

All words from all languages in one dictionary

September 18 2015, by Cécilia Carron



Credit: Ecole Polytechnique Federale de Lausanne

The universal online dictionary Kamusi has just added 1.2 million terms from several databases in its quest to translate all the meanings of every word in all the world's languages. Three African languages and 200,000 words of Vietnamese will soon follow.

Kamusi, which means [dictionary](#) in Swahili, aims to translate all the

meanings of [words](#) from 7,000 languages from around the world into all other languages, and it will include definitions and usage examples. This vast project, which began twenty years ago, is growing at an exponential rate. Some languages, like English and Swahili, are already largely available.

More than a translator, this dictionary delves into the words so that two meanings of the same term can no longer be confused. Take the English word light: in French, the automatic translator will translate it as léger or lumière depending on the context. Yet the chosen meaning is often incorrect, and the text becomes gibberish. The problem is even more pronounced when it comes to less common languages. Kamusi, which is now an NGO, intends to overcome these pitfalls. It includes both definitions and examples. In addition to being ambitious, the project is also a major technological challenge because a massive amount of data needs to be stored, organized and made accessible.

More than 1 million words in several months

This summer, EPFL's Distributed Information Systems Laboratory added 1.2 million words. Researchers imported the entire participatory translation website Open Multilingual Wordnet (OMW), making improvements to the search process and the links between the various meanings of words. The Kamusi website shows, for example, if a translation was done by a computer and if the meaning was confirmed by a human being. It also lets users see groups of words with a similar meaning in other languages along with the expanded entry for each term.

Apart from adding existing databases, the Kamusi team counts on human knowledge as well. People who have mastered several languages are translating as many words as possible. This participatory approach works especially well for little-used languages. So much so that the government of Mali has just invested 5 million CFA francs – around 78,000 Swiss

francs – in order to add words from three of its local languages. This money, a tidy sum for that country, will mostly be used locally to use any existing databases or to verify, little by little, the terms added by bilingual people.

Africa, Vietnam, Europe: an international dictionary

"This is the first time that an African government is investing its own money in this project, which is a remarkable sign of confidence," said Martin Benjamin, who founded the Kamusi project and is currently a researcher at the Distributed Information Systems Laboratory. This agreement also marks the cornerstone of a future collaborative effort to make this universal online dictionary the first electronic tool used by the African Academy of Languages (ACALAN), which is the intergovernmental body of the African Union tasked with promoting and developing the continent's languages. "Africa's growth and access to knowledge require a dictionary of this sort so that people can communicate with each other and with the outside world," said Benjamin. The dictionary is also expanding in Asia and Europe. 200,000 Vietnamese words will be added thanks to a small subsidy provided by the Swiss State Secretariat for Education, Research and Innovation. And the inclusion of other databases with Bulgarian, Greek, Croatian, Slovene and Swedish words is also in the works.

How Kamusi started: In 1995, Martin Benjamin, who now works at the Distributed Information Systems Laboratory, tried to learn Swahili in preparation for an anthropological study in Tanzania. He was unable, however, to find an English-Swahili dictionary. He then came up with the idea of creating a participatory website – at a time when the internet was in its infancy – in order to create a Swahili-English database. Work on this [language](#) pair has progressed quite far at this point. Kamusi has now become an NGO and is based in Geneva.

Provided by Ecole Polytechnique Federale de Lausanne

Citation: All words from all languages in one dictionary (2015, September 18) retrieved 19 September 2024 from <https://phys.org/news/2015-09-words-languages-dictionary.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.