

Beata Wójtowicz
University of Warsaw

Survey of Swahili Dictionaries: the Macrostructure

Abstract: Swahili dictionaries have been on the market for over a century. New publications often take into account works that have been already published and base the list of headwords on older dictionaries. While this is widely accepted practice, nowadays also new, computer-based activities may no longer be ignored. Analysis presented in this paper aims at detecting the differences and similarities among Swahili bilingual published dictionaries on the macrostructural level, and at identifying possible trends in Swahili dictionary compilation. It also takes into account corpus-based methodology and points to its superiority over intuition-based methods in dictionary compilation.

Keywords: Swahili dictionaries, bilingual lexicography, macrostructure

1. Introduction

The present article is concerned with the research into Swahili lexicography by focusing on the macrostructural level of selected general bilingual dictionaries of Swahili.

While taking under consideration issues discussed e.g. in de Schryver and Prinsloo (2000) or de Schryver (2012) that concentrate on presenting the superiority of a corpus-based approach over traditional compilation methods, this analysis aims at detecting the differences and similarities among Swahili bilingual dictionaries on the macrostructural level, and at identifying possible trends in Swahili dictionary compilation. Both the oldest as well as the newest dictionaries shall be analyzed, as solutions adopted within them are not unitary over the span of a given period. The structures used in the oldest dictionaries have been reused in subsequent ones.

Contemporary printed dictionaries available to the author¹ have been included into the analysis, as well as reprints of works published in the beginning of the 20th century that significantly impacted the Swahili lexicography, and are still available and remain at least in limited use. The analysis was conducted with focus paid to the dictionaries' usefulness. This study was concluded with full awareness that comparing such elements as lemma-sign lists of dictionaries dating over a hundred years with newer ones will naturally place the prior at a disadvantage; however, due to the still widespread accessibility of the oldest works such an analysis seems to remain expedient.

2. General remarks on the macrostructure

In the current era of corpus lexicography the technological improvements may no longer be disregarded in the field of African language dictionary compilation. Lexicographers who base their research on corpora can easily list all the typical macrostructural inconsistencies of dictionaries compiled without the use of a corpus (e.g. de Schryver and Prinsloo 2000, 2001b). De Schryver and Prinsloo (2001b: 376) cite such typical macrostructural inconsistencies found in dictionaries that were not compiled with the use of corpora:

1. inconsistencies when it comes to the relative length of alphabetical stretches, by treating certain sections of the lemma-sign list more exhaustively than others;

2. inconsistencies regarding the creation of the lemma-sign list (mostly as a result of an enter-them-as-they-cross-my-way approach to dictionary compilation) such as:

- 2.1. the omission of *words most likely to be looked for*, while

¹The list of dictionaries which were taken into consideration is not comprehensive. It has been limited to the dictionaries available at the library of the Department of African Languages and Cultures at the University of Warsaw, as well as those at the library of Helsinki University. Additional works which were facilitated from the private collection of prof. Rajmund Ohly and that of the author have also been referenced within this paper.

words less likely to be looked for are included,

2.2. the partial treatment of lexical items belonging to a *closed set* (currencies, letters of the alphabet, digits, seasons, etc.),

2.3. the unequal treatment of *various prefixes* (i.e. mostly ‘inflection’ in Bantu),

2.4. the absence of a policy to deal with *productive* versus *non-productive suffixes* (i.e. mostly ‘derivation’ in Bantu),

2.5. the blind running of each stem through *all possible verbal and nominal derivations*, simply concatenating affixes, which results in serious doubts among mother tongue speakers whether many of these derivations do exist,

2.6. the *ad hoc* handling of *transparent* versus *non-transparent derivations*;

3. inconsistencies in terms of the choice of canonical forms.

Since Swahili is one of only several Bantu languages with an access to a corpus (Helsinki Corpus of Swahili – HCS 2014) the above list seems to be a good starting point for the evaluation of the macrostructure of different dictionaries. On top of that we will also investigate other issues, as the treatment of homonymous entries or the way the entries have been arranged. We will not investigate the 3rd issue as the problem of canonical form will be addressed in a further research on the microstructural level.

3. The number of articles per page

The first commonly raised issue, concerning the increasing lassitude of the lexicographer as he moves along to the later letters of the alphabet, can be verified without any sophisticated analysis, simply by comparing the general appearance of pages in several manually compiled Swahili dictionaries. In accordance with the hypothesis - the further the letter of the alphabet, the more superficial the given entry - the length of the entries decreases and their amount on a given page increases. By repeating the experiment performed by de Schryver and Prinsloo (2001b), we discover, much as they did, examples which confirm this hypothesis, as exemplified by the *Standard Swahili-English Dictionary* by Johnson (1985/1939), henceforth *Johnson*:

<u>Letter of the alphabet</u>	<u>Page</u>	<u>Number of articles</u>
A	2	11
N	335	26
U	508	56

Nonetheless, we can also find examples, which would prove an opposing hypothesis. The example below has also been taken from *Johnson*:

<u>Letter of the alphabet</u>	<u>Page</u>	<u>Number of articles</u>
A	21	22
N	336	9
U	499	3

Similar examples are commonplace in other dictionaries. This study does not negate the legitimacy of the hypothesis (based on a survey of Sepedi dictionary). Nonetheless, it was not possible to find evident examples in the analyzed Swahili dictionaries. The differences between the length of definitions for various headwords seem to result more from the morphological characteristics of the given words. In all of the dictionaries, it is common to encounter, for example, a concentration of headwords beginning with the letters *ki*, which in general are nouns, or with *u*, which amongst others precede abstract nouns. Noun entries – as a rule in a stem-based dictionary like *Johnson* – are shorter due to the lack of possible derivative forms (unlike the case of verbs), whilst abstract words, predominantly being derivatives, are defined under the given stem. Therefore, the actual entry consists of only a cross-reference.

The above-mentioned issue may also have been the result of the niche character of Swahili lexicography. Limited demand for such dictionaries in the world market has caused them to be published as a non-commercial enterprise. Publishing houses do not expect significant profits and print the works more for prestigious purposes by entering into cooperation with the author of an already compiled dictionary. The authors, mainly Africanists, initiate work on the

dictionary more in connection with private interests than as a result of a paid commission.

4. Dictionary lemma-sign lists

4.1 Headword selection

By headwords we understand the linguistic units being defined within a dictionary. The decision what to include and what to omit in a dictionary is in itself one of the hardest aspects of a dictionary compilation process. This has been mentioned *inter alia* by Tomaszczyk (1983: 51): “One of the basic problems of lexicography is to decide what to put in the dictionary and what to exclude”. The selection of headwords, which will then be defined in a dictionary, is conditioned by several factors. These are above all: the type of a dictionary, its size, and the envisaged user group (bf. Zgusta 1971, Bańko 2001, Żmigrodzki 2003).

The bilingual dictionaries which are within the scope of this work can be divided into two groups: general dictionaries that aim at registering the largest possible amount of lexical units, such as *Swahili-English Dictionary* by Madan (1992/1903, henceforth *Madan*), *Johnson*, and *Kamusi ya Kiswahili-Kiingereza* by TUKI (2001, henceforth *TUKI*); and learners' dictionaries consisting of a smaller number of entries, focused on the learners' needs and with a simplified structure of the articles, such as *Concise Swahili and English Dictionary* of the *Teach Yourself Books* by Perrott (1965, henceforth *Perrott*), *Kamusi ya kwanza Kiswahili-Kiingereza* by Cahill (1972, henceforth *Cahill*), *Learner's Swahili-English English-Swahili Dictionary* by Jahadhmy (1981, henceforth *Jahadhmy*) or the *Modern Swahili Modern English* by Baba Malaika (1994, henceforth *Baba Malaika*). The scope of the dictionary and the goals are usually presented in the introductory part. All the dictionaries under research but one, *Swahili-Suomi-Swahili Sanakirja* by Abdulla et al. (2002, henceforth *Abdulla*), were compiled without a use of a corpus.

Due to the limited commercial viability of this type of publication, the authors attempt to reach the widest group of end users. For instance, *TUKI* (2001: viii) was created with people

learning English or Swahili language in mind: “imekusudiwa kuwasaidia watu wanaojifunza Kiingereza au Kiswahili” (‘it is aimed at assisting people learning English or Swahili’). In order to meet these assumptions, the authors have included “everyday vocabulary”, needed in basic communication. Since the dictionary was compiled without corpus research the authors based their work on other already existing dictionaries, such as *TUKI* (1981), *Johnson*, and Feeley (1990), as well as on lists published by the BAKITA² Standardization Council.

The Swahili-English part of a *Perrott* dictionary (1965: Perface) “contains all the words the compiler heard during thirty years’ residence in East Africa, together with a selection of those taken for her own use from the dictionaries of Krapf, Sacleux, and Madan and the writings of Swahili authors, and a few present-day words not yet in any dictionary”. Therefore we find such lexis as for example *malaya* ‘prostitute’ or *raia* ‘citizen’, which were not included in older dictionaries cited by the author. The English-Swahili part contains vocabulary from other dictionaries from the *Teach Yourself* series additionally “adapted to the different circumstances of a tropical country”.

Zgusta (1971: 310) notes that whilst compiling a dictionary for languages which come from very distant cultures, it is necessary to take into account the lexis of the target language, when preparing the lemma-sign list for the source language. Some concepts, objects or, for example, plants or animals can turn out to be of little importance or even be non-existent in the target language, while they remain in common use in the source language. Assuming that the users of the target language may apply the dictionary to generate texts regarding the cultural environment of the other language, it is necessary to supply the appropriate units. In accordance with this rule, it would be justified to consider including the following entries in a Swahili-Polish dictionary:

² *Baraza la Kiswahili la Taifa* – National Swahili Council of Tanzania.

mkomunisti
soseji

komunista 'communist'
kielbasa 'sausage'

And in a Polish-Swahili part:

turban 'turban'	kilemba
muszelka kauri 'kauri shell'	kauri
palma kokosowa 'cocoa palm'	mnazi

This issue was pointed out by Ohly (1967) in his review of the Polish dictionary *Mały słownik suahilijsko-polski i polsko-suahilijski* by Stopa and Garlicki (1966, henceforth *Stopa-Garlicki*). He noted that the Swahili-Polish part of the dictionary lacks entries which appeared in the Polish-Swahili part, such as *naród* 'nation', *marksizm* 'marxism', *kapitalizm* 'capitalism', *kolonializm* 'colonialism', *komunista* 'communist'. Such omission of entries may be also found in other dictionaries. *Perrott* included the entry for *Uislamu* 'Islam' in the Swahili-English part, while the reciprocal entry does not appear in the other part of the dictionary.

The lemma-sign lists probably remain the most widely criticized part of every dictionary. In each dictionary, the reviewer can always find entries which in his opinion should not have been included, as well as a significant group which was not taken into account during compilation. It is especially stressed nowadays in reviews of dictionaries that did not take into account a frequency list.

Bilingual lemma-sign lists of Swahili dictionaries have been also widely criticized (e.g. Ohly 1967, Wamitila 1997). When using dictionaries as end users, we often come upon the lack of the most basic and obvious entries. For example, the English-Swahili part of the *Jahadhmy* lacks such headwords as *Monday* or *Sunday*, whilst *Baba Malaika* contains greetings for only several apparently randomly chosen persons. In the introduction to the *Jahadhmy* dictionary, it is written that one of the main merits of the dictionary is the inclusion of vocabulary regarding sex. In search of novelty, not to be found in other dictionaries, it is possible to come upon such headwords as *penis*, *vagina* or *homosexual*, but for example it is

impossible to find the term *sexual* or the seemingly paramount word - *sex*, at a minimum in regards to gender. The introduction also bears the information that this is a learner's dictionary, albeit such entries as *biology*, *desk*, *verb*, *noun* or *homework* had not been included. The dictionary also does not give the names of months. Such examples can be reproduced on the basis of other dictionaries.

The above-mentioned lack of key headwords in both parts of the dictionary, is not only a characteristic of the *Stopa-Garlicki*. Other than omitting headwords characteristic for a given culture, the authors often forget about basic units. In *Jahadhmy*, the Swahili-English part lacks, for example, such a vital entry as *zuri* 'good, pretty', while the English-Swahili side contains the word under both *good* and *pretty*.

Irrespective of the size, each general dictionary, especially a school dictionary, should include basic language lexis. De Schryver and Prinsloo (2001b: 375) point that: "Regardless of size, any general dictionary and certainly any learners' dictionary should at least cover the basic or core vocabulary". For the English language the American, L. Thorndike, already in 1921 published the Teacher's Word Book, which was compiled for vocabulary selected on the basis of an analysis of a 4,5 million corpus, that "consists of several lists of words showing their relative frequency [...] designed to help educators and teachers determine which words are common enough to be used" (Landau 1984). On the basis of such frequency lists calculating the commonness of words in extensive, representative corpora, adequate lemma-sign lists are prepared for a given type of dictionary. The superiority of such an approach over a more traditional one has been repeatedly proven (e.g. de Schryver and Prinsloo 2001b, Verlinde and Selva 2001). Nonetheless, the implementation of the most evolved instruments in the preparation of a lemma-sign list should be accompanied by the use of basic common sense (cf. Fillmore 1992, Summers 1996). Since it is possible that the corpus may not include certain very important words. In cases when there is no corpus for a given language de Schryver and Prinsloo (2001b: 388) found that: "it remains truly surprising that a variety of manually compiled lists, each of which

poorly represents the basic vocabulary, can show so much consistency when combined with one another” and therefore they “suggest that, in the absence of an electronic corpus – which is the case for all but a few of the Bantu languages – a well-planned combination of a variety of lemma-sign lists of existing dictionaries and unpublished manuscripts, is reasonably representative of a language’s basic (and peripheral) vocabulary”.

Since a Swahili corpus exists (HCS³) it is possible to compare dictionary entries with the frequency list derived from it. The analysis demonstrates that many words very highly ranked in the frequency list have been omitted, whilst other words more seldom present or not represented at all have been defined in various dictionaries.

Table 1 illustrates the presence of randomly chosen top-ranked vocabulary from the Swahili frequency list, i.e. positions 1-200, in selected publications. The analysis also includes derivatives which have the status of a headword in all dictionaries. The analyzed works are the following: *Johnson*, *Perrott*, *Stopa-Garlicki*, *TUKI*, as well as the corpus-based *Abdulla*. *Johnson*, *TUKI* and *Abdulla* represent big general dictionaries, while the other two, *Perrott* and *Stopa-Garlicki*, are small learners' dictionaries. Due to the fact that the analyzed vocabulary is derived from the list of the 200 most commonly used words in the Swahili language, the comprehensiveness of the selected dictionaries has not been taken into account, based on the assumption that even the smallest handbook dictionary should include vocabulary of such high frequency⁴. Most of the vocabulary from the below list appears during the first year of studies at the Swahili language course for beginners at the University of Warsaw. The letter ‘Y’ in the table confirms the occurrence of a given word in

³ HCS is the biggest annotated publicly available corpus of standard Swahili texts with ca. 12 million of running words, more information at <http://urn.fi/urn:nbn:fi:lb-2014032624> [accessed 19.08.2016]

⁴ Even though the appropriateness of the corpus for linguistic research may be disputable, since it represents rather opportunistic than representative or balanced type of a resource.

the dictionary. Out of 50 lexemes 15 are missing in the oldest general dictionary by *Johnson*, 14 in *Perrott*, and 29 in the smallest *Stopa-Garlicki*, 4 entries are missing in the most up-to-date but intuition-based *TUKI*. The only missing entry, *mbalimbali* 'various', in *Abdulla* is described not in a separate entry but under *mbali* as its reduplication. The analysis proves that using frequency counts derived from corpora ensure that the most frequently used words are not accidentally omitted from a dictionary (c.f. de Schryver and Prinsloo 2000).

Table 1.

	Johnson	Perrott	Stopa-Garlicki	TUKI	Abdulla
sema	Y	Y	Y	Y	Y
la	Y	Y		Y	Y
nchi	Y	Y	Y	Y	Y
mtu	Y	Y	Y	Y	Y
ingine	Y	Y		Y	Y
fanya	Y	Y	Y	Y	Y
toa	Y	Y	Y	Y	Y
kwenye	Y	Y			Y
rais				Y	Y
kiongozi	Y	Y	Y	Y	Y
waziri	Y	Y	Y	Y	Y
mkoa		Y		Y	Y
polisi	Y	Y		Y	Y
endelea	Y	Y		Y	Y
taifa	Y	Y		Y	Y
kutokana na					Y
anza	Y	Y	Y	Y	Y
mwananchi	Y	Y			Y
kutoka					Y
mahakama	Y			Y	Y
zaidi	Y	Y	Y	Y	Y
baadhi	Y	Y		Y	Y
uchaguzi	Y	Y		Y	Y
muda	Y	Y	Y	Y	Y
wilaya	Y	Y		Y	Y
kila	Y	Y		Y	Y
taarifa	Y	Y	Y	Y	Y
mwandishi	Y			Y	Y

jiji				Y	Y
pita	Y	Y	Y	Y	Y
mbalimbali		Y		Y	
dhidi ya				Y	Y
nyumba	Y	Y	Y	Y	Y
shirika	Y	Y	Y	Y	Y
tukio		Y		Y	Y
sheria	Y	Y	Y	Y	Y
kuhusu				Y	Y
mwenyekiti				Y	Y
jeshi	Y	Y	Y	Y	Y
maendeleo	Y	Y		Y	Y
amani	Y	Y	Y	Y	Y
mpango		Y	Y	Y	Y
mkazi	Y		Y	Y	Y
umoja	Y	Y	Y	Y	Y
suala				Y	Y
ongeza	Y	Y	Y	Y	Y
mbunge				Y	Y
gazeti	Y	Y		Y	Y
kamanda				Y	Y
mfanyakazi				Y	Y

4.2 Grammatical morphemes and pronouns

In Swahili dictionaries grammatical morphemes are sometimes present on the lemma-sign list. *Madan* treats nominal, pronominal and relative prefixes, tense markers and even sounds as headwords on his lemma-sign list:

Ye, (1) relative pfx. of 1, 2, and 3 sing. referring to persons and animals. Only used independently in such phrases as *ye yote*, any one whatever, whosoever; [...]

A similar rule, albeit less thoroughly, was used in e.g. *Baba Malaika*, where alphabetically various grammatical affixes are listed:

- po 1, verbal affix: is here, is present
- po 2, verb infix: when
nilipokuja, when I came
- po- 3, verb affix: where
mahali ni.lala.po, the place where I sleep

Perrott, on the other hand, only lists the possessive particle *-a* as a headword and a few of its word forms with appropriate agreement markers, however without any commentary:

-a, of
cha, of

Newer dictionaries, such as *TUKI* or *Abdulla*, do not list any grammatical morphemes. *Abdulla* includes possessive particle in its full word form with information on class.

More common practice is the inclusion of possessive and demonstrative pronouns in their full forms, together with the class prefixes, and possessive pronouns additionally as stems. Some dictionaries introduce this selectively. In *Baba Malaika* we come across some chosen pronouns:

zangu + -pl; my/mine
funguo ~, my keys
kazi ~, my jobs/tasks
zile + -pl; those
njia ~, those paths/roads

TUKI describes possessive pronouns but not demonstrative:

-angu *kv* my, mine; possessive adjectival root for the first person singular. *Kitabu ch~* my book;

changu *kv* see -angu, mine, my: *Kitabu ~* my book; *Chungu (hiki) ni ~* this pot is mine

Perrott limits the inclusion of pronouns and only gives the stems of the possessive pronouns:

-angu, my; mine

4.3 Derivatives

As a Bantu language, Swahili is characterized by agglutinative morphology; inflection is primarily prefixal, while derivation primarily suffixal, with a small degree of stem allomorphy. Due to their complex derivational and inflectional systems, Bantu languages pose problems not experienced by lexicographers working with European languages.

One of the most important questions is how to treat derivatives in dictionaries. Should they be included in the form of sub-entries under their base forms or as separate headwords?

Derivation is a process of word formation which derive a new lexemes from roots or bases of different words by morphological rules. Kiango (2000) discusses thoroughly different derivational rules productive in Swahili and their implications for lexicography. First, the derivative normally belongs to a different grammatical category than the base from which it is derived. Second, the affix modifies or entirely changes the meaning of the base. Additionally, most of the derivational rules are semi-productive, whilst derivatives can be formed regularly or irregularly. The derivational processes which take place in Swahili can be classified into four main categories: nominal, verbal, adjectival, and adverbial derivations (cf. Polomé 1967, Ohly et al. 1998). The derivational base can be both in the form of roots as well as stems, which in themselves already possess some type of a formant. According to Kiango (2000), the above features are the key criteria for a lexicographical decision regarding including derivatives into a dictionary. Based on these factors, a decision is reached as to whether the derivative form should gain the status of a separate headword or remain as part of the entry of its base.

Due to their characteristics, derivatives can be treated in many ways in dictionaries. Lyons (1977: 524) cites two perspectives. On the one hand, derivatives should not be taken into account in dictionaries at all, due to their typically transparent morphological rules, which should be described as part of language grammar instead of in a dictionary. On the other, derivatives constitute new lexemes, derivative processes are much less productive than

inflection rules and the end result tends to deliver unpredictable meanings of the new lexemes. This justifies why derivatives should also be included in dictionaries. The latter rule is adhered to by modern lexicography.

To determine the principle of the treatment of the derivatives, they are divided into regular and irregular ones (cf. Zgusta 1971, Kiango 2000). It is recommended that regular derivatives be placed as sub-entries, whilst irregular as main headwords. Regular derivatives should be understood as those with a regular form and meaning, which can be interpreted through the derivational formant and the base. Irregular derivatives are words which have an irregular morphological structure or meaning that cannot be interpreted using the meaning of the base, from which they were formed.

Despite the fact that not all derivative processes in Swahili are fully productive, their participation in vocabulary enrichment is significant, therefore it is hard to omit derivations in dictionaries. Kiango (2000: 119) points out the need for implementing various solutions, depending on the target group of dictionary users. Native speakers of the language will be able to properly identify a much larger amount of word-forming bases than those who are still learning it.

To visualize the problem, a closer look will be taken at the following two groups of words:

1. tuma, tumia, tumika, tumikia, tumikisha, tumikiana, tumisha, tumilia, tumana, mtume, mtumishi, mtumwa, mtumwaji, tume, utume, utumi, utumishi, utumwa, matumizi, matumishi

2. la, liwa, lika, lia, lana, lisha, lishisha, kilaji, malaji, malisha, mlisha, mlishi, mlisho, mla, mlaji, mlo, ulaji, ulio.

These words constitute a problem in that a decision must be made, as to whether all of the above words should be placed within the two headwords *tuma* 'send' and *la* 'eat', from which they are derived, or treated in accordance with some different rule. The additional complication should be kept in mind that one derived word can constitute the base for further derivations, e.g.:

tuma → tumia → matumizi
la → lisha → mlishi

In Swahili lexicographical history, no single methodology regarding the treatment of derivatives has been established. Depending on the dictionary we find them in various locations, with references or without them. All published dictionaries taken into consideration within this paper, with one exception, were written by foreigners for foreigners using and learning the Swahili language. It can therefore be assumed that the decisions to apply the specific solutions used within the dictionaries were made with these end users in mind.

Kiango (2000) proposes the following rules regarding the treatment of derivatives in dictionaries for non-indigenous users of the language: regular derivatives are to be included as sub-entries, whilst irregular derivatives should be treated as separate headwords. However, due to the specificity of some derivative processes, he proposes several exceptions to the rule: regular verb-based noun derivatives should be included twice, as a headword with a reference to the base form, where the word will then be defined, e.g.:

tum.a *vt.* assign/give work to sb, dispatch a person for an errand.
mtume *n. mi-* a messenger, an emissary, apostole.
mtumishi *n. wa-* a paid servant. **mtumwa** *n. wa-* a slave.
mtumwaji *n. wa-* a messenger. **tume** *n.* a commission.
utume *n.* evangelical work. **utumishi** *n.* civil service.
utumwa slavery.
mtume *n. mi-* see **tuma**.
mtumishi *n. wa-* see **tuma**
mtumwa *n. wa-* see **tuma**
mtumwaji *n. wa-* see **tuma**
tume *n.* see **tuma**
utume *n.* see **tuma**
utumishi *n.* see **tuma**
utumwa *n.* see **tuma**

(Kiango 2000: 121)

Irregular verb-based noun derivatives should also appear twice. However, the difference lies in the fact that the definition of semantically irregular derivatives are included in the location, where the given derivative is the headword, while those morphologically irregular in accordance with the above assignment. Below an example of an entry for a semantically irregular derivative:

tum.a *vt* give work to sb., dispatch a person to a place for an errand [...]
matumizi *n.* living expenses. see **tuma**.
utume *n.* evangelical work. see **tuma**.

(Kiango 2000: 124)

The complexity of the problem can be confirmed by the fact that the entry word *utume* is mentioned in both above examples. Most likely this occurred through the inattention of the author; nonetheless, it does provoke the need to reflect on the issue of classifying specific derivatives.

The author proposes that nominal derivatives formed out of adjectival or noun bases be treated correspondingly to the rules described above. The author also suggests that adjective-based verb derivatives, due to their regularity, should only be included in the form of sub-entries to their word bases. He does, however, state that the placement of verb-based derivatives be dependent on the regularity of their forms and meanings. He then opines that semantically irregular derivatives should be placed as separate headwords with a full definition. In order to not unnecessarily cluster together unrelated meanings, he proposes resigning from referencing to the headword which is the base for the given derivative, e.g.:

pak.a *vt.* apply, lay on, spread on, smear.
pak.ia *vt.* put on board a vessel, cart, train etc.
pak.ata *vt.* hold a child on the knee, lap or shoulder.

(Kiango 2000: 129)

Kiango proposes treating adjective and adverb derivatives - correspondingly to noun derivatives - individually, depending on their regularity. Irregular forms should be introduced twice, whereas the main headword should reference to the headword of the base. This is possibly a justified concept; however, it has yet to be put into practice in any of the dictionaries. Additionally, Swahili lexicography has not introduced any guidelines regarding the inclusion of derivatives. Practically applied solutions are above all dependent on the size and purpose of the dictionary.

In small dictionaries or those intended for students, the amount of derivatives is appropriately less pronounced and they are mostly included as headwords, but without any etymological information. *Perrott* introduces sub-entries mostly for the passive form of the verb, whilst the remaining derivatives have the status of headwords. Similarly, both *Cahill* and *Baba Malaika* treat all derivatives as equal to other words and appropriate them the status of separate entries with full articles. They do not supply any information regarding the type of derivation nor give references to the base form.

Another extreme rule was used in the stem-based *Johnson* dictionary. One of the main purposes of the dictionary was to supply the etymology of each headword. As a result, the author introduced the rule to include all derivatives as sub-entries within the entry for the base, from which they were derived. Whereby he does not differentiate the division between regular and irregular derivatives, hence treats all of them with the same rule. Almost all derivatives are to be found twice; firstly, as the main headword with a reference to the headword, on which it is based, and secondly, as sub-entries, for example:

La, v. (1) eat, consume of food generally. *Watu walikula*, the people ate. (2) use, use up, require for use of efficiency (as material, time, &c.) (Cf. *tumia, chukua*) [...] Ps. **Liwa**, be eaten. &c. St. and Pot. **Lika**, be eatable, be fit for food, be eaten, be worn through. [...] Cs. **Lisha**, (1) cause to feed, feed, keep (animals), graze, pasture. [...] **Mlisha**, n. *wa-*, **Mlishi**, n. *wa-* (1) one who feeds or has care of animals or other creatures, and hence, fig. [...]

Lika, v. See under **La**, v.

Lisha, v. See under **La**, v.

Mlisha, n. *wa-*, **Mlishi**, n. *wa-*. See under **La**, v.

Mlisho, n. *mi-*. See under **La**, v.

Mlo, n. See under **La**, v.

The goal of the above-mentioned rule is to portray the relations between derivatives and their bases, which would allow for a better understanding of their meaning and formation. As a result, the overlong, expanded and complicated entries, together with the frequent necessity to double-check a single word, cannot be deemed user-friendly. Sending a person checking the headword *ogofya* 'frighten', which is the causative form of the verb *ogopa* 'be threatened', to the entry word for the adjective *oga* 'timid', from which the first two words here mentioned are derived, can be of interest to a linguist, but does seem to introduce overly detailed information for an average user.

A different rule was introduced in the Swahili-Finnish *Abdulla* dictionary. Derivative forms of the verb have been placed only within the entry for the main verb, for example:

la v I. syödä, syövyttää, kuluttaa;[...]

- **lia** *appl* syödä jtk jssk/jkn kanssa/jllk/jklta;[...]
- **liana** *appl res* syödä toisiltaan/toistensa ruoka
- **lika** *stat*[...]
- **lisha** *kaus*[...]
- **lana** *res*[...]
- **liwa** *pass*[...]

Other derivatives possess the status of headwords, whereas the information regarding their base is given in brackets, for example:

m|lo s 3/4 (la) ruoka

ma|tumizi s 6 (tumia) käyttö,[...]

The rule is quite simple and transparent, but may cause some difficulties for beginning students. It sometimes happens that the word-forming base of the second word does not itself have the status of a headword. *Tumia* is a derivative form of the verb *tuma* and

should be searched for within the latter headword. The above method was also previously used in the Swahili-French dictionary by Sacleux (1939). In the French dictionary, additional information was included regarding derivatives at the end of the definition of the base:

-Tuma [...] a. Envoyer qqn avec une mission, comme agent; [...]
Mtuma, mtumadyi, mtumwa, mtumwadyi, tuma, tume, tumi, mtume, mtumi, utume, utumi, utumo, matumo, utumwa.
mTumadyi *wa-*, [...] Celui, elle qui envoie, qui députe, qui emploie.

In the dictionary compiled by researchers from *TUKI*, all derivative forms of the verb, which is the headword, are listed at the end of the article. There are no equivalents but information is given regarding the type of derivation, e.g. *tde – tendea* ‘prepositional extension’.

l.a kt 1. eat, consume. 2. erode. (nh) (1) ~ *fedha* use money; (2) ~ *hasara* incur a loss; (3) ~ *njama* plot, conspire; (4) ~ *rushwa* take a bribe; (5) ~ *yamini* take an oath. (*tde*) **lia**; (*tdk*) **lika**; (*tdn*) **lana**; (*tds*) **lisha**; (*tdw*) **liwa**.

Should the given derivative form be widely used and possess a lexicalized meaning, it is additionally included as a headword with an equivalent. Such entries do not have references nor information regarding the verb, from which they are derived.

lish.a kt feed, nourish; maintain, support: ~*mtoto* feed a child; ~*sumu* poison sb; ~*mno* overfeed. (*tde*) **lishia**; (*tdk*) **lishika**; (*tdn*) **lishana**; (*tdw*) **lishwa**.

The remaining derivatives have the status of headwords but no information regarding their base is included:

mlo nm mi- [*u-/i-*] meal, fare

As noted by Zgusta (1971: 16), various lexicographical decisions should be made with regards to future users of the dictionary: “we

must not forget that the lexicographer is doing scientific work, but that he publishes it for users whose pursuits are always more practical, at least as regarded from his own point of view”.

Due to the above, the most user-friendly Swahili dictionary, with regards to its complexity, should do its utmost to register all existing derivations. As mentioned by Herms (1999), it is not without reason that Swahili language students give extremely high praise to the 'friendly' dictionary by *Baba Malaika*, whilst de Schryver and Prinsloo (2000) note the low popularity of those dictionaries which group words according to their roots. Snoxall (1965: 28) points out that a modern user-friendly Bantu dictionary should list verbs and their derivative forms “under the proper alphabetical position in the form in which they are used in actual speech, as words possessing meaning”. User-friendly solutions may cause a lack of information regarding semantic and grammatical relations between lexemes. But these can be maintained with the use of an appropriate referencing system.

Kiango's (2000) postulates to supply users searching for the definitions of most words with references to their word-forming bases seems to be an unnecessary complication. An alternative solution to this may be the placing of a reference to the word base at the end of the entry, following its definition. This allows more adept users to widen their knowledge, whilst the less ardent ones will be spared the necessity of needlessly browsing through the pages of the dictionary.

Placing derivations under the base from which they were derived, requires advanced knowledge of language grammar from the user. According to Herms (1999), a method of saving space within the dictionary can be the omission of most regular derivative forms. Among derivatives causing the least learning problems to students, the author classifies the regularly formed passive and prepositional, as well as the reciprocal and reflexive forms. The dictionary should not however omit derivations created by processes, which change the grammatical category of the base and those formed by prefixation.

Due to significant productivity of derivative processes in Swahili, it seems impossible to be able to include all possible derivative forms

together with their definitions, at least not as part of a traditional printed dictionary. By according derivations the status of headwords, we also include them into our dictionary lemma-sign list. We should therefore revert back to the frequency list to include those derivative forms, which possess the highest probability of being searched by the user.

In the existing dictionaries, we can observe three rules regarding the inclusion of derivatives:

- random – the lexicographer includes various derivatives randomly, e.g. *Perrott*;
- in accordance with an approved procedure, e.g. in *TUKI* the most lexicalized forms were included;
- the maximum possible inclusion of all existing derivative forms, as in *Johnson*.

Basing the decision to include various derivative forms on the frequency list gives an opportunity to feature those forms, which users have the highest likelihood to come across during their work with the language. De Schryver and Prinsloo (2001a) performed a corpus research regarding the nominal and verbal derivations of the Swahili verb *sema* ‘speak’. They extracted the derivatives from the 1.3 million corpus of the Swahili language (*Kiswahili Internet Corpus*), and then checked their omission/inclusion in two Swahili-English dictionaries: *Perrott* and *The Internet Living Swahili Dictionary*⁵ (at that time the largest electronic Swahili dictionary, which consisted of over 50000 entries). The results of the analysis are presented in Table 2, where the letter ‘Y’ confirms the inclusion of the given form in the dictionary. As the table shows, the second most frequent form was not taken into account in either of the dictionaries, whilst forms which are not at all present in the corpus were included.

⁵ The dictionary used to be available at: <http://www.kamusiproject.org/>.

Table 2.

Number of occurrences in the corpus	Form	Perrott	The Internet Living Swahili Dictionary
10.862	sema	Y	Y
137	semekana		
117	msemaji	Y	Y
29	usemi	Y	Y
26	semwa	Y	Y
20	msemo		Y
6	semea		Y
4	semezana	Y	Y
4	msema		Y
2	semana		Y
1	semesha		Y
1	semeka	Y	Y
0	semezano		Y
0	msemi		Y
0	usemaji	Y	Y

Similar examples can be observed when analyzing the inclusion of words in dictionaries from the previously mentioned list of 200 most commonly used Swahili words, which was compiled from an over 12-million HCS corpus. Among the first 200 words, it is also possible to find derivatives. The inclusion of randomly selected forms in the dictionaries as *Perrott*, *Stopa-Garlicki*, and *TUKI* has been presented in the Table 3. We did not take into account such dictionaries as *Johnson* or *Abdulla*, where derivations, at least in regards to verbal forms, are in principle included as sub-entires of their bases. The letter ‘Y’ confirms the inclusion of the given form in the dictionary.

Table 3.

	Perrott	Stopa-Garlicki	TUKI
tumia	Y	Y	Y
fanyika			
uchaguzi	Y		Y
mwandishi			Y
mchezo	Y		Y
sababisha			Y

To summarize, the existing Swahili dictionaries can be classified into two groups in regards to headword arrangement and the procedure of the treatment of derivatives. The first group are alphabetical dictionaries, where all entry words have the status of headwords (e.g. *Baba Malaika*), the second are alphabetical-nest ones, which list the derivatives as part of the given headword, as its sub-entry (e.g. *Abdulla*). The use of the second group of dictionaries requires some knowledge regarding word formation. The user passively accessing such a dictionary may have issues with locating the given word.

Similar discussion may be raised concerning phraseological units. It has been analyzed in depth in an article by A. Chuwa (1996).

4.4 Homonyms

The Concise Oxford Dictionary of Linguistics (Matthews 2007) defines homonymy as “the relation between words whose forms are the same but whose meanings are different and cannot be connected”. Typically, homonymy applies to entire units, lexemes. However, controversy surrounds the issue of differentiating homonymy from polysemy – in other words distinguishing various meanings of the same unit from various units with the same spelling. The line between homonymy and polysemy is not clear, but lexicography has always distinguished both types (Bünting 1989: 216).

A thorough analysis of Swahili homonyms was undertaken by Gibbe (1977). He introduces the distinctions between homonyms, such as homographs, homophones, and proper homonyms.

Homographs are forms identically spelled, but differing in pronunciation and meaning. This phenomenon involving the existence of such forms is called homography. On the other hand, homophones are forms identical in pronunciation, but differing in regards to spelling, etymology and meaning. This phenomenon is called homophony (bf. Matthews 2007).

Swahili homographs in the most part originate from the historical occurrence of aspiration. Nowadays aspiration is not a distinctive feature and Swahili orthography does not mark its existence in spelling. Nevertheless a host of homonymic pairs exist in Swahili, which were etymologically differentiated by its occurrence (Polomé 1967: 39f.):

paa ‘roof’
p^haa ‘little gazelle’

kaa ‘sit’
k^haa ‘crab’

Homophones described in Gibbe (1977) possess all the above mentioned attributes. They have the same pronunciation, but differ in regards to spelling, etymology and meaning. However, the difference in spelling in the below examples only consists in the use of capital and lower case letters:

ukuta ‘wall’
UKUTA ‘abbreviation for *Usanifu wa Kiswahili na Ushairi Tanzania*’

Proper homonyms identified in Gibbe (1977) are identical in both spelling and pronunciation, differing only in regards to semantic and syntactic features.

Amongst others, he identifies:

- homonyms between onomatopoeic forms, for example:

pakacha ‘night thief’
pakacha ‘wicker basket’

- homonyms, where one form is onomatopoeic, for example:

pikipiki ‘motorcycle’

pikipiki ‘a stick used to pull down fruits’

- synchronic homonyms, which exist between the same or various parts of speech.

Among nouns, he additionally distinguishes homonymic forms in singular and plural forms, for example:

mtō ‘river’, ‘pillow’

pl.: **mitō** ‘rivers’, ‘pillows’

and homonymic in either singular or plural, for example:

mganga ‘bush’, ‘medicine man’

pl.1: **miganga** ‘bushes’

pl.2: **waganga** ‘medicine men’

Paradigmatic homonymy (between various parts of speech) is not widespread. Gibbe (1977) identifies them between nouns and verbs, nouns and adjectives, nouns and adverbs, as well as amongst nouns and copula. He additionally also distinguishes homonymy with borrowed lexemes, for example:

jazi ‘jazz’

jazi ‘supplement’

jazi ‘provide’

The number of homonymic forms in Swahili varies from 2 to 5. In practical lexicography, attributing separate dictionary entries to given words constitutes an acknowledgement that they are homonyms. Traditionally, the issue of homonymy was resolved through etymological research; however, identifying homonyms on the basis

of the grammatical characteristics of the units has become an increasingly widespread practice. In lexicographical practice, homonymic entries are included one after the other and generally they are numbered:

jua¹ to know
jua² sun

On the other hand, polisemic meanings are included within one entry and their various meanings are typically listed one after the other, for example:

amini 1. believe 2. trust 3. imagine

The practice to number each separate homonym is used for instance in the following dictionaries: Snoxall (1958), *Stopa-Garlicki*, *TUKI* or *Abdulla*:

paa¹ *kt* crape (off/up) [...]
paa² *kt* rise, ascend [...]
paa³ *kt ~ moto* transfer embers [...]
paa⁴ *nm [a-/wa-]* gazelle [...]
paa⁵ *nm ma- [li-/ya-]* roof [...] (TUKI)

Lack of numeration can be observed in *Madan*, *Johnson*, *Perrott*, *Jahadhmy*, amongst others :

paa (-), small gazelle
paa (ma), roof of native house
paa, *1* to ascend; *2* to scrape (Perrott)

The above examples illustrate not only the formal methodology of marking homonyms, but also more importantly the various rules regarding their classification. Four of the homonyms included for *paa* in the *TUKI* dictionary (the third meaning is entirely omitted) correspond with the three given in the *Perrott* dictionary. This is a result of referencing different attributes during the process of

identification. The *TUKI* dictionary classifies homonyms based on etymology, whilst the *Perrott* dictionary based on categories and grammatical features. The below examples also illustrate the above:

TUKI:

mto¹ *nm mi-* [u-/i-] river

mto² *nm mi-* [u-/i-] pillow

Perrott:

mto(mi), 1 river; 2 pillow

4.5 Arrangement of entries

Two basic orders of headwords can be distinguished in a dictionary – alphabetical and non-alphabetical. The alphabetical arrangement can be *a fronte*, with an order based on the first letters of the headword and *a tergo* (reverse dictionary), with an order based on the end letters of the headword. The non-alphabetical arrangement can be based on semantic or conceptual criteria, e.g. thesaurus.

As noted by Bańko (1987), the *a tergo* headword arrangement seems to be logical in the case of languages, where the inflectional or derivational morphemes predominantly occur in the form of prefixes. Swahili is mentioned as an example of such a language. However, amongst the Swahili lexicography the *a tergo* arrangement seems not to be popular and no publication of this kind was available to the author. This most likely results from several key issues. Irrespective of the commercial viability of such an endeavor, the factors mentioned by Bańko, such as the psychological impact or the influence of European lexicography, seem to sufficiently explain the *status quo*. Another reason may also stem from the issue of describing Swahili as a prefixal language, given the amount of derivational suffixes used.

Alphabetical arrangement seems to be the most natural and it is widely used in all bilingual Swahili dictionaries, which are published in the *a fronte* version. The majority of dictionary authors assume knowledge of the alphabet by users and do not include it. The alphabet is only to be found in dictionaries, where the second language is written down in a non-Latin alphabet, e.g. the Swahili-Russian dictionary by Ol’derogge (1961), or if the author distinguished language-specific sounds, as in the Swahili-French

dictionary by Sacleux (1939). An alphabet (excluding the letter *f*) was also introduced in the Swahili-English part of the *Perrott* dictionary, in order to explain the pronunciation rules for specific letters in the Swahili language.

All dictionaries, which maintain spelling of headwords with lower-case or capital letters (in *Johnson* and *Madan* all headwords start with a capital letter), consistently ignore the differences between them, treating them as textual variants of a given letter. For example, the following order will be used for the headwords: *uimbaji-Uingereza-uingiliaji*. Due to the differing spelling variants of the same word, e.g. *uislamu* (Ol'derogge 1961) and *Uislamu* (amongst others in *Perrott*), the only sensible option seems to be the rule to ignore differences in spelling. In the case of homonyms written down with capital or lower-case letters, the order in which they are defined is dependent on the dictionary, e.g. *TUKI* has *mzungu-Mzungu*, whilst *Abdulla: Mzungu-mzungu*. The order of these words is also not always consistently applied within a single dictionary, e.g. *TUKI* lists *mzungu-Mzungu*, but *Pemba-pemba*.

In accordance with Bańko (1987), there are two possible versions of an alphabetical order, namely the 'letter by letter' order (henceforth LBL) and the 'word by word' order (henceforth WBW). The difference between the two results from the different treatment of non-letter symbols which belong to the headword. In LBL, symbols such as spaces, hyphens, full stops, apostrophes are ignored, whilst the WBW order treats them as equal to the letters of the alphabet.

In Swahili, non-letter symbols, i.e. apostrophes and sporadically spaces, frequently appear in headwords. Apostrophes can be found in headwords such as *ng'aa* 'to shine', *ng'ombe* 'cow'. In all dictionaries with Swahili as the source language, apostrophes are classified according to the LBL order, i.e. they are ignored. The following order, therefore, is used for the below headwords:

ng'aa
ngabu
ngadu
.....

ng'amu
ngano

In the case of headwords which differ from each other only in regards to the apostrophe, the headword order is dependent on the rules agreed upon in the given dictionary:

ng'oa
ngoa (Perrott)

ngoa
ng'oa (TUKI)

Due to the gradual blurring of pronunciation differences for headwords with apostrophes, the adoption of such an alphabetical order makes the search for words easier for those learning the language, irrespective of their orthographic competence level.

Headwords which include spaces remain very rare. There is, however, no generally approved rule regarding their treatment. The space can be ignored by adopting LBL or it can be assigned a place in the alphabet by using the WBW order. Adopting the LBL rule causes the headwords to be scattered and separated by other paradigmatically alien words. By using the second rule, multiword headwords, where the first word is the one which is common to all elements of the entry, will occupy adjacent places in the dictionary. Logically speaking, the space can be placed in the alphabet, preceding all the letters or following them.

The space appears as the last symbol of the alphabet in *Abdulla*. Taking into account the example below, it can be concluded that this arrangement seems quite unnatural and unintuitive. It is more likely that the user of the headword *na kadhalika* will initiate his search under the word *na* and not where it was placed in the dictionary – after *nazi*:

na
naam
.....
nazi
na kadhalika
ncha

By applying the above rule, slightly similar to the LBL order, headwords with one common first word end up scattered throughout the dictionary. Implementing the LBL order brings about an arrangement which at least at first sight seems to be more natural:

na
naam
.....
na kadhalika
nazi
ncha

By applying the WBW order with the space as the first symbol of the alphabet, we end up with the order which is most intuitive:

na
na kadhalika
naam

Such an order was introduced amongst others in *Perrott*:

mbweha
mcha Mungu
mchaguo

Another key issue is the adherence to the alphabetical order of headwords. Such impermissible lapses do, for instance, occur in *TUKI*:

malkia
malumbano
maaluni
mama

mabano
mabavu
mabadilishano
mabishano

uchale¹
uchongeaji
uchale²

In the last example given, the headword *uchongeaji* lacks a definition, while the definition itself appears in its appropriate place.

5. Conclusion

In this article the macrostructural level of various bilingual dictionaries of Swahili has been analyzed. It showed the ways in which such issues as the treatment of derivatives or homonyms has been dealt with in some bilingual printed dictionaries. It revealed that no general decisions have been agreed upon by lexicographers so far and that editorial decisions are undertaken regardless other resources. The superiority of corpus-based approach has been identified and thus the need of urgent use of electronic corpora in Swahili lexicography has to be recognized by present-day dictionary compilers who should use the electronic data more effectively, as it has been already done by e.g. English language lexicographers.

References:

A. Dictionaries

- Abdulla, A., Halme R., Harjula L., Pesari-Pajunen M., 2002, *Swahili-Suomi-Swahili sanakirja*, Helsinki: Suomalaisen Kirjallisuuden Seura.
- Baba Malaika (see: Feeley G.), 1994, *Modern Swahili Modern English*, Arusha: Danish Training Center for Development Co-operation.
- Cahill, W. F., 1972, *Kamusi ya kwanza Kiswahili-Kiingereza. A first dictionary Swahili-English*, Nairobi: Nelson.
- Feeley, G. (see: Baba Malaika), 1990, *Swahili-English Dictionary*, Denmark: MS-tryk.
- Jahadhmy, A. A., 1981, *Learner's Swahili-English English-Swahili Dictionary*, London: Evans Brothers Limited.

- Johnson, F., 1985 (first edition 1939), *A Standard Swahili-English Dictionary (founded on Madan's Swahili-English Dictionary)*, Oxford: Oxford University Press.
- Madan, A. C., 1992 (first edition 1903), *Swahili-English Dictionary*, New Delhi: Asian Educational Services.
- Ol'derogge, A. D. (ed.), 1961, *Suahili-Russkij Slovar*, Moskva: Gos. Izdatel'stvo inostrannyh i natsional'nyh slovarei.
- Perrott, D.V., 1965, *The E.U.P. Concise Swahili and English Dictionary Together with students' notes and a short grammar*, London: The English Universities Press LTD.
- Sacleux, C., 1939, *Dictionnaire Swahili-Français*, Paris: Institut d'Ethnologie.
- Snoxall, R. A., 1958, *A Concise English-Swahili Dictionary. Kamusi ya Kiingereza-Kiswahili*, Nairobi: Oxford University Press.
- Stopa, R., Garlicki B., 1966, *Mały słownik suahilijsko-polski i polsko-suahilijski*, Warszawa: Wiedza Powszechna.
- Taasisi ya Uchunguzi wa Kiswahili (TUKI), 1996, *English-Swahili Dictionary*, Dar es Salaam: Chuo Kikuu cha Dar es Salaam.
- Taasisi ya Uchunguzi wa Kiswahili (TUKI), 2001, *Kamusi ya Kiswahili-Kiingereza. Swahili-English Dictionary*, Dar es Salaam: Chuo Kikuu cha Dar es Salaam.

B. Other literature

- Bańko, M., 1987, „Porządek alfabetyczny jako narzędzie leksykografa”, in: Z. Saloni (red.), *Studia z polskiej leksykografii współczesnej*, t. 2, Białystok: Dział Wydawnictw Filii Uniwersytetu Warszawskiego.
- Bańko, M., 2001, *Z pogranicza leksykografii i językoznawstwa. Studia o słowniku jednojęzycznym*, Warszawa: Uniwersytet Warszawski.
- Bünting, K.D., 1989, *Wstęp do lingwistyki*, Warszawa: PWN.
- Chuwa, A., 1996, “Problems in Swahili Lexicography”, *Lexikos* 6, 323-329.
- De Pauw, G., de Schryver, G-M., Wagacha P.W., 2009, “A Corpus-based Survey of Four Electronic Swahili-English Bilingual Dictionaries”, *Lexikos* 9, 340-352.

- De Schryver, G-M., Joffe, D., 2004, “On how electronic dictionaries are really used”, in: G. Williams, S. Vessier (eds.), *Euralex 2004 Proceedings Computational lexicography and lexicology*, Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.
- De Schryver, G.M., 2012, “Tools to support the design of a macrostructure (nomenclature)”, in: R.H. Gouws, U. Heid, W. Schweickard and H.E. Wiegand (eds.), *Dictionaries. An international encyclopedia of lexicography. Supplementary volume: Recent developments with special focus on computational lexicography*, Berlin: Walter de Gruyter.
- De Schryver, G.M., Prinsloo D.J., 2000, *Electronic corpora as a basis for the compilation of African-language dictionaries, Part 1: The macrostructure. South African Journal of African Languages* [<http://tshwanedje.com/publications/Macro.pdf>]
- De Schryver, G.M., Prinsloo D.J., 2001a, “Towards a Sound Lemmatization Strategy for the Bantu Verb through the Use of Frequency-based Tail Slots – with special reference to Cilubà, Sepedi and Kiswahili”, in: J.S. Mdee and H.J.M. Mwansoko (eds.), *Makala ya kongamano la kimataifa Kiswahili 2000. Proceedings*, Dar es Salaam: TUKI, Chuo Kikuu cha Dar es Salaam.
- De Schryver, G.M., Prinsloo D.J., 2001b, “Corpus-based activities versus intuition-based compilations by lexicographers, the Sepedi lemma-sign list as a case in point”, *Nordic Journal of African Studies* 10(3), 374-398.
- Fillmore, J., 1992, “Corpus Linguistics” or “Computer-aided armchair linguistics”, in: J. Svartvik (ed.), *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991*, Berlin: Mouton de Gruyter.
- Gibbe, A.G., 1977, *Homonymy, Synonymy, and Antonymy in Kiswahili*, unpublished M.A. dissertation, University of Dar es Salaam.
- (HCS) *Helsinki Corpus of Swahili*, 2004, Compilers: Institute for Asian and African Studies (University of Helsinki) and CSC – Scientific Computing Ltd.

- Herms, I., 1999, "Swahili Lexicography: The Swahili-German Dictionary", in: H.E. Wolff (ed.), *Contributions to Bantu Lexicography. Languages and Literatures*, 10, Leipzig: University of Leipzig Papers on Africa.
- Kiango, J.G., 2000, *Bantu Lexicography: a critical survey of the principles and process of constructing dictionary entries*, Tokyo: Tokyo University of Foreign Studies.
- Landau, S.I., 1984, *Dictionaries. The Art and Craft of Lexicography*, New York: The Scribner Press.
- Matthews, P.H., 2007, *The Concise Oxford Dictionary of Linguistics*, Oxford University Press.
- Lyons, J., 1977, *Semantics*, Cambridge: Cambridge University Press.
- Ohly, R., 1967, „Roman Stopa, Bolesław Garlicki, Mały Słownik Suahilijsko-Polski i Polsko-Suahilijski, Kamusi Dogo ya Kiswahili-Kipolanda, Kipolanda-Kiswahili”, *Przegląd Orientalistyczny*, 3(63), 265-268.
- Ohly R., Kraska-Szlenk I., Podobińska Z., 1998, *Język suahili*, Warszawa: Dialog.
- Polomé, E.C., 1967, *Swahili language handbook*, Washington, D.C.: Center for Applied Linguistics.
- Snoxall, R.A., 1965, "Some Problems and Principles of Lexicography in Luganda", *African Language Studies* 6, 27-31.
- Summers, D., 1996, "Computer lexicography: the importance of representativeness in relation to frequency", in: J. Thomas and M. Short (eds.), *Using Corpora for Language Research. Studies in the Honor of Geoffrey Leech*, London: Longman.
- Tomaszczyk, J., 1983, "On bilingual dictionaries. The case for bilingual dictionaries for foreign language learners", in: R.R.K. Hartmann (ed.), *Lexicography: Principles and Practice*, London: Academic Press.
- Verlinde, S., Selva T., 2001, *Corpus-based vs intuition-based lexicography: defining a word list for a French learners' dictionary. Proceedings of the Corpus Linguistics 2001 conference*, Lancaster: Lancaster University.

- Wamitila, K.W., 1997, “Kamusi ya awali ya sayansi na teknolojia”, *Afrikanistische Arbeitspapiere* 51, 196-199.
- Zgusta, L., 1971, *Manual of Lexicography*, Praha: Academia.
- Żmigrodzki, P., 2003, *Wprowadzenie do leksykografii polskiej*, Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- De Pauw, G., de Schryver, G-M., Wagacha P.W., 2009, “A Corpus-based Survey of Four Electronic Swahili–English Bilingual Dictionaries”, *Lexikos* 9, 340-352.
- De Schryver, G-M., Joffe, D., 2004, “On how electronic dictionaries are really used”, in: G. Williams, S. Vessier (eds.), *Euralex 2004 Proceedings Computational lexicography and lexicology*, Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.