

A New Mass Elicitation Technique: The Dictionary Development Program

Susan Shore and René van den Berg

In a normal field work situation, lexical elicitation is typically a slow and time-consuming process, generating only a limited section of the total vocabulary of a language. The Dictionary Development Program (DDP), a new elicitation tool developed by Ron Moe (SIL) in East Africa, addresses this issue by employing the technique of rapid mass elicitation based on lists of questions arranged by semantic domain. These lists (the domain templates) are exhaustive and universal, and when used in an efficiently organized workshop can result in the collection of 10,000 words in two weeks.

Introduction

There are many ways to build a dictionary. One can take word lists, add words from texts, interlinearize texts, note down words as you talk with people, add subentries under a root entry, think of words associated with semantic domains, use Melchuk's lexical functions, networks of meaning, Murdock's *Outline of Cultural Material* (2005), etc. No matter what we do to compile a dictionary, it is a mammoth, time-consuming task.

There are 5 basic stages in dictionary building.

1. Collect words
2. Generate additional fields
3. Define words
4. Edit each entry for publication
5. Prepare front and back matter

What is the DDP?

Ron Moe of SIL has developed a way to quickly gain a large repertoire of words for **starting** a dictionary and a thesaurus. In a two-week workshop it is possible to collect somewhere between 9,000 and 12,000 words. He has called this the Dictionary Development Program (DDP). The DDP greatly helps with stages 1-2, (collecting words and adding more fields) and material to help with stage 3 (writing good definitions) is currently being developed. There is much more to producing a dictionary than merely collecting words, however that is a necessary first step, or can be used later in the project to fill in some gaps.

In order to help **systematically** collect a large number of words in a short time, Ron Moe has developed a set of comprehensive questions and example words based on **1700 semantic domains** in the following areas:

1. The physical universe
(world, water, plants, animals)
2. Person
(body, the senses, health, life stages)
3. Language and thought
(soul, mind, volition, emotion, communication, teaching)
4. Social behaviour
(relationships, kinship, belonging, work, behaviour, government, religion)

5. Home
(rooms, food, clothing, fire, cleaning, sleep)
6. Work and occupation
(agriculture, animals husbandry, hunting & fishing, working with buildings, occupations, tools, finance, business organization)
7. Physical actions
(posture, movement, move something, organizing, hiding, physical impact)
8. States
(quantity, measure, quality, colors, temperature, material composition, appearance, time, speed, location)
9. Grammar and discourse
(general, part of speech, propositions, semantic constituents related to verbs, logical relations, proper nouns)

The list of words is intended to be both **exhaustive and universal**, although there will need to be a few tweaks, additions and subtractions to make it a good fit for each language.

An example of domain template:

3.4.1.2.1 Laugh

(Categorized under 3 Language and thought, 3.4 Emotions, 3.4.1 Good emotions)

Use this domain for words referring to the expression of good feelings, including laughing--the sounds a person makes when he is enjoying himself or thinks something is funny.

Related domains: Humor; Mock

- (1) What general words refer to laughing? *laugh, laughter, laugh at (something), have a laugh*
- (2) What words describe the way a person laughs? *laugh aloud, laugh out loud, chortle, heehaw, guffaw, cackle, yuck it up, squeal with laughter, canned laughter*
- (3) What words refer to laughing a little or quietly? *laugh silently, chuckle, giggle, ripple of laughter, titter*
- (4) What words refer to laughing a lot or loudly? *roar with laughter, laugh uproariously, howl with laughter, shriek with laughter, hearty laugh, peals of laughter, laugh helplessly, laugh your head off, laugh hysterically, have hysterics, roll in the aisle, almost die laughing, double up, convulsed with laughter, split your sides*
- (5) What words refer to laughing for a long time? *can't stop laughing, have the giggles, fall over/about laughing, have a good laugh, laugh till the tears roll down your cheeks*
- (6) What words refer to laughing at someone because they did something wrong? *laugh at, mock, titter, snicker, snigger*
- (7) What words refer to beginning to laugh? *start laughing, burst out laughing, burst of laughter, dissolve into laughter*
- (8) What words refer to making someone laugh? *make someone laugh, amuse, crack someone up, have someone in hysterics, have someone in stitches, tickle, ticklish*
- (9) What words describe someone or something that causes people to laugh? *laughable, amusing, comic, comical, absurd, facetious, humor, humorous, sense of humor*
- (10) What words refer to the sound of a person makes when he is laughing? *ho ho ho, ha ha ha*
- (11) What words refer to trying not to laugh? *(try to/can't) keep a straight face, stifle a laugh*

(12) What words refer to smiling? *smile, grin, wide grin, big smile, half smile, slight smile, smirk, crack a smile, beam*

(13) What other ways do people show they feel good? *clap, applaud, applause, cheer, jump up and down, jump for joy, dance for joy, shout for joy, make merry, revel, weep for joy, ululate*

(14) What do people say when they are happy? *Hurrah! Hooray! Yeah! Three cheers! Hallelujah! Yippee! Yahoo!*

Each domain has a basic set of questions as illustrated above, and some suggestions. The suggestions are not intended to be words which are translated word for word, but are to help people think through what words are in their languages.

More about the program

Ron Moe has spelt out detailed **step-by-step procedures** which help in collecting data. He has many practical ideas in what to prepare beforehand, instructions for how to lead a workshop, instructions for the participants who will collect the words, instructions for people doing data entry into the computer, and also a file on pitfalls to avoid. A basic lexicography course is also available on powerpoint.

If using the Toolbox program for data entry, cc tables have been developed which help automate some of the functions to take the tedium out of data entry. You are able to do bulk edits, e.g. to automatically generate CV patterns.

A sample entry will look like:

\lexeme	udá
\cv pattern	VCV
\part of speech	n
\definition English	rain
\index semantic	1.1.3.3
\semantic domain	Rain

Use of the material

Several languages in Africa, Malaysia, PNG and Indonesia have used the DDP to begin building dictionaries. People have used the material in various ways.

Scenario 1: The entire list of words can be collected in about two weeks by working full-time with 12 to 20 speakers of a language. Unless there are people trained in inputting data into the computer, this could leave you with a huge corpus of unprocessed data.

Scenario 2: A core group of people can be trained to collect the data, working with resource people from the community. This will take longer than in scenario 1, but experience shows that people are enthusiastic and desire to continue adding words and improving definitions.

Scenario 3: The words are collected over a two year period in conjunction with workshops on various topics, e.g. writers' workshops, worldview, kinship and relationships, health, grammar discovery, etc.

The data can also be used in very creative ways. It is intended that people first list all the words they can think of in a particular domain, before reading through the list of suggestions and

adding to their list. This is to prevent people trying to come up with translations for words which do not exist in their languages. There can be competitions between groups, e.g. I got people from four languages to list all the words they could think of to do with carry. Each group was able to add to its list by listening to the words the other groups had written. Or a weak group's folder can later be given to a stronger group. Word games can also be played. Brainstorming or doing spider diagrams (networking) is another way to collect words. The material can be supplemented with pictorial encyclopedias, charts, and other reference material.

I (Susan) led a one-week introductory workshop for four languages, between one and three people per language. In that time we covered 73 domains. By the beginning of the second day, the men were able to work independently, although I think we would have had better definitions if I had overseen the results better and done more teaching, as well as getting them each day to go over what they had done the previous day and improve their definitions.

Limitations/Problems

1. **Time:** Despite being able to collect data in a very short time, it still takes many hours of computer processing. If 15,000 words are collected, at one minute per entry, it would take at least 250 hours to write simple entries with a basic gloss, and domain categories. Definitions still need to be revised, notes added on usage of words, example sentences need to be added.
2. **Training:** It is good to spend **at least one day** training people in how to fill in the domains. Their level of education will also determine their level of independence. It also takes time to train people to understand about **parts of speech**, or to write good, complete **definitions**. Usually what they do is a basic wordlist. But my own dictionary is not much better. Invest in the training. Lead them through different domains. Play games, have competitions to see who can write the best definition. We need to systematically train in strategies for writing good definitions. Ron Moe is currently developing material which will help people think through some of the categories which need to be thought about to write good definitions. (See his samples in the Lexicography workshop material.) Everything takes time as we make several passes over a dictionary. It takes time to train people to do **data entry**.
3. **Listing words out of their context:** Some people have objected to the use of the DDP as words are listed out of context. But let's remember that although they are listed out of their syntactic context (and most dictionaries start off as wordlists), they are still being listed in their semantic context. And in the process of discussing definitions, people are using the words in example sentences (their syntactic context). Writing words associated with semantic domains is not going to help people note **how** a particular word is used. Seeing these words in context through their usage in texts and oral literature is the only way we can truly distinguish shades of meaning.
But having something written down is a place to start and is better than nothing.
4. **Dictionaries should be organized emically, not etically:** True. So use what you can from this material, throw out what you can't, and adapt the rest. E.g. you may need to add a sub-domain for stimulants such as tea, coffee, betelnut. Taxonomies may not be the same, e.g. animal classification. Nothing is ever set in concrete with dictionaries. Ron has attempted to

make this material **universal**, and has had input from many languages around the world. English is the medium, not just a western worldview. However it will still need to be adapted to some extent for each specific language. But don't throw the baby out with the bath water!

5. **Some domains are not relevant, or others need to be expanded:** So change, adapt, expand. This will be true no matter what list we use. Some domains will be totally irrelevant, e.g. snow, anteaters, flour. Some domains are not very productive, and the same word will be generated over and over, but use the English words to help with building a complete definition, e.g. Bariai (PNG) has a word *kadonga* which means 'behavior, habit, action, work'.
6. **The DDP is in English** But part or all has now been translated into French, Spanish, Portuguese, Swahili, Chinese, Thai, Nepali, Tamil, Amharic, Malay and Indonesian, and possibly some other languages.
7. **The DDP is still a work in progress.** Ron Moe recently released version 4 of the basic domains. Stage three is still under development, which will be a tool to help with writing good definitions.

Practical Advice

It is helpful if the orthography is established before beginning to do a lot of data entry in a dictionary, otherwise there will be many inconsistencies.

Definitions should be written for the words as they go, or at least gloss all the words. Some have just listed words, without adding glosses. Difficult!

Do an initial printout fairly soon as it is encouraging to people to see the results of their work. And it becomes a learning experience as they think about other words which could be added, and check what has been written as a definition.

Conclusion

Compiling a dictionary is a series of successive approximations. As we go over and over the material from various angles, we polish, expand and develop it further. The DDP is only **one of several ways** in which a dictionary can be compiled. It is a very useful tool. It is a very small part of a several year project. For beginners it will give a tremendous jump-start to your collection of words. For those in the middle, it will help fill in some holes, and add a systematic framework of semantic domains. The **stage 1 dictionaries** which will be produced will be flawed, definitions for many words will still be inadequate, yet they will still constitute a useful resource. Example sentences will still need to be added. Words will still need to be studied in their context and definitions expanded, different sense added, etc.

From the **sociological perspective**, holding a dictionary workshop is also a great way to build **synergy** among the people themselves, and to give them a feeling of success, particularly if there is a felt need to have a dictionary to preserve the language. It can improve public relations among groups where leadership is not very supportive of academic activity. People who are not interested in working to develop their language will become interested, especially when they see

the results, and there are new tools which can help in dictionary compiling and printing. Check out *Lexique Pro* for printing dictionaries by semantic domains, or *Toolbox* for data entry which also includes *Multi-Dictionary Formatter* for printing dictionaries.

Of those who have used the program, it has been given a resounding "thumbs up" and deemed "an excellent tool." It is another tool which is now available to us in the overwhelming task of compiling a dictionary.

The DDP, Lexique Pro and Toolbox can be downloaded from:
<http://www.sil.org/computing>

The preceding document was presented at the Tenth International Conference on Austronesian Linguistics (10-ICAL). To properly reference this work, please use the following format:

<LastName>, <FirstName>. 2006. <PaperTitle>. Paper presented at Tenth International Conference on Austronesian Linguistics. 17-20 January 2006. Puerto Princesa City, Palawan, Philippines.
<http://www.sil.org/asia/philippines/ical/papers.html>

For other papers that were presented at 10-ICAL, please visit <http://www.sil.org/asia/philippines/ical/papers.html>.