Studies of the Department of African
Languages and Cultures, No 46, 2012
ISSN 0860-4649

Beata Wójtowicz
Depatament of African Languages and Cultures
University of Warsaw

Piotr Bański
Institute of English Studies
University of Warsaw

Swahili lexicography in Poland: its history and immediate future

Abstract

The interest in Swahili lexicography at the University of Warsaw has a long tradition and was initiated by the first lecturer of Swahili – prof. Rajmund Ohly. He was not only an observer but his name has been indelibly written into the history of Swahili lexicography. His passion inspired the next generation and some projects aiming at creating Swahili dictionaries have been undertaken in the Department. That resulted in a state-financed grant on Swahili-Polish dictionary that is to be delivered at the end of 2012.

1. African studies in Poland

African studies in Poland date back to the beginning of the 20th century, when research and teaching on Africa began at the Jagiellonian University in Kraków. At the same time, the Department of Oriental Studies and Sociology was opened at the Polish Academy of Sciences. The well-known Africanist Roman Stopa, author of many works on Bushman languages, lectured on African languages at the Jagiellonian University for many years. Today, research on African languages is carried out at the Department of Afro-Asiatic Linguistics (*Katedra Językoznawstwa Afroazjatyckiego*), at the Institute of Oriental Philology of the Philological Faculty.

Warszawa joined Kraków as another African research centre in the mid 50's. It was then that the Department of Semitic Studies was launched within the Institute of Oriental Studies, which had been established already in 1922. In 1969, the initial scope of its interest was expanded to Sub-Saharan Africa and the name was changed to the Department of African and Semitic Studies. The Department of African Languages and Cultures (*Katedra Języków i Kultur Afryki*)[1] has existed since 1977 as an individual unit of formerly the Institute, and currently the Faculty of Oriental Studies.

At the same time, the former interdisciplinary Institute of African Studies, which had existed at the Faculty of Geography since 1962, transformed into the Institute of Developing Countries and Regional Studies. The Institute, formerly under the name *Studium Afrykanistyczne*, focused on economical and geographically inclined matters; it also offered two-year courses in African languages intended for prospective workers in Africa.

The present Department of African Languages and Cultures of the Faculty of Oriental Studies is divided into three sections: Ethiopian, Hausa and Swahili studies. Since October 2005, it offers a three-year B.A. programme and a two-year M.A. (postgraduate) programme. Intensive language training in Hausa, Swahili or Amharic is compulsory within both programmes.

2. Professor Rajmund Ohly and Swahili lexicography

The research on Swahili lexicography in Poland was initiated by Rajmund Ohly, who began teaching in the Department of Semitic Studies in Warsaw in 1961 as the first teacher of Swahili. His interests focused mainly on languages and literatures of Africa and he quickly became an eminent expert in the field.

Earlier, during his M.A. studies in Oriental philology in Kraków, he had concentrated on Arabic studies, but got also acquainted with African Khoisan languages, and Ewe, Hausa and Swahili, as a student of Roman Stopa (cf. Piłaszewicz 2004). He soon turned his attention towards Swahili and in his Ph.D.

[1]See:http://www.orient.uw.edu.pl/web-kjika/eng/.Until 2009, the Department was referred to in Polish as "Zakład" rather than "Katedra".

dissertation defended at the University of Warsaw in 1967, he investigated the development of abstract nouns and then continued his research in the field of Swahili terminology, which lead to his 1978 *Habilitationsschrift* under the title "The development of common political terminology in Kiswahili (1885-1974), with special reference to modern Tanzania", defended at the University of Marburg in Germany. Devoted to the teaching of Swahili, he published numerous language and literature textbooks, some of which are in use until today, not only in Poland, but even in Africa.

In 1972, Ohly left Warsaw for Africa and spent the following 20 years in Tanzania and Namibia. His lexicographic career began at the Institute of Kiswahili Research of the University of Dar es Salaam, where he initiated a project of compiling an English-Swahili dictionary that was successfully finished and published only in December 1996 (TUKI 1996). In 1975, he became a member of the editorial team working on the first monolingual Swahili language dictionary that was published in 1981 (TUKI 1981). After having finished that work, the team returned to the English-Swahili dictionary project, but in 1982, Ohly left Tanzania for Namibia. In the meantime, in 1976, he was appointed professor of the University of Dar es Salaam. He also avidly participated in the work of the governmental Commission on the Swahili Language. He compiled two specialized bilingual Swahili-English dictionaries: of slang (1987) and of technical terms (1987), the latter at the request of the *Gesellschaft für Technische Zusammenarbeit* of the Federal Republic of Germany.

Professor Ohly always recognized the importance of lexicographic study in the process of language development and was an ardent supporter of promoting Swahili. In his research he kept track of the development of Swahili terminology and his passion for Swahili lexicography inspired other members of the Department to take interest in it.

In 1992, the Department of African Languages and Cultures at the University of Warsaw hosted the Catalysis summer school that was dedicated to computational linguistics and lexicography, with a focus on African languages. Research on lexicography and terminology was also conducted for M.A. theses by Polish and

foreign students, such as Bento Sitoe from Mozambique and also in Ph.D. dissertations of Albina Chuwa from Tanzania, who under supervision of prof. Ohly explored phraseological units in relation to lexicography (Chuwa 1995), and Beata Wójtowicz, who investigated Swahili lexicography and framed the outline of a new Swahili-Polish dictionary (Wójtowicz 2004; see section 5).

3. Swahili-Polish dictionaries

Even though Rajmund Ohly was the author or editor of several Swahili dictionaries, he never compiled any for Polish. Possibly, part of the reason was that one – a small Swahili-Polish and Polish-Swahili dictionary by Stopa and Garlicki (1966) – had already existed. That first (and so far the only) published Swahili-Polish dictionary had 126 pages that contained around 3500 entries in each direction. As the authors themselves wrote in the introduction, the dictionary "[...] is meant to be usable by Poles for everyday contacts with African speakers of Swahili, for comprehension of simple texts, and for the study of [Swahili]. It should also be usable by African speakers of Swahili in similar circumstances" (translation ours). Even though the dictionary was compiled by a leading Polish Africanist, it was not a success (Ohly 1967). Criticized for the selection of headwords, erroneous translations, and outdated grammatical terminology, the dictionary was never reprinted or revised.

Until the late 90's no attempt to produce a new Swahili-Polish dictionary was undertaken. Then Beata Wójtowicz, during her Ph.D. studies under the supervision of Janusz S. Bień, a computer scientist and a linguist, investigated various dictionary formats. During that time, she produced an electronic DjVu version of Stopa and Garlicki's (1966) dictionary. Preserved for historical reasons, this electronic version is not distributed due to the copyright issues. Additionally, a small Swahili-Polish Student Dictionary with over 1300 entries was compiled and published as a PDF file on the Internet, under a free license (Wójtowicz 2003)[2]. This dictionary has also been released to the FreeDict project (cf. section 6).

[2] The dictionary was originally a private glossary compiled by Anna Pytluk,

In October 2009, a new state-financed project was launched aiming at creating a new Swahili-Polish electronic dictionary; this project is described in more detail in section 5 below.

4. Dictionaries of Swahili among Polish students

Currently, Polish learners of Swahili are forced to use dictionaries in which Swahili is paired with a language other than Polish, most typically English or German. In order to gather some insight into the situation of an average student of Swahili at the University of Warsaw, in the year 2003 and then in 2009, two dictionary-usage surveys have been conducted, each of them on 30 students of the Department of African Languages and Cultures.

The surveys revealed that in 2009, only 25% of students owned and used printed versions of various Swahili-English and English-Swahili dictionaries, as opposed to 90% in 2003. Nowadays, all students use the Internet Living Swahili Dictionary (ILSD) by the Kamusi Project, which was true of only 70% of students in 2003 (50% of the latter used the downloaded offline version in the form of text files; nowadays, everyone uses the online HTML interface).

The main advantages of ILSD mentioned by the respondents are fast access to translations, vast coverage (over 60 thousand entries[3]) and free availability. The dictionary employed a morphological analyser, but interestingly, no one noticed that. The disadvantage that was understandably listed in the first place by Polish users was problems with understanding of some of the English translations. Other perceived disadvantages included unsorted senses (sometimes those that come up first are the least frequent), the lack of consistent grammatical information, the lack of explicitly shown

---

a student of the Oriental Institute who collected Swahili words and described them for the purpose of her own study. Beata Wójtowicz chose Anna's dictionary from among other private dictionaries she solicited from students for her project, edited it, introduced minor corrections and typeset it for electronic publication.

[3] The dictionary is available at http://www.kamusi.org/. The number of entries does not correspond to the number of lemmas, which was computed by De Pauw and De Schryver (2009) to be 17 thousand. Nevertheless, this is the largest Swahili-English online dictionary, built by an online community under the supervision of Martin Benjamin.

derivational families, lack of information on the pronunciation and, in some cases, an insufficient number of examples of usage. Students also complained that spelling mistakes in ILSD search terms are announced as failures and no suggestions for similar words are offered. Despite the complaints, the dictionary is regarded as a very good, largely sufficient and reliable source of lexical information[4].

In recent years, students have begun to also use the online version of the TshwaneDJe Swahili-English Dictionary (http://africanlanguages.com/swahili/). The dictionary contains less headwords but it's interface and presentation of the equivalents is regarded as more user friendly.

5. New Swahili-Polish Dictionary[5]

For many years, an increase in the interest in learning Swahili has been observed among students of the University of Warsaw. Since 2002, the Department of African Languages and Cultures has offered open courses to all students from outside the Department (the membership is limited to around 30 per semester). All these students have to cope with the lack of Swahili dictionaries on the Polish market. Therefore, in order to meet the increasing demand for a lexical resource complementing the course, we have decided to launch a new project. The aim of this project is to create a new Internet-accessible Swahili-Polish dictionary designed primarily as a didactic tool for the students of Swahili, but at the same time suitable for Polish tourists, businessmen and the like.

The three-year project, financed by the Ministry of Science and Higher Education (N104 050437), was officially launched in October 2009. The dictionary is going to have over 5000 entries as the first deliverable, and it shall then be further expanded and eventually published. We consider the electronic version as more appropriate for the beginning, because it will be accessible for free, able to be easily searched, it may provide visualization of derivational hierarchies and inflected forms, and it is easy to

[4] It is worth mentioning as a *signum temporis* that, as opposed to the 2003 survey, in 2009 no one pointed to the problem with Internet access and the costs that it incurs.

[5] At the end of 2012 the dictionary will be available at http://www.kamusi.pl.

maintain and expand on the basis of active user feedback or passive monitoring of user queries (cf. De Schryver and Joffe, 2004). For the editors of dictionaries of languages considered non-commercial from the local perspective, it is very important to be able to fix all errors and add the most pressing enhancements before the first edition gets printed – publishers are hardly willing to publish revised versions of such dictionaries, as they are not profitable enough.

The new dictionary is going to be a translation/learner dictionary – a representative of the growing trend to furnish bilingual dictionaries with features that until recently have been primarily associated with monolingual learner 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 derivational hierarchies that will provide extra lexical information and make navigation across the dictionary easier (cf. Bański and Wójtowicz 2008).

The dictionary will be encoded in XML, that will make the resource easy to maintain and expand, allow output of almost any kind and various visualization strategies. The macrostructure of the new dictionary is based on a Swahili-English dictionary skeleton derived automatically from the Helsinki Corpus of Swahili (HCS 2004). Therefore, one of the phases of the building process was a switch from a Swahili-English to a Swahili-Polish dictionary. We performed concatenation (crossing) of the HCS-derived Swahili-English dictionary with an English-Polish dictionary, hoping that such a move might speed up dictionary creation and provide a useful test case for other lexicographic projects of this kind.

During the two years that we spent waiting for funding, we have already started working on the dictionary structure and the techniques for deriving it. With the concatenation step in mind, we decided to temporarily substitute a freely available Swahili-English dictionary for the one which, we expected, would be derived from the HCS. This is what turned our attention to a resource previously created by Beata Wójtowicz for the FreeDict project, which we look at in the next section.

6. Swahili dictionaries at FreeDict.org

The FreeDict project, founded by Horst Eyermann in the year 2000 and hosted by SourceForge.net, is home to numerous (over 70) bilingual dictionaries available on open-source licences (primarily the GNU General Public License, ver. 2.0 and later). It initially hosted bilingual dictionaries produced by concatenating (crossing) the contents of the dictionaries in the Ergane project (http://download.travlang.com/Ergane/), with Esperanto as the interlanguage. It was meant to complement the DICT project, responsible for making text resources available and searchable on the Net[6] (Faith and Martin 1997).

At the very beginning, the data was kept as plain text suitable to be processed by DICT tools, but soon it was converted to the SGML format advocated by the Text Encoding Initiative, called TEI P3. Later on, the databases were transduced to the XML-ised version of TEI P3: TEI P4. The Swahili-English xFried/FreeDict Dictionary was the first FreeDict dictionary encoded according to the most recent TEI P5 XML standard.

6.1. Swahili-English xFried/FreeDict Dictionary

The first version (0.0.1) of the Swahili-English FreeDict Dictionary was published in 2000 by Horst Eyermann as a product of concatenation of Swahili-Esperanto and Esperanto-English Ergane dictionaries. It had 650 headwords and the database was encoded as TEI P3 SGML.

In 2004, the maintenance of the dictionary was taken over by Beata Wójtowicz and the first Swahili-English xFried/FreeDict Dictionary[7] was published. It was based on the dictionary derived from *Swahili-Kiswahili to English Translation Program* by Morris D. Fried (available from http://www.dict.org/links.html), which has been supplemented by entries from version 0.0.1. The entries from 0.0.1 were then enriched with POS (part-of-speech) information.

---

[6] The SourceForge.net addresses of the two projects are, respectively http://sourceforge.net/projects/freedict and http://sourceforge.net/projects/dict.

[7] *xFried* in the name of the dictionary stands for *extended Fried* – M. D. Fried being the author of the source dictionary.

This new version of the dictionary, 0.0.2, contained 1542 headwords. It was created in text format and then transduced into TEI P4 XML with tools offered by FreeDict. The following is an example entry for *ndege* 'bird, airplane' from that version.

```
  <entry>
<form><orth>ndege</orth></form>
<def>bird(s)</def>
<gramGrp><pos>n</pos></gramGrp>
  </entry>
```

In December 2008, Beata Wójtowicz and Piotr Bański created version 0.3 of the dictionary, encoded in TEI P5 XML and verified lexicographically. The number of entries increased and their contents were extended with additional translation equivalents, definitions and usage hints.

The version current at the moment of submission of the present paper, 0.4.2, was published in April 2009: it contains ca. 2650 entries, all of them described grammatically with parts-of-speech and sub-categorization information, cf. the entries below, in the working view of the dictionary[8], where *hayo* is a referential demonstrative pronoun that agrees with nouns of class 6, *hazina* is either a noun of class 9, where the singular form is identical to the plural, or an inflected possessive verb that displays agreement with class 10 and refers the user to *wa na* 'be with = have'.

[8]The working view is generated by the web browser applying a CSS (Cascading Style Sheet) that accompanies the source of the dictionary. It presents the information as it appears in the XML file, with the CSS adding some text. This view is very browser-dependent and works best in the standards-compliant Firefox (other browsers, such as Opera or Internet Explorer, do not support all of the CSS directives). A better way to query the dictionary is either via a DICT client or a WWW gateway, e.g. at http://dict.uni-leipzig.de/dictd.

hayo *pron dem ref* (agrees with cl. 6)
- these, those, the ones referred to previously or close to the hearer

hazina[1] [sg=pl] *n*
- treasury

hazina[2] *v infl* (agrees with cl. 10)
- it does not have

See also: wa na

In this version, all nouns that occur in singular-plural pairings either contain a reference to the plural form (listed as separate entries and referring back to the singulars), or indicate the fact that the singular and the plural forms are identical ("[sg=pl]"). Irregular verbal inflections (e.g. irregular imperative forms, see further below for an example of *ja*), some classes of vocabulary have been added or expanded (e.g. names of countries and their inhabitants; the present-tense irregular paradigm of *wa* and *wa na*), an expanded system of references has been added, senses are better organized and usage notes are added where appropriate. At this moment, the entry for *ndege* is as follows.

```
<entry xml:id="ndege">
  <form type="N">
    <orth>ndege</orth>
  </form>
  <gramGrp>
    <pos>n</pos>
  </gramGrp>
  <sense xml:id="ndege.1" n="1">
    <def>bird</def>
  </sense>
  <sense xml:id="ndege.2" n="2">
```

```
    <def>airplane, plane</def>
    <xr type="syn">(synonym: <ref
      target="#eropleni">eropleni</ref>)</xr>
  </sense>
</entry>
```

The publication of version 0.4.2 was accompanied by changes in the FreeDict build system, now fully adjusted to TEI P5. This allowed for the TEI dictionary source to be converted to a format readable by DICT servers for the purpose of dissemination, which means that the dictionary can now be accessed via any DICT-aware client[9]. Clients typically render the XML example above as follows.

```
 ndege <n> [sg=pl]
1. bird
2. airplane, plane
  Synonym: eropleni
```

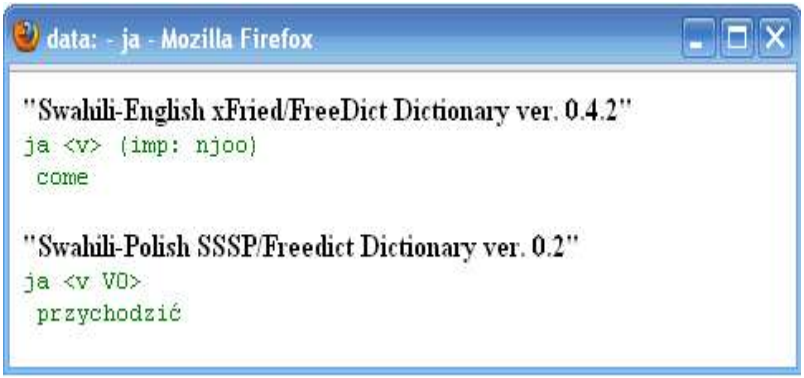
Although the dictionary is small, we were happy to note that it received a largely positive review from De Pauw et al. (2009), which we took as a signal that the dictionary can be used for the purpose of initial tests of concatenation with another resource that has been submitted to FreeDict, namely a pocket English-Polish translating dictionary by Tadeusz Piotrowski and Zygmunt Saloni.

6.2. Swahili-Polish SSSP/FreeDict Dictionary

In December 2009, a small Swahili-Polish SSSP/FreeDict Dictionary was added to the FreeDict repository. It has over 1300 entries accompanied by POS and other grammatical information, depending on the category of the headword. It mainly contains vocabulary covered during the first year of Swahili language course at the Department of African Languages and Cultures of the University of Warsaw. The dictionary was compiled in 2003 and then published in the form of a PDF file. In 2009, it was converted to

[9]A growing list of DICT clients is available at http://www.dict.org/w/software/software.

XML and re-edited, so that now it can be accessed via various DICT clients, e.g. a Firefox add-on, dict[10], as shown in the screenshot below, on the example of *ja*. The DICT protocol makes it possible for the client to search in many databases simultaneously. In our example, the results are found in the Swahili-English and Swahili-Polish dictionaries and illustrate the way the former handles irregular verbal forms (in this case, the irregular imperative form of *ja*).



The addition of a little dictionary such as the Swahili-Polish SSSP/FreeDict Dictionary illustrates something that Bański and Wójtowicz (2009) have argued for: that the FreeDict project appears ideal as a repository for this kind of small resources that might otherwise get discarded as non-publishable or only get disseminated among a small group of people, e.g. course participants. FreeDict has resources to make them accessible and usable even in their original form, and by doing so, to encourage others to expand them or use them as basis for projects targeting other languages.

[10]This Firefox add-on is available from http://dict.mozdev.org/. After it is installed in Firefox, in the options window, the user has to choose "dict.uni-leipzig.de" from the list of servers in order to guarantee that the most recent FreeDict dictionaries are queried.

7. Conclusion

The present paper surveys the origins and development of Swahili lexicography in Poland, and sketches our vision of its immediate future.

While one of us is proud to have been a student of Rajmund Ohly, we note that his lexicographic legacy has not yet been fulfilled in his native country – practically, no Swahili-Polish-Swahili dictionary exists. We would like to make the next step on the way to creating a large modern lexicographic resource of that kind in the nearest future.

At the same time, we believe we have opened the path towards a small collectively-built Swahili-Polish dictionary in the FreeDict project (after the fashion of the ILSD, though understandably on a much smaller scale). It may be useful for organizing student work and also as an example for other resources of this kind, especially those concerning non-commercial languages with small speaker and research communities.

References

Bański, P., Wójtowicz, B., 2008, "Multi-level reference hierarchies in a dictionary of Swahili", in: E. Bernal, J. DeCesaris (eds.), *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*, Barcelona: Institut Universitari de Lingüística Aplicada - Universitat Pompeu Fabra, 269-275.

Bański, P., Wójtowicz, B., 2009, "A Repository of Free Lexical Resources for African Languages: The Project and the Method", in: G. De Pauw, G-M. de Schryver, L. Levin (eds.), *Proceedings of the EACL 2009 Workshop on Language Technologies for African Languages, 31 March 2009, Athens, Greece*, Greece: Association for Computational Linguistics, 89-95.

Chuwa, A., 1995, Phraseological units and dictionary: the case of Swahili language. University of Warsaw (unpublished Ph. D. dissertation).

De Pauw, G., de Schryver, G-M., and P. W. Wagacha., 2009, "A Corpus-based Survey of Four Electronic Swahili–English Bilingual Dictionaries", *Lexikos* 9, 340–352.

De Schryver, G-M., Joffe, D., 2004, “On how electronic dictionaries are really used”, in: G. Williams, S. Vessier (eds.), *Euralex 2004 Proceedings Computational lexicography and lexicology*, Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud, 187-196.

Faith, R., Martin, B., 1997, “A Dictionary Server Protocol, Network Working Group Request for Comments #2229”, available from http://www.ietf.org/rfc/rfc2229.txt

(HCS) Helsinki Corpus of Swahili. 2004. Compilers: Institute for Asian and African Studies (University of Helsinki) and CSC – Scientific Computing Ltd. Patrz: http://www.aakkl.helsinki.fi/cameel/corpus/intro.htm

Ohly, R., 1967, Roman Stopa, Bolesław Garlicki, „Mały słownik suahilijsko-polski i polsko-suahilijski, Kamusi Dogo ya Kiswahili-Kipolanda, Kipolanda-Kiswahili”, *Przegląd Orientalistyczny* 3(63), 265-268.

Ohly, R., 1987, *Primary Technical Dictionary English-Swahili*, Dar es Salaam.

Ohly, R., 1987, *Swahili-English Slang Pocket Dictionary*, Wien: Beiträge zur Afrikanistik.

Piłaszewicz, S., 2004, „Professor Rajmund Ohly (1928-2003)”, *Africana Bulletin* 52, 157-178.

Stopa, R., Garlicki, B., 1966, *Mały słownik suahilijsko-polski i polsko-suahilijski*, Warszawa: Wiedza Powszechna.

TUKI, 1981, *Kamusi ya Kiswahili Sanifu*, Dar es Salaam: Chuo Kikuu cha Dar es Salaam.

TUKI, 1996, *English-Swahili Dictionary*, Dar es Salaam: Chuo Kikuu cha Dar es Salaam.

Wójtowicz, B., (ed.), 2003, *Studencki Słownik suahilijsko-polski*. Available in the XML version at http://sourceforge.net/projects/freedict/files/ or online at http://dict.uni-leipzig.de/dictd .

Wójtowicz, B., 2004, Ogólnolingwistyczne i leksykograficzne podstawy organizacji słownika suahilijsko-polskiego (Linguistic and lexicographic basis of a new Swahili-Polish dictionary), University of Warsaw (unpublished Ph. D. dissertation).