

# **Problems of Citation Forms in Dictionaries of Bantu Languages**

JOHN G. KIANGO

*University of Dar es Salaam, Tanzania*

## **ABSTRACT**

The headword of a dictionary entry serves different functions. First, it is a form which the user will first look up in the dictionary before reading other phonological, morphological, syntactic, semantic, and etymological information associated with it. Second, it is a form, which the user looks up to check for the spelling of the word. Third, for most entries, it serves as a representative or more technically a stem-from which a set of inflected forms of the same word could be generated by way of affixation. Fourth, it is a form that represents one meaning or several related senses.

In order to achieve these functions, the 'headword' should be entered into the dictionary in its conventional spelling, i.e. its visual representation as in ordinary writing system and not its actual sequence of acoustic sounds as in pronunciation. In lexicography, the word in its conventional spelling (the orthographic word) is also referred to as a 'citation-form'.

However, there are a number of problems relating to citation-forms that are encountered by Bantu lexicographers in the course of compiling dictionaries of natural languages. These problems are such as the problems of alphabetizing headwords, especially verbs, adjectives, and words with pre-prefixes, the problem of entering unnatural forms, the problem of entering obsolete/unactivated stem-forms, and the problem of entering unidentifiable stems.

This article discusses such problems faced by Bantu lexicographers and suggests ways of solving them.

*Keywords: Alphabetization, affixation, Bantu lexicography, citation-form, complex morphological structures, dictionary entry, headword, infinitive forms, obsolete stem-form, inflected forms, lexeme, prefixes, and unnatural forms*

## **1.0 INTRODUCTION**

The headword of a dictionary entry serves different functions. First, it is a form which the user will first look up in the dictionary before reading other phonological, morphological, syntactic, semantic, and etymological information associated with it. Second, it is a form, which the user looks up to check for the spelling of the word. Third, for most entries, it serves as a representative or more technically a stem-from which a set of inflected forms of the same word could be generated by way of affixation. Fourth, it is a form that represents one meaning or several related senses. Fifth, in old tradition, the headword was used also to indicate the pronunciation of the word. But nowadays most dictionaries prefer to indicate pronunciations immediately after the headwords.

In order to achieve the first two functions, the ‘headword’ should be entered into the dictionary in its conventional spelling, i.e. its visual representation as in ordinary writing system and not its actual sequence of acoustic sounds as in pronunciation. For example, the word ‘cat’ should be written as CAT not as /kaet/. In linguistics, the word written in its conventional spelling, e.g. CAT is what is called the ‘orthographic word’, and the word written in its actual sequence of acoustic sounds, e.g. /kaet/ is what is called the phonetic word. In lexicography, the word in its conventional spelling (the orthographic word) is also referred to as the ‘citation-form’.

The headword is also supposed to serve as **a stem**. It is a stem by which a set of inflected forms of the same word could be generated by way of affixation. In addition, it is a stem for which a set of related senses could be said to originate from it. For example, the word ‘support’ is the stem for inflected forms, such as: supports, supporting and supported. In addition, it is a form which represents multiple related senses such as: to help, to assist, to promote the interest or cause of, to uphold or defend as a valid right, to pay the cost of, to hold up or serve as a foundation or prop for, to keep something going (c.f. Webster New Collegiate, 1979). In order to make sure that the headword in a dictionary is a form that is both **a stem** and **a form that represents only one meaning or several related senses**, linguists distinguish different concepts of word forms. On the one hand they distinguish between a lexical unit and a lexeme, and on the other hand between a lexical form and a canonical form.

A ‘lexical unit’ is defined by Cruse (1986: 77) as ‘the union of a lexical form and a single sense’. Therefore, in our example given above, the word ‘support’ in unity with each of the enumerated senses forms a lexical unit. Hence our example above of the word ‘support’ has six lexical units. This is contrasted with a ‘lexeme’, which is defined by Cruse (1986: 76) as ‘a family of lexical units’. Based on this definition, we can establish that the word ‘support’ given in our example above in unity with the six related senses form **a single lexeme**. The emphasis is but here on ‘related senses’. Since if the senses are not related, they don’t form a family and hence belong to different lexemes, although they share the same stem/form. For instance, the meanings of the word ‘case’ are (a) a box or receptacle for holding something, and (b) a law suit. These two meanings are totally unrelated. Therefore, although the same lexical form represents them, they belong to different lexemes. They need to be entered into the dictionary as two different entries under two different headwords, *viz.* case1 and case2.

In summary, the ‘headword’ is sometimes referred to as a ‘canonical form’ as it serves as **the stem** of the word from which other inflected forms of the same word could be generated. Secondly, it is also referred to as **a ‘lexeme’** since **it represents a family of lexical units**, i.e. a unity of the word and several related senses. Finally, it is referred to as **a ‘citation form’** as its form represents a conventional spelling (orthography) of the word (lexeme) in question.

However, due to the multifunctional nature of **the headword** as discussed above and other lexicographical constraints such as alphabetic listing of

headwords in the dictionary, Lyons (1977) didn't hesitate to make the following observation:

This may not be the same as the citation form that is used in the everyday reflexive use of a language in a particular language community; and there may be alternative conventions in operation, e.g. the use of the infinitive form of the verb vs. its first-person singular form in Latin. It is important to realise that the citation-form is indeed a form of the lexeme (being used for a particular reflexive or metalinguistic purpose); it is not to be identified with the lexeme itself (Lyons 1977: 19).

The observation made by Lyons (1977) in this quotation is very valid and is one of the main lexicographical problems in Bantu lexicography and other lexicographies dealing with agglutinating and inflecting languages. His observation could be more appreciated in Section 2.3 below, dealing with problems of Bantu lexicography associated with citation-forms.

## 2.0 THE MAIN FEATURES OF CITATION-FORMS

After we have identified what is meant by the term 'citation-form' (c.f. 2.1), it is important to round down its main features. This will give us a basis from which problems of Bantu lexicography associated with citation-forms could easily be identified and discussed. In doing this, I will adopt what Newell (1995) called 'guidelines for choosing citation-forms' as the main features of citation-forms. Newell (1995) acknowledges that he gathered the information from a number of literature on lexicography. Thus, the main features of citation-forms are as follows:

1. Non-inflected forms, which occur naturally as (phonological) words within oral and written forms.
2. Forms with minimal affixation, i.e. that are closest to the basic root or stem.
3. Forms that are free from morphophonemic features.
4. Forms that are regarded by native speakers as 'natural' citation-forms, i.e. the ones that come quickly in their mind when they want to look them up in the dictionary.
5. Forms that are easily identifiable in isolation.
6. Forms that occur most frequently.
7. Forms that represent the basic meanings of the lexical items.
8. Stem-forms from which other inflectional forms of a paradigm could be constructed.
9. Forms that could be used to derive sub-entries. (Newell 1995: 245)

At this point we have established a meaning of a citation-form and have identified the main features of such a form. In the following, we discuss the problems of Bantu lexicography that are associated with citation-forms.

### 3.0 LEXICOGRAPHIC PROBLEMS RELATING TO CITATION-FORMS

There are a number of problems relating to citation-forms that are encountered by lexicographers in the course of compiling dictionaries of natural languages. These problems normally manifest themselves when lexicographers apply lexicographical theories and methods in entering different aspects of lexicographical information.

Lexicographical theories and methods are applied to natural languages, which by their nature have different phonological, morphological and syntactical structures. Due to this fact, there are some occasions where lexicographers experience less lexicographical problems or no problems at all. This happens when the demands put forward by such theories and methods match considerably well with the structures of the languages being compiled. But if there is a great deal of mismatch between the demands put forward by the lexicographical theories and methods and the structures of the languages being compiled, some lexicographical problems are encountered.

Lexicographers normally look for solutions to the mismatch by adopting alternative conventions. But in my view, this shouldn't have created problems, as at any event the solution to the mismatch has to be worked out. But why is it that it creates problems. The practice of looking for alternative conventions creates problems when lexicographers start to manipulate the structures of languages just for the sake of meeting the demands of the lexicographical theories and methods. It shouldn't have created problems if the practice would have meant modifying the theories and methods for the sake of meeting the demands dictated by the structures of the languages being compiled.

What we should understand at this point is that, languages are natural phenomena, but theories and methods are just human artefacts. To be more objective and scientific, we should give prime concern to the forms and structures of languages whose data cannot be disputed rather than leaning so much on theories and methods which could easily be proved wrong. It is important to make sure that the process of dictionary making does not distort or lose information of our natural languages. The objective should be to record the natural languages completely and correctly. If certain theories and methods lead us to distorting or losing the data, we had better give them a second thought.

### 3.1 PROBLEMS OF ALPHABETIZING CITATION-FORMS (HEADWORDS)

Ever since the introduction of the practice of compiling dictionaries of Bantu languages, the tradition of alphabetizing headwords in the dictionary has been one of the problematic methods in Bantu lexicography. As mentioned in Section 1.0, headwords are supposed to be entered as citation-forms of the lexemes. To be a citation-form of a lexeme, the word has to have specific features as enumerated in Section 2.0. One of the features says citation-forms should be ‘forms that are regarded by native speakers as ‘natural’ citation-forms, i.e. the ones that come quickly in their mind when they want to look them up in the dictionary’. In most cases, due to complex morphological structures of Bantu languages, certain features of citation-forms like the one we have just cited, hinder effective alphabetical listing of headwords.

#### 3.1.1 Alphabetizing Verbs

In Bantu languages, the basic natural forms of the verbs are the infinitive forms. These forms are formed by prefixing an infinitive marker to the verb stems. This makes their morphological structure to be as follows: **inf. marker** + **verb-stem** + **verb ending**.

Examples (1) through (3) below illustrate verbs from four Bantu languages:

(1) Infinitive verbs of Swahili language spoken in Tanzania.

**ku** = infinitive marker

<b>kuamka</b>	‘to wake up’
<b>kubeba</b>	‘to carry’
<b>kucheza</b>	‘to play/dance’
<b>kudaka</b>	‘to catch’
<b>kuendesha</b>	‘to drive’
<b>kufagia</b>	‘to sweep’
<b>kuguna</b>	‘to groan’

(2) Infinitive verbs of Ngombe (Lingombe) language spoken in Lisala, Zaire.

**bo** = infinitive marker

<b>bop’a</b>	‘to give’
<b>bos’omba</b>	‘to buy’
<b>boh’eleja</b>	‘to help’
<b>bos’ala</b>	‘to work’
<b>boteke</b>	‘to sell’

(ILCAA 1992: 272)

(3) Infinitive verbs of Luba (Ciluba) spoken in Kasai, Zaire.

**ku** = infinitive marker

<b>kupa</b>	‘to give’
<b>k’uvuul’a</b>	‘to undress’
<b>k’ubuk’il’a</b>	‘to call’
<b>k’utaand’ish’a</b>	‘to threaten’
<b>k’uy’a</b>	‘to go’
<b>k’ush’ip’a</b>	‘to kill’
<b>k’uny’e’em’a</b>	‘to run’
<b>k’ul’o’ong’esh’a</b>	‘to teach’

(ILCAA 1992: 306)

Examples (1) through (3) show that to list Bantu verbs in alphabetical order by their basic citation-forms, implies entering all verbs under the same letter of the alphabet. For example, Swahili verbs will be entered under ‘k’, Yasamba verbs under ‘g’, Ngombe (Lingombe) verbs under ‘b’ and Luba (Chiluba) verbs under ‘k’. This would create an imbalance in the dictionary as such letters will be overcrowded. Although this problem is encountered when compiling dictionaries of most Bantu languages, there are some exceptions. For example, the Kaka (Koko) language spoken in the southeastern part of Cameroon, has lost its infinitive marker (c.f. ILCAA 1992: 47). In this situation, the verb-stems function as infinitive verbs as illustrated in example (4) below:

(4) Infinitive verbs of Kaka (Koko) language spoken in Cameroon.

<u>to</u> lo	‘to read’
nye	‘to give’
to	‘to pinch’
sa	‘to look for’
<u>so</u> mo	‘to beat’
yok <u>we</u>	‘to return’
gbund <u>o</u>	‘to bark’

(ILCAA 1992: 49)

Example (4) above shows that, by having no infinitive markers prefixed to the verb-stems, the verbs begin in different letters of the alphabet. This enables the verbs to be distributed in different letters of the alphabet, and hence causing no problem in alphabetizing them. This observation will also be true in other languages, which have also lost their infinitive markers, like the Kaka language.

However, in order to solve the problem of imbalance in listing basic verb forms that have infinitive markers, an alternative convention had to be adopted.

The lexicographers adopted the practice of listing verbs by stems as illustrated in examples (5) through (7) below:

(5) Swahili (Kiswahili) verb-stems

amka	‘to wake up’
beba	‘to carry’
cheza	‘to play/dance’
daka	‘to catch’
endesha	‘to drive’
fagia	‘to sweep’
guna	‘to groan’

(6) Ngombe (Lingombe) verb-stems

h’eleja	‘to help’
p’a	‘to give’
s’ala	‘to work’
s’omba	‘to buy’
te <u>ke</u>	‘to sell’

(7) Luba (Chiluba) verb-stems

buk’il’a	‘to call’
l’o’ong’esh’a	‘to teach’
ny’e’em’a	‘to run’
pa	‘to give’
sh’ip’a	‘to kill’
taand’isha	‘to threaten’
vuul’a	‘to undress’
y’a	‘to go’

This method of using verb-stems as citation-forms rather than using infinitive forms, solve the problem of imbalance mentioned above. By using this method, the verbs of Bantu languages could be distributed into different letters of the alphabet. However, this method is not entirely free from problems. If taken wholesale, unnatural forms could also be entered. This danger is further explained in Section 3.2 below.

### 3.1.2 Alphabetizing Nouns

Nouns in Bantu languages are also the victims of the tradition of alphabetical listing of headwords due to their complex morphological structures. In Bantu

languages nouns are grouped into noun classes based on their noun prefixes they carry, and their concordial agreement they manifest. Each noun class has two specific noun prefixes; one indicating a singular form of the noun and a different one indicating a plural form of the noun. Examples (8) through (11) illustrate a few nouns from four Bantu languages.

(8) Swahili Nouns

<u>Singular</u>	<u>Plural</u>
mtoto	watoto
mti	miti
kiti	viti
chumba	vyumba
jumba	majumba

(TUKI 1981)

(9) Lingala Nouns

<u>Singular</u>	<u>Plural</u>
<b>moto</b>	<b>bato</b>
<b>moto</b>	<b>mito</b>
<b>libaya</b>	<b>mabaya</b>
<b>elamba</b>	<b>bilamba</b>
<b>lolaka</b>	<b>ndaka</b>

(Kaji 1992: 5–6)

(10) Ngombe (Lingombe) Nouns

<u>Singular</u>	<u>Plural</u>
<b>mwana</b>	<b>bana</b>
<b>molangiti</b>	<b>milangiti</b>
<b>lipa</b>	<b>mapa</b>
<b>esenja</b>	<b>bisenja</b>
<b>lesa</b>	<b>besa</b>
<b>bele</b>	<b>male</b>

(ILCAA 1992: 271)

(11) Luba (Chiluba) Nouns

<u>Singular</u>	<u>Plural</u>
<b>muuntu</b>	<b>baantu</b>
<b>muketa</b>	<b>miketa</b>
<b>dilaandi</b>	<b>malaandi</b>
<b>cilaamba</b>	<b>bilaamba</b>

**lukasu**  
**buta**

**nkasu**  
**mata**  
(ILCAA 1992: 305)

Bantu nouns are different from their verbs. They carry different noun prefixes, as exemplified in examples (8) through (11). At this level, they seem to be unproblematic in terms of alphabetizing them, as they can be easily distributed into different letters of the alphabet. However, there exists a problem and this problem lies within their singular and plural forms. Examples (8) through (11) show that singular and plural forms of the same noun take different prefixes. This means, although they are semantically related, they will have to be entered under separate letters.

Notice that singular and plural forms such as *mtu/watu* are word-forms of the same lexeme. According to the citation principle stated in Section 2.0, which says 'citation-forms should be the stems from which other inflected forms could be produced', word-forms cannot be entered as a headword. The stem from which they are produced is the one which should be entered as headword. The word-forms are entered in the entry of their stem. But, this decision which is based on lexicographical principles, yield unnatural forms, i.e. forms that would not be easily identified or recognized by the users, as illustrated in (12) below:

#### (12) Swahili Nouns

<u>Singular</u>	<u>Plural</u>	<u>Stem-form</u>
<b>mtu</b>	<b>watu</b>	-tu
<b>mti</b>	<b>miti</b>	-ti
<b>kiti</b>	<b>viti</b>	-ti
<b>chura</b>	<b>vyura</b>	-ura
<b>jiicho</b>	<b>macho</b>	-cho

Example (12) shows that certain nominal stems are unnatural forms. Therefore, they cannot be taken as citation-forms. To enter them as citation-forms, could cause difficulties on the part of the user in retrieving the nouns. If a lexicographer decides to enter them as citation-forms, he has to add an index at the end of the dictionary that would guide the users to look up nouns in the main body, as illustrated in (13) below.

#### (13) Swahili nouns index in alphabetical order

chura	<i>see</i>	-ura
kiti	<i>see</i>	-ti1
miti	<i>see</i>	-ti2
mti	<i>see</i>	-ti2
mtu	<i>see</i>	-tu

viti	<i>see</i>	-ti1
vyura	<i>see</i>	-ura
watu	<i>see</i>	-tu

This method of entering nouns into the dictionary is not favoured and not recommended. It makes the dictionary cumbersome to use, as the user will have to check the word in the index before looking it up in the main body of the dictionary. Therefore, in order to solve such problems, Bantu lexicographers adopted an alternative convention. In most dictionaries of Bantu languages, singular forms are entered as citation-forms. Immediately after the citation form, the appropriate plural prefix is shown. If the plural form is irregular, that is, it does not follow the regular inflected rule, the entire plural form is entered. Nouns like jimbi/**majimbi**; jiko/**majiko**; jini/**majini**; jiwe/**mawe**; jino/**meno**, will be entered as in (14) below:

(14) Swahili nouns

**jimbi** n (ma-)  
**jiko** n (ma-)  
**jini** n (ma-)  
**jiwe** n (mawe)  
**jino** n (meno)  
[n = noun]

To the native speakers, and advanced language learners, this method is very convenient and not cumbersome. By using their competence, they can easily identify or construct singular and plural forms. What needs to be done is for the lexicographer to explain the system in the front matter.

This method is not convenient for beginners. Most learners at this level do not have sufficient knowledge of grammar to enable them to identify easily singular and plural forms. Notice that, plural forms carry different noun prefixes and are hidden in the entries of the singular forms. Therefore, the insufficient knowledge of the language learners will make them have difficulties in knowing where to retrieve plural forms.

In solving this problem, which could be faced by language learners, some pioneer lexicographers of Bantu languages, opted for a different convention. They adopted the method of entering both singular and plural forms as separate independent entries. Dictionaries that used this method are like: *The Standard Dictionary of Nyanja* and *Kikongo-English Dictionary* (c.f. Benson 1964).

Although this method is very convenient for language learners, it violates the principle governing lexemes (c.f. 2.1). Singular and plural forms are not two different lexemes, therefore they cannot be entered as two separate entries. But considering the complex morphological structures of Bantu languages, this argument should not be taken as valid all the time. Normally, the above-

mentioned principles governing lexemes could be violated, especially in handling irregular forms. But since most singular and plural forms are regular, to enter all forms as independent entries violates severely the principle of economy. Due to this argument, most dictionaries of Bantu languages, did not follow this method.

### 3.1.3 Alphabetizing Adjectives

The morphological structure of Bantu adjectives could also cause problems in alphabetizing them. The structure of the adjectives is that of [**prefix + stem**]. Since adjectives qualify nouns, the prefixes have to agree with the nouns they qualify. In this adjectival structure, the function of the prefixes is just to bring concordial agreement with the nouns they qualify. The actual senses of the adjectives are expressed by the stems. Example (15) illustrates examples from Swahili and Bondei spoken in Tanzania.

#### (15) Swahili adjectives

Singular		Plural	
<u>Noun</u>	<u>Adjective</u>	<u>Noun</u>	<u>Adjective</u>
Mtoto	<u>mzuri</u>	watoto	<u>wazuri</u>
Mwalimu	<u>mzee</u>	walimu	<u>wazee</u>
Mti	<u>mkubwa</u>	miti	<u>mikubwa</u>
Mnara	<u>mrefu</u>	minara	<u>mirefu</u>
Kiti	<u>kibaya</u>	viti	<u>vibaya</u>
Kisu	<u>kichafu</u>	visu	<u>vichafu</u>
Jino	<u>bovu</u>	meno	<u>mabovu</u>
Jiko	<u>safi</u>	majiko	<u>masafi</u>

Notice that adjectives have to agree with the nouns they qualify. For this reason, they carry two different prefixes; one prefix concords with the singular form of the noun and another one concords with the plural form of the noun. Although the two forms are semantically related, by having two different prefixes, they could be entered into two different letters of the alphabet.

In order to solve this problem, lexicographers have decided to enter adjectives by stem. By using this method, adjectives can be listed into different letters of the alphabet. In addition, regular inflected forms could be entered under their relevant stem-forms. Example (16) illustrates adjectival stems-from Swahili and Bondei languages respectively.

#### (17) Swahili adjectival stems

zuri	‘good, pleasant, nice, beautiful’
zee	‘old’

refu	‘tall, long’
baya	‘bad, unpleasant, ugly’
chafu	‘dirty, filthy’
bovu	‘in bad condition’
safi	‘clean, neat’

There are a number of arguments that can be put forward in support of this method. Firstly, the function of prefixes is basically limited to bringing concordial agreement; the meanings of the adjectives are expressed by the stems. Therefore, the method of entering adjectives by stem does not lose the core meanings of the adjectives. Secondly, to native speakers, adjectival stems sound natural. When native speakers name adjectives without attaching them to any particular nouns, the stem-forms automatically come into their mind. Thirdly, the two forms of an adjective (singular and plural) belong to the same lexeme. Therefore, both forms cannot be entered as two independent entries.

### 3.1.4 Alphabetizing Words with Pre-prefixes

There are a few Bantu languages that have the pre-prefix phenomenon (also known as *augment*). These are, for example, the Haya spoken in Tanzania, the Luganda spoken in Uganda and the Kinande spoken in the Democratic Republic of Congo (DRC). In such languages, the pre-prefixes manifest themselves in nominal and verbal structures. The pre-prefix element is always a vowel, and varies due to vowel harmony. The second element, which follows the initial vowel in the structure, is a prefix. In nouns, the said element is a class prefix and in verbs it is an infinitive marker. Examples 18 (a) and (b), illustrate pre-prefixes found in nouns of the Haya and Kinande languages respectively:

#### (18) The pre-prefix phenomenon

##### (a) The Haya nouns

<b>a</b> ma l'ila	‘tears’
<b>a</b> ka yaga	‘air’
<b>a</b> ma yaga	‘festival’
<b>e</b> ki babi	‘leaf’
<b>e</b> m pambo	‘seed’
<b>e</b> n fuma	‘sweet potato’
<b>o</b> lu babi	‘banana leaf’
<b>o</b> mu twe	‘head’
<b>o</b> mu gongo	‘back’

## (b) The Kinande nouns

a ma ghetsi	‘water’
e mi liki	‘rope’
e ki tiri	‘basket’
o mu kali	‘woman’
o mo ti	‘tree’

Example (18) shows that, if nouns were to be entered by initial vowel prefixes (pre-prefixes), all nouns would be listed under the three vowel letters (a, e, o). This method has been used in Turvey (1977) *Kwanyama-English Dictionary* in entering nouns. All nouns are found under letters ‘O’ and ‘E’.

In order to solve this imbalance, other Bantu lexicographers decided to drop the initial vowel prefixes (pre-prefixes) and list the nouns by the first letter of the singular class prefix (c.f. Benson 1964). Therefore, the nouns in (18) could be entered as: (Haya) *kayaga* ‘air’, *lubabi* ‘banana leaf’, *mayaga* ‘festival’, *mpambo* ‘seed’, *mugongo* ‘back’, and *nfuma* ‘sweet potato’; (Kinande) *kitiri* ‘basket’, *maghetsi* ‘water’, *miliki* ‘rope’, *moti* ‘tree’, and *mukali* ‘woman’. Although this method complies well with the alphabetical method of listing dictionary entries, it loses the pre-prefix data. In order to recover the pre-prefix data, *The Luganda-English Dictionary* (Murphy 1972), decided to enter the initial vowel (pre-prefix) immediately after the citation-form (headword).

This is one way of handling nouns with pre-prefixes, and a convenient method as far as it goes. It enables nouns to be distributed into different letters of the alphabet, and it does not create unnatural forms. After dropping the initial vowels, the remaining parts could still be recognized as natural forms. However, in order to recover the pre-prefix data, which is lost by using this method, it is advised to adopt the method used by *The Luganda-English Dictionary* of indicating the pre-prefix immediately after the headword. The indication of a pre-prefix is important (especially to learners) as they are not randomly assigned but ‘systematically determined by the vowel of the following noun prefix’ (Valinde 1990: 430).

The method of dropping the initial vowel pre-prefix, and entering a word by the first letter of the following prefix does not yield favourable results in handling verbs. We mentioned earlier that in verbs, the initial vowel is a pre-prefix and what follows the prefix is an infinitive marker as illustrated in (19) below:

## (19) The Haya verbs (with pre-prefixes)

o ku <b>binga</b>	‘to chase’
o ku <b>chwa</b>	‘to spit’
o kw <b>esaimula</b>	‘to sneeze’
o ku <b>fumo:la</b>	‘to talk’
o ku <b>goba</b>	‘to arrive’
o kw <b>i:ja</b>	‘to come’

o kw <b>i:luka</b>	‘to run away’
o ku <b>luga</b>	‘to leave’
o ku <b>shu:ka</b>	‘to descend’
o ku <b>zo:la</b>	‘to show’

Example (19) shows that if the initial vowel pre-prefix was dropped, one would end up with an infinitive marker. Recall that in Bantu languages infinitive markers do not vary (c.f. 3.1.1). Example (19) shows that in Haya, the infinitive marker is ‘ku’. Hence, if verbs were to be entered by the infinitive marker, the Haya verbs will be entered under ‘k’. Therefore, although the method of dropping the pre-prefix works very well with nouns, it doesn’t work very well with verbs. The only solution to this problem is to enter the stems as citation forms as suggested in (3.1.1). The data in example (19) attests the fact that to enter verbs by stems would fit very well with the convenient alphabetical method. With regard to the loss of the pre-prefix data, it is suggested to adopt the method used by the *Luganda-English Dictionary*, as explained above.

### 3.2 THE PROBLEM OF ENTERING UNNATURAL FORMS

In dictionary making, lexical units that are morphologically and semantically related form a family of a single lexeme. These lexical units are entered in the dictionary under their stem (c.f. 3.1). The process of identifying basic stem-forms (canonical forms) could yield unnatural forms. By unnatural forms we mean forms that have no meanings in the language concerned (unactivated forms) or cannot easily be recognized by the native speakers when in isolation.

The other lexicographical process that could also yield unnatural forms is the process of alphabetization. As discussed above, due to the nature of derivational and inflected structures of Bantu languages, words could be distributed into very few letters of the alphabet. For example, in most Bantu languages verb-stems of a particular language are prefixed by the same infinitive markers (c.f. 3.2.1). This means that if verbs were entered by the first letter of the infinitive marker, all verbs will be placed under the same letter of the alphabet. This causes an imbalance in the dictionary. In solving this problem, words are entered by the first letter of the stems. This convenient method for alphabetizing headwords, if applied wholesale, could yield unnatural forms.

So far we have identified two lexicographical processes that could yield unnatural forms: the process of identifying a stem for a family of lexical units, and the process of alphabetizing words in the dictionary. These processes could yield two types of unnatural forms: stems that are obsolete or unactivated and stems that cannot easily be identified by the users.

## 3.2.1 Obsolete/Unactivated Stem-forms

In natural languages it is a common phenomenon to find words or a family of lexical units whose stems are obsolete, unactivated or difficult to trace their semantic origin. By unactivated we mean forms which comply with the phonological structure of the language but have not yet been assigned meanings. So, when words or lexical units are entered under obsolete or unactivated stem-forms, a problem of retrieval occurs. The problem is that such stems hardly come naturally into the minds of the users, when they look up the words in a dictionary. Example (20) illustrates stems from Swahili.

(20) Swahili obsolete/unactivated stem-forms

- |  |   |
|--|---|
| <p>(a) [<b>bony.a</b>]<br/>           ?bony.ana<br/> <u>bony.ea</u> be dented<br/> <u>bony.eka</u> able to ~<br/> <u>bony.esha</u> cause to ~<br/>           ?bony.wa</p>                          | <p>(b) [<b>bubuj.a</b>]<br/>           ?bubuj.ana<br/> <u>bubuj.ia</u> bubble out into sth<br/> <u>bubuj.ika</u> bubble out<br/> <u>bubuj.isha</u> cause to bubble out<br/>           ?bubuj.wa</p> |
| <p>(c) [<b>chachat.a</b>]<br/>           ?chachat.ana<br/>           ?chachat.ia<br/> <u>chachat.ika</u> (of blood) tingle<br/> <u>chachat.isha</u> cause to tingle<br/>           ?chachat.wa</p> | <p>(d) [<b>ing.a</b>]<br/>           ?ing.ana<br/> <u>ing.ia</u> go in, enter<br/> <u>ing.ika</u> able to be entered<br/> <u>ing.iza</u> cause to go in/enter<br/>           ?ing.wa</p>            |

Example (20) shows that Swahili verb-stems can derive derivational forms by way of suffixation. The verb root and the derived forms constitute a family of morphologically and semantically related lexical units. However, there are cases where in the morphological paradigms, derived forms have been activated but the basic stems have not. For instance, all the stems in example (20) have not yet been activated and there are no records, which show that they were being used in the past. Therefore, if such hypothesized stems (to borrow Newell's term) are entered as citation forms, the users will have difficulties in retrieving words in the dictionary.

In order not to enter such unnatural forms in the dictionary, TUKI's (1981) *Kamusi ya Kiswahili Sanifu*, decided to pick one derived form from the paradigm as a basic form. The other derived forms in the paradigm are then entered under it. In example (21), derived forms that have been chosen as basic forms are *bonyea* for paradigm (a); *bubujika* for paradigm (b); *chachatika* for paradigm (c); and *ingia* for paradigm (d). The choice of derived forms to be entered as citation forms for dictionary entries could be motivated by a number of criteria. Yet, the most reasonable one would be to choose derived forms that bear the most basic

meanings in their respective paradigms. The choice made by the authors of *Kamusi ya Kiswahili sanifu* reflects this criterion.

### 3.2.2 Unidentifiable Stems

Unidentifiable stems are another type of unnatural forms, which a lexicographer could end up with when identifying citation forms which are also canonical forms (stems). Example (21) illustrates such forms taken from Swahili.

(21) Swahili unidentifiable stems

<b>Stem-form</b>	<b>Full form</b>	<b>Gloss</b>
-nara	mnara	tower
-nara	kinara	chairperson
-dani	kidani	necklace
-atu	kiatu	shoe
-chi	mchi	pestle
-tu	mtu	person
-tu	kitu	a thing
-ti	mti	tree
-ti	kiti	chair
-to	mto	river
-to	kito	precious stone
-we	jiwe	stone
-nu	kinu	wooden mortar

The stems shown in (21) when in isolation (even by using the native speaker's competence), cannot tell what they refer to. In other words, they don't mean anything. And to make matters worse, these forms could be used as stems for a number of words. That makes attaching meanings on them more difficult. In order not to enter unnatural forms into the dictionary, three methods have been used so far.

(a) Nouns are entered by the first letter of both singular and plural class prefixes, and verbs are entered by the first letter of the infinitive prefix. This method was used in the *Standard Dictionary of Nyanja* (Benson 1964: 74). Example (22) illustrates the point by using Swahili words.

(22) Examples from Swahili

Askofu n bishop  
Bata n duck  
**Chandarua** n mosquito net

Dimbwi n pond  
**Kiti** n chair  
**Kuapa** v take an oath, swear a formal oath  
**Kubomoa** v break down, break through  
**Kuchimba** v (1) dig (2) harm  
**Kudunga** v pierce, bore  
**Kufuta** v (1) wipe (2) abolish  
**Maaskofu** n bishops  
**Mabata** n ducks  
**Madimbwi** n ponds  
**Mtu** n person  
**Viti** n chairs  
**Vyandarua** n mosquito nets  
**Watu** n persons

(b) Words with their prefixes on, are entered by the first letter of the stems. But, prefixes *in small letters* and stem-forms *in capital letters*. This method was used in Turner's *Tumbuka-Tonga-English Dictionary* (1952) as exemplified in (23) below:

(23) Entries from Tumbuka-Tonga-English Dictionary

Nouns: mwAKA,mi- year  
mwAMBI,mi- the scriptures  
chiBAMA,vi- bread  
mCHENGA,mi- sand  
Verbs: kuBABA to give birth to a child  
kuCHA to be ripe  
kuDAFYANA to argue  
(Turner 1952)

(c) Nouns are entered by the first letter of the singular class prefixes and verbs by the first letter of the stem-forms. This method has been used in dictionaries such as *Kitabu cha Maneno ya Kiswahili* (Johnson 1935), *A Swahili-English-Dictionary* (Johnson 1939), *Swahili-English-Dictionary* (Rechenbach 1967) and *Kamusi ya Kiswahili Sanifu* (TUKI 1981). Example (24) illustrates entries from Rechenbach (1967).

(24) Entries from Swahili-English-Dictionary

**Chimbi** (ma-) *dial.* cook  
**Chipukizi** (ma-) sprout, shoot  
**Darasa** (ma-) lesson, class, classroom  
**Kinu** (vi-) 1. mortar, press, mill 2. hub (of a wheel, bicycle, etc.)

**Mchicha** (mi-) edible plant, spinach

**Mchikichi** (mi-) oil palm

(Rechenbach 1967)

Of the three methods, the first method is the least favoured method. Although it enters natural forms in the dictionary, it consumes a lot of space. In addition, since it enters both singular and plural forms as independent entries, it violates the principle of lexemes. Recall that lexical units of the same lexeme should be entered under the same stem (c.f. 3.1).

The second method, though complying with the demand for natural forms, it does not make the process of retrieval lesser problematic. Headwords are still entered by the first letter of the stems. Therefore, the users will still have to figure out the stems in order to look up the words. To make matters worse, this method registers cumbersome visual representation, which could cause difficulties in identifying the words easily and quickly.

The third method is the most favoured method, as it is flexible enough to accommodate problematic features. Firstly, it complies with the principle of lexemes, and the convenient method of alphabetization. Secondly, it can handle the problem of unnatural forms in a more acceptable way. Notice that many unnatural forms could be found in nouns and very few in verbs. Therefore, in order to handle appropriately the problem of unnatural forms, nouns are entered by the first letter of their class prefixes, and the plural prefixes are shown immediately after the headwords. Verbs are entered by the first letter of their stems. As we observed in Section 3.2.1 above, most stems appear and sound natural to native speakers. However, it is suggested that the few unnatural verb-stems (obsolete, archaic, unidentifiable) that could also be encountered, should be entered as independent entries.

## REFERENCES

Bauer, L. 1983.

*English Word-formation*. Cambridge: Cambridge University Press.

Benson, T.G. 1964.

*A Century of Bantu Lexicography*. **African Language Studies** 5: 64–91.

Cruse, D.A. 1986.

*Lexical Semantics*. Cambridge: Cambridge University Press.

Hartmann, R.R.K. (ed.) 1983.

*Lexicography: Principles and Practice*. London, New York: Academic Press.

ILCAA 1992.

*Studies in Cameroonian and Zairean Languages*. **Bantu Linguistics** 3.

- Johnson, F. 1935.  
*Kamusi ya Kiswahili, yaani Kitabu cha Maneno ya Kiswahili (Swahili Dictionary, i.e. a Swahili Wordbook)*. London: Sheldon Press.
- 1939 *A Standard Swahili-English Dictionary*. Nairobi/Dar es Salaam: Oxford University Press.
- Kaji, S. 1992.  
*Vocabulaire Lingala Classifie*. Tokyo: ILCAA.
- Landau, S.I. 1989.  
*The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.
- Lyons, J. 1977.  
*Semantics 2*. Cambridge: Cambridge University Press.
- Murphy, J.D. 1972.  
*Luganda-English Dictionary*. Washington: The Catholic University of American Press.
- Newell, L.E. 1995.  
*Handbook on Lexicography for Philippine and Other Languages*. Manila: Linguistic Society of the Philippines. Monograph Issue No. 36.
- Rechenbach, C. 1967.  
*Swahili-English-Dictionary*. Washington DC: Catholic University of America Press.
- TUKI 1981.  
*Kamusi ya Kiswahili Sanifu*. Dar es Salaam: Oxford University Press.
- Turner, W.Y. 1952.  
*Tumbuka-Tonga-English Dictionary*. Blantyre: Hetherwick Press.
- Turvey, B.H.C. 1977.  
*Kwanyama-English Dictionary*. Johannesburg: Witswatersrand University Press.
- Ulmann, S. 1957.  
*The Principles of Semantics*. Oxford: Basil Blackwell.
- Valinde, Nzama Kambale 1990.  
*The Structure of Kinande*. PhD Dissertation. Ann Arbor: UMI.
- Webster New Collegiate 1979.  
*Webster's New Collegiate Dictionary*. Springfield, Mass.: G. & C. Merriam Co.

**About the author:** Dr. John G. Kiango is a Senior Research Fellow in the Institute of Kiswahili Research of the University of Dar es Salaam, Tanzania. He holds a Diploma in Education of Morogoro Teacher Training College, a BA (Ed) (Hons) degree of the University of Dar es Salaam, an MA (Linguistics) degree of the University of Kansas, USA and an EdD (Linguistics) degree of Teachers College, Columbia University, USA. His areas of specialization are general linguistics and Bantu lexicography.