
Linguistic Variation in Shona with Special Reference to Monolingual Dictionaries

Emmanuel Chabata, *African Languages Research Institute, University of Zimbabwe, Harare, Zimbabwe (emmanuelchabata@yahoo.com)*

Abstract: This article focuses on the problems lexicographers of monolingual dictionaries face when dealing with a language characterised by differences associated with geographical variation. The article specifically seeks to explore problems with which Shona lexicographers are confronted when working on monolingual Shona dictionaries. It does so by looking at some lexicographic aspects such as tone marking, sense ordering and treatment of synonyms and variants. Linguistic variation has always been a problem for writers of reference works, especially those normative in nature and function. The challenge is that of representativeness, that is, the production of a work which incorporates all the information it is supposed to provide. If the work fails to represent the language of the total population it is assumed to cover, then it is prone to criticism, usually by those whose dialect is not well represented. The article contains the writer's experience in the African Languages Lexical (ALLEX) Project, at present the African Languages Research Institute (ALRI). The ALRI, accommodated at the University of Zimbabwe, has the compilation of monolingual dictionaries for the indigenous languages of Zimbabwe as one of its major objectives. Although illustrative examples are drawn from Shona alone, it is, however, believed that the problems highlighted in this article may not be peculiar to Shona, but can also apply to other Bantu and world languages.

Keywords: ALLEX PROJECT, ALRI, CORPUS, DIALECT, MONOLINGUAL LEXICOGRAPHY, SHONA, SUBDIALECT, SYNONYM, TONE, VARIANT, VARIATION

Opsomming: Linguistiese variasie in Sjona met spesiale verwysing na eentalige woordeboeke. Hierdie artikel konsentreer op die probleme waarvoor leksikograwe van eentalige woordeboeke te staan kom wanneer hulle te doen kry met 'n taal gekenmerk deur verskille wat verband hou met geografiese variasie. Die artikel wil veral probleme ondersoek waarmee Sjonaleksikograwe te doen kry wanneer hulle aan eentalige Sjonawoordeboeke werk. Dit word gedoen deur na 'n aantal leksikografiese aspekte soos toonaanduiding, betekenisordening en die behandeling van sinonieme en variante te kyk. Linguistiese variasie was nog altyd 'n probleem vir die skrywers van naslaanwerke, veral dié wat normatief van aard en funksie is. Die uitdaging is dié van verteenwoordigheid, dit wil sê die totstandbrenging van 'n werk wat al die inligting bevat wat dit veronderstel is om te verskaf. As die werk in gebreke bly om die taal van die hele bevolking te verteenwoordig wat dit aangeneem word om te dek, dan is dit vatbaar vir kritiek, gewoonlik deur diegene wie se dialek nie goed verteenwoordig is nie. Die artikel bevat die skrywer se ervaring in die African Languages Lexical (ALLEX) Project, tans die African Languages Research

Institute (ALRI). Die ALRI, gehuisves by die Universiteit van Zimbabwe, het as een van sy belangrikste doelwitte die samestelling van eentalige woordeboeke vir die inheemse tale van Zimbabwe. Alhoewel verduidelikende voorbeelde slegs aan Sjona ontleen word, word daar egter geglo dat die probleme wat in hierdie artikel na vore gebring word, nie net eie is aan Sjona nie, maar ook geldig is vir ander Bantoe- en wêreldtale.

Sleutelwoorde: ALLEX PROJECT, ALRI, KORPUS, DIALEK, EENTALIGE LEKSIKOGRAFIE, SJONA, SUBDIALEK, SINONIEM, TOON, VARIANT, VARIASIE

1. Introduction

Researchers who carry out quantitative and qualitative research in various disciplines often refer to the term *variation*. However, this term means different things in different fields. In this article it is asserted that in the field of language, *variation* refers to systematic differences existing between different languages or between varieties of the same language. No language can be described as homogeneous; every language is characterised by a variation in expression. In fact, a simple observation of any language would show that no two speakers of a language speak in exactly the same way.

The issue of language variation can be looked at from different perspectives and at different levels. At the individual level, such factors as mood, health, age, attitude and context can lead a person to use different forms of language to refer to the same thing at different times. At group level, differences can, among others, be determined by social grouping or geographical location. Some factors such as differences in sex, age, social status and region of origin are critical in explaining the differences existing between a variety used by one group of people and that used by another. Whatever the dominant factor, the differences present can be realised at all levels of linguistic structure, that is, phonological, lexical and grammatical levels.

For a fuller understanding of the forces at work, it is usually more appropriate to discuss the joint factors leading to language variation. However, such a large-scale discussion is beyond the scope of this article which narrowly focuses on regional variation, that is systematic differences among groups of speakers of the same language coming from different geographical locations. According to Francis (1983: 15), regional variation results from differential change. Francis' view is that language change is inevitable and that when changes occur, they characteristically affect the speech of only part of the total language community. For example, different regions of a language community, by virtue of differences in their geographical location, would experience different linguistic changes at different times in the history of their language.

Fromkin and Rodman (1993: 277) also note that dialectal diversity develops when speakers of the same language are separated from each other geographically or socially. In this case, the changes that occur in the language spoken in one area or within one group of people do not spread to another. When

some communication barrier, be it a physical barrier such as an ocean or a mountain range, or a social barrier of a political, racial, class or religious kind, separates groups of speakers, linguistic changes are not easily spread and dialectal differences are reinforced. A change that occurs in one region and fails to spread to other regions of the same language community gives rise to dialectal differences which also tend to increase proportionately to the degree of communicative isolation among the groups of people. Fromkin and Rodman (1993: 277) argue that when such differences give the language spoken in a particular region its own flavour, this version of the language is referred to as a regional dialect.

The issue of regional dialects is one of many aspects that lexicographers have to treat with great care. This is because their products have a readership that usually cuts across different varieties or dialects. The challenge for the lexicographer is to produce a work that would accommodate the differences among the respective varieties without a bias towards one or the other variety. In the subsequent sections of this article, we will examine the dialect situation in Shona and see how it has added to the general problems Shona lexicographers of monolingual dictionaries encounter in their work.

2. Dialect Situation in Shona

Shona is a Bantu language spoken in Zimbabwe by approximately 80 percent of the country's total population. The language is mainly spoken in the central, southern, northern and eastern parts of the country. As in any other language, there is dialectal variation in Shona. In fact, taking geographical distribution as the determining factor, Shona is understood to consist of five major regional dialects: Zezuru, Karanga, Manyika, Ndaou and Korekore, each of which can also be divided into numerous subdialects. Zezuru is spoken in the central region of the country, in areas surrounding Harare, the capital city. Karanga is spoken in the southern parts, Manyika in the east, Ndaou in the southeastern parts and Korekore in the northeastern parts of the country. Although these dialects have a high degree of mutual intelligibility, they show systematic differences. The varieties differ in pronunciation, vocabulary and, at times, even syntactically.

The regional names under which the respective varieties of Shona are known today are labels that came with the colonisation of Zimbabwe by British settlers at the end of the 19th century. In order to appreciate the present dialectal situation in Shona, one has to understand the history of the Shona language from the arrival of missionaries in Zimbabwe to the present day. Because a more comprehensive discussion of this topic is given in Chimhundu (1992), only a brief overview would suffice here. But before we consider the Shona situation, it may be necessary to note a general observation about Africa made by Chimhundu (1992: 88) who says:

An exaggerated multi-ethnic, multilingual and multi-tribal picture of the African colonies has been painted through misinterpretation and inadequate study on the part of the early missionaries and through manipulation for administrative convenience on the part of the colonial governments.

This point is important because it tries to capture the processes that brought Africans (including the Shona) into their present sociopolitical and economic situation. The exaggeration of differences rather than similarities among people has tended to separate rather than unite people with a common language, history and destiny. Divisions created during the colonial era have caused people to identify themselves as belonging to particular groups, a development which makes co-operation and compromise among these groups more difficult and at times even impossible. These divisions have also tended to complicate the task of those who strive towards unifying the different groups or those who work on projects cutting across the established boundaries.

To return to the case of Shona: Before the arrival of missionaries, there were no known distinctions based on language in Mashonaland, that is, in the area of Zimbabwe where Shona is spoken. The people of this region defined themselves politically as subjects of particular chiefs coexisting with other chiefdoms in unlimited zones of a common culture (Chimhundu 1992: 91). At that time, they did not identify themselves according to language or dialect because, since they considered themselves as speaking the same language despite the minor differences that existed, the linguistic factor was not important. However, when the missionaries came to Zimbabwe during the last quarter of the 19th century, they tended to view all chiefdoms as tribes and also simplistically equated these to tribes speaking different languages or dialects. As noted by Chimhundu (1992: 91), this identification was more convenient for the missionaries, and later for the colonial government since it meant they had to deal with fewer chiefs with territorial claims over wider areas. For the missionaries, it was easier to have dealings with chiefs who would, in turn, become agents in spreading the gospel to their respective subjects rather than taking their own initiatives without the chiefs' backing. These missionaries, therefore, naturally helped in the creation of wider chiefdoms or groupings of chiefdoms, with tribal labels more or less corresponding with the spheres of influence of different missionary societies on the one hand, and the administrative districts and provinces of the settler government on the other (Chimhundu 1992: 91).

One important point to note about the missionaries who came to Zimbabwe is that they were not a single society operating in one area. According to Kahari (1996: 3), different missionary groups strategically deployed themselves to all four corners of the region: the American Methodists among the Ndau at Chirinda and the Manyika at Umtali (present-day Mutare); the Anglicans among the Manyika at St Augustine; the British Methodists among the Zezuru at Waddilove; and the Roman Catholics among the Zezuru at Chishawasha and Kutama. The Dutch Reformed Church of South Africa established itself among the Karanga at Zimuto, while the Salvation Army settled among the

Zezeru at Howard in the Chiweshe tribal area. Operating from these different mission stations, the missionaries set up presses for printing church literature, Bibles and educational books to create literature for the new converts who had become literate (Kahari 1996: 3).

As in most parts of Africa, formal education was seen as a strategic instrument in spreading the Christian gospel in Zimbabwe. By bringing children to mission schools, missionaries were able to teach their Christian message to them at an early age, away from interference by elders who were more difficult to convince and convert. At most mission stations in different regions of the country, mission schools were established. Students were recruited and converted. Later on, the same students were used as agents in spreading the religious teachings to their parents and other people in their local communities. However, the important point to note is that the missionary efforts did not lead to a co-ordinated educational policy. The result was that each missionary society established a different system of education for its schools. They could not agree on a common system of writing. Each group, therefore, developed its own form of Shona orthography based on the dialect of the region in which it operated. In the end, different orthographies were thus developed in the different regions. The result of these developments was the creation of dialect clusters which we have already identified as Zezeru, Karanga, Manyika, Ndau and Korekore.

The Shona people who, before the coming of the Europeans, identified themselves collectively, began to see themselves as belonging to specific dialect areas or regions. This state of affairs was later reinforced by the colonial government employing the divide-and-rule tactic as one of its strongest tools for separating the Africans who could otherwise be a formidable force were they allowed to continue identifying themselves collectively. The colonial regime strengthened the artificially created differences at the expense of the overriding and self-evident similarities. This situation was allowed to continue up to the late 1920s.

From 1903 to 1928 there was a feeling among different missionary societies that a common Shona orthography, which they all needed, should be developed. This issue was kept on the agenda of the Southern Rhodesia Missionary Conference (S.R.M.C.), an association of the various denominations working in the country. Because they could not reconcile the orthographies their constituent members had developed independently of one another, efforts by members of the S.R.M.C. in effecting a common orthography were not successful. Church representatives participating in the discussions took the stance of provincial and tribal defenders (Chimhundu 1992: 99). Following a request by the S.R.M.C., the colonial regime in Southern Rhodesia (now Zimbabwe), in 1929, invited C.M. Doke, a language expert from South Africa, to come and study the language situation in the country and then to make recommendations for the development of the country's languages, including Shona. After a survey of the language situation, Doke in 1931 published his *Report on the Unification of the*

Shona Dialects in which he recommended the creation of Union Shona by combining the varieties identified in the Shona language. Of interest with regard to this article is the fact that some dialects were recognised but the use of others for creating Union Shona were discouraged. For example, one of Doke's recommendations was that Union Shona was to draw its vocabulary from Zezuru, Karanga, Manyika and Ndaou. Korekore words were supposed to appear only sparingly whilst the inclusion of words from Budya and all the other smaller dialects were discouraged (Doke 1931: 81). For political, economic and geographical reasons, Zezuru was elevated to become the point of reference or the core when creating Union Shona. It was chosen, for example, as the norm for phonetic analysis. In written Shona, Zezuru, whose subdialects were fewer, was to be preferred above all the other varieties. This development elevated Zezuru to a position where it gained prestige, prominence and eventually dominance over the other dialects. The result was a diglossic situation where Zezuru was the high variety and all other dialects were treated as low varieties.

After Doke's recommendations, a standard Shona orthography based on Zezuru was encouraged and even promoted. It is, however, important to note that this development, paralleled with the discouragement of the use of vocabulary from other dialects, did not go without opposition from speakers of other dialects. Chimhundu (1992: 83) notes that the recommendation by Doke to create Union Shona based on Zezuru was opposed by speakers of other dialects from as early as the 1930s. Even today, some people resent Zezuru dominance which is still evident in literature written in Shona. For reasons of ethnic identity, non-Zezuru speakers would want to see their vocabulary reasonably represented in all forms of literature which are national in nature and character. On the other hand, Zezuru speakers are quite happy to see a continued promotion of their variety to the extent of having it as the only variety for standard Shona. Evidence for such feelings can be found in Kahari (1996: 7), who had this to say soon after the publication of Chimhundu's *Duramazwi ReChiShona* in 1996: "I have fears that the pendulum has swung from one extreme, Zezuru, to the other, Karanga, and the other dialects in between have suffered." Kahari here regrets the attempt by the compilers of *Duramazwi ReChiShona* to try and balance inputs from Zezuru and the other dialects. For Kahari, Zezuru's dominance should stay. Any attempt to put the dialect on the same level with other dialects he considers retrogressive. On the other hand, a field research trip carried out by the ALLEX Project in Manicaland in 1996 showed that speakers of Manyika and Ndaou dialects felt the kind of domination that would result from the continued use of Zezuru in written Shona literature at the expense of the other dialects, would not only be linguistic, but also social, economic and even political. Because of this, they strongly resent this domination.

From the discussion above, we can see that the Shona language situation as it exists today is heterogeneous and the social situation is fluid (Chimhundu 1979: 78). It is in this situation that the lexicographer of monolingual Shona dictionaries finds himself. An analysis of this situation would show that Shona

lexicographers are bound to have problems, some of which emanate from the structure of the language and some from the social situation described above. In the next section, we will look at some of the problems the team of lexicographers compiling Shona dictionaries has so far encountered.

3. Challenges for the Lexicographer of Monolingual Shona Dictionaries

Dictionary users consider dictionaries as authoritative reference works. They are believed to be accurate representations of the languages they describe. They are also seen as instruments for describing languages as wholes and not as parts of wholes. Perhaps this is the reason why they have an authoritative influence on aspects of language growth and development. One of the requirements a dictionary, especially a monolingual one, should strive to fulfil is capturing the way a specific language is used by its total speech community. In this case, the role of the lexicographer of a monolingual dictionary would be that of producing a dictionary that unites rather than separates people. A dictionary which excludes vocabulary from other dialects or which emphasises vocabulary from a selected dialect or dialects cannot serve this very important function. The reason is that speakers from ill-represented varieties would find it difficult to identify with the dictionary. This would also have a negative impact on the marketing or acceptance of the dictionary, as well as on the general development of the language involved. Given the nature of the responsibilities of the lexicographer of a monolingual dictionary, the process of compiling a dictionary which balances variations from different dialects of a language is not an easy task. In this section, we will explore the problems lexicographers face in compiling monolingual Shona dictionaries which are representative of the varieties of the language, especially given the fluid social situation characterising the language.

Monolingual Shona lexicography started with the launch of the ALLEX Project in September 1992 when a group of lexicographers undertook the compilation of monolingual dictionaries for the indigenous languages of Zimbabwe. Their efforts have so far culminated in the publication of *Duramazwi Re-ChiShona* (Chimhundu 1996) and *Duramazwi Guru Re-ChiShona* (Chimhundu 2001), both of which are medium-sized general monolingual Shona dictionaries. During the compilation process, the team has encountered a number of problems rooted in dialectal variation. One such problem concerns the treatment of variants. Variants are taken here to refer to headwords meaning the same but differing slightly in pronunciation. The pronunciation usually differs with respect to one syllable of the word forms concerned, and is usually correlated with dialectal differences. Examples of Shona variants include the following pairs:

<i>shuro</i>	vs	<i>tsuro</i>	(hare)
<i>nzara</i>	vs	<i>zhara</i>	(hunger)
<i>-famba</i>	vs	<i>-hamba</i>	(walk/travel)

<i>-dya</i>	vs	<i>-ja</i>	(eat)
<i>chii</i>	vs	<i>chinyi</i>	(what?)

Both words forming a pair of such variant forms would appear as dictionary entries. However, because the word forms basically mean the same, they cannot both be given definitions for this would occupy a lot of space without adding anything to the explanation of the entries involved. A more economic way of dealing with such headwords would be to make one of each pair the main headword entry carrying the sense(s). The second word form would not carry a definition, but would just be cross-referred to the main entry. We can illustrate this with the example of *shuro* and *tsuro* given above. If, for example, we decide to take *tsuro* as the main entry, we would present the definition(s) under it. We would also indicate that *shuro* is a variant form of *tsuro*. When we come to *shuro*, information about tone and word class would be given but for the definition(s), the dictionary user would be referred to *tsuro*. The problem with this kind of presentation is to decide which word form to make the main entry, since by making one of the pair the main entry it would appear as if it has been given some superiority over the other.

The principle so far adopted in Shona dictionaries is to use the more common form as the main headword. However, the immediate question that arises is how to determine the more common form of the two. Relying on one's intuition in making decisions on such cases does not always work, for what may be common to one person or a group of persons may not necessarily be common to another person or group. In such cases corpora would be useful. The form with more occurrences in the corpus would automatically become the main entry. However, this is not as simple as it appears on the surface. The reality is that the usefulness of a corpus depends on a number of factors, for example, on whether the corpus being used can be relied on as truly representative of language use as it applies to all the regional dialects of the language concerned. If the corpus used is biased towards one or more dialects, the picture that would emerge may be misleading.

The challenge of making an informed choice between two word forms that should be cross-referred to each other is also encountered when dealing with synonyms. Synonyms are taken here to refer to two or more different words which mean the same, or rather, which refer to the same object or event. Unlike in the case of variants where differences in terms of form are minimal, synonymous words differ greatly with regard to form. Their relationship lies in their similarity of lexical meaning. Examples of Shona synonyms include the following:

<i>gudo</i>	vs	<i>dede</i>	vs	<i>bveni</i>	vs	<i>diro</i>	(baboon)
<i>-mhanya</i>	vs	<i>-rumba</i>	vs	<i>-gogoma</i>			(run)
<i>-konza</i>	vs	<i>-nyenga</i>	vs	<i>-pfimba</i>			(declare love)
<i>hwahwa</i>	vs	<i>doro</i>	vs	<i>mhamba</i>			(beer)
<i>chibage</i>	vs	<i>barwe</i>	vs	<i>bonore</i>	vs	<i>gwere</i>	(maize cob)

The treatment of synonyms in Shona dictionaries is almost the same as that of variants. The more or most common of the synonymous words would carry the meaning and would become the main entry. The other(s) would be cross-referred to it. However, as noted in the case of variants, the problem is that of choosing the main entry and the ones to be cross-referred. We can take as example the Shona verbs *-mhanya* and *-rumba* given above. Whilst *-rumba* might be common to a Ndaou speaker, it may not be common to speakers of the other dialects. Although *-mhanya* seems to be more common to speakers of more than one dialect (also according to the Shona corpus where it appears 874 times as compared to the 29 occurrences of *-rumba*), making it the main entry may meet resistance from speakers who do not use it in their everyday language. In fact, the concept of making the more common word form the main entry is not acceptable to speakers of smaller dialects whose representation in the corpus is usually limited. This can be explained mainly by the fact that a large part of the Shona corpus comes from written materials, and the written literature in these dialects is very small. To speakers of these dialects, a strict following of what is in the corpus would appear to be just a way of consolidating the prestigious position of those dialects that have for a longer time been used in writing. Despite the fact that their vocabulary is included in the dictionary, they may feel that it is playing a secondary role if most of the headwords are not made main entries.

The arrangement of senses also poses a challenge, especially if the senses come from different dialect areas. Often there are cases where a headword has more than one sense. As far as the arrangement of the senses is concerned, the principle adopted for Shona dictionaries has been to list the more commonly used or the literal sense first and the rarely used or the metaphorical sense(s) subsequently. To illustrate this, we can take the Shona example *-bikira* which has two senses, that is, (a) prepare food for someone and (b) prepare a love potion for a beloved. When the above principle is applied, the first of these two senses would be listed as the first sense in the dictionary because its meaning can easily be deduced by combining the meanings of the main verb *-bik-* (cook) and that of the applied extension *-ir-*. The second sense, which is metaphorical, would consequently be listed as a second sense. It is important, however, to note that some senses tend to be more common in some areas and in some contexts than others. Also, what may be regarded as a metaphorical sense in one region may actually be a general and basic sense in another region. As illustrative example we can use the verb *-mamisa* which can mean (a) cause someone to defecate or (b) beat someone thoroughly. Whilst the second sense of this verb may be regarded as metaphorical by speakers of other dialects, to Ndaou speakers it is the general and basic sense. In such cases, lexicographers would end up using their intuition, a practice which may not reflect the language use of most Shona speakers.

Another problem is linked to the marking of tone. One function of a dictionary is to provide the correct way the words of a language are pronounced.

In fact, researches carried out in several languages have shown that tone marking in dictionaries is not only important to the mother-tongue speakers of a language but is also important to second-language speakers, since they have to learn how to pronounce individual words correctly. To Shona lexicographers, tone marking is a big challenge because there often are cases where one word form is pronounced differently across regional dialects. The following can be used as illustrative examples:

<i>sekuru</i>	(HHH/LHL)	(grandfather/uncle)
<i>-fashaidza</i>	(L/H)	(boil water or something else)
<i>barwe</i>	(HH/HL)	(maize cob)

As can be seen from these examples, a word form can be pronounced with more than one tone. In trying to accommodate this kind of variation, the Shona team has resorted to the policy of marking the different tones that can be realised with a particular headword. The problem of ignorance, however, sometimes arises. Although a lexicographer is expected to know everything about the language with which he is working, this is not the case. Sometimes he may not be aware of the second or third pronunciation of a certain word. To give such a word a single tonal realisation may not accurately capture the situation that pertains in all dialect areas, resulting in a deficiency in the dictionary.

In this section, we have taken examples from Shona to try and show how linguistic variation within a language can be problematic in the compilation of a dictionary. Issues such as synonymy and tone marking have been explored and some problems Shona lexicographers encounter when dealing with these issues, have been singled out.

4. Conclusion

In this article, we have tried to show some of the difficulties Shona lexicographers face as a result of working with a language characterised by variation. The major problem that has been highlighted is that of trying to compile a dictionary balancing inputs from different dialects. Besides the technicalities involved in the compilation process, the article has shown that some problems concern the attitude towards one dialect or another. The article has also indicated the need for well-informed principles when making choices between pairs or among groups of words.

Space has also been devoted to the discussion of the dialectal situation in Shona. By using the example of Shona, it has been shown that differences existing between speakers of the same language are not always natural. Some have been imposed from outside. To emphasise such differences can only impede research such as dictionary compilation which cuts across different regions or dialects. Although examples have been drawn from one language, the problems discussed and the arguments put forward in this article can apply to almost all languages since almost all languages show variation.

References

- Chimhundu, H.** 1979. Some Problems Relating to the Incorporation of Loan Words in the Lexicon. *Zambezia* 7(1): 75-91.
- Chimhundu, H.** 1992. Early Missionaries and the Ethno-linguistic Factor during the 'Invention of Tribalism' in Zimbabwe. *Journal of African History* 33: 87-109.
- Chimhundu, H. (Ed.)**. 1996. *Duramazwi ReChiShona*. Harare: College Press.
- Chimhundu, H. (Ed.)**. 2001. *Duramazwi Guru ReChiShona*. Harare: College Press.
- Doke, C.M.** 1931. *Report on the Unification of the Shona Dialects*. Hertford: Stephen Austin.
- Francis, W.N.** 1983. *Dialectology: An Introduction*. London/New York: Longman.
- Fromkin, V. and R. Rodman.** 1993. *An Introduction to Language*. New York: Harcourt Brace Jovanovich
- Kahari, G.P.** 1996. A Review of *Duramazwi ReChiShona*. Harare.