

CHALLENGES IN ENDANGERED LANGUAGE LEXICOGRAPHY

ASIALEX 2013 August 20th 2013

Antonia Cristinoi, Laboratoire Ligérien de Linguistique,
Université d'Orléans

François Nemo Laboratoire Ligérien de Linguistique, Université
d'Orléans

Background

When it comes to endangered minority languages, lexicography is faced with **specific limitations and challenges**.

Based on our field lexicography experience and on the writing of a **Palikur (Arawakan, French Guyana)** dictionary, this paper aims to present some of the shortcomings of dictionary production, and what should and could be done to address the specific challenges which have to be met.

Main topics

1. Palikur language and people
2. What is so special about endangered language dictionaries?
3. Frequent limitations of small language dictionaries
4. Collecting vanishing words
5. Translation/description issues

1. Palikur language and people



Palikur
communities
in the world



Palikur
communities
in French
Guyana

Palikur language and people

- A population living in French Guyana and Brazil
- Between 850 and 1000 speakers in French Guyana
- Total population : about 2000

Palikur language

- Linguistic family: Arawakan
- Complex grammar (classifiers)
- Multilingual environment

2. What is so special about endangered language dictionaries?

ELDs vs “classic” dictionaries

1. Authors

- one (or in the best of cases a small group) non-native author, most often a linguist, an ethnographer or a missionary working in a multilingual context
- professional staff of trained native speakers

☞ Bilingual dictionaries

ELDs vs “classic” dictionaries

2. Funding

- one time shot with no update possibilities
- hardly ever any new improved versions

☞ **Complicated methodological choices**

- Dictionaries for widely spoken and thoroughly described languages are definitely a profitable commercial enterprise, and are systematically re-printed and enriched

☞ **Easy funding**

ELDs vs “classic” dictionaries

3. Users and uses

- small number of users, most of them *scholars* but **more recently also**
- *indigenous people*
- *non-native speakers* interacting with indigenous people

-
- distributed world-wide **and recently**
 - very specific audiences (learner’s dictionaries, specialized dictionaries, historical ones etc.)

Endangered language dictionaries will thus be used:

- for research purposes;
- to document a specific language and more importantly a specific culture;
- to preserve a linguistic and cultural heritage which would disappear without written material;
- to help indigenous people communicate in a foreign (often dominant) language by finding the proper equivalents for indigenous words;
- to help non-native speakers understand native-speakers and their cultural background;
- to provide a stable orthography for the whole lexicon.

ELDs vs “classic” dictionaries

4. Ideological stakes

- transition from oral to written language practices
- making orthographical choices
- standardizing language
- deciding which language variety will constitute THE standard
- preserving language and culture
- keeping track of language evolutions
- documenting a specific field (i.e. astronomy)
- enforcing terminology and language use
- defending national languages and obviously
- encourage and facilitate cross-cultural communication

 **political consequences**

ELDs vs “classic” dictionaries

5. Data collection

- scarcity of available corpora for indigenous languages
 - elicitation methods derived from ethnographical work
 - Need for linguistic and encyclopaedic competence which will drastically influence both field practices and subject choice (one will more likely work on familiar themes)
-
- infinity of possible corpora
 - selection of a representative corpus sample
 - multiple authors = more data = work on various fields according to available corpora

ELDs vs “classic” dictionaries

6. Information formatting

All the elements mentioned before have a crucial influence on the amount and the type of information selected for lexical descriptions.

Financial issues : length of a dictionary, number of entries and layout.

The number of authors and linguistic and extra-linguistic competence (and for endangered language lexicographers the time they spend with indigenous people) will also play a role in **data selection**, as well as **political issues** (words that one can or cannot include the dictionary) and **user profile** (which may be the most delicate aspect).

3. Limitations of small language dictionaries

Limitations of small language dictionaries

All the issues discussed above have a noticeable impact on dictionary making and consequently on dictionaries.

Even if the language documentation perspective is increasingly present in lexicographical practices on endangered languages, the resulting dictionaries still have a variety of flaws, deriving from:

- **data collection**
- **data description**

Limitations of small language dictionaries

Data collection

- Partial coverage of the lexicon, usually general vocabulary (often reduced to mere lexical lists) due to the use of an onomasiological approach for word collection, through a kind of enlarged Swadesh list methodology;
- lexical artifacts, i.e. words that are not used by the speakers, whose referents are absent from the cultural background;

Ex. *ihpaki ku ka aynsima uhokriviyevevwi ay - hinduismo [s.m.]*

(Green 2000 :160)

- few specialized words;
- few examples of language in use if any.

Limitations of small language dictionaries

Data description

- complex structure, which makes dictionaries hard to use by untrained users;
- complicated writing systems which make it impossible for non-specialists to find information;
- equivalents or equivalent explicative structures (as we are dealing bilingual dictionaries) that are either too precise or too vague and thus exclude some user categories;
- little cultural/encyclopaedic information on important subjects.

3. Collecting vanishing words

i. A plea for lexicographically relevant corpora

While corpus absence is a self evident, the presence of some forms of corpora raises several questions like **lexical variety** and **corpus relevance** for a given language. Since most available corpora are the result of grammar descriptions, they will certainly fail to show the lexical complexity of a language.

Ex. In Palikur, the only available corpus (when we began our work on the Palikur-French dictionary) was the bible translated into Palikur by American missionaries (who also did a considerable linguistic work on the language). Using such a corpus as dictionary basis is problematic, as we are dealing with translation from a source text with a completely distinct historical, cultural and geographical background, produced by a non-native speaker.

i. A plea for lexicographically relevant corpora

For poorly documented languages, corpus doesn't seem to be the best way to achieve an exhaustive lexical inventory. However, active elicited corpus can be a really interesting tool for lexical description as it can provide three types of data:

- examples illustrating language in use,
- information on word polysemy (to some extent)
- valuable encyclopaedic and cultural information which is vital for a “documenting” dictionary.

i. A plea for lexicographically relevant corpora

The method we used for the creation of a corpus for bird, plant and insect names involves long term fieldwork and interdisciplinary specific missions.

The informants would be asked to give as much information as possible on one object (plant, animal, etc.):

- physical description (for a plant leaf and stem textures, sap colour and consistence, taste),
- identification criteria,
- different uses,
- symbolic role etc..

i. A plea for lexicographically relevant corpora

The final corpus will be in fact a collection of individual corpora that would provide a constellation of new words and valuable information to facilitate description.

Note: The presence of a specialist who can provide additional scientific information makes this kind of corpus even richer. This method can be combined with the use of a specific type of word list.

ii. Word lists for lexicographers?

Using the onomasiological approach for lexical inventories, through an enlarged Swadesh list methodology leads to the reduction of dictionaries to mere bilingual lists of words.

The word list methodology has three other major inconveniences:

- the consequent risk of missing the lexical specificities of a language by using an ethnocentric method (indeed adapted for short-term fieldwork),
- the creation of lexical artefacts,
- the omission (and eventually loss) of a great part of the lexicon, concerning highly specialized concepts.

ii. Word lists for lexicographers?

However, the main flaw of the word list method is that **it focuses on the most common and used parts of the vocabulary, which are actually the least endangered.**

The vanishing words, those corresponding to very specific cultural realities and only known by the most experienced speakers are the most threatened ones.

These words name **biological entities (animals, plants), ritual practices, mythical entities, traditional medication and objects which are no longer used in everyday life.**

ii. Word lists for lexicographers?

The same process of **lexical erosion** applies not only to words but also to word meanings and word uses.

So, linguists are clearly doing things in the wrong order, for obvious reasons: short time field work and thus lack of confidence from native speakers, no multidisciplinary competence, lack of funding, etc..

One way of avoiding at least the ethnocentric character of word lists is to **elaborate context specific lists with specialists of other fields.**

iii. Inter-active eliciting

The best way to grasp little used, specialized words, is observation on a daily basis, which involves:

- learning the language,
- building a long-term relationship with the community
- working with the right informants.

This can only be done through long-term field work which allows the linguist him/herself to acquire a deep knowledge of the context.

iii. Inter-active eliciting

However, as far as the bio-lexicon is concerned, we have developed an alternative methodology which does not require extreme field work (in our case forest expeditions and so forth), the ***multi-stimulus approach***.

Ex. For bird names, use of images and recorded bird songs combined (and sometimes information on habitat and behaviour that can be found in good bird guides) for a thorough identification, which lead to more effective results (completed by random *in situ elicitation*).

Also effective with frogs and plants.

Less effective with insects.

iii. Inter-active eliciting

The multi-stimulus method (involving this time touching or feeling things) is also interesting when working on adjectives concerning physical attributes (soft, stingy, etc.), tastes or textures.

4. Translation/description issues

Finding the right equivalent(s) in bilingual dictionaries

Finding the most accurate equivalent structure means taking into account the variety of users mentioned before:

- providing information for scientists (via scientific names),
- providing a direct equivalent in the other language for native speakers
- providing some sort of description where the equivalent is not self-evident for non native speakers (which even classic monolingual dictionary fail to do).

Political stakes for accurate descriptions are also crucial.

Finding the right equivalent(s) in bilingual dictionaries: not to do list

1) Narrow equivalents/descriptions

.ipɔkasilisili N

◆ Liane (sp.), *Mesechites trifida* (Jacq.) Müll. Arg.

et *Condylocarpon guianense* Desf. (Apocynaceae)

◇ Etym. || liane *Odontadenia grandiflora*/ fine ||

□ La tige de cette espèce est très fine. Grenand (1989 :198)

Finding the right equivalent(s) in bilingual dictionaries: not to do list

2) Hypospecific equivalents/descriptions

khalise, n. ♂ ZOO., espèce de chenille qui mange les restes de cassave (JPB)

Patte (2011 :143)

3) Missing descriptions

SABUATETE, n. ♂ ZOO., (Melanerpes cruentatus) pic à chevron d'or.

Patte (2011 :191)

Finding the right equivalent(s) in bilingual dictionaries: not to do list

4) Hyperonymy (the use of a generic term as an equivalent for a very specific.

Ex. the Palikur word *kasis* which we translated by *fourmi*, the French equivalent of ant and thought to be used as a generic term for all ant-like species.

However, during field sessions on ethnoentomology with a trained specialist we realized that what the Palikur called *kasis* was in fact a very specific ant, that the term does not have a generic use and that the equivalent was more difficult to find (in this case, two out of three types of information were possible: scientific name and defining gloss but no vernacular name).

Conclusion

ELDs are extremely complex projects and the challenges they have to meet are a huge task for one researcher. Consequently, they can hardly be flawless.

It is possible improve them considerably by:

- i) bearing in mind when working on endangered languages that it is crucial to focus on the most vulnerable part of the lexicon which is the most likely to vanish rapidly;
- ii) targeting more than one category of users and thus working not in a strictly bilingual but an encyclopaedic perspective;

iii) working if possible in multidisciplinary teams;

iv) creating areal (e.g. Amazonian) multi-stimulus tools that could be made available for the individual lexicographer in order to facilitate his/her work on specialized portions of the lexicon.