

Lexicographical Developments in the Shona Language as Reflected in the Making of the *Duramazwi Guru ReChiShona (DGS)*¹

N. MPOFU

African Languages Research Institute, University of Zimbabwe

Abstract

This article traces lexicographical developments in Shona, one of the major languages of Zimbabwe, with particular focus on corpus building and the role the corpus has played in Shona lexicography in the past hundred or so years and recent developments as reflected in the making of Duramazwi Guru ReChiShona by the African Languages Research Institute (ALRI) team of the University of Zimbabwe.

Background

Lexicography in Shona is not a new discipline. It dates as far back as the 1850s when missionaries began constructing orthographies for Shona speakers in the areas in which the missionaries were stationed. These early orthographies were to be used to construct vocabularies that would enable the translation of religious texts from English into Shona. From then until the 1990s, several glossaries and dictionaries were produced. As Fortune (1979, 1992) correctly observed, Shona dictionaries compiled in this period were all bilingual in nature. Their primary purpose was to provide a written basis for the lexical items of the language as a whole (Fortune 1992: 18) and were targeted at foreign mission workers, settlers, miners, and prospectors in order to aid them in their interactions and contacts with the local people. Most of these early publications were essentially grammar texts that merely described the nature of the language to non-Shona speakers.

According to Fortune, these early publications revealed both the compilers' very limited knowledge of the language and of the techniques of dictionary making (1992: 17). The fact that compilers of these early publications were describing a language that had not been written before, often worked in isolation in their remote mission stations and relied mainly on their own Bible translations for headwords or lexical items to use in the

1. Herbert Chimhundu ed 2001, *Duramazwi Guru ReChiShona*, Harare: College Press.

glossaries or manuals, led to the production of different orthographies, some of which distorted the Shona language. Moreover, the compilers did not give adequate coverage to cultural items mainly because they were either not aware of them, did not understand Shona culture, and, therefore, could not explain it, or they regarded African culture as inferior and in need of replacement by 'civilised' cultural items and notions.

Soon after Doke developed a Shona orthography in 1931, Reverend Barnes published *A Vocabulary of the Dialects of Mashonaland* (1932) which laid the foundation for subsequent serious lexicographic work in Shona. Barnes took the initiative to order entries alphabetically and to organise the words and their meanings in such a way that he was able to break away from the tradition of explaining words by providing examples of sentences in which they could be used.

Hannan and Dale published the two most known bilingual dictionaries in Shona in 1959 (revised and expanded in 1974) and 1981, respectively. These dictionaries demonstrated the compilers' knowledge of the techniques of dictionary making and showed that they had a more profound grasp of the language than their predecessors. Dale's 1981 dictionary, *Duramazwi*, though comprising a mere 249 pages and, therefore, not comprehensive in its coverage of Shona words, proved to be a useful record of the Shona lexicon.

Dale's dictionary was different from its predecessors in that it gave headwords and definitions in Shona before translating the definitions into English, 'thus paving the way for entirely monolingual Shona dictionaries' (Fortune 1992: 20). Unlike earlier publications, Dale's dictionary also provided synonyms, antonyms, and variants of the headwords, as well as illustrations to complement the given definitions.

More recently, with the publication of two mono-lingual Shona dictionaries, Shona lexicography has developed from merely being a means through which a non-Shona speaker can learn the language to being a record of the language in its own right. Compiled by Shona speakers, using modern techniques, the new dictionaries differ from all previous Shona dictionaries in that they treat Shona as both the object and the instrument of description.

The Corpus in Shona Lexicography

A corpus is a collection of texts, collected to facilitate the study of a language or part of a language. In order to construct a dictionary for any language, it is necessary to build up and analyse a corpus in order to establish which words are actually used by the native speakers and how they are used (Ore 1992: 20).

Previous Shona dictionaries were handicapped by the fact that they relied heavily on the Biblical literature that the compilers thought was

relevant to the exclusion of other important literatures and sources. Moreover, the lack of the appropriate technology to generate a reliable electronically processed corpus compounded the problems of dictionary making. Hannan and Dale used an index-card system to order and process the corpus they were using and relied mainly on biblical literature and Bible translations, whose language was not always reliable or appropriate. Not surprisingly, Hannan's dictionary contains some obscure and unrepresentative words, such as: *angere* (angel) instead of the known form *ngirozi*, *hafubhaki* (half-back), and *endekezi* (a volume of the Bible). Had he been using an electronically generated and processed corpus, it would have been very clear to him that such words were very uncommon and he might have entered them for historical interest only.

The great leap forward in dictionary making in recent times has been made possible by the use of information technology in both lexicographical research and the production and presentation of lexicographical material. Since the 1980s, there has been massive investment in the construction and exploitation of computerised corpora of naturally occurring language, both spoken and written (Singleton 2000: 198). The African Languages Lexical Project (ALLEX) of the University of Zimbabwe, now the African Languages Research Institute (ALRI), pioneered the use of electronic corpora in dictionary making in Zimbabwe and developed an electronically processed Shona corpus, generated from both written and oral sources, which, at 2002, has over 2.2 million running words. Oral sources include interviews, informal conversations, church services, classroom lessons and debates, while written sources include fictional material, ranging from prose to poetry and plays, and non-fictional material such as school textbooks, other Shona non-fiction literature, including literature in foreign languages that have been translated into Shona, such as *Tsanga Yembeu (A Grain of Wheat)* by Ngugi wa Thiong'o). All this material was then encoded or scanned, tagged, proofread and parsed and then included in the corpus.

The advantage of utilising computers in dictionary making is that compilers of dictionaries are easily able to identify the instances and contexts in which different words are used. Furthermore, with an electronic-processed corpus, it is possible to make concordance files, which record relevant information about words that the definer can then use to construct definitions.

The ALLEX team gathered its data from the various Shona-speaking areas of Zimbabwe with the help of research assistants who tape-recorded interviews and activities at churches, schools, sporting events and at individual homesteads. Thus the team was able to collect materials on different topics and issues in varied settings so as to capture as complete a range of regional variations as possible. This was consistent with Kipfer's (1984: 32) observation that 'the primary source of data for the dictionary

maker is the utterances of the speakers of that language ... the differences and changes that occur in a language must be recorded by the lexicographer'.

The ALLEX team also developed a concordance programme in order to identify the frequency of headword occurrences in the corpus and the different contexts in which particular words are used. Concordances were found to be useful also because they enabled the team to deduce the various meanings and styles of use associated with each word, making it easier to recognise what words were used in what contexts. The ALLEX team used the corpus it had built up in the compilation of two Shona dictionaries, *Duramazwi reChiShona* (1996) and *Duramazwi Guru reChiShona* (2001). The dictionaries are slightly different in scope and coverage because they are aimed at different audiences, the former being a general sized monolingual dictionary, while the latter is a much more advanced dictionary.

The corpus was used as a source for headwords, for identifying the different senses of each word and for citations. However, in both *Duramazwi reChiShona* (DRC) and *Duramazwi Guru reChiShona* (DGS), although material was collected in the various dialectal regions of the country, the headwords were not marked for dialect, nor is there any indication in these books to suggest the area(s) in which any of the words are spoken. All terms were treated equally and neutrally. With the help of the corpus, it was possible to identify the most common form of a word, and it was this form that was given as the main entry, while the less common forms were entered as variants under that entry. These variants were also entered and cross-referenced to the main entry. The dictionaries, thus, give a range of the variants and synonyms of the word, regardless of the districts in which they are used. For example:

nzara [zhara] D- z9 Nzara kunzwa kuda kudya . . . (Hunger)

zhara D- z9 Ona nzara 9. (See *nzara*)

-ngandudza [-ngandutsa] D it Kungandutsa kutungidza . . . (To set alight)

-ngandutsa D it Ona -ngandudza. (See *-ngandudza*)

Main entry definitions in *Duramazwi reChiShona* and *Duramazwi Guru reChiShona* are in sentence form, as well as the citations or examples of usage to give the user the context in which the particular word has been used. Not all citations were generated from the corpus, for as those involved in dictionary making will know, a corpus can never contain everything that is to be found in a language. Moreover, common day-to-day words, sometimes referred to as 'toothbrush words', have very low frequency counts or may not even be there at all. Nevertheless, they can not be left out simply because they do not appear in the corpus.

The Use of Databases in Shona Lexicography

An innovation in DRC and DGS was that they were developed through databases and not through word processing. The database used in the compilation of *Duramazwi reChiShona* had fields for the headword, the variants, tone, word class, noun class, verb information, namely, whether a particular verb was transitive or intransitive, the plural form of the word, the definition field, and synonym and antonym fields. Global definition and 'compare' fields were added. The dictionaries were produced without using any word-processing programme. The advantage of this method is that there are no shifts in the structure of the entries, thus making the final production process easier.

There are many advantages of using a database programme in dictionary making, including the ease of movement from one entry to another in the defining and editing stages and the greater consistency in handling words that are in the same category. More accurate cross-referencing between variants and synonyms and consistency checks for words of the same syntactic or semantic type are also possible. In addition, it is possible to suppress information that should not appear in the final manuscript but which can be recalled later for use in future dictionaries.

Other Developments

Shona lexicography has developed, not only in terms of the techniques of dictionary making, but also in respect of the innovations that have been introduced in the dictionaries themselves. *Duramazwi Guru reChiShona* was innovative in terms of lemma status and the structure of the dictionary itself. The dictionary is in two parts: Part 1 being the A-Z section of the dictionary including idioms, while Part 2 is the section with proverbs and figures of speech. Proverbs and idioms were both given lemma status, while the tradition with other Shona dictionaries has been to give them as run-on entries under the most dominant noun or verb. The issue of how to handle or where to place multi-word lexical units posed problems to the editors. The decision to enter idioms and proverbs as headwords was designed to compile a more user-friendly dictionary.

Another innovation was that, for lemmas with more than two senses, a global definition, in the form of a paraphrase and not a complete definition, was given. The global definition helps the user, who wants to get a sense of the general meaning of a word, to do so. For instance, for the verb *-bata*, the global definition is:

-bata K it Kugunzva noruoko kana kuisa muchanza . . . (to touch with the hand or to hold in your hand).

Other words related to headword were indicated by the use of the *tarisa* [TAR] (compare) marker.

Several style markers were also incorporated to mark special senses. These style markers include *manje* (*chimamdamani*) for slang or colloquial uses; *kare*, for an archaic sense of the word; *kapato*, for trademarks; *tiku* (*chituko*), for swear or offensive words; *nyatso* (*chunyatso*), for taboo words; and *nhanha*, for baby talk. These markers were designed to indicate the various nuances of particular words.

Through its contact with other languages, Shona has acquired and naturalised many non-Shona terms, mostly from the English language. These loanwords posed challenges in the compilation of DGS because of their orthographic make-up, as they contained some letters that are not in the current orthography. The problem with the current Shona orthography is that it does not contain the letters *l*, *q*, *x* and the digraphs *th* and *rh*. One thus finds that some loanwords only exist in speech, though they may be everyday usages, but are not accepted when written down. Examples are *thiyeta* (operating room and movie hall), *thiyori* (theory), *themomita* (thermometer), *losheni* (body lotion), and *laibhurari* (library).

The problem that faced the compilers of DGS was whether to leave out such words completely or to include them and, if so, how to handle them. Previous dictionaries had completely left out such forms. The compromise that was finally reached was to enter *l* words as variants of *r*, thus *losheni* was entered under *l* with an asterisk (*) to indicate that this form was not acceptable in the current orthography, and then cross-referenced to *rosheni* defined under *rosheni*. The *th* words were also entered with an asterisk (*) and defined because there was no other way of handling them. This problem brought to light the discrepancy that exists between speech and writing in Shona, which requires urgent attention in order to cater for such loanwords, which have no Shona equivalent but which are increasingly used by Shona speakers.

There are some loanwords in Shona that are phonetically aspirated such as *resipi*, *ragibhi*, *vege*. These have continually been written without an *h* because of the current orthography, yet in speech they are aspirated. In DGS, both aspirated and unaspirated forms were entered, with the aspirated form being spelt with *rh* and carrying an asterisk (*) and cross-referenced to the unaspirated form.

Duramazwi Guru reChiShona also went beyond previous Shona dictionaries by providing a comprehensive back matter section, comprising no less than 41 pages and providing information on names of African countries, scales of measurement, judiciary terms, colour terms, times of the day, days of the week, months of the year, seasons of the year, names of chiefs, their areas of jurisdiction and their totems and clan names, as well as literary and grammatical terms.

Conclusion

With the making of DGS, Shona lexicography has indeed come a long way, from being simply a tool which a non-Shona speaker could use in his/her quest to learn a new language, to the current situation in which Shona is presented and defined in Shona, rather than English.

References

- BARNES, B. H. 1932, *A Vocabulary of the Dialects of Mashonaland*, London: The Sheldon Press.
- CHIMHUNDU, H. ed 1996, *Duramazwi reChiShona*, Harare: The College Press.
- ed 2001. *Duramazwi Guru reChiShona*, Harare: The College Press.
- DALE, D. 1975, *A Basic Shona-English Dictionary*, Gweru: Mambo Press.
- DALE, D. 1981, *Duramazwi: A Shona-English Dictionary*, Gweru: Mambo Press.
- FORTUNE, G. 1979, 'Shona lexicography', in *Zambezia*, 11 (i), University of Rhodesia.
- FORTUNE, G. 1992, 'General lexicographic experiences — Zimbabwe', in H. Chimhundu ed, *African Languages Lexical Project First Workshop Report*, Harare: University of Zimbabwe.
- HANNAN, M. 1959, *Standard Shona Dictionary*, Salisbury: The Literature Bureau.
- 1974, *Standard Shona Dictionary*, Salisbury: The Literature Bureau.
- HARTMANN, A. M. 1893, *An Outline Grammar of the Mashona Language*, Cape Town: Juta.
- 1894, *English-Mashona Dictionary*, Cape Town: Juta.
- KIPFER, B. 1984, *Workbook on Lexicography*, Vol. 8, University of Exeter.
- LOUW, C. S. 1930, *Manual of ChiKaranga*, Bulawayo: Philpott and Collins.
- MARCONNES, F. 1929, *A Manuscript Grammar of Karanga*, Johannesburg: Witwatersrand University Press.
- SINGLETON, D. 2000, *Language and the Lexicon*, UK: Arnold Publishers.