

## So you want to produce a dictionary?



You have a target audience in mind and an idea about what should be in it, but how do you go about making it a reality? You know that the backbone of a dictionary is a corpus of words, but how do you go about collecting hundreds (or more likely, thousands) of “vertebrae” that form that backbone? Do you start with a list of words in one language and have someone translate them? Do you record stories told by local speakers and then analyze them? Or is there another approach?

## Consider the Rapid Word Collection method

One main objection to starting a dictionary project with a wordlist in another language is that the words are not naturally generated. The problem with a text-based approach is that it takes a long time to accumulate a large number of words. Rapid Word Collection (RWC), on the other hand, is both fast and natural.



### 2.1.1 Head

Use this domain for the parts of the head.

- (1) What words refer to the head?
  - *head*
- (2) What are the parts of the head?
  - *feature, scalp, back of the head, chin, temple, poll, skull, brain*
- (3) What are the parts of the face?
  - *face, forehead, cheek, chin, cleft (in chin), dimple (in cheek), jaw*
- (4) What are the parts of the neck?
  - *neck, nape (back of the neck), throat, Adam's apple, wrinkle, fold in skin of neck*
- (5) What words describe the parts of a person's face?
  - *facial, thin (face), high (forehead), thick (eyebrows), long (eyelashes), beady (eyes), stubby (nose), fleshy (cheeks), narrow (mouth), double chin*

### How it works

Speakers from the focus language community—at least 25 to 30—gather together in the same location for two weeks. Each day these individuals work as a team to collect as many words as possible on paper and then enter them into a computerized database. This is done with a series of word-association exercises, using a questionnaire that is organized according to meaning into nearly 1800 different categories, like the one illustrated to the left.

The team is divided into several smaller groups—six groups of three or four people who do the actual word-collection, two or three people who write short definitions (glosses) in a language of wider communication for the newly-collected words, two to four typists who type the words and glosses into the computer, and several individuals with managerial roles. Each word-collection team has a leader who reads the questionnaire and translates it into the focus language for the benefit of the other team members, one or two language experts who say the words that come to mind as the word-association prompts are spoken by the group leader, and a scribe who writes down the words his team members suggest.

## The Results

At the end of the two-week period, a lexicon of all the words collected (typically 10,000-15,000) can be printed. After the workshop, a few individuals are selected to clean up the raw data—correcting typos, eliminating duplicate entries, and so on. Adding grammatical information, fleshing out definitions, including example sentences, etc., is done prior to publishing the desired dictionary. The amount of time needed to accomplish all of the above will depend on the scope of those efforts.



While either *WeSay*<sup>1</sup> or *Fieldworks Language Explorer*<sup>2</sup> (*FLEX*) can be used for data entry, currently *FLEX* is the software of choice for the preparation of a lexical database for publication as a dictionary. It is designed to facilitate the selection of the exact subset of words that should be included in a particular publication, with the ability to filter out everything else. Whatever the scope of the dictionary that is targeted initially, the database can serve as the core of any future publications as well, probably addressing a different audience, without the need to ever again raise the question of how to gather the “vertebrae.”

## The Next Step

For more information on the Rapid Word Collection methodology, see [RapidWords.net](http://RapidWords.net) or contact Kevin Warfel at [kevin\\_warfel@sil.org](mailto:kevin_warfel@sil.org).



*You can't preserve the culture without knowing about the language.* —Abdul Razak Sulemana, Master of Philosophy-Linguistics, University of Ghana-Legon

*One of the important things is the dictionary. So if the dictionary is able to come into existence, Wow! So with a dictionary, it will give easy access for us to actually teach our children in our own language.* —Peter Adaawen, Project Coordinator, Buli Literacy Project

*To be able to do this in ten days is revolutionary and I will urge languages that would like to make dictionaries to adopt this methodology.* —Dr. George Akanlig-Pare, Professor of Linguistics, University of Ghana-Legon

---

<sup>1</sup> *WeSay* helps non-linguists build a dictionary in their own language, <http://lingtransoft.info/apps/wesay>

<sup>2</sup> *FLEX* is the lexical and text tools component of SIL *FieldWorks*, <http://lingtransoft.info/apps/flex>