

# **CHANGING WORLDS, VANISHING WORDS: LEXICOGRAPHY ON THE EDGE**

**Antonia CRISTINOI & François NEMO**  
**Université d'Orléans. LLL. UMR7270**

**AUSTRALEX. Adelaide. July 25<sup>th</sup> 2013**

# SOME FACTS ABOUT ENDANGERED LANGUAGE LEXICOGRAPHY

In the context of endangered languages, the main reason for the disappearance of words, apart from language extinction itself, are:

- life-style changes affecting the relation with the environment,
  - medicine,
  - religion,
  - schooling,
  - language contact.
-

In this situation, the words which are the most likely to disappear are the ones that name:

- biological entities (animals, plants),
- ritual practices,
- mythical entities,
- traditional medication,
- objects which are no longer used in everyday life.

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

---

Even though it is self-evident that **dictionaries are memorizing devices of language use and word meaning**, most traditional bilingual dictionaries (and word lists) unfortunately fail to save this most fragile part of the lexicon and provide little help in the lexicographical preservation of disappearing words.

The main reason for this failure is that eliciting these words is indeed a difficult task which requires specific skills and techniques, far from universal word-list procedures and onomasiological methods.

Field lexicography of minority/endangered languages, is a difficult task which faces many limitations:

- time-consuming and under-funded;
- absence of available data and corpora (and its numerous consequences like for example an under representation of lexical polysemy);
- shortcomings of the various existing techniques of lexical elicitation;
- lack of expertise of the lexicographer in many fields of knowledge, and sometimes in those aspects of life which are culture or area specific.

# VANISHING WORDS

---

In all but written languages (i.e. in written registers of written languages), words (or word meanings) may disappear if they are no longer used.

This process becomes more important whenever changes in terms of

- ✘ “way of living”
- ✘ “way of doing things”

lead to the disappearance of a whole section of the lexicon.

# VANISHING WORDS

---

In rapidly changing societies such as the contemporary Palikur society, even if the language itself is not in immediate danger of extinction, it is obvious that the ongoing changes affecting life-style are threatening large bodies of traditional knowledge and its lexicon, and that dictionaries can become a new way of transferring this knowledge to younger generations.



# FIELDWORK & METHODOLOGY

# MAIN OBJECTIVES

---

Based on our field-experience on **Palikur**

an Arawakan/Maipurean language spoken in French Guyana and Brazil by more than 2000 people,

we shall describe some of the lexicographical methods that have proved to be efficient for lexical recovery, among which we count :

- ✘ the multi-stimulus method *and*
- ✘ the inter-disciplinary approach.

# WORKING ON THE BIO-LEXICON

Remembering the frequent coincidence between lexicographical work and life-style changes, we shall focus on biological knowledge and the “bio-lexicon” as an illustration:

- ✘ of the way lexical erosion and disappearance may occur in a small period of time in such circumstances, with little or no warning signs;
- ✘ of the necessity to adopt to some degree an “emergency” approach to field lexicography in order to “save” these sections of the lexicon before it is too late.

# WORKING ON THE BIO-LEXICON

We shall hence discuss as completely interdependent the following issues :

- ✘ detailed nature of lexicographical descriptions of biological lexical items;
- ✘ methodology and techniques in the collection of this bio-vocabulary;
- ✘ saving words (and their uses) from being forgotten.

# LEXICAL DESCRIPTIONS

---

A way to understand the nature of this inter-dependency is to realize that ultimately what is at stake in the three cases is the satisfaction of what is known in Pragmatics as Grice's Maxim of quantity (even though it was in fact first formulated by Mats Furberg, 1963), which states that a contribution must :

*“provide no less information than is required”.*

# LEXICAL DESCRIPTIONS

---

It is indeed this general constraint which makes somehow unsatisfactory:

- ✘ the lack of exhaustivity of a dictionary in terms of coverage of a lexical domain;
- ✘ the lack of exhaustivity of a lexicographical entry in terms of the information it provides contrastively with the information which is (or could be) expected;

the second issue having strong consequences on the first one.

# SOME EXAMPLES

We shall illustrate this second issue by examining some examples of partial satisfaction of this constraint of completeness in a *Yukuna Dictionary* :

**iñapimí** s. piconcito: picón. *V.apén.* aves. (=?)

**aro'ojómaji, ajo'ojómaji** s. piconcito: picón. *V.apén.* aves. (=?)

**choro** s. golondrina. *V.apén.* aves. (= swallow)

**juripícha'a** s. golondrina. *V.apén.*aves. (= swallow)

**jutá, utá** s. golondrina negra. *V.apén.* aves. (=black swallow)

**juwiche** s. azulejo (especie de tângara). *V.apén.* aves.  
blue-ejo

**kapana** s. gavián, especie más grande. *V. perí.*  
*V.apén.* aves.

**kalapichi perí** s. gavián especie. *V. perí.* *V.apén.* aves.

**kaijmeru** s. mariposa. (= butterfly)

**kaijméruna** (término genérico) mariposas.

**lachamaru** mariposa nocturna.

**pa'ajrú** mariposa nocturna.

**phichí panami** mariposa grande, color café.

**pina** mariposa grande, morfa azul.

*V.apén.* insectos.



from a **Yine/Castellano Dictionary:**

**pushchopu** s. victor díaz (especie de pájaro)

**pushropushro** s. pichihuichi (especie de pájaro)

**sumpa** s. copal (especie de árbol) **Gashgaji nanwaka**

**tuspatlu sumpa mama.** Mi mamá usa resina de copal para aplicar en el interior de la tinaja

**shaniyaka** s. huanayo (especie de garza)

**shawashka** s. paucarcillo (especie de pájaro)

**Serolnikta nikanata shawashka.** El paucarcillo estaba comiendo maduro.

**taki** s. tibe (especie de gaviota) *Ksatu gajerni taki.* El tibe vive en la playa.

and from a **Piapoco Dictionary**:

**àapi s. culebra** (en general).

**àapi àicu minali s. macabrel** (lit., culebra árbol morador); *Corallus enydris enydris*.

**àapi íinu wirìichu s. saltón del monte** (lit., culebra su esposa saltamontes); *Tettigoniidae*.

**àapi íiwitami s. vinegarone** (lit., culebra cabeza muerta); *Thelyphonidae*.

**àapi wéetéeriwa iyú s. culebra venenosa** (lit., culebra por medio de la cual morimos).

# FACT ANALYSIS

---

A specificity of such terms, which are not only **nouns** but also **names** is that their description requires no less than three types of information :

- ✘ information about what we shall call the “label” itself, i.e. about the word(s) used to name a bird or a plant;
- ✘ information about the specific (or generic) living beings which they name, whose label name in the target language or in scientific nomenclature must be provided;

# FACT ANALYSIS

---

and, since the bio-lexicon is not common ground for many if not most speakers, especially when scientific names are provided, and thus cannot by itself satisfy the maxim of quantity:

- ✘ basic (or more than basic) information about the nature of what is named.

# INFORMATION TYPES

---

- ✘ The first type of information (“label information”) is morphological or etymological and can thus be troublesome whenever word-formation is not transparent.
- ✘ The second type of information (“denominative value”) is a matter of identification and is arguably the most difficult part of information to obtain
  - + i) when it comes to non spectacular or familiar species;
  - + ii) in terms of hyperonymy status.

# INFORMATION TYPES

---

- ✘ The third type of information, which is crucial to the readability of the dictionary for the ordinary reader, is often rather easy to obtain but almost always insufficient to allow identification (and strongly dependant on the informant's level of expertise).

# COMMENTS

---

Once acknowledged the ultimate necessity to provide all these types of information, it must be noticed that:

- ✘ there are specific problems in obtaining each of them;
- ✘ there is an ambiguity problem whenever one doesn't know which information is provided, as in the “golondrina negra” example, which can be interpreted in terms of each of the three types of information;

-

# COMMENTS

---

- ✘ systematicity is at stake if the nature of the information provided is different in distinct entries.
- ✘ all these issues are constantly present when the lexicographer him/herself works with his/her informants, *asking*, *obtaining* and *interpreting* the information he/she needs.



# METHODS FOR ELICITATION AND IDENTIFICATION:

---

## Stimulus based:

- ✗ using handbooks (pictures, drawings, etc.);
- ✗ using sound files;
- ✗ using biological collections;
- ✗ field experience;
- ✗ field collaboration with specialists;

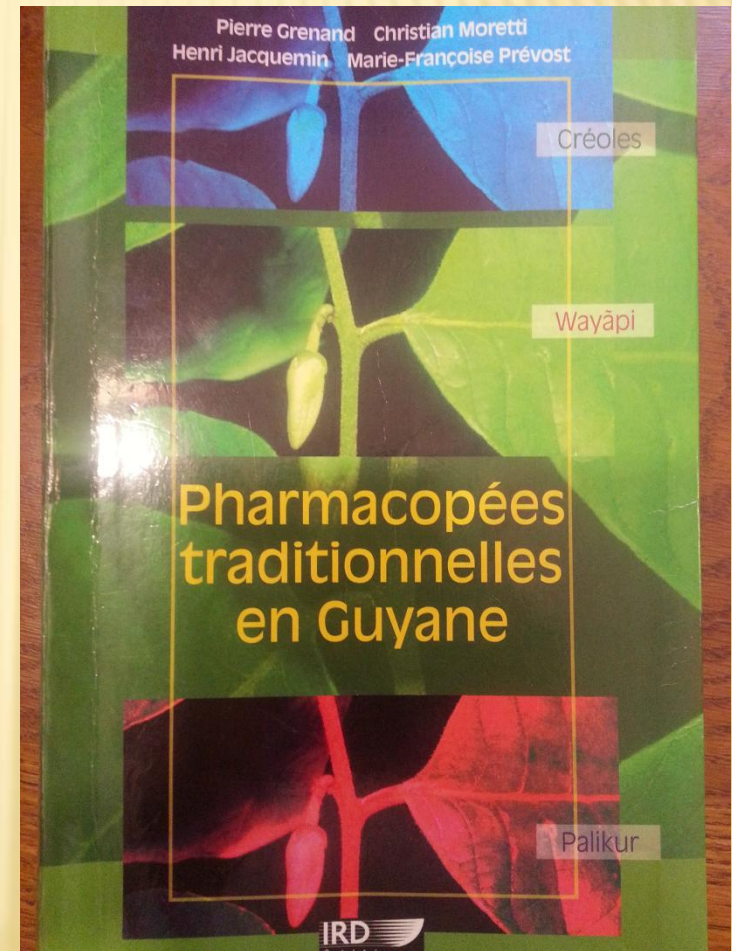
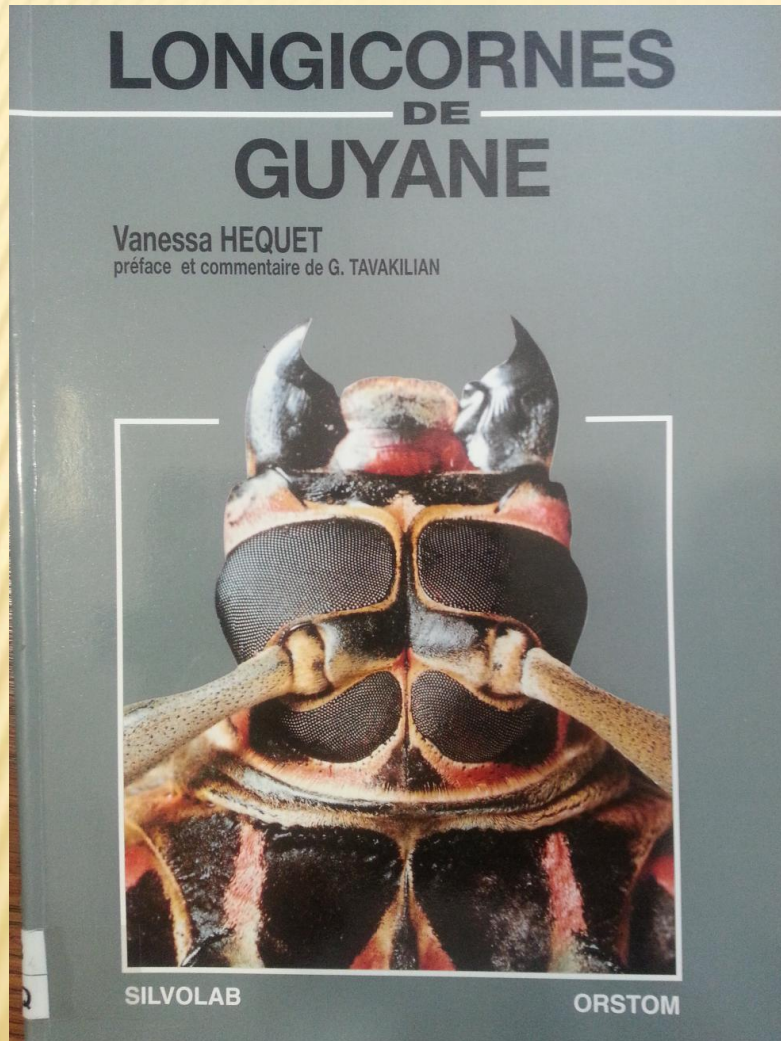
## Knowledge based:

- ✗ obtaining lists of names;
- ✗ corpus data.

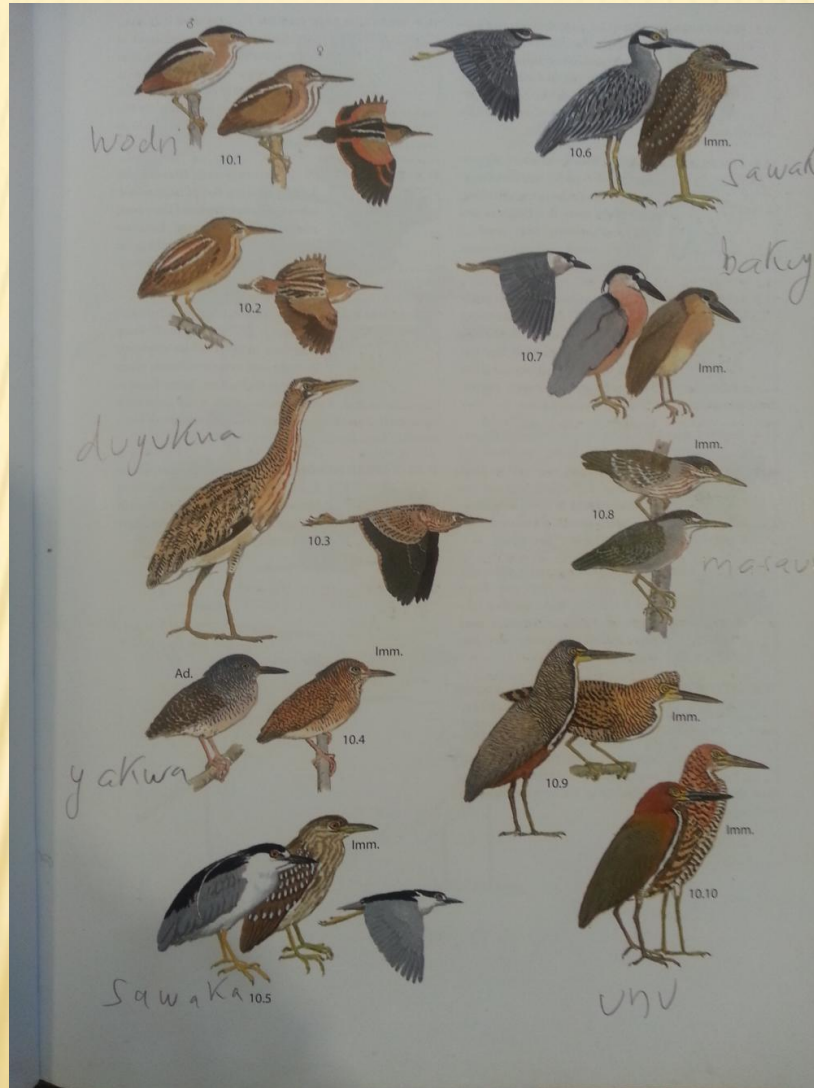
These methods are not mutually exclusive, because:

- some of them can be used only for specific sections of the biolexicon (e.g. typically drawings);
- elicitation of a given name may be extremely easy with one technique and virtually impossible with another one;
- identification is either straightforward (or seemingly so) or a long and uncertain task depending on the technique used (and on the word's status) ;
- combining techniques ensures the accuracy of the identification or reveals significant discrepancies (e.g. sound and pictures in raptor's classification).

# USING DRAWINGS AND HANDBOOKS



# USING DRAWINGS AND HANDBOOKS



# USING DRAWINGS AND HANDBOOKS

Using drawing or pictures to obtain names appears to be a technique which, despite its obvious advantages for the lexicographer in terms of exhaustivity and precision, faces strong limitations, among which:

- ✘ the fact that form is often not the main criterion for animal identification in the real life experience of the speakers;
- ✘ the fact that size is virtually inaccessible with images;
- ✘ the fact that for huge parts of the biolexicon (insects, trees, snakes to a large degree, frogs to a significant degree, etc.), this technique appears impracticable;

This technique may also be misleading when it comes to defining the exact scope of a name due to the fact that using handbooks forcefully introduce a scientific bias in classificatory systems and that using local names known by the informants also introduces a translation bias - from which it follows for instance:

- that the fact that A may be called N should never mean that N actually applies only to A;
- that the fact that a western hyperonym H is called N by an informant should never mean that N has indeed an hyperonymic status in his/her language;
- that presenting in a L2/L1 order what has been obtained by a L1/L2 question can be misleading.

But also:

- the fact that the reality of the technique is that questioning based on images *is always associated with verbal explanations from the questioner* which can deeply affect the nature of the answers are obtained (and for instance can produce artifacts and/or mere translations of what the linguist says);
- the fact that among polyglot informants, there are interferences of naming systems.

Similarly, it appears that people are often able to identify animals by **sound only** (or **traces**), especially for secretive species or in rainforest environments, and more generally on **behavior** or perspective, and simply cannot recognize animal images.

Identificational clues are extremely various and completely unrelated with the meaning of a name (Nemo, 2003): one doesn't need to see a *sakaska* scratching the ground to recognize it as a giant snipe, despite the fact that the label *sakaska* is derived from the verb "to scratch". Generally speaking, labels are indeed only "characterizers" (Cadiot & Nemo, 1997) and do not define conditions that would have to be satisfied for class membership.



This is why field experience may prove to be crucial: after two complete “non-field” studies of Palikur birds by two distinct researchers (including sound for one of them):

- a simple 15 minutes walk provided no less than 4 unknown names of familiar birds;
- a one hour trip to distinct biotopes can provide dozens more;
- field experience also provides crucial clues on ambiguous denominations, thus clarifying questions which couldn't be solved otherwise.


# USING SOUNDS



# SOUND FILES

---

Stimuli such as songs and calls can now be easily obtained for any region of the world, providing extensive data bases for the lexicographer (<http://www.xeno-canto.org/>).

For instance, since we are in Adelaide, we can get a sound file of the red wattled bird: [redWattleBird21072012.mp3](#) 

For night, secretive or forest species, these stimuli are of course crucial because hearing them is often the standard form of contact with humans.

# COLLABORATION WITH BIOLOGISTS AND ETHNOLOGISTS

---

We have developed collaborations centered on field work with specialists in areas such as archeology, botany, ethno-botany and ethno-pharmacology,.

Such collaborations allow both sides to obtain:

- ✗ much more data;
- ✗ more precise data (in both the biological and the orthographical/linguistic dimensions of the issue);
- ✗ in a much shorter time.

# INTERDISCIPLINARY MISSIONS

---

A one week mission in 2011 with archeologists and Palikur informants which consisted in:

- ✘ the tagging of 800 trees of a “crowned mountain” archeological site;
- ✘ the identification and naming by Palikur informants of those 800 trees;
- ✘ the recording of short (2-3 minutes) ethno-linguistic comments on each tree-type and its uses, allowing for the constitution of a large corpus which can be made available in online versions of the dictionary;
- ✘ the use of this corpus to extract other unknown lexical items or unknown meanings of known ones.

# INTERDISCIPLINARY MISSIONS

---

Another illustration of this kind of work is a one week mission in 2012 with an ethno-pharmacologist centered on insects and insect uses.

# CONCLUSION AND PERSPECTIVES

As mentioned earlier, it is both true that all aspects of the lexicon which are associated with vanishing life-styles are in danger of disappearing, and that saving the bio-lexicon is a strongly time-consuming activity which requires far more dedication than the normal lexicographer can normally engage.

Apart from adopting and combining some of the techniques which have been described, and accepting to invest time on saving those words before it's too late, we would like in conclusion to open a double perspective.



# ADOPTING AN AREAL PERSPECTIVE

---

Endangered language lexicography would gain considerably if we could provide for each linguistic area a kit of tools and knowledge which could be used by each lexicographer to work much more efficiently and limit the amount of time which is required to obtain satisfying results in the description of the bio-lexicon.

For instance creating a **sound data base** for Amazonian lexicography would allow for dozens of languages a far more efficient work.

Similarly, given the **areal nature of many names**, a collaborative compilation of existing knowledge would simplify considerably etymological issues.

# ADOPTING A MULTIDISCIPLINARY PERSPECTIVE

---

This areal effort, if combined with a multidisciplinary perspective, would also make available for each area a set of questions which could be relevant in the study of each group of live-beings, providing the individual lexicographer with both a basic and expert knowledge of what can be questioned about each of them.

Simultaneously, as we have seen, adopting a joint work approach with locally available specialists is probably the only way to reach certain goals.

# ADOPTING A NON-GRICEAN PERSPECTIVE

---

A last and paradoxical issue concerns the necessity to adopt and accept to some degree a Grice-free approach to endangered languages lexicography.

- ✘ When working with informants, the lexicographer is often exposed to what we can call the “partial knowledge syndrome” (PKS), in other words the fact of having incomplete information about a word.
- ✘ For obvious reasons, (s)he is then tempted not to mention the information at all because of the information (s)he lacks, in order to avoid criticism or even in some cases looking ridiculous.

For also obvious reasons, vanishing words and word uses are much more exposed to being concerned by PKS than other words.

The paradox then is that in many cases, concerning such “on the edge” words, the little we know will finally not be mentioned at all.

Our final suggestion would thus be to adopt explicitly an “emergency” approach to endangered language lexicography, in terms of which a specific section of the dictionary would be dedicated to words whose knowledge has remained too limited to allow full integration but whose existence has to be documented in order to make possible later recovery.

# REFERENCES

---

Grenand, Françoise (Chief-Editor) (2009) *Encyclopédies palikur, wayana, wayãpi : langue, milieu et histoire*, fascicule Encyclopédie des Amérindiens de Guyane, PUO- CTHS, Paris.

Launey, Michel (2003) *Awna parikwaki. Introduction à la langue palikur de Guyane et de l'Amapa*. Centre d'études des langues indigènes d'Amérique. Paris : IRD Éditions.

Green, Diana (compiler), Green, Harold (compiler). (2010) *Yuwit kawihka dicionário Palikúr - Português*. SIL.

Klimpp, Deloris A (1995). *Vocabulario Piapoco-Español*. Bogotá: Asociación Instituto Lingüístico de Verano. ix, 190 pages.

Patte, Marie-France (2012) *La langue arawak de Guyane. Présentation historique et dictionnaires arawak-français et français-arawak*. Paris : IRD.

Schauer, Junia G. (compiler); Schauer, Stanley (compiler); Yukuna, Eladio (compiler); Yukuna, Walter (compiler). (2005) *Meke kemakánaka puráka aloji: wapura akó chu, eyá kariwana chu (Diccionario bilingüe: Yukuna - Español; Español - Yukuna)*. . Editorial Fundación para el Desarrollo de los Pueblos Marginados.

Urquía Sebastián Rittma & Wagner (Compiladores)  
(2008) *Tokanchi Gikshikowaka – Steno Yine  
Yami. Diccionario de la lengua Yine* (parcial).