

Lexicography and language learning of Swahili L2 at UNIOR: the Swahili-Italian online dictionary project

F. AIELLO, M. TOSCANO, R. TRAMUTOLI*
Università degli Studi di Napoli “L’Orientale”

ABSTRACT

The present essay outlines the progress of a lexicography project, namely a Swahili-Italian online dictionary, which was developed at the University of Naples “L’Orientale” (henceforth UNIOR) and conceived as a useful digital tool for Italian L1 learners of Swahili in a context of renewal of the teaching/learning of Swahili at the UNIOR, encouraged by web-based distance tools and platforms.

As will be explained in detail, the software of the online lexical database, first developed in the period between 2003 and 2009, is in the process of being updated with some technical and lexicographical improvements, aiming for a resource that will be increasingly user-friendly to Swahili language students, and in general to Italian-speaking learners.

* The authors discussed the content of this article in strict cooperation and agreement. However, for academic purposes, Flavia Aiello is responsible for the sections “Introduction” and “Swahili L2, lexicography and web-based learning at Unior”; Maddalena Toscano for the sections “The online dictionary UWAZO” and “Swahili word structure and dictionary entry in UWAZO”; Rosanna Tramutoli for the sections “From UWAZO to KIU: software access and management tools” and “New learner-oriented features in the Swahili-Italian online dictionary KIU”.

KEYWORDS

Swahili L2 language learning; Swahili-Italian online dictionary; lexicography.

INTRODUCTION

This contribution aims to chart the developments of the Swahili-Italian online dictionary project at UNIOR, which was led by the three authors, and discusses some of the inherent challenges in Bantu languages/Swahili lexicography which targets foreign learners, along with the solutions recently adopted in the course of the dictionary updating.

The paper is organised in five sections, the first of which provides a background to the development of the dictionary's project by giving an overview of the teaching of Swahili language and literature at UNIOR. It also touches upon the related need for lexical resources for Italian speakers and the progressive use of distance learning and web-based didactic materials and activities by the teaching staff. The subsequent two parts are devoted to the origins of the project, namely the online lexical database UWAZO: the history of the project, its participants and the initial design of the database, which originates from lexicographical research applied to Swahili language and word structure. Owing to the old software being out-dated, UWAZO is currently in the course of being reprogrammed from an IT point of view by an IT specialist. This new phase has provided an opportunity to redesign some features of the dictionary, both in terms of software access and management tools, and of learner-oriented characteristics of the entries, as will be discussed in the last sections of this paper.

SWAHILI L2, LEXICOGRAPHY AND WEB-BASED LEARNING AT UNIOR

The teaching of Swahili language and literature at UNIOR was established in 1969 by Elena Bertoncini Zúbková, an internationally renowned scholar of Swahili language and literature, whose educational activity was closely connected with research on the Swahili lexicon¹, resulting in the production of lists of words and vocabularies². These were (and still are) much needed tools for the Italian students of Swahili, given the ongoing rapid expansion of this

¹ See, for instance, Bertoncini Zúbková 1973.

² The work of Bertoncini Zúbková in this field remains largely unpublished. The only publication, out of print, is a small vocabulary published in 1977, *Vocabolario swahili-italiano e italiano-swahili* (1977).

Bantu language, called *Kiswahili* by its speakers, in Eastern and Central Africa. From the Indian Ocean coast and its neighbouring islands, where the Swahili people (*Waswahili*) live, the Swahili language has spread in the continent due to various historical factors, such as the caravan routes of the Arabo-Swahili merchants, the activity of missionaries, the language policies of colonial and postcolonial governments and the flows of migrants and refugees. It is currently used by approximately 99,000,000 people, mainly in Kenya, Tanzania and in the eastern regions of the Democratic Republic of Congo (DRC)³. It is also spoken in Rwanda, Burundi, Uganda and, to a lesser degree, in Somalia, the Comoro Islands, Mozambique, Malawi, Southern Sudan and Zambia⁴. As a consequence of the post-independence development of Swahili in a wide geo-cultural area and in multiple contexts of language use, especially in Tanzania, where it has become the dominant medium of communication in the public sector (basic education, administration, courts, national assembly, media etc.), the Swahili lexicon has been constantly growing, also due to a huge lexicographical initiative undertaken by 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 Kiswahili⁵, TUKI)⁶. These latter institutions, besides producing general dictionaries, have been creating and disseminating terminologies for many domains of language use, such as law, science, and IT glossaries⁷.

Regarding bilingual lexicographical works aimed at Italian learners, apart from E. Bertoncini Zúbková's work which, as remarked above, remains substantially unpublished or out of print, 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, Milano 2004) and Gianluigi Martini's *Dizionario swahili. Swahili-italiano, italiano-swahili* (Hoepli, Milano 2016). Furthermore, a terminological work has appeared, namely a Swahili-Italian linguistic glossary by Rosanna Tramutoli (*Kamusi ya isimu Kiswahili – Kiitaliano*, TUKI, Dar es Salaam 2018), published by the university press of the Institute of Swahili at the University of Dar es Salaam within

³ See <https://www.ethnologue.com/>.

⁴ Swahili has been declared national/official language in Kenya, Tanzania, DR Congo, Uganda, and more recently in Rwanda. As underlined by M. Mulokozi (2003), the Swahili language has also an international status, being amongst the official languages of the African Union since 2005. Furthermore, since 2019 Swahili is an official working language of the Southern Africa Development Community, an organisation composed of 16 states of Central, Eastern and Southern Africa.

⁵ In 2009 the Institute was renamed TATAKI (*Taasisi ya Taaluma za Kiswahili – Institute of Swahili Studies*).

⁶ See Sewangi 2007.

⁷ See Aiello & Toscano 2017.

the sphere of a long standing cooperation agreement with UNIOR. All these useful works for Italian speakers, are however in print format, and partly only available at the library of UNIOR. Thus, in the early 2000s, the development of an online lexicographical resource for Italian learners was initiated by M. Toscano, as will be explained in detail in the next paragraph, in a context of renewal of the courses of Swahili at UNIOR, encouraged by Internet and digital technologies, which have revolutionised the teaching of foreign languages by introducing web-based distance tools and platforms.

Over the years, different digital materials and activities have been developed at UNIOR by the teachers of Swahili (E. Bertoncini Zúbková, M. Toscano, F. Aiello), in cooperation with mother-tongue collaborators, researchers and technicians⁸, to be used for “blended” or “hybrid” teaching/learning, i.e., to be combined with traditional face-to-face language courses⁹. These include a list of online resources for autonomous learning, available on the website of UNIOR’s Language Centre, CLAOR (Centro Linguistico di Ateneo Università L’Orientale)¹⁰, as well as e-learning modules which offer teaching materials and exercises integrating the contents of the Swahili language courses. The e-learning resources and activities are offered through the Moodle digital platform managed by CLAOR, and consist of two courses, one for beginners (“Swahili livello iniziale”) and one for intermediate-advanced level (“Swahili livello avanzato”)¹¹, which can be accessed by registered students who are given the course login by the teacher. The basic-level Swahili e-learning course targets students, enrolled in BA or MA programs at UNIOR¹², who are starting the study of one or two African/Oriental languages; the intermediate-advanced course is offered to 3rd year BA students and MA students (4th/5th year). The learning objectives of the first course are to consolidate Swahili basic grammar knowledge and develop listening and comprehension skills. The intermediate-advanced course has been designed for students learning Swahili language

⁸ All the individual credits are available on the web pages quoted in footnote n. 11.

⁹ Smyrnova-Trybulska, Eugenia. 2009.

¹⁰ Swahili is the sub-Saharan language most widely taught at university level, inside and outside Africa, therefore a great deal of academic and non-academic teaching materials is available online, alongside Swahili-language media (information portals, musical videos, films etc.). The resources are classified in different sections, such as “integrated abilities”, “information portals”, “culture and society” etc., see <https://www.unior.it/ateneo/7644/1/swahili-sul-web.html>.

¹¹ See respectively <https://elearning.unior.it/course/view.php?id=84>; <https://elearning.unior.it/course/view.php?id=5>.

¹² Undergraduate degree courses: “Oriental and African Languages and Cultures”, “Political Science and International Relations”, “Comparative Languages and Cultures”; Master’s Degree Courses: “Languages and Cultures of Asia and Africa”, “International Relations and Institutions of Asia and Africa”, “Comparative Literatures and Cultures”. See https://www.unior.it/index2.php?content_id=17159&content_id_start=2&titolo=lauree-e-lauree-magistrali&parLingua=ITA.

and literature, mostly to further develop the written comprehension of literary texts and the awareness of translation strategies. The switch to “blended” teaching/learning has been facilitated by the user-friendly features of Moodle, for which basic computer literacy is sufficient, and has involved the need to redesign course materials, as one simply cannot shift content from one medium to another, i.e. from textbook to digital, asynchronous technology¹³. Rather, it is important to acknowledge the characteristics and potentials of the new medium. Therefore, fair amount of new content has been introduced, based on audio/picture/video resources (such as vocabulary quizzes, cloze tests for listening comprehension etc.), which expose the students to “realia” and/or to language use in full contexts. Moreover, an advantage of Moodle-based activities is that they promote student-centred learning since the students can self-evaluate their achievements step by step.

Furthermore, from a partnership of UNIOR with Federica Web Learning, a platform created by the University of Naples Federico II to offer high-quality MOOC (Massive Open Online Courses), a MOOC for Swahili language and culture has been created by F. Aiello and published online in 2020, entitled “Karibuni! Introduzione alla lingua e cultura Swahili” (Welcome! An introduction to Swahili language and culture)¹⁴. The course is structured in 10 teaching units, containing slides (with text, pictures and links), short video-lessons and auto-evaluation quizzes meant both for students and a wider public.

Finally, a digital corpus of Swahili texts was collected by M. Toscano in collaboration with a number of students and researchers. It is a small (one million words) Swahili untagged raw corpus consisting close to fifty full texts, taken mainly from contemporary written literature, with the addition of some oral narratives and non-literary works (socio-political essays, handbooks about agriculture, media studies, information technology). It is available to researchers and MA students for their dissertation research on Swahili language, literature and linguistics.

AINI, a Swahili lemmatizer based on MSDOS¹⁵, was used for extracting lemma from the corpus.

The corpus was available online for a short while and was also searchable through the use of regex (regular expressions)¹⁶ incorporated in Shika Neno¹⁷. For reasons of copyright restrictions only two lines of the retrieved contexts text were accessible to the user.

¹³ See Azaryad Shechter, Deborah 2015.

¹⁴ See https://www.federica.eu/mooc/c/karibuni_introduzione_alla_lingua_e_cultura_swahili.

¹⁵ By Professor T. Schadeberg and P. S. E. Elias, Leiden (no more available).

¹⁶ For info on regular expressions see <https://www.sketchengine.eu/guide/regular-expressions/>.

¹⁷ SHIKA NENO was a Linux environment, language dependent software for context retrieval. It allowed search for Swahili bases (by M. Sorrentino and M. Toscano, no more available).

The corpus is now only accessible offline. The digital corpus is used for searching examples and quotes to be included in the Swahili-Italian online dictionary.

The main tool used for information retrieval is CONCORDANCE¹⁸, a software for context retrieval which operates in a Windows environment. It is a language independent software which allows search for forms. CONCORDANCE also allows use of regular expressions and provides access to full text when the need for disambiguation occurs. It is for offline use only.

Here below are some screen shots from CONCORDANCE, the main software used by UWAZO operators to provide reliable examples of lemma usage.

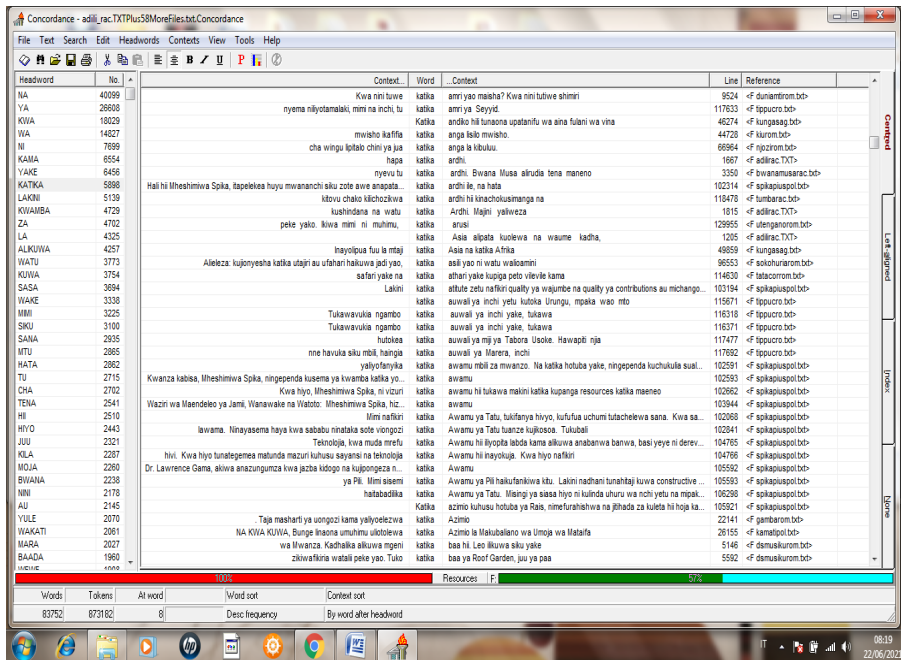


Figure 1. Concordance - Context sort of KATIKA, by word after headword

¹⁸ Concordance: <http://www.rjcw.freemove.co.uk/>.

2 right		3 right		4 right	
No.	Word	No.	Word	No.	Word
183	Ya	663	Ya	203	Ya
161	Nchi	312	Na	175	Na
74	Maisha	192	Wa	80	Wa
53	Nyumba	179	Kwa	70	Katika
50	Dunia	164	Za	67	Kwa
50	Mj	161	Yake	55	Za
50	Sehemu	130	Chi	43	Ni
48	Hababa	106	Lakini	38	Yake
47	Bunge	100	Yetu	37	Watu
46	Kia	93	Hii	30	Chi
46	Mambo	79	Huo	25	La
45	Jani	73	Wata	24	Hii
45	Kpindi	66	Sasa	24	Kwamba
45	Ushari	60	Sana	23	Hata
44	Maka	41	Kwanza	19	Kuwa
42	Kiji	38	Watu	19	Nchi
38	Chumba	37	Maka	17	Chi
38	Shule	37	Kiwahili	16	Hii
37	Maeneo	36	Huo	15	Wake
35	Kundi	34	ili	15	Ameyo
35	Safari	30	ila	15	ili
34	Mj	29	Kwa	15	Maisha
34	Umwingu	29	Zote	15	Alikuwa

Figure 2. Concordance - Collocations of KATIKA

Headword	No.	Context	Word	Context	Line	Reference
WA	40599	nipokuwa nimekaa, Wambie naseba	katika	amani ya vurumi. Staki	58074	<F mzinglerom bdt>
YA	26508	pokuwa kweli, wewe u	katika	amani, maana baba yako kana	117558	<F sippocr bdt>
KWA	18029	Nawe are	katika	amara yake, wambie na wasenzi wote.	117197	<F sippocr bdt>
WA	14827	nyema nilyotamaki, nimi na nchi, tu	katika	amri ya Seyyid.	117833	<F sippocr bdt>
NI	7899	Kwa nini luwe	katika	amri ya masha? Kwa nini luwe shirini	9524	<F dunantrom bdt>

View - FA...C:\chivetta\SWAHILI\Archive_testi_swahili\CORPUS\adili_rac.TXT\Plus58MoreFiles.txt

Clipboard Search Options Window Help

Manayema.

Alipoambwa khabari ile Mirambo akamwita Sef: Waambie wale Waarabu, kesho waende zao, maana zimekuja khabari amri kadha wa kadha. Nami nasema khabari hizo uwongo. Zija-pokuwa kweli, wewe u **katika** amani, maana baba yako kana ndugu yangu. Lakini nataka, waondoke hawa Waarabu, waende zao, wasipate khabari wapagari wao, watu wa Tabora, watacupa misigo wakimbwa. Itakuwa khabara kuu kwao.

144. Akenda Sef bin Hamed akawambia wale Waarabu na watu wangu: Hesho safari, mwendu zemu. Nami nataka kurudi Tabora. Wakasafiri wale Waarabu pamoja na watu wa Mirambo.

Na siku ya pili akapewa Sef bin Hamed pembe tano kwa frasiila tisa wa robo akapewa na nguo, madeule na bushi na vitambi bure vinne. Aktoa na watu wake sita akampa Sef bin Hamed, Menda nao hawa watu. Watatu awachukue Hamed bin Muhammed pwan, akawambie Seyyid Barghash: Sasa baada ya kuja Hamed bin Muhammed sitaki tena vita naye. Hamed

117558.35 Total? Top 117554 Bytes: 6258005

Figure 3. Concordance - Access to full text, for needs of disambiguation

THE ONLINE DICTIONARY UWAZO

UWAZO¹⁹ is a Swahili-Italian online bilingual lexical database for Italian speakers. It contains about 5000 headwords selected from a Swahili frequency list²⁰ and from various other sources, including, in addition to the above quoted works of Merlo-Pick and Toscano, the lexicon used in the teaching materials of Swahili courses (by E. Bertoncini Zúbková) and some specific lexicons (e.g. immigration, body parts) resulting from various unpublished works.

Based on T.E.I. guidelines²¹ and developed over a period of time between 2003 and 2009²², UWAZO came into being mainly thanks to IsIAO²³ funding and the collaboration of various experts²⁴.

Alongside the T.E.I. guidelines, reference was made also to the guidelines contained in the materials of the lexicography and terminology courses²⁵, acquired by the staff team (M. Toscano, F. Aiello, R. Tramutoli) through participation in specific training courses, as well as reference works for lexicog-

¹⁹ Originally available at the link http://old.iuo.it/diprapa/swahili_leksikoloja/uwazo/default.asp, temporarily hosted at the link <http://www.siamoinsieme.org/mawazo/>.

²⁰ See Bertoncini Zúbková 1973.

²¹ T.E.I. (Text Encoding Initiative): <http://www.tei-c.org/index.xml>.

²² UWAZO was presented at the DWS 2004. Maddalena Toscano, Giuseppe Marzatico, Salvatore La Gala and Massimiliano Sorrentino: *Building a corpus based Kiswahili-Italian online lexical data base*, Third International Workshop on DICTIONARY WRITING SYSTEMS (DWS 2004), Brno, Czech Republic, 6-7 September 2004 (<https://nlp.fi.muni.cz/dws2004/pres/#17>).

²³ Funding came from the bilateral agreement IsIAO-UNIOR, 2002-2003 and from the bilateral agreement UNIOR-Univ. of Dar es Salaam, 2002-2005.

²⁴ M. Toscano for project planning, Swahili language consultancy from E. Bertoncini Zúbková and S. Sewangi (Univ. of Dar es Salaam). Inserting data by F. Aiello, M. Toscano, with occasional cooperation from C. Marzio and G. Aquaviva. As for the software, the first version was by Giuseppe Marzatico and Salvatore La Gala (MARS, Napoli). The second version, now in use, was built by Tommaso Borrelli and Luciano Piedimonte.

²⁵ June 1996. *Stage Inter-regional de formation à l'utilisation du logiciel MARIAMA*, a 10 days training course organized by the GDRE 1172 of CNRS, at the Centre de Formation de Luminy, Marseille, FR.

September 7-18 1998. *Intensive course on lexicography*, by AFRILEX-SALEX, Univ. of Pretoria, R.S.A.

August 5-8 1999. *16th Terminology Summer School, Module 1 – Terminology Teaching and Training*.

August 9-11 1999. *1st Terminology Summer Campus*, by TermNet and Infoterm, Univ. of Donau, Krems, AU.

July 16-20 2001. *Training workshop in lexicography and lexical computing*, by Information Technology Institute, University of Brighton and Lexicography Masterclass, Univ. of Brighton, (UK).

June 8-12 2015. *The 2015 Lexicom Workshop in Lexicography and Lexical Computing*, Masaryk University's Centre in Telč, Czech Republic.

raphy and publications on lexicographic standards. In the project phase other available DWSs were considered²⁶.

UWAZO was specifically tailored for a list of needs which include flexibility in setting masks and labels, controlled on-line access, search facilities, output in .doc and .xml, and a user friendly and corpus based approach. Unlike a traditional printed dictionary, the online dictionary supports the transfer of knowledge from research to teaching/learning of Swahili (teaching module of computer applications to African languages), while also enhancing tools for distant learning and cooperation.

The UWAZO general framework is based on the aims and objectives of target users, and on language aspects which are relevant to the project. These elements were taken into account while defining the data base structure. Resources and tools available at the time were also considered.

UWAZO target users are Italian speakers/learners of Swahili, namely language students also majoring in (literature, political science, etc.) who are required to learn grammar (morphology and syntax) up to advanced level, to read and translate literature and political texts; to produce proper Swahili texts. Potential interested users are also non-language students (other individuals from NGOs, tourists, etc.) who need to know grammar at beginner/intermediate level.

UWAZO's basic aim is to support Italian mother tongue speakers in the autonomous learning of Swahili by providing detailed information on grammar/morphology and clues to syntactic structures. It also aims to provide updated lexical information, i.e., loanwords from English (e.g. *skrini* 'screen'); new meanings assigned to already existing words (e.g. *kibarua* 'temporary work') or neologisms (e.g. *simu ya mkononi*, 'mobile phone'; *tovuti* 'website'; *mtandao* 'internet').

All entries contain an indication of the grammatical category 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. All examples were taken from the above-mentioned Swahili corpus.

Following T.E.I. guidelines for printed dictionaries which include free and structured entries, the UWAZO database structure is based on various groups which include a fixed list of elements with free position. Sub-class elements are also available. Main T.E.I. groups list used in UWAZO include: Gruppo grammaticale, Traduzione, Esempio, Etimologia, Confronta, DictScrap. Open lists of labels can also be created by the operator.

²⁶ MARIAMA, by R. Nicolai, Univ. of Nice - (historical linguistics, restricted access) LinguaLinks; Shoe Box, by S.I.L. (field work, commercialised); DBT - LEXXIKO, by E. Picchi, ILC-CNR, Pisa (old version in MSDOS environment, new WINDOWS version not commercialised).

UWAZO management tools allow parallel login by various operators; it is possible to identify different log in identities but it has only two levels of management (control on giving new access only).

UWAZO does not allow off-line work (e.g. lexicographers working at home with their notebooks); it allows online work only. The corpus, presently available only offline, has to be made available to each operator.

Apart from the simple search for lemma, various advanced search options are available to the user: by PoS (part of speech), by type of word (form, morpheme, stem). It is also possible to search for words, Swahili and Italian, used in the examples, as shown in the following screen shots from UWAZO:

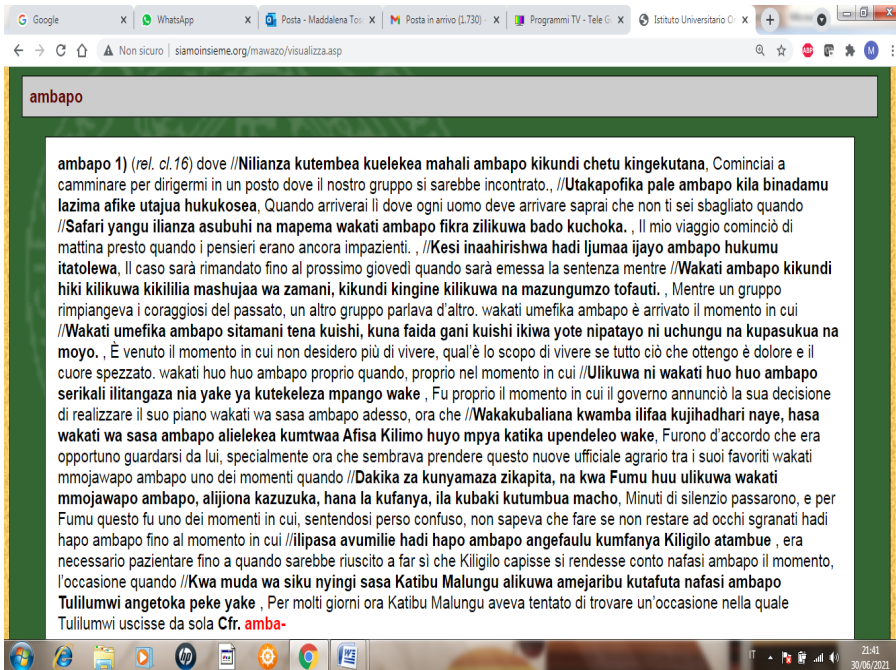


Figure 4. UWAZO - Basic search for lemma ‘ambapo’

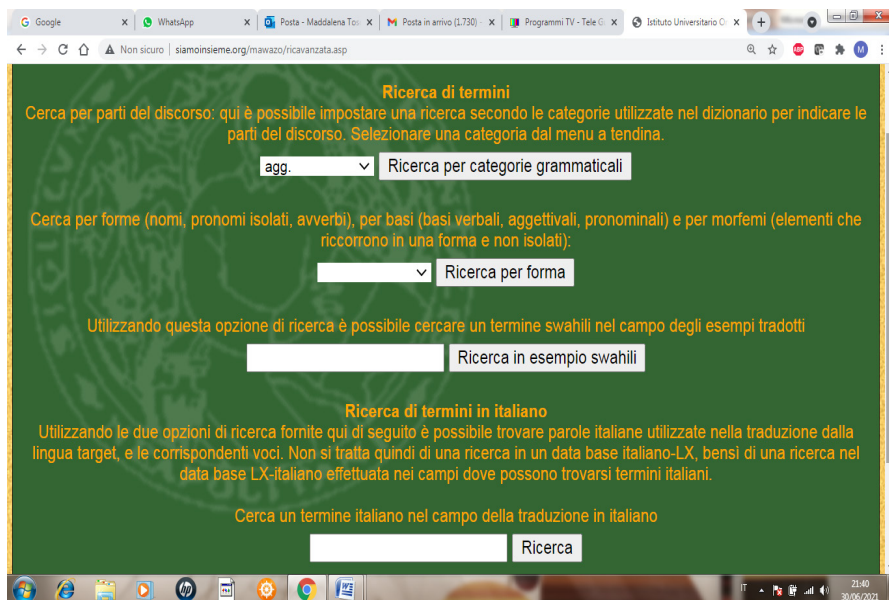


Figure 5. UWAZO - Advanced search options

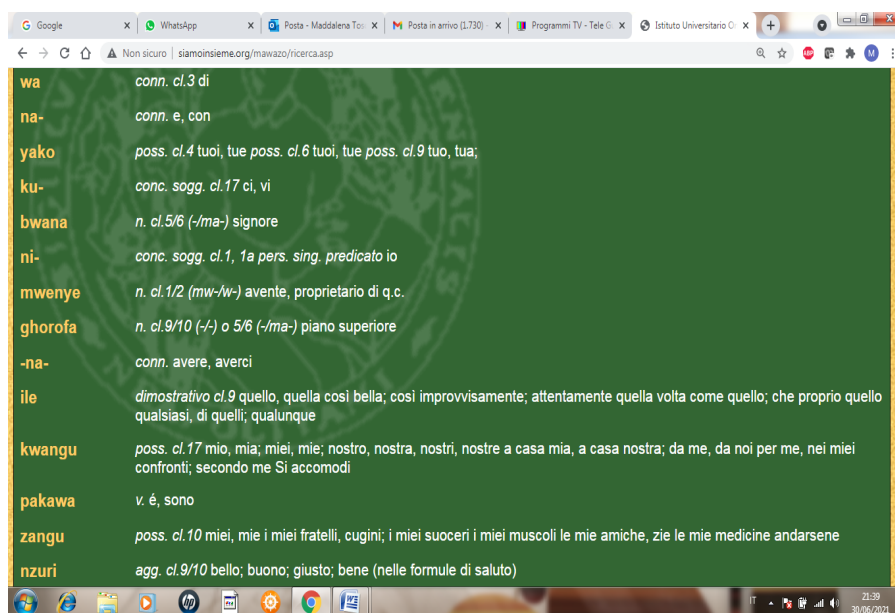


Figure 6. UWAZO - Search for 'nyumba' (casa) in Swahili examples. By clicking on each lemma shown in the list, the user will see different Swahili examples containing the word 'nyumba'

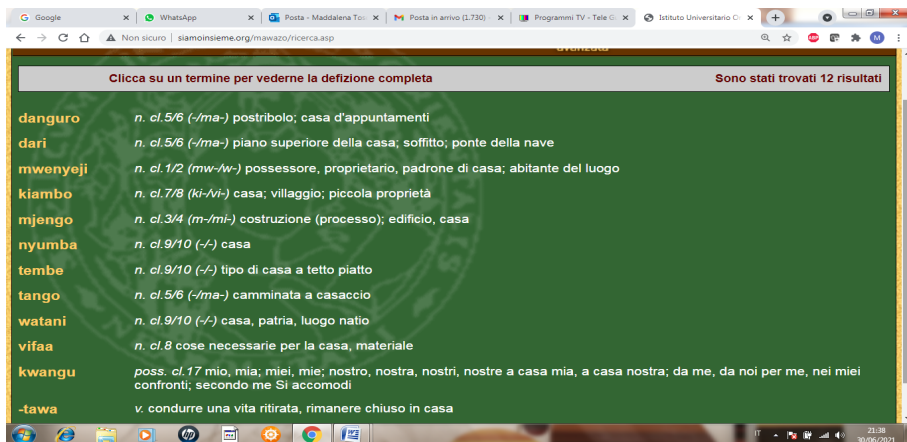


Figure 7. UWAZO - Search for word ‘casa’ in the Italian translation of examples in Swahili

SWAHILI WORD STRUCTURE AND DICTIONARY ENTRIES IN UWAZO

Bantu languages are characterized by a complex system of noun classes, represented by noun prefixes. With the exception of classes 12 and 13 (which are no longer productive in the language), and class 11-14 which have merged into a single class, the Swahili noun class system includes 15 noun classes. Swahili words follow a main basic order: *Morpheme(s) + Stem*:

- nouns: class concord + 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);
- adjectives: adjectival class concords (cl. 1...18) + adj. stem;
- pronominals: pronominal concords (cl. 1...18) + stem; stem + pr. conc. (1...18);
- verbs: subject class concords (cl. 1...18) + tense marker + (relative conc. cl.1...18) + (object concords cl. 1...18) + verbal stem + final vowels/extensions.

Swahili is an agglutinative language, which means that morphemes are juxtaposed to form words. Within the Swahili lexicographical tradition, the accepted lemmatization strategy is to list nouns in their full forms with class prefixes, whereas the prefixes of verbs, numerals, pronouns and inflected adjectives are ignored, and the stems alone are listed²⁷. Thus, in standard printed dictionaries, we find the following types of Swahili entries:

²⁷ See Wójtowicz 2016: 410.

- nouns: singular form (e.g.: *moyo* ‘heart’ → *mioyo*; *moyoni*, *mioyoni*²⁸);
- adjectives: stem (e.g.: *-baya* ‘bad’ → *mbaya*, *wabaya*, *mibaya*, *baya*, *ma-baya*, *kibaya*, *vibaya*, *kubaya*, *pabaya*);
- pronouns: stem (e.g.: *h-* ‘this’ → *huyu*, *hawa*, *huu*, *hii*, *hili*, *haya*, *hiki*, *hivi*, *hizi*, *huku*, *hapa*, *humu*; *huyo*, *hao*, etc.);
- verbs: stem (e.g.: *-soma* ‘to read’ → *-somea*, *-someka*, *-somesha*, *-somwa*, *-somana*, etc.);
- invariable forms (some prepositions, conjunctions, adverbial forms, etc.).

Considering that Swahili students are used to this lexicographical tradition, for the online Swahili-Italian dictionary we have decided to choose solutions adopted by almost all Swahili dictionaries in regard to the process of lemmatization. Consequently, we have listed only the stems alone for verbs, numerals, and inflected adjectives, ignoring subject concord and agreement prefixes.

On the other hand, in some other cases we have decided to reject traditional lexicographic solutions usually adopted in printed dictionaries, in response to beginner learners’ needs and we have listed closed sets of grammatical words, such as pronouns, in their full forms (with the stems also included as separate entries). We have also included all grammatical morphemes as separate entries, adding an explanation about their function.

Thus, the basic structure of an entry in the online Swahili-Italian dictionary includes the following elements:

- headword;
- variant/variants of the headword;
- poS (part of speech), that is, the grammatical category;
- noun class;
- translation (gloss and/or description);
- examples (quotations, collocations, idioms, proverbs, etc.);
- example translations;

Moreover, each entry is 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.

²⁸ *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’ (Bertoncini Zúbková 2009: 7).

In considering the Swahili word structure, it is evident that, while 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 root²⁹. 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.

FROM UWAZO TO KIU: SOFTWARE ACCESS AND MANAGEMENT TOOLS

The UWAZO software is presently being redesigned from an IT perspective owing to the old software now being obsolete. The upgrade is carried out by an IT specialist on the basis of the lexicographical indications provided by the three authors of this contribution. It includes some improvements aimed at the creation of a resource that will be increasingly useful to Swahili language students, and in general to Italian-speaking learners, as outlined in the next paragraph.

In the first phase of assembly of the new dictionary, the main effort was devoted to redesigning an updated version of the software, importing and editing previous data by tidying inconsistent entries, adding relevant missing information, and completing sets of grammar inflected forms. Closed sets belonging to semantic and grammatical categories, like days of the week, months and adjectives were verified, and additional vocabulary was also collected by students and researchers who worked on chosen sets they found useful in their studies, such as body parts or the Swahili COVID lexicon.

The new version of the dictionary, called KIU (Kiswahili-Italian-UNIOR), will be published online on an ad hoc page of the UNIOR website, and will be freely accessible to university students and the general public. The number of entries in the database (currently about 5,000) will also be increased.

The KIU dictionary, accessible by users from computers or smartphones, supports two different levels of access depending on the role: administrators, having the highest level of access to the database, can implement and edit data, and set up and manage the accounts of students and collaborators. Students and learners can have full access to database and software tools except for data publishing, which needs administrator approval.

UWAZO management tools also allowed parallel login by various operators, where it was possible to identify different login identities. However, unlike UWAZO, which only provided control on creating new login access, the

²⁹ Wójtowicz 2016: 410.

updated software enables the administrator to control both access and data editing as well as publishing of all other operators.

Another implementation which characterizes the new software consists in the export/import of full or selected data to facilitate editing and updating by remote access.

Moreover, in addition to the Swahili-Italian dictionary, the new software was designed to be extended to other Bantu languages; in particular a Zulu-Italian lexical database (including a collection of body vocabulary and a small literary corpus) is being developed by R. Tramutoli.

NEW LEARNER-ORIENTED FEATURES IN THE SWAHILI-ITALIAN ONLINE KIU DICTIONARY

New features for the entries have been introduced in the updated version of the Swahili-Italian online dictionary, in order to address specific Swahili L2 learner needs.

The new dictionary is going to be a learner's dictionary, that is, a bilingual dictionary with features that until recently have been primarily associated with monolingual learner's dictionaries: "extended grammatical information (meant to make the creation of real sentences easier, by providing hints for constructing the proper agreement patterns) and with visualisation of derivative forms that will provide extra lexical information and make navigation across the dictionary easier"³⁰.

As regards to Swahili nouns, the entire plural form is entered immediately after the class prefixes and the searches can be carried out on headwords and plural forms of headwords; thus, if a user looks up for *maembe* 'mangos', he/she will be directed to the entry for *embe* 'mango'.

Differently from paper dictionaries and other online lexical databases, the online Swahili-Italian dictionary offers the possibility to search both for regular and irregular forms of the plurals, such as *jiko* (sg. cl. 5 'kitchen') - *meko* (= *ma* + *iko*; pl. cl. 6 'kitchens'); *jino* (sg. cl. 5 'tooth') - *meno* (= *ma* + *ino*, pl. cl. 6 'teeth'), where the noun class prefix of cl. 6 *ma-* coalesces with the vowel *-i* changing to *me-*. Indeed, students at a beginner level might not be familiar with similar plural forms of class 6 (e.g. *meko*, *meno*), which change consistently compared to the corresponding singular forms, following the Bantu rules of vowel coalescence. The possibility to search for full plurals is thus very convenient for learners at a beginner level who do not have sufficient knowledge of grammar to enable them to identify easily singular and

³⁰ Bański & Wójtowicz 2008: 269-275. Bański & Wójtowicz 2012: 60-72.

plural forms which carry different noun prefixes and are hidden in the entries of the singular form³¹.

Another important issue with respect to macrostructure is related to the handling of derivation, which in Swahili, like in other Bantu languages, is very robust and typically creates dozens of complex lexemes from a single root, especially with regard to verbal roots³².

Verbs are described by the type of derivation and references to their base/root in the case of derivatives, or to derivatives in the case of roots. Through a mechanism of cross-entry references, derivatives are inserted as searchable sub-entries showing both sides of the derivational process (derivative→ root/root→ derivative)³³. In standard paper dictionaries, only frequently used derivative verbs (e.g. *-pendeza* ‘be attractive’; *-endelea* ‘continue’) or those extended forms of the verb which have become lexicalized or fossilized, and which meaning is independent from the derivational process, (such as *-sima-ma* ‘stand’; *-kumbatia* ‘hug’; *-patikana* ‘be available’; *-wezekana* ‘be possible’) are usually searchable as separate entries and do not require learners to be aware of their morphological structure.

The Swahili dictionary for learners provides the possibility to look up all verbal derivative forms, i.e. all extended verbs, which have been listed as searchable sub-entries linked to the corresponding verbal bases. Thus, learners can either directly search for extended verb stems (e.g. *-fundishana* ‘teach each other’; *-jibizana* ‘answer each other’) even without being familiar with the derivational process. Alternatively, they can look for the basic form of the verb and consult the corresponding possible derivations. The extended mechanism of cross-entry references, which also characterizes other types of learner-oriented dictionaries designed for Bantu languages (see T-Lex; isiZulu.net)³⁴, helps “to maintain a system whereby derivatives have the status of headwords, while simultaneously the derivational and semantic relationships between forms are preserved”³⁵.

Swahili diminutive and augmentative forms are generally created by adding noun class prefixes of class 5/6 and 7/8 respectively to the nominal stem, for instance:

mbuzi cl. 3 (goat) → *kibuzi* cl. 7 (little goat) → *buzi* cl. 5 (big goat).

Thus, augmentatives and diminutives have also been entered as searchable derivatives forms linked with the corresponding nominal base, except for those forms which have acquired independent meaning in Swahili and thus

³¹ Kiango 2005: 264.

³² Wójtowicz 2016: 410.

³³ Wójtowicz 2016: 411.

³⁴ TshwaneLex: <https://tshwanedje.com/dictionary/swahili/>; <https://isizulu.net/>.

³⁵ Wójtowicz 2016: 410.

constitute separate headwords (e.g. *m-ji* cl. 3 ‘town’ → *ki-ji* cl. 7 ‘village’; → *ji-ji* cl. 5 ‘metropolis’)³⁶.

Moreover, the updated Swahili-Italian dictionary (KIU) includes different entry variants (orthographic, dialectal/continental, allomorphs), searchable as cross-reference entries and visualized as part of the headword. Thus, entries are searchable not only in the plural form, but also according to the different orthographic and phonetic variants. For instance, both alternative forms *asante*, *ahsante* ‘thank you’, have been inserted in the entry.

Similarly, the search of noun class prefixes includes all allomorphs, that is the possible variants occurring in different phonological contexts, such as:

cl. 1 **m-**: *mw-* and *mu-* are two different variants which occur in different phonological contexts; *mw-* before vowels except before vowel “u”, where the prefix *mu-* does not undergo vowel assimilation -, e.g. *muungwana* ‘gentleman’)

cl. 2 **wa-**: *w-* before vowel (e.g. *mwalimu*, pl. *walimu* ‘teacher/s’);

cl. 3 **m-**: *mu-* and *mw-*; similar situation as class 1 prefix;

cl. 7 **ki-**: *ch-* before vowel;

cl. 8 **vi-**: *vy-* before vowel.

The Swahili noun class system is quite standardised and homogeneous; each noun class is represented by noun prefixes which mark all elements of a Swahili sentence and thus encodes the grammatical information necessary for the grammatical agreement. Nevertheless, in some cases, apart from noun class prefixes, other semantic and grammatical skills, which can hardly be included in the entry of standard dictionaries, are required in order to construct a correct Swahili sentence.

In order to enrich the learning tools and support beginner learners in the acquisition of Swahili grammatical knowledge, the design of the updated Swahili-Italian dictionary provides a tool for adding grammatical comments or usage notes where appropriate. This option allows us to enhance the quality of information contained in the entry with the aim of supporting the acquisition of grammar skills and expanding the learner’s vocabulary through the addition of:

- 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; *mama*, ‘mother’, cl. 9; *waziri* ‘minister’ cl. 5); possessive agreement with animate nouns in class 9/10 referring to close relationship (e.g. *bibi yangu* ‘my grandmother’: possessive agreement is 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/use of a term in translations

³⁶ Class 12, including the diminutives in *ka-*, has disappeared in standard Swahili and has been reintroduced by some Tanzanian authors from the hinterland under the influence of other Bantu languages; in a few cases, also the plural *tu-* of class 13 was imported.

and oral/written production³⁷. For this purpose, semantic explanations can also be accompanied by a number of labels/tags 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.

Swahili learners will rely on phonological tools as well, such as audio recordings (and possibly the transcriptions) of difficult words to pronounce. Since Italian L1 Swahili learners do not usually face difficulties in Swahili pronunciation due to the high degree of similarity with the Italian phonological system, pronunciation has not been inserted by default for each entry; nevertheless, a few terms containing difficult sounds (not present in the Italian phonological system) are supported by audio recordings with pronunciation (e.g. words containing the sound *ng'*, such as *ng'ombe* or the sound *j*, such as *jambo*). This learner-oriented feature is also useful when adding hints for Swahili words with different accents and meaning, e.g. *barabàra* 'highway', *baràbara* 'perfectly'. Since in Swahili accent regularly falls on the penultimate syllable, the dictionary offers the possibility of inserting an audio support for those exceptional cases where an accent falling on the third last syllable of a word is distinctive in that it produces difference in meaning.

Furthermore, the dictionary provides another useful learner-oriented tool through a function which allows cultural terminology to be supported by images/pictures, together with a description/definition which substitutes the gloss/translation.

The possibility of adding images and descriptions to an entry is particularly helpful in order to clarify specific terms, which cannot be easily translated into Italian or for which the Italian gloss is not exhaustive enough to explain a cultural concept (e.g. *ugali*: "typical Swahili food similar to 'polenta'"; *kanga*³⁸: "coloured women's textile"). This additional information is generally avoided in printed dictionaries and cannot be included due to printing size restrictions.

³⁷ A cross-reference system is also exploited to link synonymous entries, such as *kinywa - mdomo* 'mouth'.

³⁸ "The kanga is a widely spread printed cloth, mainly used by women as a dress. The cloth measures ca. 110 cm in height and 150 in length. It is defined by a border (*pindo*), a central field (*mji*) and usually contains on the lower third a printed proverbial inscription (*jina*). We know that it was "invented" around the 1880s in Zanzibar and imported from Europe in this form, i.e. already imprinted with patterns and inscription. It played an important role in the emancipation of slaves and their integration into the Muslim Swahili community of the East African coast. [...] the kanga allows for the communication of the unspeakable, whereby the interactants cannot be held responsible for their interaction." Beck, 2001: 157, 166.

REFERENCES

- Aiello, Flavia & Toscano, Maddalena. 2017. "On some Swahili ICT terms". In Batic, G. C. & Leger, R. (eds.), *Studia Africana, Papers in Honour of Sergio Baldi*, 19-34. Köln: Rüdiger Köppe Verlag.
- Azaryad Shechter, Deborah. 2015. "Blended Learning Course Format on Moodle: A Model for Beginner Level Foreign Language Courses in Higher Education" n: Gómez Chova, L., López Martínez, A., Candel Torres, I. (eds.), *INTED2015 Proceedings* (9th International Technology, Education and Development Conference Madrid, Spain. 2-4 March, 2015), 183-209 Madrid: IATED.
- Bański, Piotr & Wójtowicz, Beata. 2008. "Multi-level reference hierarchies in a dictionary of Swahili". In Bernal, E. & De Cesaris J. (eds.), *Proceedings of the XIII EURALEX International Congress* (Barcelona, 15-19 July 2008), 269-275. Barcelona: Institut Universitari de Lingüística Aplicada – Universitat Pompeu Fabra.
- Bański, Piotr. & Wójtowicz, Beata. 2011. "New XML-encoded Swahili–Polish Dictionary: Micro- and Macrostructure". In Goźdz-Roszkowski, S. (ed.). *Explorations across Languages and Corpora. PALC 2009*, 497-514. Frankfurt a. Main: Peter Lang.
- Bański, Piotr. & Wójtowicz, Beata. 2012. "Swahili lexicography in Poland: its history and immediate future". *Studies of the Department of African Languages and Cultures* 46, pp. 60-72.
- Beck, Rose M. 2001. "Ambiguous signs: the role of the kanga as a medium of communication." *Swahili Forum* 8, pp. 157-169.
- Bertoncini Zúbková, Elena. 1973. "A tentative frequency list of Swahili words", *Annali dell'Istituto universitario Orientale di Napoli* 33, pp. 297-363.
- Bertoncini Zúbková, Elena. 1977. *Vocabolario swahili-italiano e italiano-swahili*. Napoli: Opera Universitaria, IUO.
- Bertoncini Zúbková, Elena. 1985. "Quantitative analysis of Swahili vocabulary", *Studies in African Linguistics*, supplement 9 (présis from the 15th Conference on African Linguistics, UCLA, March 29-31, 1984), pp. 22-26.
- Bertoncini Zúbková, Elena. 2009. *Kiswahili Kwa Furaha*. Corso di lingua Swahili (Tomo I). Roma: Aracne.
- Kiango, John G. 2005. "Problems of Citation Forms in Dictionaries of Bantu Languages". *Nordic Journal of African Studies* 14 (3), pp. 255-273.
- Mulokozi, Mogyabuso. 2003. "Kiswahili as a national and

- international language”,
Kiswahili 66, pp. 66-80.
- Sewangi, Selemani. 2007.
“Uundaji wa istilahi za
Kiswahili kileksikografia na
athari zake”, *Nordic Journal
of African Studies* 16 (3), pp.
333-344.
- Smyrnova-Trybulska, Eugenia.
2009. “About Some
Basic Aspects of Distance
Learning”. In Eugenia
Smyrnova-Trybulska (ed.),
*Theoretical and practical
aspects of distance learning:
collection of scholarly papers*,
13-35. Cieszyn: University of
Silesia in Katowice.
- Wójtowicz, Beata. 2016.
“Learner Features in a
New Corpus-based Swahili
Dictionary”. *Lexikos* 26, pp.
402-415.