Linguistics Terminology and Neologisms in Swahili: Rules vs. Practice

Gilles-Maurice de Schryver

Ghent University and University of Pretoria

gillesmaurice.deschryver@UGent.be

ABSTRACT

In this article we discuss the use of Swahili terminology in the field of linguistics. In particular, we are interested in finding out whether the rules laid out by scholars in the scientific literature for the creation of terminological neologisms in Swahili correspond with actual practice. In order to do this, three steps are taken. In Step 1 we undertake the semi-automatic extraction of linguistics terminology, by comparing occurrence frequencies in a special-purpose corpus consisting of ten recent Swahili language/linguistics textbooks, with their corresponding frequencies in a twenty-two-million-token general-language reference corpus of Swahili. In Step 2 we study the source languages and actual word formation processes of the terms and neologisms with the highest keyness values obtained during the previous step. This discussion is divided into several sections, one section per source language. In Step 3, the terms and neologisms that have been found are compared with their treatment (or absence thereof) in two existing reference works, a general dictionary and a linguistics terminology list. These three steps are preceded by brief introductions to (i) the Swahili language; (ii) its dictionaries and terminology lists; (iii) its metalexicographical, terminological and neologism studies; and (iv) our use of

the term "neologism." The three steps are followed by a discussion of our findings and a conclusion.

Keywords: Bantu, Swahili, corpora, semi-automatic term extraction, linguistics terminology, terminological neologisms, lexicography, digital dictionaries

THE SWAHILI LANGUAGE

The Bantu language Swahili (or Kiswahili in the language) is *the* lingua franca of East Africa, spoken by up to 100 million first- and second-language speakers, especially in Tanzania and Kenya, but also in the neighbouring countries to the west and south (Mohamed 2009, iv–v). It is one of the most well-known African languages and is taught at virtually every African-language department in the world.

Although grammatically a Bantu language, a large part of its vocabulary is actually borrowed from Arabic, with regards to which the Encyclopædia Britannica (2014) simply states that "there are an enormous number of Arabic loanwords in the language." These loanwords, in turn, have found their way into neighbouring languages (e.g., Baldi 2011; 2012).

DICTIONARIES AND TERMINOLOGY LISTS FOR SWAHILI

The existing lexicographic output, the result of a century-and-a-half-old craft rather than a modern science, consists of several hundred dictionaries and terminology lists. Benson (1964, 65) traces the origins of Swahili lexicography back to the earliest manuscript the missionary Krapf sent home from the field, "a vocabulary which became quite an extensive work," written in 1844. Krapf's first published dictionary, *A Dictionary of the Suahili Language*, dates from 1882. The next important lexicographic work is the *Dictionnaire français-swahili* by Sacleux (1891), followed half a century later by its counterpart, the *Dictionnaire swahili-français* (Sacleux 1939). Around the turn of the nineteenth century, Madan (1894; 1902; 1903) produced his set of bilingual English-Swahili-English dictionaries. Subsequently, Swahili was successfully paired with German by Velten (1910; 1933) in his *Suaheli-Wörterbuch*. All of

these dictionaries are consulted to this day, as are those of Johnson (1939a; 1939b), which are based on those of Madan (1902; 1903).

A welcome addition to the output of the early missionaries and the trade are the Swahili dictionaries produced by academic institutes. Chief among them has been TUKI (*Taasisi ya Uchunguzi wa Kiswahili*), also known under its English acronym IKR (Institute of Kiswahili Research), at the University of Dar es Salaam. In addition to their three flagship dictionaries – a monolingual Swahili dictionary (TUKI 1981; 2004), a bilingual English-Swahili dictionary (IKR 1996; 2000; 2006) and a bilingual Swahili-English dictionary (TUKI 2001; 2014) – TUKI has been engaged in the compilation of numerous terminology lists. These include terms for linguistics (Massamba 1990; 2004).

The most popular online dictionary for Swahili remains the one produced fifteen years ago by Hillewaert and de Schryver (2004).

METALEXICOGRAPHICAL, TERMINOLOGICAL, AND NEOLOGISM STUDIES FOR SWAHILI

Despite the size and importance of the language as well as the large number of practical lexicographic products, the number of metalexicographical, terminological and neologism studies for Swahili is underwhelming, at least in the English academic literature (cf. de Schryver 2009, 387). Exceptions include the insightful works of Mdee (1990), Tumbo-Masabo (1990), and Kiango (2000). The main reason for this state of affairs is that most research in this field has actually appeared *in* Swahili, published by TUKI. Excellent examples are the edited collections by Tumbo-Masabo and Mwansoko (1992), Kiango and Mdee (1995), and Mwansoko and Chuwa (1995).

GOAL OF THE PRESENT ARTICLE

In this article we wish to study the use of Swahili terminology in the field of linguistics, and in particular to find out whether the most important terms and neologisms extracted from a small purpose-built Swahili language/linguistics corpus are adequately covered in two of the existing reference works, one the general online bilingual Swahili-English dictionary by Hillewaert and de Schryver (2004), the other

the special-purpose paper semi-bilingual Swahili-Swahili-English linguistics terminology list by Massamba (2004). In passing, we will also test whether a mostly automated approach to doing this works (well) for Swahili.

Our definition of a neologism is the one proposed by the father of lexicography as a scientific discipline: "neologism is a term which can refer to any new lexical unit, the novelty of which is still felt" (Zgusta 1971, 179). Language and especially modern linguistics being a relatively recent field of knowledge for which terminology has had to be developed in Swahili, the assumption in this article is thus that quite a number of Swahili language/linguistics terms are still novel enough to be called neologisms. With Zgusta (1971, 181ff.) we may call these "terminological neologisms." As part of this study, we will thus also need to look into how new concepts for language/linguistics have been and are currently coined in Swahili, through borrowing and/or through language-internal processes. The various results will then be brought together in order to offer suggestions for the compilation of future reference works for Swahili.

In a way, one may view the present study as picking up where Tumbo-Masabo (1990) left off. In her PhD dissertation, she had analysed the development of Swahili neologisms in mathematics, combing through all the then existing terminology lists (and dictionaries more generally), but specifically leaving out textbooks:

> This study does not deal with the use of coined terms in mathematical text books. According to Haugen's fourfold model [(Haugen 1966)], the use of neologisms in text books is at another level of language planning, the implementation level. Implementation is a very important process, as it determines the appropriateness and acceptability of the terms coined. Therefore, it warrants separate research. (Tumbo-Masabo 1990, 62)

In the present study, our starting point is a corpus of recent textbooks (for language/linguistics rather than mathematics). By extracting the terms and neologisms found in these textbooks, and by then comparing those with the terms and neologisms suggested in earlier reference works, we will be operating on the implementation level.

SEMI-AUTOMATIC EXTRACTION OF SWAHILI LINGUISTICS TERMINOLOGY

In order to semi-automatically extract Swahili linguistics terminology, we compare the occurrence frequencies in a special-purpose corpus consisting of ten recent Swahili language/linguistics textbooks, with their corresponding frequencies in a twenty-two-million-token general-language reference corpus of Swahili. That reference corpus is an amalgamation and extension of earlier Swahili corpora (e.g., de Schryver and Prinsloo 2001; de Schryver et al. 2006; Devos and de Schryver 2016). The contents of the special-purpose corpus are shown in Table 1. (Full bibliographic details, including the original titles in Swahili, may be found in the References.)

Interestingly, the values in Table 1 suggest that the range of vocabulary used in the (two) Kenyan textbooks is wider than in the (eight) Tanzanian textbooks, as can be deduced from the STTRs (which indicate the number of "new" words entering a textbook for each additional 1,000 words). In all, the total number of running orthographic words in the language/linguistics textbook corpus is just over 400,000 (the tokens), about 35,000 of which are unique orthographic words (the types). To compare these types with the more than 500,000 types in the reference corpus, the methodology presented in Taljard and de Schryver (2002) is employed. In short, the KeyWords tool of WordSmith Tools 6 (Scott 1996–2015) was set to extract all the so-called positive "keywords." According to the manual of WST6:

> The "key words" are calculated by comparing the frequency of each word in the word-list of the text you're interested in with the frequency of the same word in the reference word-list. [...] To compute the "key-ness" of an item, the program therefore computes
>
> - its frequency in the small word-list
> - the number of running words in the small word-list
> - its frequency in the reference corpus
> - the number of running words in the reference corpus
>
> and cross-tabulates these.

TABLE 1 Swahili Language/Linguistics Textbooks in the Special-purpose Corpus (KN = Kenyan, Nat. = nationality, std. dev. = standard deviation (of STTR), STTR = standardized type-token ratio, TZ = Tanzanian)

Nat.	Author	Year	Translation of book title into English	Tokens	Types	STTR	std. dev.
TZ	Khamisi	2008	Relational Grammar	36,209	4,713	35.24	62.27
TZ	Khamisi	2009	Proficiency in Swahili: Grammar 1	8,867	1,846	35.34	56.46
TZ	Kihore	2009	Proficiency in Swahili: Grammar 2	14,490	2,335	36.18	58.12
TZ	Kihore, Massamba and Msanjila	2003	Structural Grammar of Standard Swahili: Secondary and Tertiary	43,347	5,480	37.16	60.13
TZ	King'ei	2010	Basics of Sociolinguistics	55,959	8,546	43.51	55.38
TZ	Massamba, Kihore and Hokororo	1999	Syntax of Standard Swahili: Secondary and Tertiary	36,927	5,147	38.45	58.58
TZ	Massamba, Kihore and Msanjila	2004	Phonology of Standard Swahili: Secondary and Tertiary	23,044	3,057	33.93	62.09
KN	Matei	2008	The Binoculars of Grammar: The Exact Reading of Swahili Grammar. Questions and Exercises	61,331	9,986	45.56	54.33
TZ	Nyangwine and Masebo	2002	Swahili: Secondary 3 and 4. Development of Swahili, Literature, Grammar, Use, Formation and Knowledge	91,717	12,678	43.35	56.56
KN	Waihiga	1999	Explanatory Grammar of Swahili	39,684	7,494	44.15	55.94
			SUM	411,575	34,768	41.00	61.03

Statistical tests include:

- the classic chi-square test of significance with Yates correction for a 2 x 2 table
- Ted Dunning's Log Likelihood test, which gives a better estimate of keyness, especially when contrasting long texts or a whole genre against your reference corpus. (Scott 1996–2015, 245)

In other words, orthographic words which are unusually frequent in the language/linguistics corpus when compared to their frequency in the reference corpus were considered to be keywords (with the additional requirement that these keywords had to occur at least ten times across the special-purpose textbooks). To calculate the "keyness values," WST6 was set to employ Dunning's (1993) log-likelihood test.[1]

The assumption is therefore that most of the keywords offered by WST6 will also be (specialized) terms from the field of language/linguistics, and that this list will also contain some neologisms. Here we are presented with an intriguing conundrum: given that Swahili has borrowed vocabulary so heavily from Arabic in the course of so many centuries, when does a borrowing stop being a borrowing and become part of the Swahili language?

The result of this exercise is presented in the online Addendum.[2] WST6 offered over 2,000 keywords – too many to study in detail for the purposes of this article. We therefore decided to go down the list in order of decreasing keyness value, until we reached 100 language/linguistics terms. The first observation that can be made concerns the effectiveness of the methodology used. To reach 100 terms, we had to go down to item 142, which means a *precision* of seventy percent. This value is much higher than the one obtained for the Bantu language Northern Sotho (Taljard and de Schryver 2002, 54–56), which is good

[1]The methods offered in WST6 to calculate "keyness values" are admittedly rather limited; in WST7 (Scott 1996–2019) the standard chi-square has been dropped, while the newer log ratio (Hardie 2014) and BIC score (Gabrielatos 2018) have been added as statistical measures. If our Swahili corpora were available in the Sketch Engine (Kilgarriff 2003–2019; Kilgarriff et al. 2004; Kilgarriff et al. 2014), we could also have made use of the ranking options available in that corpus query package.

[2]https://dictionarysociety.com/journal-supplements

news for the methodology proposed by Taljard and de Schryver (2002). Secondly, hardly any of the (42) discarded items are actually junk, as they are simply typical for the kind of writing found in textbooks, with material such as *huweza* 'usually can', *zifuatazo* 'the following', *huwa* 'always is/are, usually is/are', *hutumiwa* 'is/are usually used', *huonyesha* 'usually show', *eleza* 'explain', ... *yaani* 'in other words'. Other items are used to number material (i, ii, iii, iv, …), to contrast material (*au* 'or'), or even to stress material (*hiki* 'this (one)', *hizi* 'these (ones)'). For at least three items it could also be argued that they are terms: *kitabu* 'book' (< Ar), *Kiingereza* 'English' (< En) and *andika* 'write' (Sw). In point of fact, Massamba (2004, 33) includes *Kiingereza* 'English' as a lemma sign in his linguistics terminology list.

Conversely, and moving to the actual 100 terms, one could argue that *tunga* 'compose, make' and *hudondoshwa* 'is usually dropped' should not have been kept as language/linguistics terminology. But these are details. The overall picture as to the provenance of the top 100 keywords is shown in Figure 1: thirty-six percent have been adopted from other languages, especially Arabic (26%) and English (9%) with one word from Latin (1%), namely *nomino* 'noun'.

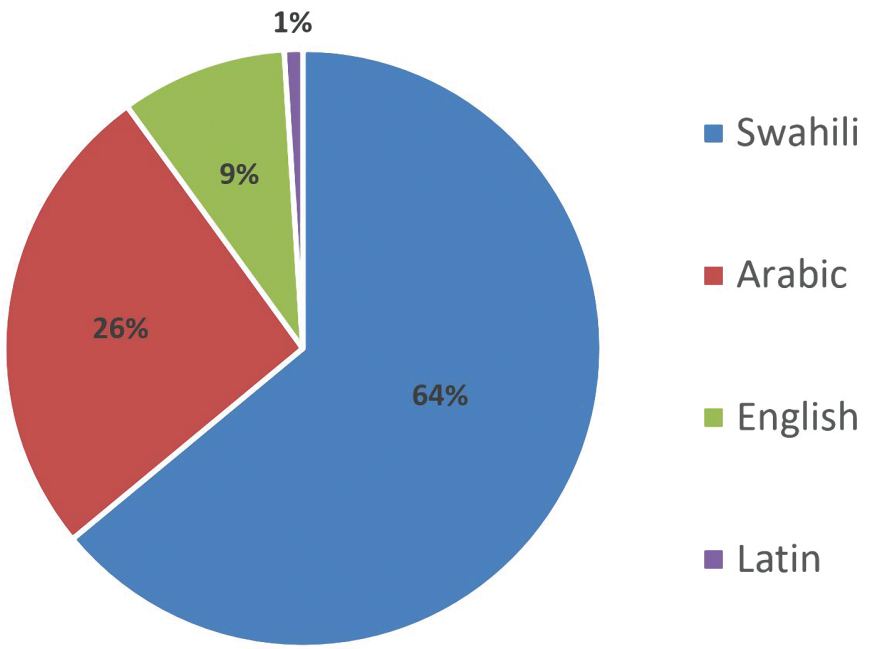

FIGURE 1 Source Language of the Top Swahili Language/Linguistics Terms

NEOLOGISMS IN SWAHILI LINGUISTICS TERMINOLOGY

We now wish to study the actual word formation processes of the terms and neologisms with the highest keyness values obtained during the previous step, namely, of the "top 100." The analysis follows the framework illustrated in Gauton et al. (2003) for the Bantu languages Zulu and Northern Sotho.

When it comes to the terms with an English source, these are: *sentensi* 'sentence', *konsonanti* 'consonant', *mofimu* 'morpheme', *silabi* 'syllable', *fonolojia* 'phonology', *kisintaksia* 'syntactic(al); syntactically', *sintaksia* 'syntax', *nazali* 'nasal' and *mofu* 'morph'. Four aspects are immediately clear when Swahili borrows from English: (i) when there is no final vowel, one is added (here *-u* following *m* and *f*; *-i* elsewhere), (ii) non-native consonant clusters are replaced by (mostly) single consonants, (iii) some morphemes may be made 'lighter' by adding *-ia*, and (iv) part-of-speech mismatches are handled both morphologically and syntagmatically (e.g., the English adv. 'syntactically' takes the prefix from class 7 (*ki-*), while the adj. 'syntactic(al)' uses the adverbial *ki-* in a connective construction (*-a* ~)).[3] Therefore, rather than making use of a *loanword* in which the original (English) spelling has been retained, Swahili terminologists prefer *transliteration*, whereby the phonological structure of a loanword is adapted to the sound system of the borrowing language. In simple terms: Swahili likes sequences of CVCV…, where Cs may also be prenasalized. These observations from actual Swahili *practice*, as found in textbooks, correspond with the Swahili *rules*, as for example laid down by Tumbo-Masabo (1992, 26–31).

With regard to the terms with an Arabic source, the first problem, of course, is to be able to confidently state that a certain term is indeed originally Arabic. Here, we were guided by the etymological information found in the online monolingual Swahili dictionary which is part of the Oxford Global Languages project (OGL 2017), in addition to cross-checks in the language portal bab.la (2017). These showed that fully a quarter of the top 100 language/linguistics terms in Swahili seem to originate from Arabic. For these, it is possible to make a number of observations. Firstly, some terms have been borrowed straightforwardly, both with regard to form and meaning, such as *lugha* 'language' < لُغة [luġa]

[3]For more on this, see de Schryver (2006; 2008).

'language', or *maana* 'meaning' < مَعْنًى [ma'nā] 'meaning'. Secondly, in the case of words in the previous category without a final vowel in the original Arabic, such a vowel is added in Swahili: *sauti* 'sound; voice' < صَوْت [ṣawt] 'voice', or *kamusi* 'dictionary' < قاموس [qāmūs] 'dictionary'. Thirdly, extending this further, vowels may be interspersed in order to obtain the open syllable structure (CVCV…) typical of Swahili: *lahaja* 'dialect' < لَهْجة [lahja] 'dialect; tone', *sahili* 'simple' < سَهْل [sahl] 'simple', or *herufi* 'letter' < حَرْف [ḥarf] 'letter'. Fourthly, for each of the three previous groups, the original meanings may be extended to obtain the linguistic terms, thus respectively: (i) صِفَة [ṣifa] 'quality, trait' > *sifa* 'feature', or عَلامة ['alāma] 'marker' > *alama* 'symbol'; (ii) إِعْرَاب [i'rāb] 'the system of nominal, adjectival and verbal suffixes of Classical Arabic' > *irabu* 'vowel', فَصيح [faṣīḥ] 'eloquent' > *fasihi* 'literature', or نَفْس [nafs] 'soul' > *nafsi* '(grammatical) person'; and (iii) صَرْف [ṣarf] 'inflection; declension' > *sarufi* 'grammar'. Unlike the neologisms sourced from English (discussed in the previous paragraph), these borrowings from Arabic are likely to be many centuries old. This is confirmed by looking up the twenty-six Arabic loans from the top 100 in Sacleux's (1939) dictionary from eighty years ago: as many as two thirds are covered, with a direct link to the Arabic in that dictionary's etymological slot.[4] Those that are not covered in Sacleux's (1939) dictionary may indeed be of more recent origin, such as: *lahaja* 'dialect' < لَهْجة [lahja] 'dialect; tone', *istilahi* 'terminology' < مُصْطَلَح [muṣṭalaḥ] 'terminology', or *methali* 'proverb, saying' < مَثَل [maṯal] 'parable, epigram, example'.[5]

The remaining sixty-four terms in the top 100 all seem to be inherently Swahili. Ten of these are readily available as "plain nouns," meaning that they are not derived from words in other word classes. These are: *ngeli* 'noun class', *mzizi/mizizi* 'root(s)', *kishazi/vishazi* 'clause(s)', *yambwa* 'object', *kibantu* 'Bantu ... (e.g., language)', *msamiati* 'vocabulary', *nduni* 'feature', and *kaakaa* 'palate'. (Note that both singular and plural nouns may be returned by WST6, given that our Swahili corpora are not POS-tagged.) A total of forty-one terms are nouns derived from verbs, for which template-like 'constructions' are employed. These are found in class pairs 3/4, 5/6, 7/8, 9/10 and 14, for which the distribution is shown

[4]See the online Addendum, column "Sacleux 1939," for the details.

[5]More detailed studies of Arabic loanwords in Swahili in general may be found in Tucker (1946), Mwita (2009), Harvey (2014), and Fattakhova and Mingazova (2015).

in Table 2. Examples for each are: *muundo/miundo* 'structure(s)' [3/4] < *kuunda* 'to make, to construct'; *matamshi* 'pronunciations, speech' [5/6]< *kutamka* 'to pronounce, to articulate'; *kiambishi/viambishi* 'affix(es)' [7/8] < *-ambika* 'be said'; *tungo* 'construction' < *kutunga* 'to construct, to form' [9/10]; and *utamkaji* 'articulation' [14] < *kutamka* 'to pronounce, to articulate'. Other terms are more adjective-like, such as *ambatani* 'compound' < *kuambatana* 'adhere to one another, joined by one another', or *nyambulishi* 'derivational' < *kunyambulika* 'to be extended'. The last group ('other' in Table 2) is heterogeneous, and for instance contains the verb stem *tendeka* 'stative, neuter' < *kutendeka* 'be accomplished'.

TABLE 2 Types of Constructions Underlying the Inherently Swahili Language/Linguistics Terms: Cl. = class(es), ext = extension, R = root

Cl.	Construction	N
plain noun	—	10
3/4	m-/mi-R(+ext)-o	6
5/6	ø-/ma-R-i	3
	ø-/ma-R-o	4
7/8	ki-/vi-R(+ext)-i	13
	ki-/vi-R-o	3
	ki-/vi-stem	4
9/10	ø-R-o	1
	ø-stem	1
14	u-R(+ext)-i	5
	u-stem	1
adj.	R-i	5
other	…	8
	SUM	64

With regard to the term formation processes observed here, it is clear that we are mainly dealing with what Mtintsilana and Morris (1988, 110) call *semantic specialization*, whereby words from the general vocabulary acquire additional, more technical meanings. The other strategies which Mtintsilana and Morris (1988, 110–12) discuss, such as *compounding*, whereby a term is coined by combining existing words, or the exploitation of the *synonym* richness of a vocabulary, are hardly

found here.[6] Similarly, the translation strategies listed by Baker (1992, 26-42), such as the use of a more general word or *superordinate*, the use of a more *neutral* or less expressive word, the use of *cultural substitution*, a *paraphrase* using a related word, or a paraphrase using an *unrelated* word, are also not seen for the top 100 Swahili language/linguistics terms.

To bring this section to a close, Figure 2 shows the distribution of the word classes of the top 100 Swahili language/linguistics terms, according to source language. Unsurprisingly, the great majority of terms are nouns, followed by adjectives and a few verbs.

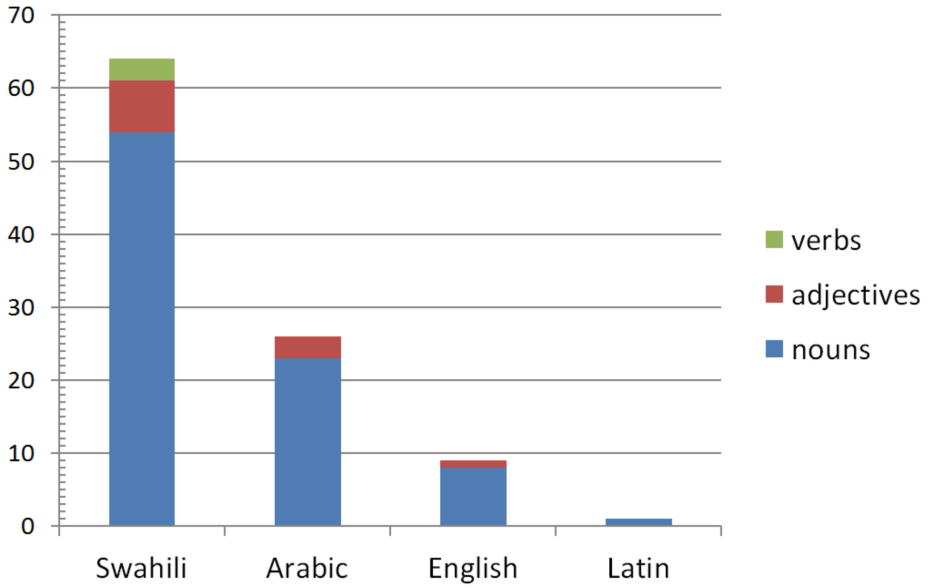

FIGURE 2 Word Classes of the Top Swahili Language/Linguistics Terms, per Source Language

LINGUISTICS TERMINOLOGY AND NEOLOGISMS IN EXISTING DICTIONARIES

How well is terminology covered in a general-language dictionary? One obviously does not expect to find all terminology of all fields in a broad reference work, but some ought to be covered. The top 100 language/linguistics terms were searched for in the general online Swahili-English dictionary by Hillewaert and de Schryver (2004). As

[6]See the online Addendum, columns "Strategy" and "Construction."

many as forty-seven were found, eighteen of which were also specifically labelled as belonging to the field of linguistics. A further five were found, but without the specialized meaning, while forty-eight were not found at all. This is shown in Figure 3.

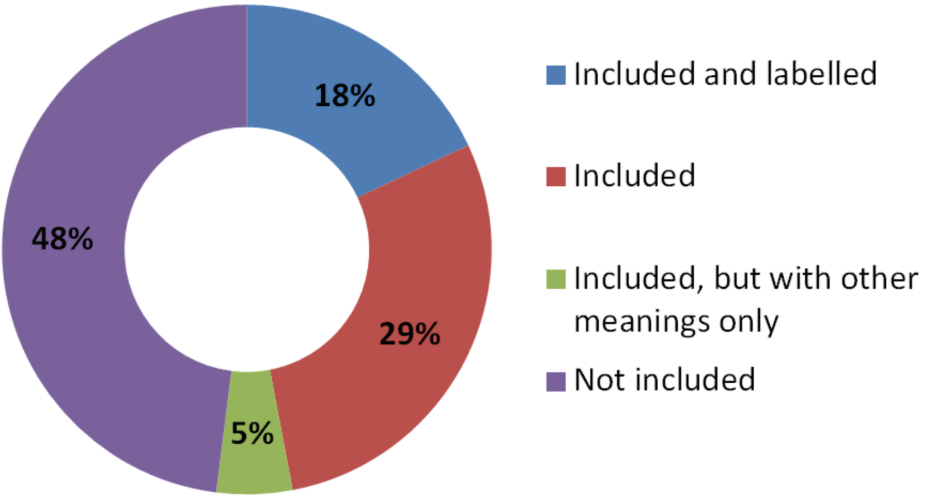

FIGURE 3 Top Swahili Language/Linguistics Term in a General Online Swahili-English Dictionary (Hillewaert and de Schryver 2004)

Of the forty-seven terms that were found, eighteen were sourced from Arabic, English and Latin. Of the fifty-three terms that were not found, another eighteen were sourced from Arabic and English. Exactly half the top borrowed language/linguistics terms were thus covered in this dictionary; the other half were not.

We repeated this exercise for the special-purpose paper semi-bilingual Swahili-Swahili-English linguistics terminology list by Massamba (2004). Given that a manual count of the number of lemma signs in this terminology list revealed a total of 1,398 entries, it would not be too much to expect that the top 100 terms (which are actually used in textbooks) are also covered. While the outcome is better than for the general-language dictionary, nonetheless only sixty-five of the terms were found, while thirty-five were not. Details of the different categories of found vs. not-found are shown in Figure 4. For the not-founds, two categories are slightly problematic: seventeen of the missing terms are actually used throughout the microstructure of the terminology list itself (e.g., *KN = kikundi/kirai nomino* 'noun phrase', *KT = kikundi/kirai kitenzi* 'verb phrase', *ambatani* 'compound', *elekezi* 'transitive; prescriptive', ...),

while another six are only used as the second part of compounds and can thus not reasonably be found (e.g., *tendeka* 'stative, neuter' can only be found at *kauli* ~ 'stative voice' and at *kitendeka* 'stative marker', or *nyambulishi* 'derivational' can only be found at *kiambishi* ~ 'derivational affix' and at *ambishaji* ~ 'derivation'). In terms of loanwords from Arabic and English, nine have not been lemmatized. For instance, on the macrostructural level, *sanifu* 'standard' can only be found at *lahaja* ~ 'standard dialect' and at *lugha* ~ 'standard language', while *sahili* 'simple' can only be found at *njeo* ~ 'simple tense', at *sentensi* ~ 'simple sentence' and at *wimbi* ~ 'simple wave', while it is also imbedded in the lemma sign *usahilishaji sheria* 'rule simplification'. Conversely, both *sintaksia* 'syntax' and *kisintaksia* 'syntactic(al); syntactically' are used throughout the terminology list but have not been lemmatized at all. Finally, the absence of entries for *fasihi* 'literature', *methali* 'proverb, saying' and *ufasiri* 'translation' may perhaps be explained if it is assumed that only terms from *core* linguistics were meant to be included in this terminology list.

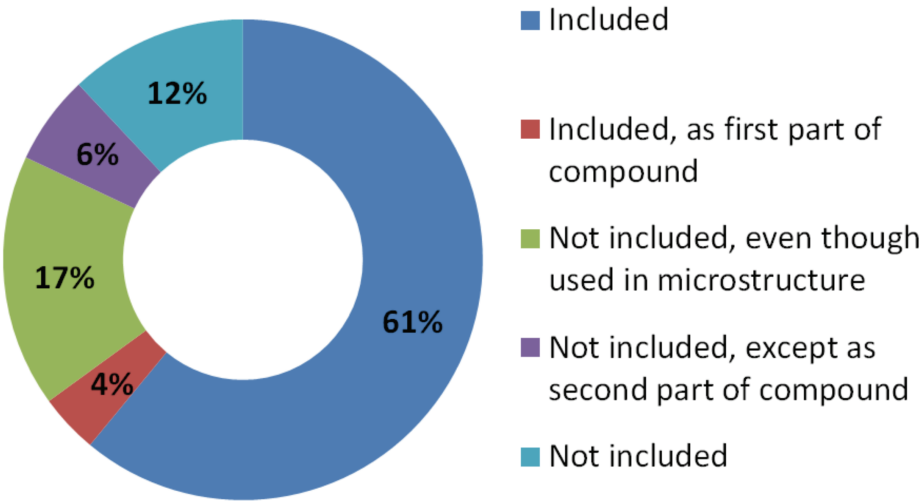

FIGURE 4 Top Swahili Language/Linguistics Terms in a Special-Purpose Paper Semi-bilingual Swahili-Swahili-English Linguistics Terminology List (Massamba 2004)

DISCUSSION AND CONCLUSION

With reference to terminological neologisms in the field of Swahili language/linguistics, it may thus be concluded that most loanwords from Arabic have ceased to be neologisms, given that they entered the language many decades, sometimes even centuries, ago. Two-thirds of the Arabic-sourced terms in our study were for instance found in Sacleux's (1939) dictionary of eighty years ago. However, some entered recently enough from Arabic to continue to cause confusion: see for instance the discussion in Gibbe (2008, 79–80) regarding *istilahi* (before standardization *sitilahi*, < Ar, now: 'terminology') vs. *msamiati* (Sw, now: 'vocabulary'). Most present-day loanwords to express concepts in linguistics are sourced from English, and in this, practice follows suggested theory (Mwansoko 1990; 2001, 325–27). While all of the found English-sourced terms were also included in Massamba's (2004) fifteen-year-old terminology list, for some users these may still feel 'new enough' to be seen as neologisms (e.g., *mofimu* 'morpheme', *silabi* 'syllable', *fonolojia* 'phonology', …). Lastly, for the terms that were (recently) coined through language-internal processes, most may well still be considered to be neologisms as they are the result of *semantic specialization*, whereby general Swahili vocabulary is extended to acquire additional, more technical meanings (e.g., *utamkaji* 'articulation' (< Sw *kutamka* 'to pronounce, to articulate'), *kiunganishi* 'conjunction' (< Sw *kuunganisha* 'to connect' < Sw *kuunga* 'to join, to unite'), *kiimbo* 'intonation' (< Sw *kuimba* 'to sing'), …).

In the conclusion to her PhD dissertation, Tumbo-Masabo (1990) pointed out:

> It has been predicted that since Swahili society is not technologically advanced and relies on imported technology, the main source of neologisms in its language would be through adoption and loan translation (Ohly 1978). Contrary to this hypothesis, derivation has been the most productive process of coining new words in mathematics, which is a basic field in technology. Ohly (1978) underestimated the vitality of Kiswahili and the input of the speakers of the language in generating new words appropriate to [the] contemporary world. (Tumbo-Masabo 1990, 98)

By and large, our results for Swahili terminology and neologisms in the field of language/linguistics are similar to those found for the field of mathematics, that is, Tumbo-Masabo's 'derivations' correspond to the language-internal coinages which are most frequently found for our more recent terminological neologisms. However, Tumbo-Masabo (1990) arrived at her findings after studying terminology and neologisms in mathematics in the then available reference works, but she did not study whether these neologisms were also being picked up. Our research started at that other end and now reveals that the Swahili coinages in language/linguistics that are found at the implementation level are also overwhelmingly found in contemporary reference works – and these coinages must thus, by and large, have been found appropriate and acceptable.

Finally, future terminographers would be well advised to extract terms from the materials that are actually used. Ideally, this could even be partly automated, as was shown in the present article for Swahili language/linguistics terminology. Automating this process may also facilitate the uncovering of new concepts, either freshly coined through language-internal processes, or borrowed from a language of wider diffusion.

ACKNOWLEDGMENTS

Heartfelt thanks are due to Jutta De Nul, who prepared the special-purpose corpus and did the draft tagging of the keywords.

REFERENCES

bab.la. 2017. Bilingual Arabic-English dictionary. https://en.bab.la/dictionary/english-arabic/ (accessed May 1, 2019).

Baker, Mona. 1992. *In Other Words: A Coursebook on Translation*. London: Routledge.

Baldi, Sergio. 2011. Swahili: A donor language. *Lingua Posnaniensis* 53(1): 7–24.

Baldi, Sergio. 2012. Arabic loans in East African languages through Swahili: A survey. *Folia Orientalia* 49: 37–52.

Benson, Thomas G. 1964. A century of Bantu lexicography. *African Language Studies* 5: 64–91.

de Schryver, Gilles-Maurice. 2006. Compiling modern bilingual dictionaries for Bantu languages: Case studies for Northern Sotho and Zulu. In *Atti del XII Congresso Internazionale di Lessicografia, Torino, 6-9 settembre 2006 / Proceedings XII Euralex International Congress, Torino, Italia, September 6th-9th, 2006*, edited by Elisa Corino, Carla Marello, and Christina Onesti, 515–25. Alessandria: Edizioni dell'Orso.

de Schryver, Gilles-Maurice. 2008. Why does Africa need Sinclair? *International Journal of Lexicography* 21(3): 267–91.

de Schryver, Gilles-Maurice. 2009. Lexikos at eighteen: An analysis. *Lexikos* 19: 372–403.

de Schryver, Gilles-Maurice, David Joffe, Pitta Joffe, and Sarah Hillewaert. 2006. Do dictionary users really look up frequent words? – On the overestimation of the value of corpus-based lexicography. *Lexikos* 16: 67–83.

de Schryver, Gilles-Maurice, and D.J. Prinsloo. 2001. Towards a sound lemmatisation strategy for the Bantu verb through the use of frequency-based tail slots – with special reference to Cilubà, Sepedi and Kiswahili. In *Makala ya Kongamano la Kimataifa Kiswahili 2000. Proceedings*, edited by James S. Mdee, and Hermas J.M. Mwansoko, 216–42, 372. Dar es Salaam: Taasisi ya Uchunguzi wa Kiswahili, Chuo Kikuu cha Dar es Salaam.

Devos, Maud, and Gilles-Maurice de Schryver. 2016. From usually going to epistemic possibility. Origin and development of an epistemic sentence adverb in Swahili. In *6th International Conference on Bantu Languages, Workshop on the expression of mood and modality in Bantu languages*, edited by Anon., 11. Helsinki: University of Helsinki.

Dunning, Ted. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics* 19(1): 61–74.

Encyclopædia Britannica. 2014. Swahili language. https://www.britannica.com/topic/Swahili-language (accessed May 1, 2019).

Fattakhova, Aida R., and Nailya G. Mingazova. 2015. Arabic loanwords in Tatar and Swahili: Morphological assimilation. *Journal of Sustainable Development* 8.4: 302–08.

Gabrielatos, Costas. 2018. Keyness analysis: Nature, metrics and techniques. In *Corpus Approaches to Discourse: A Critical Review*, edited by Charlotte Taylor, and Anna Marchi, 225–58. Oxon: Routledge.

Gauton, Rachélle, Elsabé Taljard, and Gilles-Maurice de Schryver. 2003. Towards strategies for translating terminology into all South African languages: A corpus-based approach. In *TAMA 2003 South Africa: Conference*

Proceedings, edited by Gilles-Maurice de Schryver, 81–88. Pretoria: (SF)² Press.

Gibbe, Abel G. 2008. Maendeleo ya istilahi za Kiswahili [Development of Swahili terminology]. In *Ukuzaji wa Istilahi za Kiswahili*, edited by John G. Kiango, 79–99. Dar es Salaam: Taasisi ya Uchunguzi wa Kiswahili, Chuo Kikuu cha Dar es Salaam.

Hardie, Andrew. 2014. Log Ratio – an informal introduction. http://cass.lancs.ac.uk/log-ratio-an-informal-introduction/ (accessed May 1, 2019).

Harvey, Andrew. 2014. Epenthetic vowels in Swahili loanwords. *Journal of Linguistics and Language in Education* 8(2): 17–45.

Haugen, Einar I. 1966. *Language Conflict and Language Planning: The Case of Modern Norwegian*. Cambridge, MA: Harvard University Press.

Hillewaert, Sarah, and Gilles-Maurice de Schryver. 2004. Online Kiswahili (Swahili) – English Dictionary. http://africanlanguages.com/swahili/ (accessed May 1, 2009).

IKR. 1996. *TUKI English-Swahili Dictionary / Kamusi ya Kiingereza-Kiswahili*. Dar es Salaam: Institute of Kiswahili Research, University of Dar es Salaam.

IKR. 2000. *TUKI English-Swahili Dictionary / Kamusi ya Kiingereza-Kiswahili. Second Edition* [2nd rev. edn.]. Dar es Salaam: Institute of Kiswahili Research, University of Dar es Salaam.

IKR. 2006. *TUKI English-Swahili Dictionary / Kamusi ya Kiingereza-Kiswahili. Third Edition* [3rd rev. edn]. Dar es Salaam: Institute of Kiswahili Research, University of Dar es Salaam.

Johnson, Frederick. 1939a. *A Standard English-Swahili Dictionary (Founded on Madan's English-Swahili Dictionary)*. Nairobi: Oxford University Press.

Johnson, Frederick. 1939b. *A Standard Swahili-English Dictionary (Founded on Madan's Swahili-English Dictionary)*. Nairobi: Oxford University Press.

Khamisi, Abdu M. 2008. *Maendeleo ya Uhusika [Relational Grammar]*. Dar es Salaam: Taasisi ya Uchunguzi wa Kiswahili, Chuo Kikuu cha Dar es Salaam.

Khamisi, Abdu M. 2009. *Stadi za Lugha ya Kiswahili: Sarufi I [Proficiency in Swahili: Grammar 1]*. Dar es Salaam: Taasisi ya Uchunguzi wa Kiswahili, Chuo Kikuu cha Dar es Salaam.

Kiango, John G. 2000. *Bantu Lexicography: A Critical Survey of the Principles and Process of Constructing Dictionary Entries*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

Kiango, John G., and James S. Mdee, eds. 1995. *Utafiti na Utungaji wa Kamusi [Research on the Making of Dictionaries]*. Dar es Salaam: Taasisi ya Uchunguzi wa Kiswahili, Chuo Kikuu cha Dar es Salaam.

Kihore, Yared M. 2009. *Stadi za Lugha ya Kiswahili: Sarufi II [Proficiency in Swahili: Grammar 2]*. Dar es Salaam: Taasisi ya Uchunguzi wa Kiswahili, Chuo Kikuu cha Dar es Salaam.

Kihore, Yared M., David P.B. Massamba, and Y.P. Msanjila. 2003. *Sarufi Maumbo ya Kiswahili Sanifu (SAMAKISA): Sekondari na Vyuo [Structural Grammar of Standard Swahili: Secondary and Tertiary]*. Dar es Salaam: Taasisi ya Uchunguzi wa Kiswahili, Chuo Kikuu cha Dar es Salaam.

Kilgarriff, Adam. 2003–2019. Sketch Engine. https://www.sketchengine.eu/ (accessed May 1, 2019).

Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, and Vít Suchomel. 2014. The Sketch Engine: Ten years on. *Lexicography: Journal of ASIALEX* 1(1): 7–36.

Kilgarriff, Adam, Pavel Rychlý, Pavel Smrz, and David Tugwell. 2004. The Sketch Engine. In *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6-10, 2004*, edited by Geoffrey Williams, and Sandra Vessier, 105–15. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.

King'ei, Kitula G. 2010. *Misingi ya Isimujamii [Basics of Sociolinguistics]*. Dar es Salaam: Taasisi ya Taaluma za Kiswahili, Chuo Kikuu cha Dar es Salaam.

Krapf, Johann L. 1882. *A Dictionary of the Suahili Language, with Introduction Containing an Outline of a Suahili Grammar*. London: Trübner and Co., Ludgate Hill.

Madan, Arthur C. 1894. *English-Swahili Dictionary, Compiled for the Use of the Universities' Mission to Central Africa*. Oxford / London: Clarendon Press / Society for Promoting Christian Knowledge.

Madan, Arthur C. 1902. *English-Swahili Dictionary* [2nd rev. edn.]. Oxford: Clarendon Press.

Madan, Arthur C. 1903. *Swahili-English Dictionary*. Oxford: Clarendon Press.

Massamba, David P.B. 1990. *Kamusi Sanifu ya Isimu na Lugha [Standard Dictionary of Linguistics and Language]*. Dar es Salaam: Taasisi ya Uchunguzi wa Kiswahili, Chuo Kikuu cha Dar es Salaam.

Massamba, David P.B. 2004. *Kamusi ya Isimu na Falsafa ya Lugha [Dictionary of Linguistics and the Philosophy of Language]*. Dar es Salaam: Taasisi ya Uchunguzi wa Kiswahili, Chuo Kikuu cha Dar es Salaam.

Massamba, David P.B., Yared M. Kihore, and J.I. Hokororo. 1999. *Sarufi Miundo ya Kiswahili Sanifu (SAMIKISA): Sekondari na Vyuo [Syntax of Standard Swahili: Secondary and Tertiary]*. Dar es Salaam: Taasisi ya Uchunguzi wa Kiswahili, Chuo Kikuu cha Dar es Salaam.

Massamba, David P.B., Yared M. Kihore, and Y.P. Msanjila. 2004. *Fonolojia ya Kiswahili Sanifu (FOKISA): Secondari na Vyuo [Phonology of Standard Swahili: Secondary and Tertiary]*. Dar es Salaam: Taasisi ya Uchunguzi wa Kiswahili, Chuo Kikuu cha Dar es Salaam.

Matei, Assumpta K. 2008. *Darubini ya Sarufi: Ufafanuzi Kamili wa Sarufi ya Kiswahili. Maswali na Mazoezi [The Binoculars of Grammar: The Exact Reading of Swahili Grammar. Questions and Exercises]*. Nairobi: Phoenix Publishers.

Mdee, James S. 1990. *Theories and Methods of Lexicography in the Standard Swahili Dictionary*. unp. PhD diss. Karl Marx University, Leipzig.

Mohamed, Amir A. 2009. *Kiswahili for Foreigners* [3rd rev. edn.]. Zanzibar: Goodluck Publishers.

Mtintsilana, Priscilla N., and Rose Morris. 1988. Terminography in African languages in South Africa. *South African Journal of African Languages* 8(4): 109–13.

Mwansoko, Hermas J.M. 1990. *The Modernization of Swahili Technical Terminologies: An Investigation of the Linguistics and Literature Terminologies*. unpub. PhD diss. University of York, York.

Mwansoko, Hermas J.M. 2001. Uboreshaji wa mfumo wa uingizaji wa istilahi za kimataifa katika Kiswahili [Improvement of the system for inserting international terminology into Swahili]. In *Makala ya Kongamano la Kimataifa Kiswahili 2000. Proceedings*, edited by James S. Mdee, and Hermas J.M. Mwansoko, 318–31. Dar es Salaam: Taasisi ya Uchunguzi wa Kiswahili, Chuo Kikuu cha Dar es Salaam.

Mwansoko, Hermas J.M., and Albina R. Chuwa, eds. 1995. *Tahakiki na Uchapishaji wa Kamusi [Analysing and Publishing Dictionaries]*. Dar es Salaam: Taasisi ya Uchunguzi wa Kiswahili, Chuo Kikuu cha Dar es Salaam.

Mwita, Leonard C. 2009. The adaptation of Swahili loanwords from Arabic: A constraint-based analysis. *The Journal of Pan African Studies* 2(8): 46–61.

Nyangwine, Nyambari C.M., and J.A. Masebo. 2002. *Kiswahili: Kidato cha Tatu na Nne. Maendeleo ya Kiswahili, Fasihi, Sarufi, Matumizi, Utungaji na Ufahamu [Swahili: Secondary 3 & 4. Development of Swahili, Literature, Grammar, Use, Formation and Knowledge]*. Dar es Salaam: Nyambari Nyangwine Publishers.

OGL. 2017. Oxford Global Languages – Swahili. https://sw.oxforddictionaries.com/ (accessed May 1, 2009).

Ohly, Rajmund. 1978. *Language + Revolution = KISWAHILI*. Inaugural Lecture Series 15. Dar es Salaam: Dar es Salaam University Press.

Sacleux, Charles. 1891. *Dictionnaire français-swahili*. Zanzibar: Mission des P.P. du St. Esprit.

Sacleux, Charles. 1939. *Dictionnaire swahili-français*. Travaux et mémoires de l'Institut d'ethnologie 36–37. Paris: Institut d'ethnologie.

Scott, Mike. 1996–2015. WordSmith Tools, version 6. https://www.lexically.net/wordsmith/ (accessed May 1, 2019).

Scott, Mike. 1996–2019. WordSmith Tools, version 7. http://www.lexically.net/wordsmith/ (accessed May 1, 2019).

Taljard, Elsabé, and Gilles-Maurice de Schryver. 2002. Semi-automatic term extraction for the African languages, with special reference to Northern Sotho. *Lexikos* 12: 44–74.

Tucker, Archibald N. 1946. Foreign sounds in Swahili. *Bulletin of the School of Oriental and African Studies* 11(4): 854–71.

TUKI. 1981. *Kamusi ya Kiswahili Sanifu [Dictionary of Standard Swahili]*. Dar es Salaam: Oxford University Press.

TUKI. 2001. *Kamusi ya Kiswahili-Kiingereza / Swahili-English Dictionary*. Dar es Salaam: Taasisi ya Uchunguzi wa Kiswahili, Chuo Kikuu cha Dar es Salaam.

TUKI. 2004. *Kamusi ya Kiswahili Sanifu. Toleo la Pili [Dictionary of Standard Swahili. Second Edition]* [2nd rev. edn]. Nairobi: Oxford University Press.

TUKI. 2014. *Kamusi ya Kiswahili-Kiingereza. Toleo la Pili / Swahili-English Dictionary* [2nd rev. edn.]. Dar es Salaam: Taasisi ya Uchunguzi wa Kiswahili, Chuo Kikuu cha Dar es Salaam.

Tumbo-Masabo, Zubeida N.Z. 1990. *The Development of Neologisms in Kiswahili: A Diachronic and Synchronic Approach with Special Reference to Mathematical Terms*. unpub. PhD diss. Teachers College, Columbia University, New York, NY.

Tumbo-Masabo, Zubeida N.Z. 1992. Uundaji wa istilahi za Kiswahili [Formation of Swahili terminology]. In *Kiongozi cha Uundaji wa Istilahi za Kiswahili [Guidelines for the Formation of Swahili Terminology]*, edited by Zubeida N.Z. Tumbo-Masabo, and Hermas J.M. Mwansoko, 21–42. Dar es Salaam: Taasisi ya Uchunguzi wa Kiswahili, Chuo Kikuu cha Dar es Salaam.

Tumbo-Masabo, Zubeida N.Z., and Hermas J.M. Mwansoko, eds. 1992. *Kiongozi cha Uundaji wa Istilahi za Kiswahili [Guidelines for the Formation of Swahili*

Terminology]. Dar es Salaam: Taasisi ya Uchunguzi wa Kiswahili, Chuo Kikuu cha Dar es Salaam.

Velten, Carl. 1910. *Suaheli-Wörterbuch, I Teil, Suaheli-Deutsch*. Berlin: Selbstverlag d. Verf.

Velten, Carl. 1933. *Suaheli-Wörterbuch, II Teil, Deutsch-Suaheli*. Berlin: Selbstverlag d. Verf.

Waihiga, Gichohi. 1999. *Sarufi Fafanuzi ya Kiswahili [Explanatory Grammar of Swahili]*. Nairobi: Longhorn Publishers.

Zgusta, Ladislav. 1971. *Manual of Lexicography*. Janua Linguarum Series Maior 39. Prague / The Hague: Academia / Mouton.