

KIU (Kiswahili–Italiano–UniOr): The UniOr online Dictionary for Italian L1 Swahili Learners

Flavia Aiello, *Department of Asian, African and Mediterranean Studies,
University of Naples 'L'Orientale', Italy (faiello@unior.it)*

and

Rosanna Tramutoli, *Department of Asian, African and Mediterranean
Studies, University of Naples 'L'Orientale', Italy
(rosanna.tramutoli@gmail.com)*

Abstract: KIU is an online bilingual Swahili–Italian dictionary with about 6000 entries aimed primarily at Italian L1 Swahili learners and which has been developed at the University of Naples 'L'Orientale'.

The project was started in 2003 by M. Toscano and developed with the collaboration of language experts and young researchers until 2009 with the aim of offering online lexical resources from Swahili to Italian for learners of the language. After a long interruption, the work was resumed in 2019 by the authors of this article in cooperation with M. Toscano and a team of expert IT technicians. The current work consists of the development ex novo of the dictionary software, which had become obsolete, along with a redesign of some lexicographic features.

In this report we will show how the upgraded version of the dictionary software has been implemented, with relevant learner-oriented features, by taking into consideration the standard lexicographic characteristics of Swahili–Italian bilingual dictionaries. This dictionary represents a valuable support for L2 learners and is the only on-line Swahili–Italian dictionary expressly built for university students and Italian users at large.

Keywords: SWAHILI, ITALIAN, ONLINE DICTIONARY, LEARNERS, LEXICOGRAPHY

Opsomming: KIU (Kiswahili–Italiano–UniOr): Die UniOr aanlyn woordeboek vir Italiaanse L1-Swahili-aanleerders. KIU is 'n aanlyn tweetalige Swahili–Italiaanse woordeboek met ongeveer 6000 inskrywings wat hoofsaaklik gerig is op Italiaanse L1-Swahili-aanleerders en wat deur die Universiteit van Napels 'L'Orientale' ontwikkel is.

Die projek is in 2003 deur M. Toscano van stapel gestuur en is met die samewerking van taalkundiges en jong navorsers tot 2009 ontwikkel met die doel om aanlyn leksikale hulpbronne uit Swahili na Italiaans vir aanleerders van die taal aan te bied. Ná 'n lang onderbreking is die werk in 2019 deur die outeurs van hierdie artikel in samewerking met M. Toscano en 'n span kundige IT-tegnici hervat. Die werk bestaan tans uit die ex novo-ontwikkeling van die woordeboeksagteware wat verouderd geraak het asook uit 'n herontwerp van sommige leksikografiese kenmerke.

In hierdie verslag sal ons aantoon hoe die opgegradeerde weergawe van die woordeboeksagteware, met inagneming van die standaard leksikografiese eienskappe van Swahili–Italiaanse tweetalige woordeboeke, geïmplementeer is met relevante aanleerder-georiënteerde kenmerke. Die

woordeboek verleen waardevolle steun aan L2-aanleerders en is die enigste aanlyn Swahili-Italiaanse woordeboek wat spesifiek vir universiteitstudente en vir Italiaanse gebruikers oor die algemeen geskep is.

Sleutelwoorde: SWAHILI, ITALIAANS, AANLYN WOORDEBOEK, AANLEERDERS, LEKSIKOGRAFIE

1. Overview of the existing lexicographic resources for Italian L1 Swahili learners

Swahili lexicography, including bilingual dictionaries, began in the last decades of the 19th century, and during colonial times was practised by foreigners who came to East Africa either as missionaries or colonial administrators (Chuwa 1996: 324).¹

After independence was gained by East African states, Swahili lexicography continued to grow in a context of development of Swahili in a wide geocultural area and in multiple contexts of language use. In Tanzania this was especially so. There, the language became the dominant medium of communication in the public sector (basic education, administration, courts, national assembly, media etc.) thanks to the commitment of experts from the National Swahili Council (Baraza la Kiswahili la Taifa, BAKITA) and the Institute of Swahili at the University of Dar es Salaam (Taasisi ya Uchunguzi wa Kiswahili², TUKI). These institutions, besides producing general dictionaries, monolingual and bilingual (SWA-ENG and ENG-SWA), created and disseminated terminologies for many domains of language use, such as law, science, and IT (Sewangi 2007; Aiello and Toscano 2017). Bilingual dictionaries were also produced over the years in countries where Swahili is taught at university level (such as France, Germany, Russia, etc.).³ Throughout the history of Swahili lexicography, most new dictionaries were based on their predecessors. From the 2000s, however, innovative corpus-based dictionaries also started to appear (Wójtowicz 2016: 402), such as the Swahili-Finnish dictionary (Abdulla et al. 2002) and the Swahili-Polish dictionary by Wójtowicz (2003), which is available also online (see Wójtowicz and Bański 2012 for more details).

Since the 1990s, indeed, several bilingual electronic dictionaries for Swahili L2 learners have been developed, the majority of them targeting English speakers. A community-based SWA-ENG online dictionary was started in 1994 at Yale University under the name of Internet Living Swahili Dictionary, later run by the association Kamusi Project International with the aim of incorporating other languages (www.kamusi.org). The SWA-ENG and ENG-SWA dictionaries produced by TUKI in print form are also accessible online as basic databases listing the entries by alphabetical order via the links <https://swahili-dictionary.com/swahili-english> and <https://swahili-dictionary.com/english-swahili> respectively. The TshwaneDJe Swahili-English Dictionary (<https://tshwanedje.com/dictionary/swahili/>) is a corpus-driven dictionary of Swahili, downloadable upon purchase. It is based on web-based corpus data, with over 16,000 entries

and phrases and over 36,000 translation equivalents. Presently, the most popular SWA–ENG / ENG–SWA dictionary available for free online is the *Glosbe — Swahili–English dictionary* (<https://it.glosbe.com/sw/eng>) which is part of the Glosbe multilingual dictionary (6000 languages) created by a Polish team and which offers translations created by users or automatically generated on the basis of a database of translated examples.

All these numerous lexicographical works, both printed and online, can be used by the students of Swahili L2 of the University of Naples 'L'Orientale' (henceforth UniOr), who can find many bilingual dictionaries in our library and have free Wi-Fi access to the Internet. Many Italian students, though, find it difficult to use a bilingual dictionary which is not in their native language for their exercises (drills, comprehension tests, translations, oral production, etc.), especially during the first years of their studies when they do not have a perfect mastery of English or other foreign languages. Also, advanced students of Swahili have a great need of resources in Italian as the translation of literary texts and other specialist writings (for instance essays or newspaper articles on politics) is one of the focusses of the teaching of Swahili at UniOr. This is the way it has been since the teaching was established in 1969 by E. Bertoncini Zúbková, an internationally renowned scholar of Swahili language and literature, whose educational activity was closely connected with research on the Swahili lexicon, resulting in the production of lists of words and vocabularies. This huge work has remained largely unpublished, apart from a small vocabulary (*Vocabolario swahili–italiano e italiano–swahili*, Opera Universitaria I.U.O., Naples 1977), which is now out of print.

Concerning bilingual lexicographical works aimed at Italian learners, three dictionaries have been published so far. These are Vittorio Merlo Pick's *Vocabolario kiswahili–italiano e italiano–kiswahili* (EMI, Turin 1961, re-edited in 1978, currently out of print), Maddalena Toscano's pocket-size *Dizionario swahili. Swahili–italiano, italiano–swahili* (Vallardi, Milan 2004) and Gianluigi Martini's *Dizionario swahili. Swahili–italiano, italiano–swahili* (Hoepli, Milan 2016). Furthermore, a terminological work, namely a Swahili–Italian linguistic glossary by Rosanna Tramutoli (*Kamusi ya isimu Kiswahili–Kiitaliano*, TUKI, Dar es Salaam 2018), was published by the University Press of the Institute of Swahili at the University of Dar es Salaam within the sphere of a longstanding cooperation agreement with UniOr. All these works for Italian speakers are, however, in print format and only partly available in the library of UniOr. With regard to online resources for Italian learners of Swahili, the following dictionaries are those available on the web:

(1) *Glosbe — Dizionario Swahili–italiano* (<https://it.glosbe.com/sw/it>), also IT–SWA, as part of the above-mentioned Glosbe multilingual dictionary, which contains many entries, but is not always reliable. It presents a number of incorrect results, such as in fig. 1, where the second translation of the word *utumishi*, 'battuta' (setback) is completely wrong, as well as the Italian translation of the example.⁴

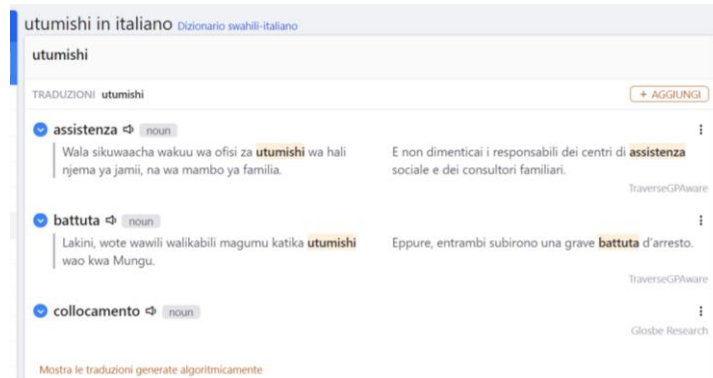


Fig. 1: Incorrect translation in *Glosbe* — *Dizionario Swahili-italiano*

In some instances, when the translation is missing, *Glosbe* proposes some 'hypotheses' by using an algorithm and the user is warned to be cautious. This is the case, for instance, when we look for grammatical forms such as possessives, demonstratives, etc. with their class concord as in fig. 2. Instead of a direct translation of *yako* (second-person possessive *-ako* + class 9 [singular] verbo-pronominal concord *y-*), that is 'tuo, tua' (your, second-person singular possessive referring to a masculine or feminine noun), there is a list of misleading results as the Italian translations are derived from the English ones, and 'your' can be translated into Italian as 'tuo, tua' but also as 'tue, tuoi' (your, second-person singular possessive referring to a masculine or feminine noun in the plural form) and 'vostro, vostra, vostre, vostri' (your, second-person plural possessive referring to a masculine or feminine noun in the singular or plural form).



Fig. 2: Grammatical forms in *Glosbe* — *Dizionario Swahili-italiano*

(2) *Online Swahili Italian Dictionary* (<http://www.etranslator.ro/swahili-italian-online-dictionary.php>) (also IT-SW), which provides automatic translations through English and seems to be even less reliable, as we see in fig. 3 where *yako* (see the explanation above) has been wrongly translated into Italian as 'il vostro' (your/yours, second-person plural possessive referring to a masculine noun in the singular form).

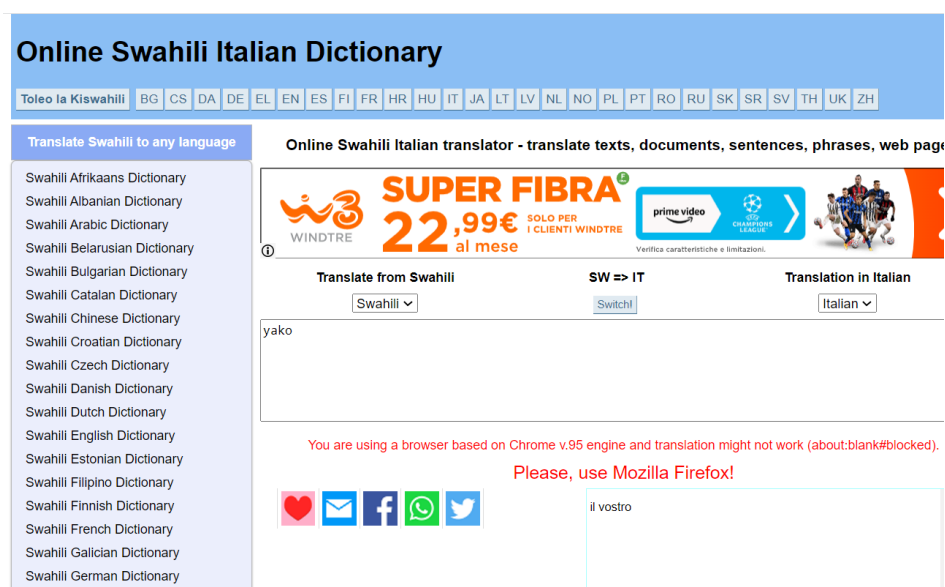
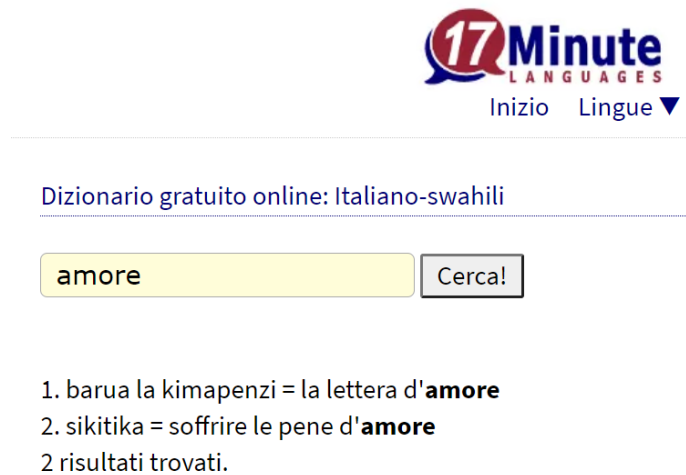


Fig. 3: *Online Swahili Italian Dictionary*

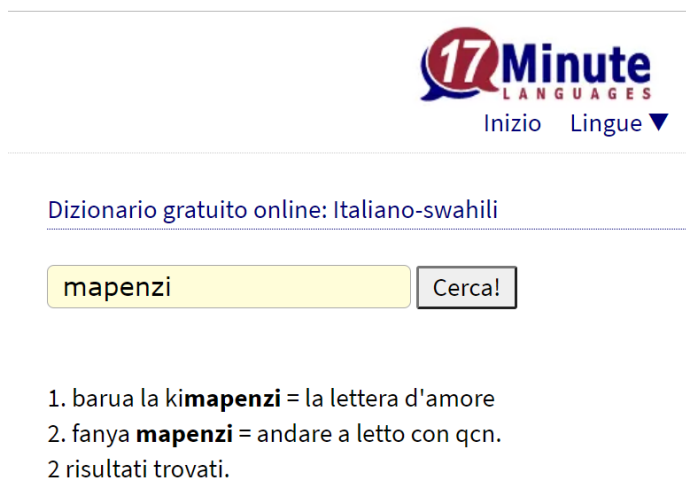
(3) *17minute-languages* — *Dizionario gratuito online: Italiano–swahili* (<https://www.17-minute-languages.com/it/dizionario-di-swahili/>), which allows the search for SWA–IT and IT–SWA and works as an automatic translator offering a list of the sentences containing (in the Italian translation) the searched word or form. The results are sometimes incorrect or lacking the main, basic translation as, for instance, in fig. 4, showing the result for the word 'amore' (love). Here we don't find the main Swahili translations for 'amore' (*mapenzi, mahaba, upendo*) but two examples: (1) *barua la (sic!) kimapenzi* = la lettera d'amore (love letter), where the Swahili word *mapenzi* is used in a particular way, namely with a ki- prefix with an attributive function; (2) *sikitika* = soffrire le pene d'amore (to suffer the pains of love). Actually, the Swahili verb *sikitika* means 'to be sorry, to be sad', which clearly was translated into Italian as 'soffrire le pene d'amore' in a specific context.



The screenshot shows the 17 Minute Languages logo at the top right, with the text 'Inizio Lingue' and a dropdown arrow. Below the logo, the text 'Dizionario gratuito online: Italiano-swahili' is displayed. A search bar contains the word 'amore', and a 'Cerca!' button is to its right. Below the search bar, two results are listed: '1. barua la kimapenzi = la lettera d'**amore**' and '2. sikitika = soffrire le pene d'**amore**'. At the bottom, it says '2 risultati trovati.'

Fig. 4: *17minute-languages* — *Dizionario gratuito online Italiano–swahili*

Similarly, when looking for the Italian translation of *mapenzi* (see fig. 5), the first translation given in the Swahili–Italian printed dictionaries, 'amore', is not given, instead we see two examples: (1) *barua la kimapenzi*, the same expression as commented on above, which, furthermore, is grammatically incorrect, as it presents *barua* as a noun belonging to class 5/6, whereas it belongs to class 9/10; 2) *fanya mapenzi* (to make love).



The screenshot shows the 17 Minute Languages logo at the top right, with the text 'Inizio Lingue' and a dropdown arrow. Below the logo, the text 'Dizionario gratuito online: Italiano-swahili' is displayed. A search bar contains the word 'mapenzi', and a 'Cerca!' button is to its right. Below the search bar, two results are listed: '1. barua la kimapenzi = la lettera d'amore' and '2. fanya mapenzi = andare a letto con qcn.'. At the bottom, it says '2 risultati trovati.'

Fig. 5: *Incorrect translation of mapenzi in Dizionario gratuito online*

(4) *Vocabolario Italiano–Swahili*, on the webpage 'Changamano Onlus/Karibu!' by Nino Vessella (<https://swahili.it/glossword/index.php?a=index&d=1>; latest update in 2008). This dictionary is a useful resource, but it only allows the search IT–SWA. The entries are mainly based on the above-mentioned Vittorio Merlo Pick's *Vocabolario kiswahili–italiano e italiano–kiswahili*, for instance, the following entry 'amore' is very similar in the two dictionaries (see fig. 6 and 7):



Fig. 6: 'amore' in *Vocabolario Italiano–Swahili*

amore *m.* upendo [pendo (ma)], mapendo, upenzi (ma), huba (ma), mahaba; (*passione*) shauku (-), uchumba, mapenzi; (*vicendevo-*le) upendano (ma); (*persona amata*) mpenzi, muhebi (-); (*proprio*) choyo, kujipenda, majisifu. *Per a. di Dio*, kwa ajili ya Mungu. *Lavorare con a.*, -fanya kazi kwa bidii. *Far l'a.*, -chunuka, -tumuka. *Per a. o per forza*, kwa vyo vyote.

Fig. 7: 'amore' in Vittorio Merlo Pick's IT–SWA dictionary

Here we find also some additions related to new lexicon, for instance AIDS (in Swahili UKIMWI, the acronym of Upungufu wa Kinga Mwilini, literally 'deficit of protection in the body'), though other contemporary vocabulary is missing, like words related to IT (e-mail, Internet, web, etc.).

From this overview, it appears that online lexicographical resources for Italian learners of Swahili are not wholly satisfying, especially with regard to the direction SWA-IT. This reality inspired the idea of developing a lexicographical tool accessible through the Internet in order to help our students in their exercises and translations when they are far from the university. This need for online resources also emerged vividly with the outbreak of the Covid-19 pandemic, since the UniOr libraries were closed, and all the teaching activities were based on digital platforms.

2. KIU: the updated Swahili-Italian online dictionary for learners

The development of an online lexicographical resource for Italian learners (UWAZO) was started at the beginning of the 2000s by M. Toscano in a context of renewal of the courses of Swahili at UniOr and was encouraged by the introduction of new digital technologies, which have revolutionised the teaching of foreign languages. The Swahili-Italian online dictionary UWAZO was based on T.E.I. guidelines⁵ and was developed at UniOr between 2003 and 2009.

The old software used for UWAZO had become obsolete and so in order to meet the increasing need of updated digital learning resources for Swahili students, especially beginners, updated lexicographic software, namely KIU (Kiswahili-Italiano UniOr), has been designed. This work was done from 2020 to 2022 by a team of experts on the basis of the lexicographical indications received by the authors of the present article, in collaboration with M. Toscano. The dictionary project aimed to create updated software that would be developed in accordance with recent lexicographic practices.

The Swahili-Italian dictionary is designed primarily as a didactic tool for Swahili students, but at the same time it is suitable for a wider audience, including people working for Italian NGOs, staff of cultural associations and institutions, embassy staff, tourists, businessmen and anyone interested in Swahili language and culture. While there are very few lexical resources available for Italian L1 Swahili students, this online dictionary will serve as a language-learning support and supplement other teaching materials and paper-dictionaries used in class. The dictionary KIU will be published online on a dedicated website linked to the UniOr website,⁶ and will be accessible for free to university students and the general public.

The website of the lexicographic project, apart from being a dictionary interface, contains an introduction, information on the language, a dictionary user guide, a grammatical sketch, and information on how to quote the lexicographic work. The interface of KIU will be accessible through a general webpage which will include a description of the lexicographic research project, academic publications of the lexicographic team, information about the software, and all contacts and credits of the subjects and institutions involved in the project.⁷

Unlike printed dictionaries, it will be possible to search KIU without limitations, and it is easy to maintain and expand. The database contains about 6,000 headwords. Most of them were selected from a Swahili frequency list (Bertoncini Zúbková 1973) and from various other sources, e.g. Merlo-Pick 1961 and Toscano 2004 (see par. 1 about other Swahili–Italian dictionaries). The sources also include the lexicon used in the teaching materials of Swahili courses by prof. E. Bertoncini Zúbková. Moreover, beside complete sets of inflected forms, closed sets of words such as days of the week, months, and general lexicon, the dictionary also contains some specific vocabulary collected by students and researchers who worked on chosen sets they found useful in their studies like body parts or immigration.

Following T.E.I. guidelines for printed dictionaries, the KIU database structure is based on various groups which include a fixed list of elements with free position. Sub-class elements are also available. The main T.E.I. groups used in KIU include: *Gruppo grammaticale* (Grammatical Group), *Traduzione* (Translation), *Esempio* (Example), *Esempio Tradotto* (Translated Example), *Confronta* (Cross-reference), *DictScrap* (additional notes regarding grammatical indications or specific usage). Also, open lists of labels, to be set by the operator, are possible.

It is freely accessible online by users and supports two different levels of access depending on the role. Administrators have the highest level of access to the database and can implement and edit data, and set up and manage the accounts of students and/or collaborators. Students and learners will have full access to database and software tools except for data publishing which needs the approval of the administrators.

Moreover, in addition to the Swahili–Italian dictionary, the new software has been designed to be extended to other Bantu languages, in particular, a Zulu–Italian lexical database (including a collection of body vocabulary) is under development by R. Tramutoli and will be accessible through the general webpage of the lexicographic project.

The following sections briefly outline the microstructure and macrostructure of the dictionary as well as the software tools used to implement learner-oriented features.

3. Dictionary entries in KIU

Bantu languages are characterized by a complex noun class system, each class being expressed by a noun prefix; classes are generally paired as singular and plural, e.g. cl. 1 is singular, + human; cl. 2 is plural, + humans. Apart from classes 12 and 13 (not productive anymore in the language) and classes 11–14 which have merged into one class, the Swahili noun class system includes 15 noun classes. Swahili words are composed in a main, basic order, i.e. *Morpheme(s)+Stem*⁸:

- nouns: class prefix + noun stem (noun classes: cl. 1/2, 3/4, 5/6, 7/8, 9/10, 11–14/10; verbal nouns: class 15; 'locative' classes: 16, 17, 18)⁹ (see Table 1);
- adjectives: adjectival concords (cl. 1...18) + adj. stem (nominal prefixes and adjectival concords are usually identical, except for the concord *my-* class 4 which occurs with adjectives beginning with the vowels e- or o-, and for class 11–14 which follows the same adjectival agreement of nouns in class 3);
- pronominals: pronominal concords (cl. 1...18) + stem; stem + pr. conc. (1...18) (see Table 2);
- verbs: subject class concords (cl. 1...18) + tense marker + (relative conc. cl.1...18) + (object concords cl. 1...18) + verbal stem + final vowels/ extensions (see Table 3).

Table 1: Noun class prefixes

cl.	sing.	pl.
1/2	m-, mu-, w- ¹⁰	wa-, w-
3/4	m-, mu-, mw-	mi-
5/6	Ø, ji-, j-	ma-, m-
7/8	ki-, ch-	vi-, vy-
9/10	Ø, ¹¹ n-, ny-, m-, etc.	n-, ny-, m-, etc.
11/10	u-, w-	n-, ny-, m-, etc.
12/13 ¹²	ka-	tu-
15	ku-, kw-	

Table 2: Pronominal Concords (possessives and demonstratives)

Class	Possessive Concord	Demonstrative Concord
1/2	w-/w-	yu-/wa-
3/4	w-/y-	u-/i-
5/6	l-/y-	li-/ya-
7/8	ch-/vy-	ki-/vi-
9/10	y-/z-	i-/zi- y-/z-
11/10	w-/z-	u-/zi-
15	kw-	ku-
16	p-	pa-
17	kw-	ku-
18	mw-	mu-

Table 3: Subject Concord (SC), Object Concord (OC), Relative Concord (RC)

Class	SC	OC	RC
cl. 1/2 (1pers.)	ni-/tu-	-ni-/ -tu-	-ye-/ -o-
cl. 1/2 (2pers.)	u-/m-	-ku-/ -wa- -eni	-ye-/ -o-
cl. 1/2 (3pers.)	a-/wa-	-m-/ -wa-	-ye-/ -o-
cl. 3/4	u-/i-	-u-/ -i-	-o-/ -yo-
cl. 5/6	li-/ya-	-li-/ -ya	-lo-/ -yo-
cl. 7/8	ki-/vi-	-ki-/ -vi	-cho-/ -vyo-
cl. 9/10	i-/zi-	-i-/ -zi-	-yo-/ -zo-
cl. 11/10	u-/zi-	-u-/ -zi-	-o-/ -zo-
cl. 15	ku-	-ku-	-ko-
cl. 16	pa-	-po-	-po-
cl. 17	ku-	-ko-	-ko-
cl. 18	mu-	-mo-	-mo-

Swahili is an agglutinative language, which means that morphemes are juxtaposed to form words. Within the Swahili lexicographic tradition, the accepted lemmatisation strategy is to list nouns in their singular full forms with class prefixes, whereas the concords of verbs, numerals, pronouns and inflected adjectives are ignored, and the stems alone are listed (Wójtowicz 2016: 410). Thus, in standard printed dictionaries, we find the following types of Swahili entries:

- nouns: singular form (moyo→mioyo; moyoni, mioyoni¹³);
- adjectives: stem (-baya→mbaya, wabaya, mibaya, baya, mabaya, kibaya, vibaya, kubaya, pabaya, mubaya);
- pronouns: stem (h-→huyu, hawa, huu, hii, hili, haya, hiki, hivi, hizi, huku, hapa, humu; huyo, hao, etc.);
- verbs: stem (-soma→somea, -someka, -somesha, -somwa, -somana, etc.);
- invariable forms (some prepositions, conjunctions, adverbials, etc.).

The problem of how to compile dictionaries for elementary learners is often presented in Bantu lexicography as a debate between 'stem lemmatization' (often considered as the 'right size') vs. 'word lemmatization' (typically the 'wrong size',

or at least looked down upon, especially by linguists) (De Schryver 2010: 162; see also Prinsloo 2009: 152-153).

The debate on the proper arrangement of the Bantu lexicon has been clearly summarized by Bennett (1986: 3-4), who highlights how the complex derivational system of Bantu languages poses difficulties:

If items are alphabetized by prefix [...] a verb will be listed far from its nominal derivations, however transparent these may be. [...] A competing school arranges the lexicon by stem or root; this usefully groups related items and saves on cross-referencing. Unfortunately, in such a system the user must be able to identify the stem, which given the sometimes complex morphophonemics of Bantu languages may not be easy.

Considering that Swahili students are used to the lexicographic 'stem tradition', in the online dictionary KIU we have decided to choose solutions found in nearly all Swahili dictionaries with regard to the process of lemmatization. Therefore, we have listed the stems alone for verbs, numerals, and inflected adjectives, ignoring agreement concords. However, it has been argued that the so-called 'stem tradition' in dictionary-making of Bantu languages is inadequate for young learners, who fail to isolate stems (see De Schryver 2010). In some cases, therefore, we have decided to reject traditional lexicographic solutions usually adopted in printed dictionaries and have responded to beginner learners' needs by opting for the lemmatization of full words for closed grammatical sets such as pronouns (including the stems also as separate entries).

Thus, each dictionary entry includes the following types of information:

- Headword
- [Variant/variants of the headword]
- [Full plural of nouns]
- PoS (part of speech), that is the grammatical category
- Noun class
- Translation (gloss and/or description)
- Examples (quotations, collocations, idioms, proverbs, etc.)
- Translations of examples
- [Cross-reference]


KIU (dizionario online Kiswahili-Italiano UniOr)

Ricerca veloce di un lemma

In questa sezione è possibile cercare termini italiani che si trovano nella traduzione dalla lingua target. La ricerca produce la lista delle voci che contengono il termine cercato. È una ricerca fruibile da chi ha una conoscenza della lingua di livello almeno iniziale-intermedio.

Risultati per: moyo

moyo

1. **moyo** *n. cl.3/4 (m-/mi-)* cuore; animo; coraggio 

moyo, (pl. **mioyo**) (*n. cl.3/4 (m-/mi-)*) cuore; animo; coraggio **kwa moyo**, volentieri; a memoria **-jipa moyo/ -(ji)piga moyo konde**, farsi coraggio **-fa moyo**, essere scoraggiato
Cfr. **fuadi**

Fig. 8: Screenshot of the entry search 'moyo' in KIU

All entries contain an indication of the grammatical category (PoS) and at least one simple gloss. Most of the items are divided into sub-items and completed with examples of use. The inflected grammatical forms in particular contain various examples of use referring to the different meanings and functions. Most of the examples were taken from the UniOr digital corpus of Swahili texts, presently available only offline, which has been collected by M. Toscano in collaboration with a number of students and researchers. It is a small Swahili untagged raw corpus (1 million words) made mainly of about fifty full texts, mostly contemporary written literature, with the addition of some oral narratives and non-literary works (socio-political essays, handbooks about agriculture, media studies, informatics). The corpus is at the disposal of researchers and MA students for their thesis research on Swahili language, literature and linguistics.

Each entry is also categorized according to the type:

- form (nouns or invariable entries);
- grammatical stem (verbs, pronouns, adjectives, variable entries in general);
- morpheme: e.g. noun prefixes, subject prefixes, object markers, derivational suffixes, etc.;
- compounds.

Given Swahili word structure and the elements of the dictionary entries, it is evident that, since we are working with a Bantu language, we have to address problems not experienced by lexicographers working with European languages.

These problems are connected primarily to two issues: the form of headwords and the presentation of the numerous derivatives of a single root (Wójtowicz 2016: 410). In the following sections, we will explore challenges and difficulties regarding the design of a new Swahili online dictionary as a learning/teaching language tool.

4. How to transfer specific grammatical knowledge into an online dictionary

KIU is a learner dictionary, thus, differently from standard dictionaries, it has among its scopes, the aim of supporting Swahili students in autonomous language learning. It does this not only by building an up-to-date and comprehensive lexicon, but also by guiding students through a search process based on a system of interrelated grammar skills. In the construction of the online Swahili-Italian dictionary we faced several challenges which highlight the difficult process of transferring all grammatical knowledge and linguistic descriptions useful for Swahili students at beginner level into a lexicographic database. Without a doubt an online dictionary is much more comprehensive and allows for more types of search for entries compared to a printed dictionary, thus it represents a valuable tool for Swahili learners at a beginner level. Nevertheless, before presenting the new features of the dictionary (see par. 5), we should elucidate the limitations and difficulties experienced in producing a Swahili dictionary for learners.

The Swahili noun class system is quite standardized and homogeneous. Each noun class is expressed by prefixes which mark all elements of a Swahili sentence and thus encode the grammatical information necessary for grammatical agreement. Nevertheless, in some cases, apart from knowledge of class prefixes, other semantic and grammatical skills, which can hardly be included in entries in standard printed dictionaries, are required in order to build a correct Swahili sentence. For instance, Swahili students at beginner level face huge difficulties while applying the rule of animate noun agreement, that is, those nouns whose meaning refers to animate entities generally follow the same grammatical agreement as these latter, even if not belonging to class 1/2. Moreover, this rule does not apply to possessives with nouns in class 9/10 indicating human beings, which, only in this specific case, follow the possessive agreement of nouns in class 9/10:

<i>*rafiki wangu</i>	class 1 agreement
<i>rafiki yangu/rafiki zangu</i>	class 9 agreement

Apart from lexical entries, the dictionary headword list also includes grammatical morphemes, such as noun class prefixes, and, since it is not possible to translate the morphemes into Italian, their function in Swahili is explained. However, class 9/10 and class 5, include a huge number of loanwords, characterized by a zero prefix, thus, in this case, no entry can be inserted in the data-

base since no morpheme corresponds to the noun prefix. Similarly, in class 5, we generally find a zero prefix before deverbatives, e.g. *ombi* 'prayer', from the verb *kuomba* 'to pray'; the class 5 prefix *ji-* is found only before most of the monosyllabic roots, e.g. *jiwe* 'stone'.

In addition to the loanwords, class 9/10 contains invariable nouns, historically characterized by the nasal prefix N- which occurs before nominal and adjectival stems, and often undergoes phonological changes according to the general Bantu rules, for instance, the morpheme of class 9 *n-* changes to *ny-* before vowel-initial stems. Swahili learners looking for entries such as '*nyumba*, pl. *nyumba*'; '*ndizi*, pl. *ndizi*' or '*njia*, pl. *njia*' in the dictionary will need to have full knowledge of the corresponding phonological Bantu rules and some examples in order to remind them of the correct grammatical agreement according to the phonological changes of prefix *n-* + consonant: e.g. *njia ndefu* (**njia nrefu*); and *n-* + vowel: e.g. *ndizi nyingi* (**ndizi ningi*).

Furthermore, Swahili locative classes (16, 17, 18) behave differently from others, and grammatical rules applying to them can't be easily represented in a dictionary system (see also Toscano and Sewangi 2005: 274-275). For instance, the noun prefix of class 16 *pa-* only occurs with very few nouns (e.g. *pahala* 'place'), while the noun prefix of class 17 (*ku-*) is only used as adjectival concord and in copula constructions; similarly, the locative prefixes *mu-* and *mw-* (cl. 18) only occurs as a noun prefix in the term *mwahala*, which is often used as plural form of *pahala* 'place' (Bertoncini 2009: 186).

Some classes, which are still present in other Bantu languages, are not productive in Swahili, and, although we can still see traces in the language, this information is not transparent from the description in a dictionary entry. For instance, class 11-14 is conventionally indicated with a hyphen to point out that we are not dealing with a class pair (class 14 is not the plural class of 11), but rather with two different classes, originally separated, and which have merged into one in a later development of Swahili language, while remaining independent in other Bantu languages (such as in Zulu or Xhosa). In Swahili, class 11 includes mostly nouns referring to parts of a mass (e.g. *unywele* 'hair'), long objects (e.g. *ufunguo* 'key') and nominal deverbatives (e.g. *wimbo* 'song'). Class 14 is the class of abstract nouns (e.g. *utu* – humanity) and does not have a plural, while nouns of class 11 have the plural in class 10.

In most cases, singular and plural classes in the Swahili noun system are organized in pairs (for example, class 1 singular – class 2 plural; class 3 singular – class 4 plural, etc.). However, in a number of cases, Swahili nouns can have the plural in a different class (like in the case of nouns in class 11-14 which have plural in class 10) or even in two classes, for instance, some Swahili nouns can have plural both in class 6 and in class 10 (e.g. *rafiki/rafiki* or *rafiki/marafiki*¹⁴). Also, a few nouns, especially loanwords, are assigned to class 5 or to class 9 in different dictionaries. For example, the words *kamusi* 'dictionary'; or *dawa* 'medicine' are indicated as belonging to cl. 5 in some dictionaries and to cl. 9 in others.

Thus, we have shown that there are crucial issues in transferring Swahili grammatical knowledge into a dictionary. Although there are some limitations,

we have tried to overcome most of them by creating an up-to-date online dictionary specifically designed to address the needs of learners, as illustrated in the next paragraph.

5. KIU as a learning support

Since the basic aim of KIU is to support Italian students in their autonomous learning of Swahili, it provides extended information beyond what is minimally necessary in a normal dictionary. Indeed, it is assumed that learner dictionaries (e.g. see also the Swahili–Polish dictionary presented in Wójtowicz 2016) aim to provide help not only in the process of text reception but also in text production, that is, the dictionary is not exclusively centred on translation. Rather, detailed information on grammar or lexical usage and information on morphological and syntactic structures are included, in order to support learning activities such as grammar exercises (drills and text comprehension), translations, and oral production/comprehension. The dictionary also aims at providing updated lexical information, i.e. loanwords from English (e.g. *skrini* 'screen') or neologisms such as those related to technology and informatics (for instance *simu ya mkononi*, 'mobile phone'; *tovuti* 'website'; *mtandao* 'internet') or those related to the COVID-19 pandemic (*barakoa* 'mask', *Korona* 'Corona virus').

Apart from the simple search for a lemma in the alphabetical order, it is also possible to search for Italian words in the translation of Swahili examples. Although KIU is not an IT–SWA dictionary, this search option is still helpful both for beginner Swahili students working on short text production or the translation of simple sentences and for advanced students dealing with oral and written production.

Grammatical stems (e.g. possessives, demonstratives) are searchable also in all inflected forms according to the noun class grammatical agreement. In this sense, the dictionary offers reliable support for Swahili learners at beginner level with specific entry features which help users to familiarize themselves with the Bantu noun class system and its morphological and phonological rules. Moreover, some irregular plural forms are entered immediately after the class prefixes, and the user has the possibility to search for both headwords and the full plural form (e.g. *jicho*, pl. *macho*; *uso*, pl. *nyuso*; *ulimi*, pl. *ndimi*).

This method is very convenient for learners at beginner level who do not have sufficient knowledge of grammar to enable them to identify easily singular and plural forms which carry different noun prefixes and are hidden in the entries of the singular form (Kiango 2005: 264). Derivatives, such as some extended verbs, are inserted as searchable entries and linked with the basic verb through a mechanism of cross-entry references, showing both sides of the derivational process (derivative→ root and root→ derivative) (Wójtowicz 2016: 411).¹⁵ Through the cross-reference system, users can also search for different phonological variants of the same word (e.g. *asante*, *ahsante*; *blanketi*, *blangeti*; *santuri*, *senturi*).

Furthermore, in order to enrich the learning tools and support beginner learners, the design of the updated Swahili–Italian dictionary provides a tool for adding grammatical comments or usage notes where appropriate (*dictScrap*). This option allows the quality of information contained in the entry to increase with the aim of supporting the acquisition of grammatical rules and expanding vocabulary. This is achieved through the addition of the following.

- Indications on the correct grammatical agreement for more complex cases: e.g. the agreement of animate nouns from non-human classes (e.g. *kijana* 'young man' cl. 7; *bibi*, cl. 9 'grandmother, lady'); possessive agreement with animate nouns in class 9/10 referring to close relationship (e.g. *bibi yangu* 'my grandmother': the possessive agrees in class 9 and not in class 1 **bibi wangu*), etc.
- Notes on the semantic features of a term in order to disambiguate meanings and facilitate the appropriate choice or use of a term in translations and oral and written production.¹⁶ For this purpose, semantic explanations can also be accompanied by a number of labels indicating status (formal, informal, slang, derogative, euphemism, vulgar, colloquial etc.); register (literary, familiar, popular, etc.); semantic field (biology; zoology; military; music; legal; medicine; religion, etc.), frequency of use (common, rare); figurative or extended meaning.

In addition, cultural terminology, referring to untranslatable concepts and things can be supported by images accompanied by a definition. The possibility of adding images and descriptions to an entry is particularly helpful in order to clarify cultural terms which can't be easily translated because they are not part of Italian culture or in cases in which the Italian gloss or definition is not exhaustive enough to explain the concept (e.g. *ugali*: 'typical Swahili food similar to polenta'; *kanga*: 'coloured women's textile'; *chapati* 'Indian unleavened flatbread'). This additional information is generally avoided in printed dictionaries and can't be included due to printing size restrictions.

Finally, unlike printed dictionaries, learners can rely also on audio recordings of difficult pronunciations, e.g. the sound 'j' in the word *jicho* or the sound 'gh' in *ghali*.

6. Endnotes

1. These are the main bilingual publications: Krapf, *Swahili English Dictionary* (1882); Sacleux, *Dictionnaire Français–Swahili* (1891) and *Dictionnaire Swahili–Français* (1939); Madan, *English–Swahili Dictionary* (1894) and *Swahili–English Dictionary* (1903); Johnson, *A Standard Swahili–English Dictionary* (1939) and *A Standard English–Swahili Dictionary* (1939). For more details see Chuwa 1996.
2. In 2009 the Institute was renamed TATAKI (Taasisi ya Taaluma za Kiswahili — Institute of Swahili Studies).

3. Some examples are A. Lenselaer, *Dictionnaire Swahili-Français* (partiellement adapté du Standard Swahili-English Dictionary de Frederick Johnson avec l'autorisation d'Oxford University Press), Karthala, Paris 1983; G. Mertens, *Dictionnaire Kiswahili-Français et Français-Kiswahili*, Karthala, Paris 2006; K. Legère, *Wörterbuch Deutsch-Swahili*, Köppe, Cologne 2006; A.I. Kutuzov, *Kratkij Suakhili-Russkij i Russko-Suakhili Slovar / Kamusi fupi ya Kiswahili-kirusi na Kirusi-kiswahili*, Izdatel'stvo Sovetskaja Enciklopedija, Moscow 1965; B.V. Trivedi, *Gujarati-Swahili Shabdapothi* (Gujarati-Swahili Dictionary), Mombasa 1955.
4. 'Lakini, wote wawili walikabili magumu katika utumishi wao kwa Mungu' has been translated as 'Eppure, entrambi/e subirono una grave battuta d'arresto' (Yet, both suffered a serious setback), instead of 'Eppure, entrambi/e affrontarono difficoltà nel loro servizio a Dio' (Yet, both faced hardships in their service to God). Utumishi in this context means 'servizio, servire (Dio)', i.e. 'service, to serve (God)'.
5. T.E.I. Text Encoding Initiative: <http://www.tei-c.org/index.xml>
6. www.lessicafrica.it. The work is in progress, and so the link is not yet active. It will be activated by the end of 2023.
7. The project is funded by the Department of Asian, African and Mediterranean Studies at UniOr and by ISMEO (International Association for Mediterranean and Oriental Studies); see the UniOr webpage KIU (Kiswahili-Italian-Unior): the online Swahili dictionary for Italian-speaking learners.
8. Except for the pronominal stem H-, followed by the vowel and the concord, e.g. *kitabu h-i-ki* 'this book', cl. 7 (stem + morpheme).
9. Swahili locative classes currently only include the nouns *mahala/pahala* 'place, location' (cl. 16) and the plural *mwahala* (cl. 18).
10. The last prefixes indicated in the table usually precede stems beginning with vowels.
11. The original prefix *n-* (class 9/10) often disappears or undergoes phonetical changes when occurring before certain initial sounds of the Bantu stems, modifying to *ny-* (before vowel) or *m-* etc.
12. Classes of diminutives used in some Tanzanian Bantu languages and imported in some Swahili non-standard varieties that are spoken in interior regions.
13. *Moyo/mioyo* (cl. 3/4): 'heart'; *moyoni*: 'in the heart'; *-ni* is a locative suffix that expresses the locative relations indicated in Italian through prepositions, such as 'in/at/from' (cf. Bertoncini Zúbková 2009: 7).
14. Cfr. *Rafiki/rafiki* (Bertoncini Zúbková 2009: 30); *rafiki/marafiki* (BAKITA 2015: 878)
15. A fundamental aspect of the Bantu languages is the possibility of extending the basic form of the verb (e.g. *-penda* 'to love'), by adding different suffixes to the root, thus modifying the meaning (and often also the morphosyntactic features) of the verb, e.g. *-pend-ana* 'loving each other', *-pend-wa* 'to be loved', *-pend-elea* 'to prefer', etc.
16. A cross-reference system is also exploited to link synonymous entries, such as *kinywa* – *mdomo* 'mouth'.

7. References

- Abdulla, A., R. Halme, L. Harjula and M. Pesari-Pajunen. 2002. *Swahili-Suomi-Swahili Sanakirja*. Helsinki: Suomalaisen Kirjallisuuden Seura.

- Aiello, F. and M. Toscano.** 2017. On Some Swahili ICT Terms. Batic, G. and R. Leger (Eds.). 2017 *Studia Africana — Papers in Honour of Sergio Baldi*: 19-32. Cologne: Rüdiger Köppe Verlag.
- BAKITA.** 2015. *Kamusi Kuu ya Kiswahili*. Nairobi/Kampala/Dar es Salaam: Longhorn.
- Bennett, P.R.** 1986. Grammar in the Lexicon, Two Bantu Cases. *Journal of African Languages and Linguistics* 8(1): 1-30.
- Bertoncini Zúbková, E.** 1973. A Tentative Frequency List of Swahili Words. *AION* 33: 297-363.
- Bertoncini Zúbková, E.** 1977. *Vocabolario swahili-italiano e italiano-swahili*. Naples: Opera Universitaria I.U.O.
- Bertoncini Zúbková, E.** 2009. *Kiswahili kwa Furaha. Corso di lingua swahili*. Rome: Aracne.
- Chuwa, A.R.** 1996. Problems in Swahili Lexicography. *Lexikos* 6: 323-329.
- De Schryver, G.-M.** 2010. Revolutionizing Bantu Lexicography — A Zulu Case Study. *Lexikos* 20: 161-201.
- Kiango, J.G.** 2005. Problems of Citation Forms in Dictionaries of Bantu Languages. *Nordic Journal of African Studies* 14(3): 255-273.
- Kutuzov, A.I.** 1965. *Kratkij Suakhili-Russkij i Russko-Suakhili Slovar / Kamusi fupi ya Kiswahili-kirusi na Kirusi-Kiswahili*. Moscow: Izdatel'stvo Sovetskaja Enciklopedija.
- Legère, K.** 2006. *Wörterbuch Deutsch-Swahili*. Cologne: Rüdiger Köppe Verlag.
- Lenselaer, A.** 1983. *Dictionnaire Swahili-Français* (partiellement adapté du Standard Swahili-English Dictionary de Frederick Johnson avec l'autorisation d'Oxford University Press). Paris: Karthala.
- Martini, G.** 2016. *Dizionario swahili. Swahili-italiano, italiano-swahili*. Milan: Hoepli.
- Merlo Pick, V.** 1961. *Vocabolario kiswahili-italiano e italiano-kiswahili*. Turin: EMI.
- Mertens, G.** 2006. *Dictionnaire Kiswahili-Français et Français-Kiswahili*. Paris: Karthala.
- Prinsloo, D.J.** 2009. Current Lexicography Practice in Bantu with Specific Reference to the *Oxford Northern Sotho School Dictionary*. *International Journal of Lexicography* 22(2): 151-178.
- Sewangi, S.** 2007. Uundaji wa Istilahi za Kiswahili Kileksikografia na Athari zake. *Nordic Journal of African Studies* 16(3): 333-344.
- Toscano, M.** 2004. *Dizionario swahili. Swahili-italiano, italiano-swahili*. Milan: Vallardi.
- Toscano, M. and S. Sewangi.** 2005. Discovering Usage Patterns for the Swahili *amba*- Relative Forms cl. 16, 17, 18: Using Corpus Data to Support Autonomous Learning of Kiswahili by Italian Speakers. *Nordic Journal of African Studies* 14(3): 274-317.
- Tramutoli, R.** 2018. *Kamusi ya isimu Kiswahili-Kiitaliano*. Dar es Salaam: TUKI.
- Wójtowicz, B. (Ed.).** 2003. *Studencki Słownik Suahilijsko-Polski*. Available online at <https://www.mimuw.edu.pl/~jsbien/BW/SSSP/SSSP.pdf>.
- Wójtowicz, B.** 2016. Learner Features in a New Corpus-based Swahili Dictionary. *Lexikos* 26: 402-415.
- Wójtowicz, B. and P. Bański.** 2012. Swahili Lexicography in Poland: Its History and Immediate Future. *Studies in African Languages and Cultures* 46: 59-72.