



The Kamusi Project

2011



Copyright 2011, Kamusi Project USA and Kamusi Project International



This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/> or send a letter to Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.



The Kamusi Project 2011

Table of Contents

Kamusi History	4
PALDO - Uniting African Languages Across the Digital Divide	8
KamusiTERMS - Tackling Contemporary Challenges of African Languages	13
Projects in the Pipeline	19
The Kamusi Project Business Model	23
User Profile	27
Kamusi People: Personnel and Partners	30



Kamusi History

The Kamusi Project arose from a student's frustrations at learning the Swahili language. It became a central resource for that language worldwide. It is now using that experience to benefit languages throughout Africa. Martin Benjamin was learning Swahili to prepare for his Anthropology PhD research. While on a Fulbright intensive language studies program in Tanzania in 1993, he was regularly stymied by the old Swahili dictionaries then available, which were confusing and incomplete. He read about a project that had used something called "the Internet" to parcel out the work of breaking supposedly uncrackable cryptographic code, and thought a similar process could be applied to writing a new Swahili dictionary. He mentioned the idea to Ann Biersteker, his Swahili professor at Yale University, who encouraged him to write a proposal to the local branch of the Consortium for Language Teaching and Learning (CLTL). The Consortium approved the proposal in autumn 1994. In December of that year, in the same week that Netscape released the first "web browser," the Kamusi Project was born.

The first step was for Benjamin to enter about three thousand terms into a spreadsheet, copied with permission from existing learners' glossaries. He then divided those terms into packs of 100 and put those files on a "gopher" server that people could access via a command line interface and dial-up modem. The intent was for volunteers to each expand one pack with new terms, and to keep subdividing the packs as contributions rolled in. That idea never really worked, however, because the process was too cumbersome and the number of Swahili enthusiasts using computers was too small. Instead, the project received copyright permission for a large out-of-print dictionary by Charles Rechenbach, was awarded a larger grant from the full CLTL, and concentrated on data entry and the development of a website (Yale's first in the social sciences or humanities) to distribute the results to the public.



In 1996, Dr. Biersteker was awarded funding for the project from the United States Department of Education’s International Research and Studies program (IRS). This grant enabled the development of the “Edit Engine,” a tool that makes it possible for anyone to help edit dictionary entries. The Edit Engine went live in 1999, a year before Wikipedia began with a similar model (and with the important difference that all Kamusi changes must be approved by an editor before becoming public). At the same time, data became available through a searchable online database, rather than having to be downloaded as text or Excel files.

A second IRS grant in 2003 supported many additional features, such as a photo uploader for users to illustrate dictionary entries with appropriate images, a parser to return useable dictionary entries from conjugated verbs, and a grouping tool to organize entries according to priority and sense. By 2006, the Kamusi Project was being used about a million times a month by 60,000 unique visitors.

2007 marked a major transition for the project, which had run out of funding. Benjamin had left Yale, where the project was still housed, and moved to Lausanne, Switzerland for family reasons. Several interesting potential partnerships were emerging around the idea of expanding the Kamusi model to other African languages. However, these projects for the international public were better housed at an institution devoted specifically to the cause of language development. It was decided to move the project to the care of the non-profit World





Language Documentation Centre, based in Wales, as an interim home while steps were taken to incorporate Kamusi independently. The online presence was established as kamusiproject.org, and then kamusio.org when that name was donated by its original registrant. There have been some technical problems maintaining a robust server using cloud-based services, and some project features are still waiting to come back online after the transition, but most functionality has remained available to most users despite the unfunded platform relocation.

Incorporating Kamusi was completed in 2010. The organization is actually two legally independent non-profit entities: Kamusi Project USA for American-based activities and Swiss-based Kamusi Project International for projects with the rest of the world. Our US status makes it possible for Americans, historically Kamusi's most generous supporters, to continue contributing to our work. At the same time, Swiss incorporation facilitates work with partners throughout Africa, due to Switzerland's special open relations with most of the world. The two organizations have independent boards and completely separate accounting. Dr. Benjamin now serves as Executive Director of both NGOs.

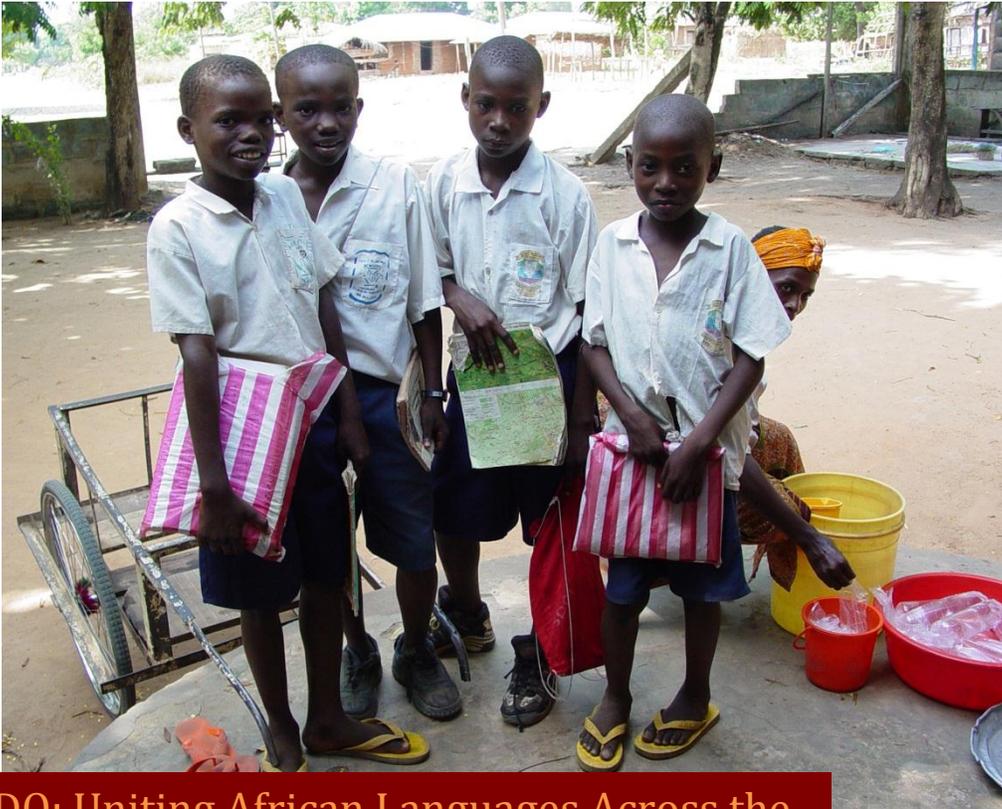
Since 2007, the Kamusi Project has embarked on several exciting new initiatives:

- The US National Endowment for the Humanities is supporting the development of a dictionary for Kinyarwanda, the language of Rwanda, based on the Kamusi model. This has spurred the development of PALDO, the Pan-African Living Dictionary Online, which will be a rich dictionary for numerous African languages. Modifying the Edit Engine for the multilingual environment has proven to be a substantial technical challenge, with programming work expected to be completed in the second half of 2011.



- With the support of the International Development Research Centre of Canada (IDRC), the Kamusi Project led an initiative to create computing locales for nearly 100 African languages.
- Also with IDRC support, we led a project to codify ICT terminology for 12 languages spoken around Africa. This project has led us into terminology development on a wider scale, through the onset of KamusiTERMS (Kamusi for Technology, Economy, Rights, Medicine, and Science). IDRC is now supporting development of enhanced collaborative terminology software, which will go live later in 2011.

The history of the Kamusi Project has been one of both innovation and struggle. Funding resources for African languages are few and far between, and the project has found that it is very difficult to make progress unless key partners can be remunerated for their time. Nonetheless, the Kamusi Project has pressed forward to become the world's most used resource for the Swahili language, and is now in a technical and regulatory position to provide similar services for many more of the languages spoken by Africa's one billion people. Many new and innovative projects are now in the pipeline, with partners from countries throughout Africa. The next chapters of this history are poised to be written.



PALDO: Uniting African Languages Across the Digital Divide

In Africa, a billion people speak 2000 languages. This linguistic diversity is a rich contribution to our human heritage, but also a persistent challenge to the continent's prosperity. Africans have long had to address the world, and even their own governments, in foreign languages such as English, French, and Portuguese. Most Africans cannot go to secondary school in their mother tongues, and they must use one of a few second languages to work or trade beyond their home regions. Few resources exist to ease the communications barriers between local and international languages, much less among the continent's many tongues. With a history that places a low value on most of the languages Africans learn from birth, the majority of Africa's people are excluded from participating equitably in global knowledge and the global economy.

The Kamusi Project is an international NGO that is dedicated to producing communications resources such as dictionaries and glossaries for African languages; "kamusi" is the Swahili word for "dictionary." With the ambition of documenting every word in Africa, the project has developed PALDO, the Pan-African Living Dictionary Online. Built on a unique hybrid model of scholarly and community contributions, PALDO uses innovative least-cost technology to extend its results as widely as possible, free to all its users. PALDO is intended as a tool to promote language equity across Africa while preserving the cultural knowledge that its languages contain.



PALDO is designed so that new languages can be added quickly, easily, and relatively cheaply. Once a language is added to the system, it is linked to all of the other languages therein. In one step, dictionaries come into being between a given language and numerous other languages spoken all around Africa. After each data item is created, it becomes permanently and freely available to the public in a variety of formats. Adding data to PALDO is similar to installing a solar panel – all the work happens at the beginning, while the benefits continue to flow for years to come.



The Kamusi Project seeks out language specialists throughout Africa and scholars at universities around the world who can take the lead for their languages. These experts commit to completing a certain number of entries, and to editing entries that are submitted by the public. Partners are trained in the specially-designed PALDO software, and receive continuous technical support along the way. Work on a language begins when funds are secured. Partners are remunerated on a per-term basis, which provides an incentive structure toward steady completion. Because each language advances independently, lexicons for many languages can be developed simultaneously.

Dictionary entries can come from many sources. The primary path for PALDO is the log records of tens of millions of lookups on PALDO's predecessor, the Internet Living Swahili Dictionary. These log records give a solid record of the most frequent terms that dictionary users seek, providing a ranked list that guides partners in producing parallel entries for the most important concepts. In addition, log records demonstrate which items are missing from the database and should be slated for addition. Existing data sets, such as out-of-print dictionaries, are sometimes available without copyright restriction, though making such works compatible with PALDO can involve a lot of effort, particularly producing definitions and matching concepts. The Kamusi Project has also planned methods, yet to find development funding, to build preliminary data sets using SMS search and response, and to harvest terms from translators using computer-assisted translation tools. Finally, the database is open to public contributions that are vetted by the language editor.



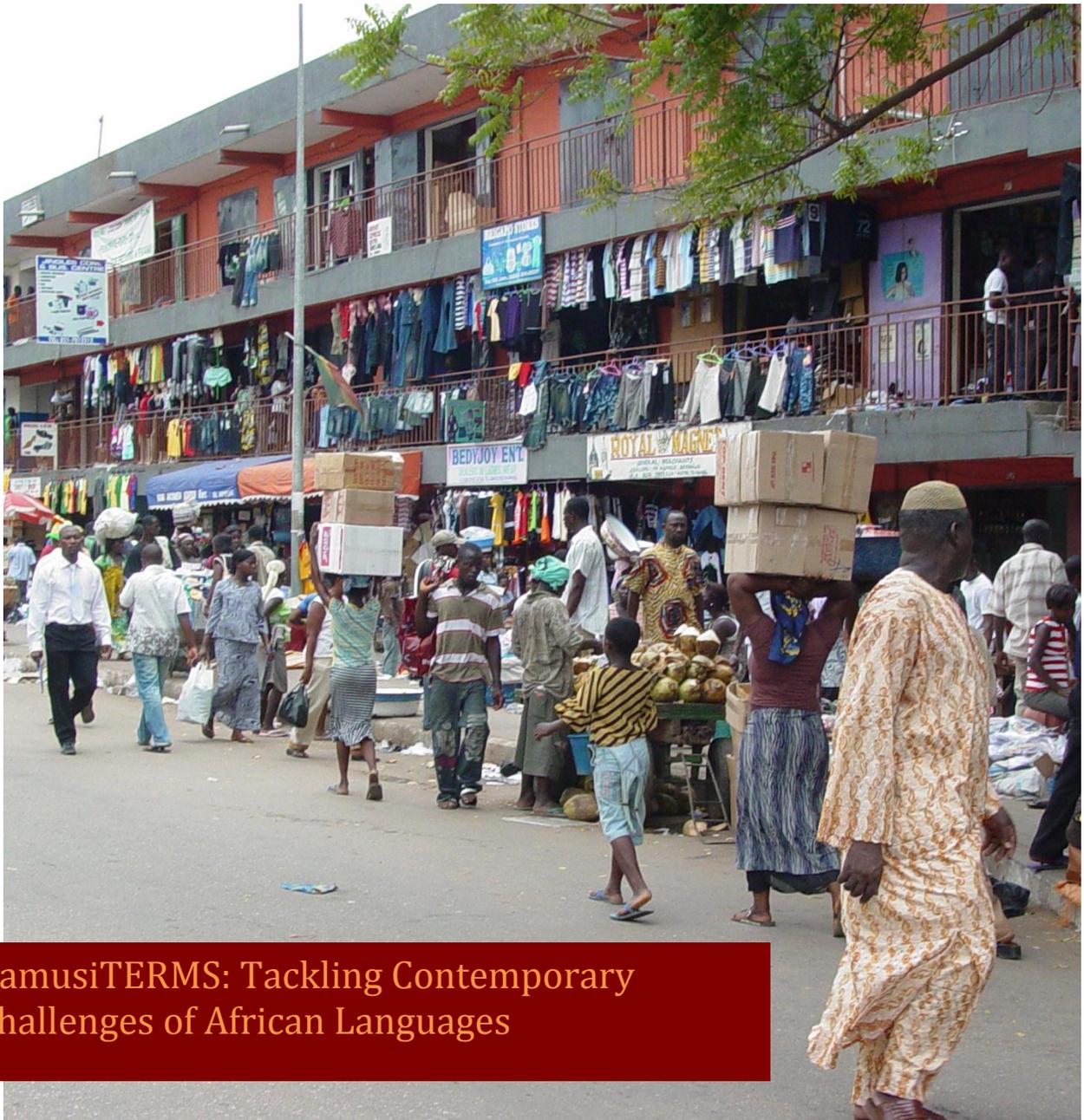
Languages are linked at the level of the concept. For example, the English word “fork” has several senses: cutlery for eating, a road that splits in two, a tool for tuning musical instruments, or software code that starts in a new direction from an existing project. The Swahili term “uma” is equivalent to the first concept, “njiapanda” to the second, “chuma cha noti” to the third, and “tawi la programu” to the last. Editors align the concepts within the PALDO software. When the term for the cutlery concept is added in a third language, all the links for that concept are attached. If the Zulu term for fork as cutlery, “imfologo,” is added in reference to English, a link is automatically created to “uma” in Swahili, and any other equivalents in the database for other languages based on existing confirmed language pairings. Such automatic links are shown as computer-predicted until a human editor provides confirmation. Through the linking tool, the one-time task of adding an entry for a concept in one language multiplies into a reference chain for many.

Through this system, good monolingual dictionaries come into existence for languages that usually do not have dictionaries of their own. Very few dictionaries exist for African languages that include definitions in those languages; for most African students, using a dictionary for their language (if one is available) is similar to looking up “fork” in English and finding the definition only in French or Russian. At the same time, dictionaries are created for language pairs that would be highly unlikely to be united otherwise. A bridge is built from one language to many others. This bridge can be used by multinational peacekeeping forces, by traders, by intergovernmental commissions, by students studying away from home, by audiences for music and movies and satellite news, or in any other circumstance where a speaker of one PALDO language interacts with a speaker of another.



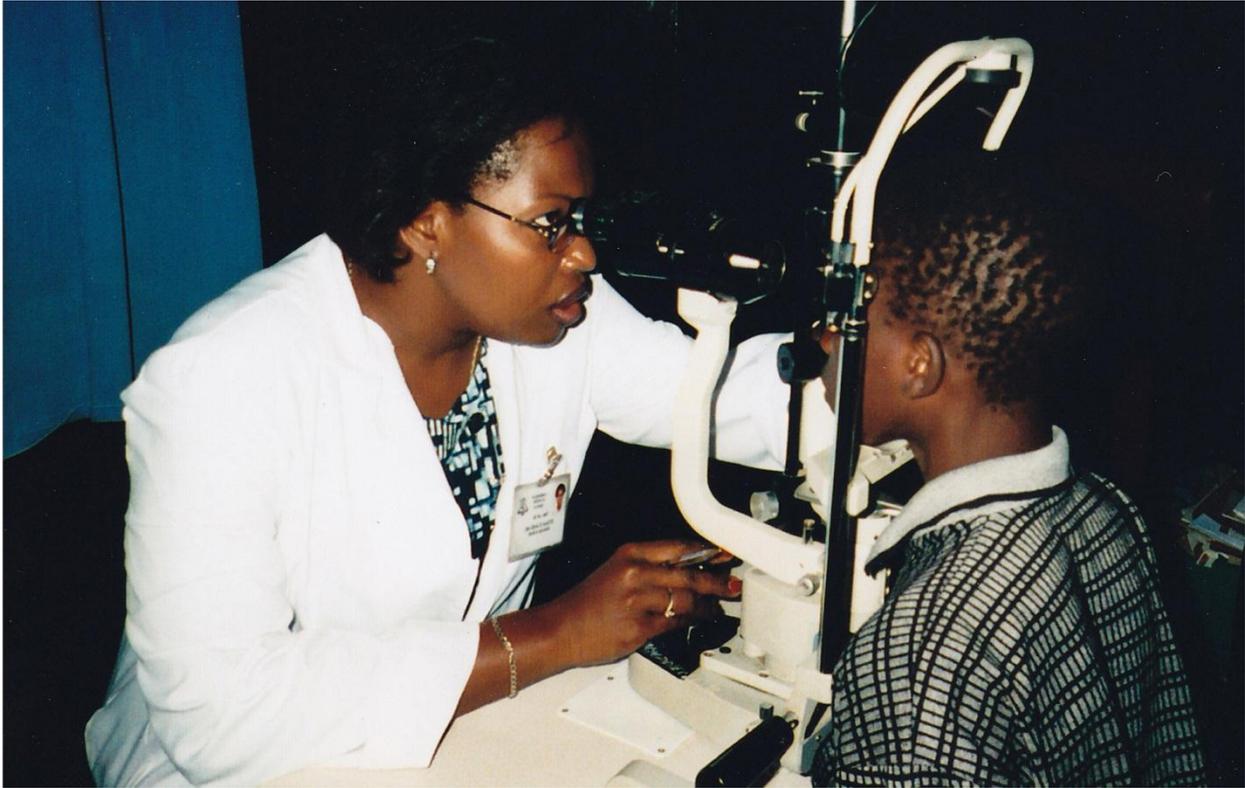
The Kamusi Project is working on innovative systems to ensure the greatest possible reach of PALDO data. Outputs are currently available via the web, web-enabled phones, and can be printed out as paper dictionaries. A stand-alone program has been developed to use the dictionaries on offline electronic devices, with all of the tools available to web users. A prototype SMS system has proven successful and is awaiting a telecom partner for implementation. With mobile devices as the future of communications in Africa, applications will also be developed for smartphones and tablets, and for new technologies as they emerge. As Africa becomes increasingly wired, PALDO will be at the ready for people to learn about their own languages and those of their neighbors.

With each language added to PALDO, Africa's linguistic richness is documented, preserved, and put to work. Educational prospects are improved, communications among people are enhanced, and opportunities open up for jobs and trade. Reminiscent of other technological tools that once seemed impossible and now seem indispensable, PALDO stands poised to open up deep knowledge about a vibrant portion of the world's linguistic heritage. With the Pan-African Living Dictionary Online, the Kamusi Project and its partners will help harness Africa's language diversity for greater prosperity and mutual understanding.



KamusiTERMS: Tackling Contemporary Challenges of African Languages

The importance of language is often overlooked when confronting the serious challenges that international organizations seek to address in Africa. For example, the contours of the AIDS epidemic in sub-Saharan Africa are well known: according to the most recent data from UNAIDS (<http://www.unaids.org/globalreport>), a 5.2% overall adult HIV prevalence, about 2 million people newly infected each year, millions of children who have lost at least one parent to AIDS, billions of dollars spent in the battle, and suffering that has affected virtually every family.



The numerous programs taking part in the world's largest health crisis response cover a wide spectrum, from drugs manufacture, through health systems, to community engagement. Arguably it is the latter that has had the biggest impact on infection rates, with HIV incidence falling in many countries during the past decade as communications outreach has led to changes in sexual behavior. Yet, thirty years into the crisis, efforts to combat HIV/AIDS continue to be slowed by the difficulties of discussing epidemiology, treatment, and prevention in Africa's many languages.

Only a fraction of Africa's billion people speaks a language of global communication. For community programs to be effective, they have to be delivered in the languages that the target groups understand. Translations of important terminology in areas such as HIV prevention and treatment must be available, authoritative and universal. Creating these linguistic bridges to Africa's 2000 languages is a major challenge.

Language often stands as an obstacle to success, whether in humanitarian programs, government or commerce. In practice, it is the local agency or medical staff who have to provide translations, with no specific training. Where dictionaries do exist for a language, they rarely include the specialist terms needed.



This practice has spawned generations of inconsistent translations, causing confusion both within the humanitarian organizations and in the communities they wish to serve. Examples abound from direct experiences of working in Africa’s linguistic *mélange*, with observations of untrained staff attempting to translate terms such as “account deficit” or “waybill” or “crop rotation” or “cost/benefit analysis,” with poor results, affecting the morale of staff and the effectiveness of programs.

With the ambition of documenting every word in Africa, the Kamusi Project has developed KamusiTERMS (Kamusi for Technology, Economy, Rights, Medicine, and Science) a comprehensive approach to producing terminology sets for any domain, for any language. The KamusiTERMS initiative is based on two premises:

- 1) Communicating with people in their own language requires much less effort, and will be much more effective, than expecting them to communicate in a foreign language that few understand
- 2) Every language has the capacity for full communication in any domain, provided key principles are followed for terminology development. For example, twenty years ago English did not have ICT terms for “web” or “browser,” yet the productive capacities of the language were employed so that today most English speakers understand exactly what is meant by that terminology.



KamusiTERMS employs a unique participatory methodology for terminology development that maximizes the likelihood that a term set will be accepted and used by a language community. By using a combination of paid experts and community volunteers, the costs of this work are kept to a minimum.

- 1) Subject specialists with expertise in a language are engaged for a preliminary translation of a domain-specific terminology set.
- 2) The specialists are encouraged to leave question marks or multiple choices in cases of uncertainty.
- 3) Then members of the public are invited to comment on the problem terms, cast non-binding votes for existing proposals, and propose their own suggestions or new ways of looking at a concept. For example, a Swahili ICT community member broke an impasse by seeing “cache” as “temporary storage” rather than the English metaphor of a hiding place.
- 4) While the experts are still called upon to make the final decision, this democratized input and review process results in term sets that are much more likely to have universal uptake than the traditional top-down approach to terminology development.





One recent example of Kamusi’s work is an IDRC-sponsored effort by the African Network for Localization to produce a glossary of information technology terms for 10 African languages. This provides translations and definitions across all 10 languages of 2500 terms, from “absolute path” to “zoom out.” This glossary is now central to software development projects in those languages.

The Kamusi Project is seeking partners in development and humanitarian aid to build language resources that will permanently improve the communications environment of those agencies. If language issues can be identified and addressed up front in the planning stage of a program, the community interaction will be improved, as will the effectiveness of local agency staff.

The Kamusi Project has identified a number of areas where a targeted glossary should be developed as a reusable foundation in humanitarian and development programs. The glossaries can be quickly expanded to provide for additional languages or common use terms specific to the situation. Some of the proposed glossaries are:

- 1) A list of common terms for disasters and emergency response. The selection of

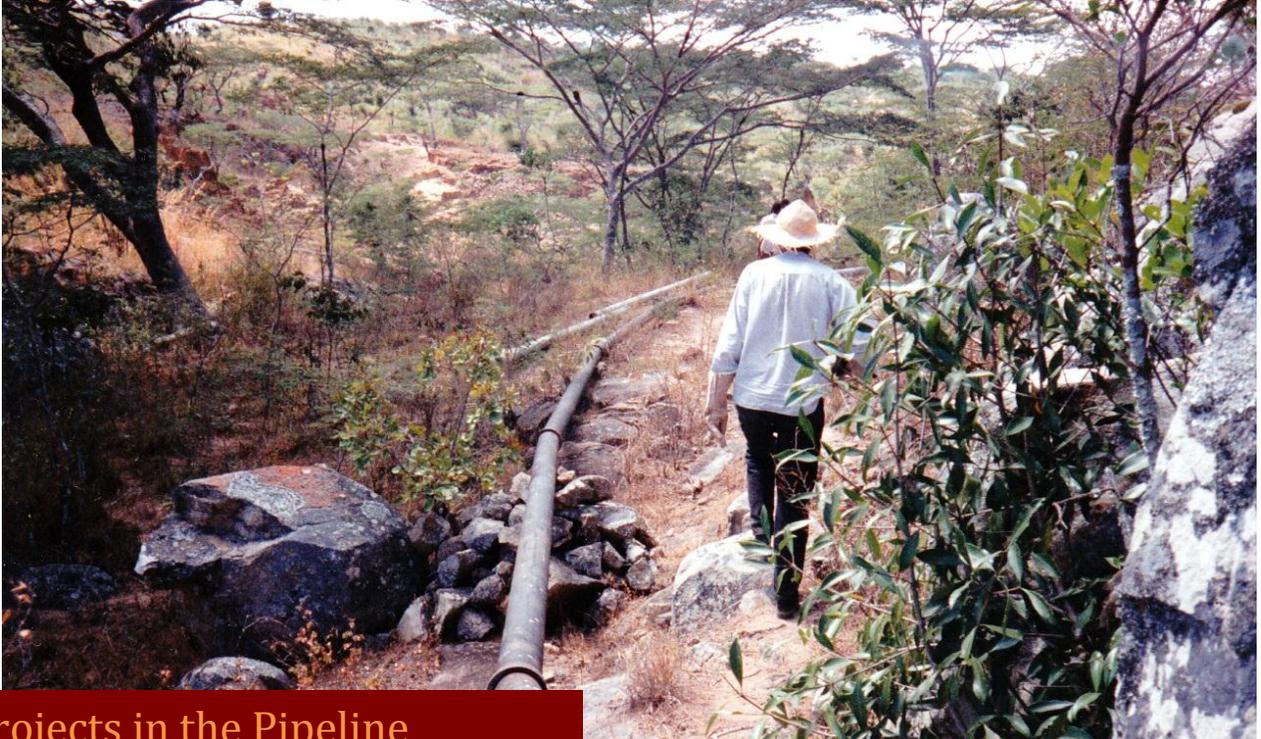


- terms and languages could sensibly be made depending on key disaster risks in any area.
- 2) Medical vocabularies with extensions targeting particular disease risks (for example human pandemic influenza and the associated terminology of hygiene and disease containment).
 - 3) Environmental terms relevant to habitat protection or sustainability issues.
 - 4) Specific vocabulary associated with the planning, implementation, evaluation and reporting of humanitarian and development programs. In some cases these terms may be specific to individual agencies.

The Kamusi Project works with agency partners to determine and define the terminology concepts within their domain. Kamusi will locate specialists who can work on the languages where those agencies are active, then guide those experts through the translation and community review processes. When the subject glossary for a language is complete, it will be made available to the public in numerous open source formats, thereby promoting its usefulness as a widely accepted communications resource.

With each project, the additional specialty terms and languages are added to an online database. This improves the ability of the international humanitarian community to communicate effectively in the regions where they are active. In this way, the Kamusi Project and its partners will speed the day when language diversity is harnessed as a tool to be used toward improved international cooperation, rather than remaining a barrier to a program's success.





Projects in the Pipeline

To achieve the Kamusi Project's ambitions of documenting every word in Africa, we are working with numerous partners around Africa and beyond, on a variety of linguistic and technical fronts. The modular design of the Kamusi Project architecture means that we can work on many components simultaneously, for different languages and different types of terminology. We have partners and plans for several new projects that can get underway rapidly once funding is available. Among these:

- **Language for Health.** This project will bring together more than two dozen partners in the production of health-related terminology, document translation, and software tools. With an initial focus on the pressing public health needs surrounding maternal health and HIV/AIDS, partners in at least ten countries will develop an open dataset that will adhere to the standards of the World Health Organization and the International Standards Organization. Results will be widely distributed through KamusiTERMS (Kamusi for Technology, Economy, Rights, Medicine, and Science), our initiative for coordinated participatory terminology development throughout Africa in numerous technical domains.



- ***Taifa Leo Corpus Mining.*** Kenya's oldest Swahili-language daily newspaper will make available digitized versions of its entire archive, dating back to 1958. Working with the University of Nairobi, we propose to establish a research database containing the full corpus. Using advanced linguistic analysis to mine the corpus, we will be able to harvest for the Swahili dictionary every word that has appeared in the newspaper during more than 50 years, along with usage examples and time trend data.
- **SALT Across the Sahara** (Songhay and Amazigh Lexicography and Terminology). We intend to join forces with The Dictionary of the Amazigh Language that is aimed at creating a comprehensive lexicon for the Tamazight language family (a.k.a. Berber) spoken by North African minorities from Morocco to Egypt, and Songhay.org that is documenting the Songhay languages of Mali and Niger in the central Sahara. We will make the systems of our three projects interoperable and synchronize the databases. This unified dictionary will follow the path of the great salt caravans that for centuries bridged the Sahara to facilitate the flow of trade and knowledge.
- **E-Gikuyu, E-Tswana, E-Kabuverdianu.** We are currently developing PALDO, the Pan-African Living Dictionary Online, which will interlink numerous languages through the dictionary building and dissemination systems of the Kamusi Project. Work on any one language can proceed independently, as partners step forward. We currently have strong partners with whom we have developed solid proposals for Gikuyu (one of Kenya's major languages), Setswana (spoken throughout Botswana, and also a major official language in South Africa) and the Kabuverdianu language of Cape Verde. E-Gikuyu will additionally involve academic collaboration with the University of Nairobi and the School of Oriental and African Studies in London on the development of computer assisted translation tools and their use in trilingual translation among Gikuyu, Swahili, and English, and will serve as a model to work with several more languages of Kenya. E-Tswana will include advanced development of morphological analysis and speech recognition tools for Bantu languages. E-Kabuverdianu will additionally provide a much-needed Portuguese component that will enable future work in Lusophone Africa. PALDO components for many other languages are also waiting in the wings.

- **SMS to Extend Reach and Quality.**

Africa had 508.6 million mobile telephone subscribers as of October 2010, meaning at least half the population has access to text messages. We have a working prototype system to deliver dictionary data via SMS. However, people would not use such a service if they risked getting poor results, so we have also devised a model whereby queries that are not in the Kamusi database will be sent to team members for rapid completion. The database will thus be improved for future queries, and team members will be rewarded with mobile airtime. Using SMS to provide access to mobile subscribers is a neat technical solution to overcome Africa's low rate of internet penetration, but involves unavoidable telecom costs.

- **The Universal Compendium of Recurring Concepts (UnCore).**

No matter the language, people repeatedly seek certain types of information. This information is often available online through databases, news outlets, or other informative websites – but usually only in one or a limited selection of languages. For example, many people search for current weather

information, and most of what they seek can be described by a finite list of terms, such as “high temperature,” “wind speed,” and “occasional rain showers.” If those concepts were stored in a multilingual database, content producers could link their data to the concept, and users would see the concept displayed in their preferred language. Taking localization to a new level, data producers will be able to provide their core data in any language without ever attempting their own translations. UnCore will use the cooperative systems developed for PALDO and KamusiTERMS to produce a range of vocabulary lists that can be queried by remote websites, for categories ranging from banking to transportation to sports. While the idea for UnCore grows from a desire to make recurring data available to speakers of African languages, linguistic walls similarly block all sorts of data flows within Europe, Asia, and the Americas. UnCore will therefore be open to any language for which participants are available to help develop the data.





- **Community Dictionary Building: Word-a-Day for Your Language.** With 2000 languages spoken in Africa, it will be a long time before we can launch formal components for each one. However, we can provide the space for languages to incubate their own dictionaries within PALDO, using the social power of the cloud. Through networks such as the African Languages Group on Facebook, we can reach thousands of speakers of African languages. We propose a system for volunteers to register for their language. They will be sent one word a day in a transnational language, and provide the equivalent term and definition in their mother tongue. If they do not reply within 24 hours, the term will be returned to the general pool.
- **Translation Tools Push/Pull.** Our partners at translate.org.za have developed sophisticated free and open software for computer assisted translation (CAT), called Virtaal and Pootle. We propose to integrate these tools tightly with our dictionary and terminology databases. When a term within their document is in our system, the translator will see the dictionary entry in a sidebar. When the term is not in our system, the translator will have the option to provide the information, which will then be processed by project editors. Documents that are translated with an open license will be included in translation memory for future users of the CAT tools, added to a corpus for machine translation development, and sentences will be harvested as usage examples of particular terms for the online dictionaries.
- Also on the horizon: Video dictionaries of African sign languages, multilingual mathematics primers, automatic machine translation among Bantu languages...



The Kamusi Project Business Model

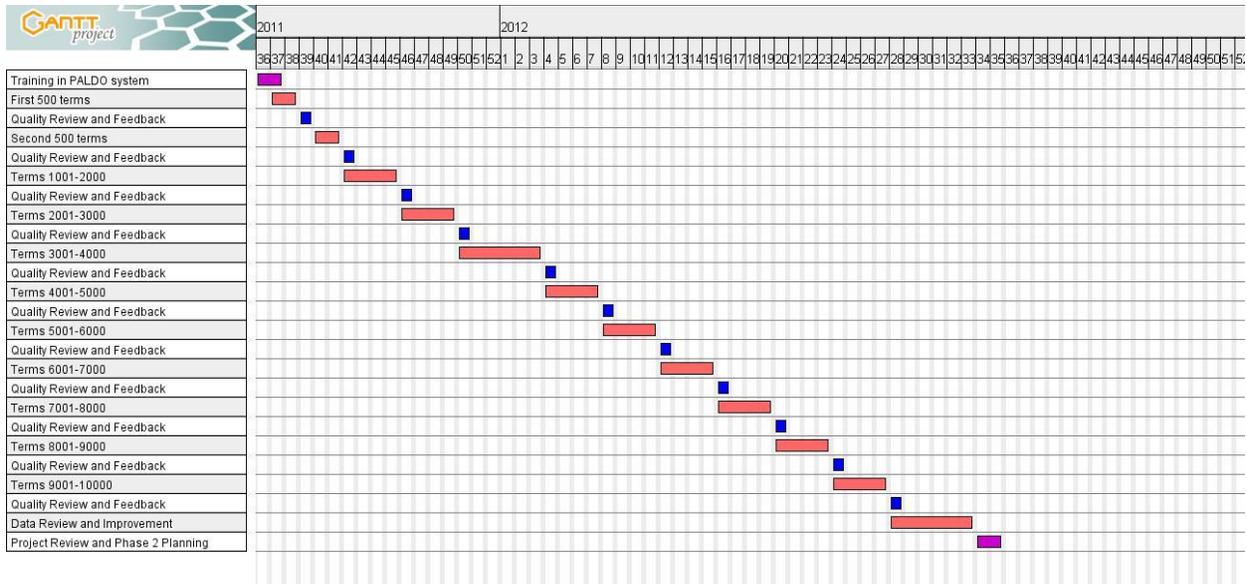
The Kamusi Project creates dictionaries, terminology glossaries, and learning tools for African languages based on two principles:

- 1) All of our resources are produced or reviewed by experts.
- 2) All of our resources are made available to the public for free.

Producing our data is expensive, because our partners need to be compensated for their time. Distributing our data is also expensive, especially the cost of keeping our programming up-to-date with emerging technologies and new developments. Computer equipment and software, servers to handle millions of queries a month, training of partners in the use of our system, communications – all of these things are essential for the project to maintain and expand its resources. And yet, we give away everything we produce, to anyone, for free.

We have implemented a number of methods of generating income:

- Advertisements on the website through Google AdWords
- Sale of merchandise with the Kamusi logo through CafePress
- Affiliate program book sales through Amazon
- Membership subscriptions
- Donations from users



Together, these efforts generate enough to keep the servers running, but no more. Because the project does not earn revenue from what it produces, the business model depends on support for the production work itself. We have a clear sense of the costs involved in creating entries for a dictionary or a technical glossary. Most of these are one-time expenses. For example, creating 10,000 dictionary entries for the most common words in a particular language costs approximately \$40,000:

- Set up costs for data model, training in software, and travel expenses*.....\$10,000
 - \$2/term for the lexical work on a single entry.....\$20,000
 - \$1/term for integrating an entry within PALDO.....\$10,000
- \$40,000

(Sample project timeline above)

Once we have the data in the system – that is, a permanent resource for a language, available to the public for free – ongoing expenses for disseminating and updating that data are low. The challenge, from a business perspective, is finding parties interested in funding the various components that constitute the Kamusi Project vision.





A combination of four types of support can sustain the project:

- 1) Revenue from the general public, as indicated above, should provide the baseline funding to keep the lights on. As the project's resources expand, so will the user base – as will, we hope, the ability of the project to attract members and other ongoing forms of community support.
- 2) Grants from agencies that fund academic and development activities are highly appropriate for much of the work of the Kamusi Project. We partner with leading scholars and specialists throughout and beyond Africa to produce high-quality lexicons suitable for academic peer review. The drawback of this funding mechanism is that research grants for African languages are scarce, so years can pass between the start of planning and the availability of funding.
- 3) Organizations and companies with specific interests can support the development of technical glossaries for their particular domains, within the KamusiTERMS initiative. For example, a health organization focused on ophthalmological treatment could finance an eye-care glossary that would improve communications between patients and medical staff



while an airline with destinations around Africa could sponsor an aviation glossary that would ensure that local maintenance crews with different linguistic backgrounds all follow the same installation procedure for the same part. Although few organizations have the capacity to produce their own glossaries, taking advantage of the Kamusi Project's systems and networks to navigate Africa's complicated linguistic environment is a low-cost way to build a resource that will be of immediate and continuing use in many specialized fields.

- 4) Foundations and other charitable entities can fund particular components, such as a dictionary for the language of a certain region, or provide general support that the Kamusi Project can allocate to emerging needs.

By seeking funding partners who can recognize the value of what we create, and invest in its realization, the Kamusi Project can cover the costs of the necessary work as we move steadily toward the goal of providing the public with every word in Africa, for free.

User Profile

The Kamusi Project receives about 60,000 visitors a month to the website of the Internet Living Swahili Dictionary, who conduct about one million lookups each month. Users arrive from virtually every country in the world, including substantial numbers from Africa. In addition, many people visit the site once but use it often, by downloading our content for free for offline use. African usership is expected to increase rapidly during the next few years as internet access on the continent is improved and as the Kamusi Project rolls out new mobile access features and new dictionaries and terminologies for languages around the continent.

When the project began in 1994, it was conceived primarily as a tool for university students in the United States studying Swahili. While they remain an important audience (with notable traffic spikes around exam time), analysis of our discussion forums and anonymous user logs reveals that we now serve a much wider range of users:





Who?

University students studying Swahili (Europe, Americas, Australasia, non-Swahili Africa)
African students at US or European universities
Students at African universities
UN, World Bank, and other multilaterals
Non-governmental organizations
Zoos, museums, botanical gardens
Hospitals, health insurers, health agencies
Elementary and secondary schools in Africa
Elementary and secondary schools out of Africa
Software manufacturers
Computer and mobile phone producers
Law firms, police, justice agencies
African government agencies
European and US government agencies

Why?

Learning vocabulary and grammatical structure, writing, reading literature
Help with general studies and daily interactions
Understanding coursework and texts in English
Official document and meeting translations
Providing field services, translating documents
Interpretive displays, research, naming
Patient interaction, disease monitoring and prevention
English learning and Swahili studies
Introducing global awareness
Localizing software for African markets
Building products for African markets
Legal cases (prosecution and defense), prisoner communications
Communicating with the public
Social and emergency services for immigrants, diplomatic and security communication



Shipping companies	Cargo and packages going to, from, or through Africa
Multinational firms	Selling to, buying from, or investing in Africa
African firms	Contacting overseas markets, sourcing foreign supplies
Name seekers	Naming a business, boat, pet, child, etc.
Pen-pals and the lovelorn	Reading and writing letters
The curious	Come across Swahili in a film or reading
African office workers	Correspondence, using localized software
African readers and internet users	Reading Swahili and foreign books and websites
Journalists (print, radio, TV, bloggers)	Interviews, transcriptions, source documents
Translators	Technical terms and nuance
Scholars	Linguistics, literature, field research
The regulars	Bilinguals or Swahili speakers who habitually look up words

Kamusi People: Personnel and Partners



Kamusi Project Staff



The **Executive Director** of the Kamusi Project is the organization's founder, Martin Benjamin. Dr. Benjamin received his PhD in Anthropology from Yale University in 2000, with a research focus on poverty, health, and development aid in Tanzania. After teaching Swahili and Anthropology at the university level, he left in 2006 to devote full time to developing resources for African languages. He is currently a research collaborator in Linguistics at the University of Lausanne, in Switzerland, where he lives with his wife and baby daughter.

The Board of Directors of *Kamusi Project USA* currently has five members:



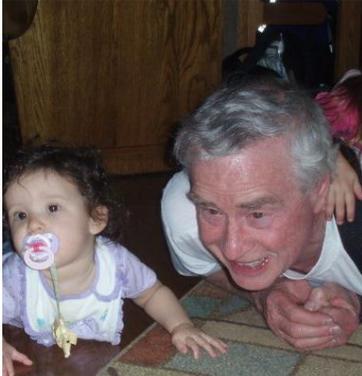
President: Warigia Bowman holds a doctorate in public policy from Harvard's John F. Kennedy School of Government. She is keenly interested in rural development issues in Africa and in America. She has significant experience consulting with African NGOs on issues around politics and technology, and has worked with NEPAD, ATPS and the Bill Gates Foundation. She has taught at Kabarak University in Kenya, and is currently teaching at American University in Cairo. She is working on a book on information technology policy in East Africa. Dr. Bowman is married and the mother of three children.



Secretary: Charles Riley is the catalog librarian for African languages at Yale University Library, and co-founder of Athinkra, LLC. A large part of his work involves integrating technical support for African languages into existing systems for information processing, through contributions toward standards development, research on fonts and scripts, and encouraging publishing activity. He is most familiar with the languages of West Africa, where he served as a Peace Corps volunteer and can still occasionally be found.



Treasurer: Susan Werner is an insurance/securities compliance consultant with Northwestern Mutual. She is the founder and director of Hearts in Unity, a humanitarian aid NGO with a mission to feed, clothe and educate orphan and other at-risk children in Tanzania, and is a board member and secretary of the Maasai Outreach Mission NGO (Kenya). She joined the Kamusi Project as an active user seeking to improve her Swahili language skills.



Director: Thomas J. Hinnebusch is Professor Emeritus of Linguistics and African Languages at the University of California at Los Angeles, where he taught Swahili and Linguistics for 32 years. He is currently Director of the UCLA Language Materials Project at UCLA's Center for World Languages. His research interests include historical/comparative Bantu studies, Swahili lexicography, and the lexicography of Swahili poetry.



Director: Don Osborn holds a PhD in resource development from Michigan State University, and has diverse experience in localization of information technology in African languages, agricultural development, community natural resource management, civil affairs, and teaching and training. His international experience includes 12 years in Africa, 2 in China, and one in Israel. Publications include a Fulfulde (Maasina)-English-French Lexicon, and African languages in a Digital Age (HSRC & IDRC, 2010). He is the founder and director of Bisharat, Ltd., a language, technology and development initiative.



The Board of Directors of *Kamusi Project International* currently has three members:



President: Thomas Bearth is Professor Emeritus in African Linguistics at the University of Zurich. He has conducted theoretical and applied linguistic research in West Africa for 40 years. His numerous publications include the Internet and CD-ROM university courses ALI-Akan and ALI-Swahili, and two books on the Tura language of Côte d'Ivoire. His current research interests focus on the interface between language diversity, communication and knowledge management for poverty alleviation, health and crisis management.



Secretary: Hugh Peterken served as Chief Information Officer of the International Federation of Red Cross and Red Crescent Societies. In this role he was responsible for developing and managing the information and communications technologies for the Red Cross in 86 countries. He has worked with a number of not-for-profit organizations at the national and international level, with a focus on gearing technology initiatives to support the outcome objectives of humanitarian programs.



Treasurer: Mohomodou Houssouba has a PhD in English, with a focus on pedagogy. He is a writer and linguist working on language policy and education reform. Resistance to the adoption of national languages has motivated him to explore the relationship between school authorities and other stakeholders in Mali. Since 2005, he has coordinated the translation of software into Songhay. He initiated the online Songhay dictionary project (Songhay.org) in 2008. He lives in Basel, Switzerland.





Our **technical development** is conducted by several organizations and individuals, depending on specific programming projects. All of our programming is done with a vision to support the growth of IT development capacity within Africa.

- *Translate.org.za*, based in Pretoria, South Africa, is developing our next-generation multilingual online dictionary editing software (PALDO) and terminology infrastructure (KamusiTERMS). Project programmers are from South Africa and Egypt.
- *Suuch Solutions*, based in Accra, Ghana, manages our website on a daily basis. Suuch and its sister organization, *Kasaboron*, were founded by Paa Kwesi Imbeah as an outgrowth of his work programming for the Kamusi Project at its original home, Yale University, where he studied.
- *Appfrica Labs* is a technology incubator based in Kampala, Uganda, with which we have worked on dictionary software and Luganda terminology development. We are currently seeking support to work together on innovative SMS applications for data editing and dissemination.
- Arthur Buliva is a freelance programmer from Nairobi, Kenya, who is developing a platform-independent tool to use and contribute to project data offline.
- *IT46* is a Swedish-based IT consultancy dedicated to knowledge transfer to promote social change. We worked with IT46 on our first-generation terminology software, Glossmaster, and the Afrigen initiative to develop computer locales for 100 African languages, among other projects.



The Kamusi Project works with numerous partners who contribute toward the ultimate objective, documenting every word in Africa. Some of the organizations with which we have worked, currently work, or have proposals under development include:

- Alt-I, the African Languages Technology Initiative in Ibadan, Nigeria
- ANLoc, the African Network for Localization, an international consortium creating technical resources for African languages
- Arabize, a leading Egyptian localization company based in Cairo
- The Center for Next Generation Localization, in Limerick, Ireland
- The Centre for Language Studies in Zomba, Malawi
- Educ, Effort de Démocratisation de l'Usage des TIC pour les Communautés locales, an NGO in Kinshasa, DRC
- The Goethe Institute of Yaoundé, Cameroon
- Infoterm, the International Information Centre for Terminology, of Vienna, with a mandate from UNESCO and the International Standards Organization
- International Development Research Centre, a Canadian government agency
- kinyarwanda.net, a dictionary project from Kigali, Rwanda
- Meedan, a social network that uses machine and human translation to build communications between Arabic and English





- The Ndimi Centre for African Languages, in Nairobi
- Ngozi University of Bujumbura, Burundi
- Open Society Institute of Southern Africa, a leading Johannesburg-based foundation that works in Angola, Botswana, DRC, Lesotho, Malawi, Mozambique, Namibia, Swaziland, Zambia and Zimbabwe
- The Rosetta Foundation, a not-for-profit translation organization based in Ireland
- Service Employees International Union 1199SEIU Funds College Preparatory
- Songhay.org, a dictionary project with collaborators in Mali, Niger, Switzerland, and Ireland
- TEDECO, Tecnología para el Desarrollo y la Cooperación, based at the Faculty of Informatics of Madrid Polytechnic University in Spain, a group that collaborates on improving ICT infrastructure with academic institutions in developing countries
- The Translation Bureau, a Canadian government agency
- Translators Without Borders, a not-for-profit translation organization based in Paris, France
- tzLUG, the Tanzania Linux Users Group, in Dar es Salaam
- University of Botswana, Department of English, in Gaborone
- University of Lausanne, Department of Linguistics, Switzerland



- University of Nairobi, School of Computing and Informatics, in Kenya
- University of South Africa, Department of African Languages
- University of Tizi-Ouzou in Algeria
- Vox Humanitatis, a not-for-profit organization headquartered in Alba, Italy, focused on less resourced cultures and their languages
- The World Language Documentation Centre, based in Wales



Photo Captions and Credits

Page	Explanation and photographer (if known)
Cover	Farmer leaving for paddy rice harvest on the Niger River, Bagoundié, Mali. Photo by Mohomodou Houssouba.
2	The Kamusi Project logo depicts a sailing ship reminiscent of those that have connected Africans for centuries.
3	Girl carrying palm nuts, Mchukwi, Tanzania. Photo by Martin Benjamin.
4	ZenAir, an experimental aircraft, in Isimikinyi, Tanzania. Photo by M. Benjamin.
5 top	Kaburo Kobia of the Kenya ICT Board addresses the 2011 ANLoc conference near Nairobi. Photo by M. Benjamin.
5 bottom	Growing bananas in Isimikinyi, Tanzania. Photo by M. Benjamin.
6	Women carry headloads of cement on a construction site in Kigali, Rwanda. Photo by M. Benjamin.
7	Martha Chaula, Chair of her sub-village, Isimikinyi, Tanzania. Photo by M. Benjamin.
8	Boys returning from school, with woman behind preparing water for sale, Ikwiriri, Tanzania. Photo by M. Benjamin.
9	Safety signs in English and Swahili aboard a Precision Air turboprop, East Africa. Photo by M. Benjamin.
10	Mandinka greetings in the Garay script, Derklé neighborhood of Dakar, Senegal. Photo by Charles Riley.
11	Members of tzLUG work on Swahili ICT terminology, Dar es Salaam, Tanzania. Photo by M. Benjamin.
12	Horseback on a farm in Orange Free State, South Africa. Photo by M. Benjamin.
13	Market in Accra, Ghana. Photo by M. Benjamin.
14	Eye care at Muhimbili Hospital, Dar es Salaam, Tanzania. Photo by M. Benjamin.
15	HIV billboard in English in Kampala, Uganda. Photo by M. Benjamin.
16 top	Village pharmacy in northern Tanzania. Photo by M. Benjamin.
16 bottom	“Vumilia” means “tolerate.” House on Chole Island, Tanzania. Photo by M. Benjamin.
17	Baskets for sale at monthly market, Tambalang’ombe, Tanzania. Photo by M. Benjamin.
18 top	Rowboat on the Nile, southern Egypt. Photo by M. Benjamin.
18 bottom	Bibliotheca Alexandrina, Alexandria, Egypt. Photo by M. Benjamin.
19	Laying a pipeline near Malangali, Tanzania. Photo by M. Benjamin.
20	Preparing to lay a water pipeline in Isimikinyi, Tanzania. Photo by M. Benjamin.
21	Communications tower in Arusha, Tanzania. Photo by M. Benjamin.
22	Kids moving a building stone with levers in Isimikinyi, Tanzania. Photo by M. Benjamin.
23	Small store in Tanzania. Photo by M. Benjamin.
24	Mat weaving on Chole Island, Tanzania. Photo by M. Benjamin.
25	Receiving payment on Chole Island, Tanzania. Photo by M. Benjamin.
26	Receiving payment in Tambalang’ombe, Tanzania. Photo by M. Benjamin.
27	National University of Rwanda, Butare, Rwanda. Photo by M. Benjamin.
28	Boy at play on Mafia Island, Tanzania. Photo by M. Benjamin.
29	Billboard in Giza, Egypt. Photo by M. Benjamin.
30	Kamusi programmers Henry Addo and Arthur Buliva with project director, Akosombo Dam, Ghana.
31 – 33	Photos of staff and directors from various sources.
32 bottom	Dhow at sunset, Zanzibar, Tanzania. Photo by M. Benjamin.
33 bottom	Small boat in Zanzibar, Tanzania. Photo by M. Benjamin.
34	Kamusi programmers Arthur Buliva, Jojoo Imbeah, and Alaa Abd El Fattah meeting in Nairobi. Photo by M. Benjamin.
35 top	Kamusi Project partner Emanuel Feruzi of tzLUG in Cairo, Egypt.
35 bottom	Net fishing on the Niger River, Bagoundié, Mali. Photo by Mohomodou Houssouba.
36	Gathering building stones in Isimikinyi, Tanzania. Photo by M. Benjamin.
37	Love taxi, Alexandria Egypt. Photo by Dwayne Bailey.
39	Dhow in Zanzibar, Tanzania. Photo by John Ho.



The Kamusi Project began in 1994 with the development of the Internet Living Swahili Dictionary, the first online resource for an African language.

The project now has two principle foci, a unified general dictionary for African languages called PALDO (the Pan-African Living Dictionary Online), and specialized terminology development through KamusiTERMS (Kamusi for Technology, Economy, Rights, Medicine, and Science).

- *PALDO will bring together numerous languages within the rubric initially created for Swahili, which will be highlighted in a study of bilingual dictionaries created with Title VI support from the US Department of Education.*
- *KamusiTERMS began with a large Information Technology terminology set for 10 African languages in addition to English and French.*

*Both elements build on a unique model of scholarly guidance and community input toward achieving our goal: **Every word in Africa.***