona-speaking community were incorporated as headwords. Those that the editors felt were still restricted to their technical fields were excluded because they were considered unfit for a general dictionary such as DGC. Also selected were some international words used in fields such as science and mathematics. Chitauro-Mawema (2000: 212) defines international words as those “technical words which carry specific unchanging and unambiguous senses in the contexts in which they occur and are used internationally”. These are terms that are usually borrowed from other languages through the process of acquiring the respective concepts. Terms such as **ajebhura** (algebra), **ikweta** (equator) and **sirabhasi** (syllabus), for example, were thought of as standard terms everywhere and therefore had to be selected as headwords if the dictionary had to be comprehensive in the true sense.

Some slang words were also selected as headwords. These are informal words that are only used by people who know each other well or those who share the same interests. Usually slang words are ephemeral, their use only lasts for a very short time. However, there are some words in this category that tend to stabilise and become part of the normal and conventional vocabulary of a language. It was because of these words that settle in a language that a decision was made sparingly to allow slang words into DGC. Words such as **mushe** (all right), **bhoo** (all right) and **kanjani** (how are you) were entered as headwords because they have been in circulation for a long time and have become part of the Shona vocabulary. However, the problem of dealing with such words is that decisions on whether a particular word has stabilised or not are rather subjective.

A few words that violate the Shona alphabet system were also entered as headwords in DGC. Examples of such words include *lita* (litre), *jeli* (jelly), *loni* (lawn) which have the letter l and *thiyeta* (theatre) which has the cluster th. The letter l and the cluster th are not acceptable in correct Shona spelling. However, these words were included in the dictionary because this is how they are said or pronounced by Shona speakers. It was felt that leaving them out would mean a loss to the language described. However, an asterisk was added to these headwords to show that they are not yet accepted in the writing sys-
Similarities and Differences between Duramazwi ReChiShona and Duramazwi Guru ReChiShona

tem. For their meanings, users were also referred to corresponding headwords with the letter r in place of l and ti in place of th. These headwords are only acceptable in the writing system but do not accurately reflect how they are pronounced by speakers. The headwords with l and th were entered with the hope that the Shona language committee would consider incorporating these and other such letters in the Shona alphabetic system since they are commonly used. However, by the time of publication of the dictionary, the decision was still pending.

The addition of headwords from various word categories naturally made DGC bigger than DRC. The DGC is more than double the size of DRC, both in terms of the number of entries and the number of pages covered. Whilst DRC occupies 504 pages, DGC consists of 1280 pages. In terms of the number of headwords, DRC contains about 16000, whilst DGC has almost 37000. On the basis of these statistics, DGC has also become the biggest monolingual dictionary in any African language.

2.2 Sense selection

Like the selection of headwords, the selection of senses for dictionary entries is also determined by the type and size of the dictionary being compiled, its purpose as well as its target users. For DRC these factors played a major role in determining the number and kind of senses included in the dictionary. The definitions provided for the words in this dictionary were basic explanations of their meaning. Subsidiary or extended senses were excluded for two main reasons. Firstly, since the dictionary was meant for learners in their early years of secondary education, the explanations were supposed to be simple and were to be those encountered in daily language use. Secondly, the issue of limited space also played a part. This dictionary was not supposed to exceed a prescribed number of pages and to make sure that it was kept within the required size, only an average of two senses per headword was allowed. This was decided after the realisation that most words in Shona, especially verbs, can carry many senses, some being principal and many more being extended, specialised or metaphorical.

The situation was different in DGC since more space was allotted to this dictionary than to DRC. Emphasis shifted from limitation of senses to making the dictionary as comprehensive as possible. In this respect, it was decided that a headword should carry all the senses it has. To illustrate this, the example of the verb -bata (touch) can be taken. This verb has a wide range of meanings, most of which are extensions of its basic sense of "touching". Examples of such senses include those of "working, tightening, catching, intoxicating, strengthening" and others that are derived from the basic sense of getting in contact with something. Because all its meanings could not be included in DRC, only four were provided. These were the ones that the dictionary editors felt were basic and would immediately come to anybody’s mind when the verb is utter-
ed. However, following the new decision of including in DGC all the senses of a word, the verb -bata was provided with a total of 18 senses (14 more than those that were provided for in DRC). Most of the senses added were those that were left out on the grounds that they were extensions of the basic sense(s) of this verb. Like -bata, most headwords were provided with more senses in DGC than in DRC. It was for this reason that the idea of adding global definitions to such headwords was introduced. A global definition can be described as a general definition that is put under a headword which has more than one specific sense. It carries the main idea expressed by a word and it takes traits from other definitions provided under that headword. It describes the basic concept to which the headword refers, that which is inherent in every sense provided under it. To illustrate this, the verb -chaira which carries three different senses, can be taken as example. A global definition, kufambisa kuenda mberi (cause to move forward), was provided to capture the general idea expressed by this verb. All three explanations that follow describe different ideas to which -chaira can refer, but all of them have the element of “moving forward”.

Another important area which is related to defining and sense selection is that of illustrative examples. Fox (1987: 137) says “the use of examples forms an integral part of the learning of a word”. Examples help in reinforcing the meaning of a word, not by acting as a reformulation of the definition, but by showing how the word is actually used in an appropriate context, a typical grammatical structure and together with words that are normally associated with it. They are usually added to illuminate those definitions that are not clear. There were fewer occasions in DRC where examples were found necessary. Besides the need for saving space, the main reason for this had to do with the kinds of senses provided in this dictionary. Since the given meanings were basic, it was felt that users would not have difficulties in understanding the explanations. On the other hand, there often were cases in DGC where examples were considered necessary, because there were a lot of subsidiary, sometimes closely related senses, that were defined in this dictionary. These senses are specialised and/or metaphoric. They usually carry hidden meanings and are also rarely used in everyday interaction. It is because of their nature that it was felt that examples were needed to illustrate the contexts in which they are used. In this respect, the examples were provided so as to aid the dictionary users’ understanding of the meanings of the headwords. The examples were also considered useful in showing the ways in which each sense is different from the other(s).

The link between DRC and DGC with the Shona electronic corpus is worth mentioning. DRC has often been described as corpus-aided or corpus-assisted. Very little evidence was drawn from the Shona corpus during headword selection and defining. This was mainly because the corpus, whose compilation started at the same time as that of DRC, was still very small for any meaningful use as a source of dictionary headwords, senses and illustrative examples. On the other hand, DGC can be described as corpus-based. This is because quite a
substantial number of headwords, senses and illustrative examples in this dictionary came from the corpus. For senses, a concordance programme was employed to provide contexts in which words are used in the corpus. A concordance consists of a list of all occurrences of each word in a text. An analysis of the different contexts in which a word is used yields different senses of that word. Because the editors of DRC usually relied on their memory for headwords and definitions, the tendency was to include only those words that were commonly used as well as providing general definitions that they could easily remember. However, a shift from heavy reliance on a few people’s memories to the use of the corpus in DGC made it possible for both common and rare headwords and senses to be recalled and included, thus resulting in a more comprehensive product.

2.3 Structure

Different books, including reference works can be arranged in a variety of ways, depending on the kinds of information that they contain. DRC and DGC differ with regard to the way information is arranged in the two dictionaries. Basically the differences emanate from the additions that were made in DGC. DRC is simply divided into the front matter and the main body of the dictionary. Svensen (1993: 16) refers to the front matter of a dictionary as "instructions of use", describing how the dictionary is organised, how it was compiled as well as how it can be used effectively. The main body here refers to the part that contains headwords and their senses, alphabetically ordered from A to Z.

Whilst the main body of DRC consists of just one section, that of DGC is divided into two, that is, Chikamu I (Section I) and Chikamu II (Section II). Chikamu I contains lexical, multiword lexical units and idioms and their senses. Besides the addition of idioms, which are phrasal in nature, this section may be paralleled to the main body of DRC. On the other hand, Chikamu II contains other phrasal headwords, that is, proverbs and pithy sayings. These were excluded from the first section of the dictionary for a number of reasons. Firstly, they were excluded because they had to appear in the dictionary as full or complete statements, starting with capital letters and ending with full-stops. It was felt that the inclusion of complete statements in the same section with other entries would yield an unattractive presentation on the page, especially given the fact that the phrases would not be evenly distributed across the dictionary. Proverbs and pithy sayings were also excluded from this section because they are not usually found in smaller dictionaries. Lastly, and arguably most importantly, it was difficult to decide under which headwords respective phrases should be placed. For idioms it was relatively easy since each one was listed under its main verb, because the main verb in an idiom is fairly predictable. However, the reverse is true of proverbs and pithy sayings where there are usually more than one word which can be regarded as the main word.

Unlike DRC, which only has the front matter and the main body, DGC
also has a back matter, that part which comes after the main alphabetical listing of the dictionary. In this section, information which is not needed for the correct use of the dictionary, but which may be useful in other ways, is appended. The appendices provide some practical information that can be utilised by the dictionary user in his/her daily life. The kinds of information appended to DGC include systematic lists of appellations of chiefs in different parts of Zimbabwe and their respective totems, names of African countries and currencies used in these countries, measurements, weights, etc. Whilst information placed under entries in the main body of a dictionary can be regarded as linguistic or as serving linguistic functions, that in the back matter is general, cultural or otherwise.

2.4 Defining formats

Words in a dictionary are not defined in a haphazard manner. Instead, lexicographers have to follow systematic ways of defining which they develop even before they start explaining or describing what the words mean. These systematic ways used have often been referred to as defining formats and have also been described as laid-down principles that “provide guidelines or paths that a definer follows when defining” (Chabata 1995: 2).

Defining formats are developed for each class of words in a dictionary and a number of defining methods or principles can be identified. However, we will only refer to two methods that are relevant to our discussion. These are the traditional method of defining and that of the Collins Birmingham University International Language Database (COBUILD). In short, in the traditional method all the information provided for each headword is contained in the definition and words are defined outside their contexts of use. In fact, statements are made about what the words mean, but very little is said about how they are used (Hanks 1987: 121). This method is usually followed for reasons of saving space. On the other hand, when the COBUILD method is used, words are defined within the contexts of their use. Hanks (1987: 118) notes that the use of this method results in explanations that consist of two parts. The first part represents a departure from the traditional method in that it actually places the word being defined in a typical structure, thus showing its use. For example, the first part of the definition for bang (knife) would start with the headword itself, that is, banga … The second part explains the meaning. Because the words being defined form part of their explanations, the practice tends to occupy a lot of space. However, the method is user-friendly.

The defining formats developed for DRC and DGC can be described as a judicious mixture of the traditional method and the COBUILD method. Although a mixture of the two defining methods was adopted for the two dictionaries, a closer look at the formats shows that those for DGC have moved closer to the COBUILD method than those used in DRC. For example, in DRC a principle was adopted that only nouns with at most three syllables were to be
part of the headword's definition(s). In other words, only these nouns were supposed to be defined using the COBUILD method, while those with more than three syllables were not supposed to be included or repeated in the definition. This measure was adopted as a way of saving space. In DRC, space saving was of paramount importance, especially if it is considered that, in Shona, for example, one can coin very long nouns by joining word forms from different word categories. Examples of such nouns include *kadende-mafuta* (small calabash of oil), *chitundu-mutsere-mutsere* (rocket) and *chibaya-mahure* (prickly plant). For such headwords, the traditional method was seen as more suitable since it would help keep definitions short. Although the idea of saving space was still upheld in DGC, it was felt that there was more space allotted to this dictionary in comparison to its predecessor. Because of this, user-friendliness was considered more important than space. This, for example, led to an increase in the number of syllables from three to four for a noun to be repeated in a definition.

It is also important to note that in addition to the defining formats used in DRC, more were created for DGC. The defining formats added are of two kinds. The first is that of formats developed to provide for new categories of headwords such as proverbs and pithy sayings introduced in DGC. The other kind is that of formats developed for categories of headwords which existed in DRC but were intended to augment the already existing ones. For example, new and more explicit formats were developed for nouns. The new formats developed can be regarded as more explicit in giving typical contexts in which particular nouns are used. To illustrate this, the noun *jobhukadhi* (job card) can be taken. This noun has different senses in different contexts. For example, in industry it is a piece of paper on which work intended for each day is written. In garages that repair machines, it is used to refer to a piece of paper on which machine defects as well as the cost of repairing them are recorded. As a way of distinguishing and describing the meanings, the defining format that was used is as follows:

1. *Mumakambani, jobhukadhi* … (In companies, a job card is …)
2. *Mumagaraji, jobhukadhi* … (In garages, a job card is …)

When this defining format is applied, one can clearly understand the different, but closely related senses that this word has when used in different contexts.

3. **Similarities**

So far the focus of this article has been on the differences between DRC and DGC. This is not to suggest that there are no similarities between them. On the contrary, they share a number of similarities which will be looked at in this section.

One of the most outstanding similarities between DRC and DGC is the fact
that they are general monolingual dictionaries. They are described as "general" because they are concerned with the Shona language as it is generally used in the communities where it is spoken. They are also monolingual dictionaries, so far the only ones in Shona. According to Zgusta (1971: 249), in monolingual dictionaries the object of description (that is, the headword) and the descriptive instrument (that is, the meaning or explanation) should be the same. Unlike bilingual dictionaries whose aim is to help in translating from one language into another, the aim of these dictionaries is to describe the Shona language in a way that enhances its development.

Another important point is that these dictionaries were compiled by native speakers of Shona who also happen to be language experts. The dictionaries tend to provide adequate descriptive and/or explanatory senses for headwords. The senses are generally better when compared to the lexical equivalents provided for in the bilingual dictionaries which preceded them. The senses describe phenomena and/or events in a way that helps the user to conceptualise a thing that he/she has not even seen before. This contrasts with dictionaries which were compiled by those we can refer to as ‘language tourists’, that is, people who were neither mother-tongue speakers nor language experts of Shona. This could be the reason why their dictionaries, for example, those by Hannan (1959) and Dale (1981), leave a lot to be desired in terms of the way they document and also describe the language.

DRC and DGC can also be characterised as synchronic. According to Zgusta (1971: 202), the task of diachronic dictionaries is to deal with the development of the lexicon, whereas the purpose of synchronic dictionaries is to deal with the lexical stock of a language at one stage in its development. Words and their senses in DRC and DGC were collected and presented as they are used or understood in the Shona language community today. There is a field for etymological information on the database for DGC, but it was suppressed during the final stages of editing the dictionary. Although very little historical information was incorporated, this is not to suggest that these dictionaries contain no such information. A few names of people and places that are historically significant, were in fact included as headwords. However, when such headwords were defined, not every detail about them was given; only the basic information was provided. For example, there are two entries for the noun guruuswa. The first is defined in general as a forest with thick and long grass. The definition for the second refers to the historically significant place in East Africa where the Shona people once settled on their way from their original abode in West Africa to the present-day Zimbabwe. No information in this definition mentions the socio-economic and political way of life of the Shona people in the guruuswa area, information which is usually given in cases where historical detail is prioritised.

Unlike the bilingual Shona–English dictionaries already referred to, in which information on the dialectal sources of headwords and senses is provided, no such details are given in both DRC and DGC. Headwords and senses
are drawn from all areas where Shona is spoken and their sources are not shown in the dictionaries. In some cases, a single headword could have different meanings in different dialectal areas. In these cases, Shona–English bilingual dictionaries would indicate that a specific headword or sense is commonly used in a particular dialectal area. This lexicographic practice, though important in giving such comprehensive details, has the disadvantage of highlighting the differences among varieties of one and the same language. Research carried out by the ALLEX Project in the eastern parts of Zimbabwe prior to the publication of DRC, showed that speakers in this region identify themselves firstly as speakers of particular dialects before being speakers of union Shona. They were very keen to know whether vocabulary from their dialect areas was included in the dictionary. They were delighted to discover that words they use in their daily lives were actually given as headwords.

If the scenario described above is anything to go by, a dictionary that marks dialectal sources of headwords and senses could be described as a catalyst for division among speakers of a language. In fact, it is against this background that the compilers of DRC and DGC decided to exclude such information. It was felt that vocabulary in the dictionaries would be given equal status and treatment. In this way, the dictionaries would become melting-pots in which all Shona words are thrown to result in one product, standard Shona. The dictionaries can therefore be viewed as agents of unification between speakers of different geographical locations in the Shona-speaking communities.

4. Conclusion

In this article, a comparison between DRC and DGC was made. The writer has tried to show the ways in which the latter is considered a more advanced dictionary than the former. DGC is more advanced in two ways, that is, in terms of its size and the presentation of its meanings. The headword and sense selection for this dictionary was more comprehensive for it includes language used in all spheres of life. DGC is also the first dictionary in the history of Shona lexicography to include phrasal headwords such as proverbs, idioms and pithy sayings. The improvement in the quality of presentation of meaning can be accounted for by the use of the Shona corpus as well as the experience the editors gained from having worked on DRC.

The article also discussed a few of the similarities shared by the two dictionaries. That DRC and DGC would have much in common can be surmised from the mere fact that the same team compiled both of them.

References


