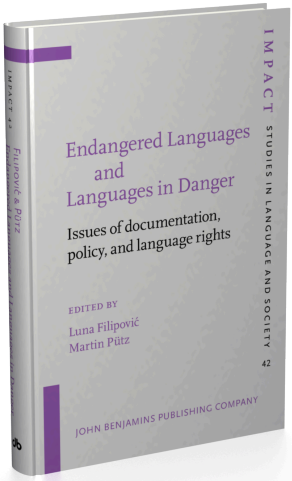
IMPACT: Studies in Language, Culture and Society
42

Redefining priorities, methods and standards in endangered-language lexicography

From lexical erosion in Palikur to areal lexicography

François Nemo | University of Orléans
Antonia Cristinoi | University of Orléans

https://doi.org/10.1075/impact.42.16nem



Pages 361–386 of
Endangered Languages and Languages in Danger: Issues of documentation, policy, and language rights
Edited by Luna Filipović and Martin Pütz
[IMPACT: Studies in Language, Culture and Society, 42]
2016. ix, 413 pp.

Redefining priorities, methods and standards in endangered-language lexicography

From lexical erosion in Palikur to areal lexicography

François Nemo and Antonia Cristinoi
University of Orléans

Based on our ten-year fieldwork on Palikur (Arawakan) in French Guyana, this paper discusses the need for researchers (assisted by native speakers) to devote more time and energy to documenting the lexicons of endangered languages (EL) and the urgency to focus on the parts of the lexicon that are subject to strong lexical erosion, in order to allow the transmission of saved knowledge to younger generations. It also describes the main difficulties in documenting these sections of the lexicon and advocates the need for new principles, methods and standards in EL lexicography so as to obtain more complete and satisfactory information for interdisciplinary work, for tool, knowledge and content sharing, and finally for dismissing an all-or-nothing approach according to which incomplete lexicographical descriptions and knowledge should not be published.

Keywords: Amazonia, biolexicon, dictionary, documentation, endangered language/s, lexical erosion, lexicography, lexicon, Palikur, transmission

1. Introduction

Based on our experience of language documentation and field lexicography in an Amazonian context, namely our work on a Palikur-French dictionary in French Guyana, our aim here is to consider what can be done and changed in order to avoid *lexical erosion*.

After defining lexical erosion in general and in the Palikur context in particular, and after stressing the importance of preserving the lexicon and its use in specific situations in a language conservation situation, we shall discuss what the priorities, methods and specificities of EL lexicography should be in such conditions and, in order to meet these particular challenges, we shall advocate

DOI 10.1075/impact.42.16nem

a necessity to reconsider lexicographical work and formats.[1] We shall build our reasoning on the following premises:

1. that more time and energy should be devoted to documenting the lexicons of ELs, which is too often considered as a minor scientific objective, and the existing dictionaries are not really formatted to truly contribute to word and word meaning preservation;
2. that we need to adopt an *emergency approach* to field lexicography and language documentation in order to save the frailest sections of the lexicon (that are the most likely to disappear);
3. that the lexical treatment of each lexicographical entry must be reconsidered in order to allow its preservation;
4. that access to the saved lexicon (and knowledge) must be provided for younger generations.

Focusing on biological and environmental knowledge and the *biolexicon*[2] as an illustration of a part of the lexicon that has to be saved before it is too late, we shall then discuss:

1. the specific difficulties in documenting these sections of the lexicon for the average linguist and/or lexicographer;
2. the need to obtain much more precise descriptions of the lexical units involved than is usually considered necessary;
3. the different techniques which may and must be used to obtain these types of information;
4. the necessity to associate the classical lexicographical approach with other methods of lexical documentation, namely collecting small oral corpora for each lexical item;
5. the need for the linguist to engage in multi-disciplinary work when dealing with sections of the lexicon like the biolexicon (plant and animal names to be

---

1. For a presentation of existing formats and standards in lexicographical work, see for instance, Hartmann (2003) and Mosel (2011).

2. As far as Palikur is concerned, the risk for the dictionary to contribute to bio-piracy is almost null, mainly because their use of medicinal plants, for instance, has already been documented in great detail (*Pharmacopées traditionnelles en Guyane*, Grenand et al. 2006). As for the general issue of knowing whether all the collected information should be made public or not, it is important to distinguish between the importance of storing all information and the fact of making it public. Whereas all our Palikur oral corpora, for instance, will be deposited and stored at the BnF (French National Library), researchers and their informants have the possibility to decide on the availability of each corpus.

exact), which cannot be collected thoroughly enough without a certain level of expertise;

6. given the scarcity of resources available and the time-consuming nature of lexical documentation in specific linguistic areas such as Amazonia, there is an urgency for linguists to share their tools and knowledge or to build these tools together;
7. the fact that since it is possible to considerably reduce the time-consuming nature of lexicographical work and to enhance the quality of basic encyclopedic glosses (like glosses describing plants, for instance) if glosses could be shared by all lexicographers within a linguistic area such as Amazonia, it would also be extremely useful to adopt new authorship standards which would allow such sharing practices.

Finally, we shall dwell on the necessity to reconsider the formatting of both micro-structures and macro-structures in EL lexicography, either by meeting the specific lexical description constraints associated with preservation issues, or on the contrary by rejecting an all-or-nothing approach to EL documentation which may lead to the loss of precious but incomplete and thus unpublishable information.

We shall further discuss using phonetic forms (written in IPA) as headwords whenever the existing word spelling may prove ambiguous or whenever it may ensure easier access to the information provided by the dictionary for the members of a given linguistic community.[3]

2. On lexical erosion

A language may disappear as a whole, but it can also disappear by subject specific fractions, through a continuous process of lexical erosion (see also Cristinoi & Nemo 2013), which takes place whenever important changes affecting one linguistic group rapidly modify the lifestyle of its members (schooling, relation to the environment, medicine, religion and language contact), thus threatening the transmission of both words and word meanings.

In the Amazonian context, the words which are the most likely to disappear are the ones that name biological entities (animals, plants) and environmental knowledge, ritual practices, mythical entities, traditional medication and objects which are no longer used in everyday life.

---

3. In our case, as the Palikur phonetic system is rather simple, the IPA transcription would be easily readable by a Palikur speaker.

2.1 The Palikur context

Even though the main goal of this paper is not to describe in detail the dynamics of an Amazonian language such as Palikur, some of its features are worth mentioning. Palikur is an Arawakan language spoken in the north of the Brazilian state of Amapa and in French Guyana, representing today a community of possibly 3000 people, most of whom still speak the language (but have a rather uneven linguistic competence). The Palikur community has a rapidly growing demography (at its smallest, it was probably reduced to less than 400 people), a tendency to spread westwards in French Guyana with new and fast growing settlements west of Cayenne (Macouria) and a high degree of population mobility between the different areas, notably from Brazil to French Guyana (for instance, in Saint-Georges de l'Oyapock, possibly half of the Palikur community has come from Brazil during the past forty years and paradoxically Palikur is to some degree becoming a kind of vehicular language between French-speaking and Brazilian-speaking Palikur people). All of this has direct consequences for language dynamics (as a result of marriages between Brazilian and Guyanese Palikur, villages such as Favard, where Palikur has been replaced by Guyanese Creole during the past fifty years, have recently experienced the reappearance of Palikur speakers, with young children speaking the language while some of their grandparents do not).

Palikur is exposed to multilingualism and has a long-term relationship with an old variety of French-based Creole (which is the official language of its Karipuna neighbors and was spoken in the area before Northern Amapa was assigned to Brazil at the start of the twentieth century), and experiences both resistance to creolization and attraction to it. It has faced a more recent but massive exposure to French and Brazilian Portuguese through media and the school system (Cristinoi & Nemo 2014 unpublished field data). The primary school system in the state of Amapa is a bilingual one (Palikur-Portuguese), which implies among other things that writing in Palikur is part of the curriculum, whereas in French Guyana Palikur and other indigenous languages have long been ignored within the education system. French has been the only language of schooling and education for a very long time and vernacular languages have only recently started to be studied in schools on a more or less regular basis.

The dynamics of the language appear to vary in the different areas, depending on the chronology of each settlement. In the oldest ones, notably in Saint-Georges de l'Oyapock, one may notice a recent tendency to adopt a language mix of two main and equally important languages, Palikur and French. In others, a language mix of Creole and Palikur is now challenged by the rising importance of French and the disappearance of Creole-speaking neighbors, such as in Régina.

2.2 Lexical erosion in the Palikur context

Generally speaking, the use and the vitality of a language may be measured in terms of the number of active users (speakers), especially in younger generations,[4] and hence it could be said that the Palikur situation is not dramatic in this respect. Nevertheless, the vitality of a language also depends on the scope of its use and its ability to be used in all areas of life. Much has been said about the capacity of languages, even strong languages, to deal with the necessity to adapt lexically to new patterns of life and to modernity and external contact. As far as the lexicon is concerned, adapting to modernity appears to be one aspect of the problem, since it is impossible to assume that the lexicon of a language is automatically transmitted to the next generations. However, it is also crucial that this happens so that the use of a language may be pursued in various domains. When non-transmission occurs, the generation gap in terms of lexical competence is a reality with which lexicographers are directly confronted. Measuring this lexical competence, however, is a larger issue which requires extensive work and specific criteria in order to be precisely defined, measured, characterized and understood in its variability.

In the Palikur context, after considering all the members of the community, it soon became obvious that lexical competence was uneven, that it was necessary to measure its variability, and that it probably included lexical erosion between generations. As a first step, we have thus developed lexical surveys aimed at measuring actual lexical competence and knowledge among the Palikur with the following methodology (Cristinoi & Nemo 2014 unpublished field data):

1. the surveys concerned a range of distinct lexical fields so as to cover all aspects of life (and the lexicon);
2. for each field, the knowledge of a set of supposedly basic, non-basic and sophisticated (i.e. quite specialized) words was tested;
3. all tests were conducted in Palikur and were bidirectional, either providing the words and asking about their meaning or by providing notions and asking for their Palikur equivalents;
4. the survey was conducted in all the major Palikur settlements in French Guyana (notably Saint-Georges de l'Oyapock, Régina and two villages in the Macouria area: the Kamuyene and Norino villages);

4. It must be noted that the reluctance of teenagers to speak the language should not be considered as definitive. In Saint-Georges de l'Oyapock, for instance, the teenagers from the last years of the twentieth century were often simply refusing to admit speaking Palikur but many of them have later become language-transmitting parents.

5. the survey was conducted for different age and sex categories in each settlement and the personal profile of each interviewee was documented, in terms of movements between villages and from Brazil to Guyana and also in terms of education and schooling.

These surveys have shown that the differences/discrepancies and commonalities between speakers and settlements follow different patterns. To start with, we noted that lexical erosion and the loss of vocabulary is not always the result of modern acculturation or language disappearance but is often the quick result of the existence or inexistence of specific techniques and practices in the various settlements of the community, since the learning of specific words appears to be greatly conditioned by the actual use of the corresponding objects. If such objects are sometimes simply not relevant in the specific environment in which a new settlement appears, the words may disappear from lexical consciousness very quickly, as could be observed in Saint-Georges de l'Oyapock for large parts of the fishing vocabulary, due to the impossibility of using the fishing techniques found in Rocawa (Amapa) and in other settlements; this is why it is crucial to document them.

Another crucial observation is that the patterns of lexical erosion which could be observed do not follow a pattern of *reduction* to a core (or fundamental) Palikur lexicon, in which all *basic words* would have remained known, followed by less-remembered *non-basic words* and finally a forgotten *specialized vocabulary*. It appears on the contrary that reduction is massively field related: for instance, in the villages in which the overall lexical erosion was the strongest, we could find young Palikur who seem to know quite well the lexicon of family life, including idiomatic expressions, but are almost unable to name any animal in Palikur, including those living on their doorstep like cats.[5]

It also appears that lexical erosion is a highly generational process, which means that at certain moments, important portions of the lexicon still exist among the older generations (50+) but could be lost for the younger ones, whose life has deeply changed, and this is independent of their global attitude toward the language itself.

As for Palikur, it appears that despite the fact that the overall number of Palikur speakers has probably never been higher during the past centuries, there is a big risk and strong probability that in the next twenty-five years half of the Palikur lexicon will be gone, and, as importantly, that the impoverishment of certain

5. It must be noted that such lexical surveys complement ethnographic work as they allow new access to the reality of language use within the diverse settings of a community. In our case, years of fieldwork has revealed surprisingly diverging lexical dynamics among these settings.

lexical fields could simply block the use of Palikur for evoking aspects of life which paradoxically have always been central to the Palikur culture.[6]

2.3 Dictionaries and lexical erosion

Without considering this issue in full detail here, it seems that such patterns of lexical erosion and lexicon endangerment have dramatic consequences for language endangerment as a whole. The implication is that beyond a language's capacity to provide vocabulary for the modern world, its capacity to resist such lexical shift/lexical replacement is crucial for its future.

In this respect, what appears crucial whenever lexical erosion has already started is the capacity to restore and revitalize a language's lexicon and in order to do so the tools which allow an indirect transmission of the lexicon need to be created.

It must be stressed that for a language such as Palikur, the generational gap associated with literacy and spending years in the school system, even though it endangers the traditional oral transmission of the lexicon, also enables dictionaries and websites to become a new form of collective memorization of words and word uses. This must stimulate researchers to find new modes to transmit knowledge and experience between generations.

However, for a paper or online dictionary to play this role, much has to be reconsidered in the elaboration of the dictionary itself, as we shall see further on in this paper.

3. Reconsidering methods in field lexicography and EL lexicography

Field lexicography for minority languages or endangered ones is in itself a difficult task, which faces many limitations. If it were to be considered in terms of preservation of the lexicon, its principles would have to be reviewed.

3.1 Limitations of EL lexicography

Some of the shortcomings of minority or EL lexicography are global, and among them we can note:

---

6. A complete presentation of the Palikur people and culture is available in Grenand (2009). For a presentation of the grammar of Palikur, see Launey (2003). The only published Palikur dictionary is Green & Green (2010).

1. the fact that it is a time-consuming task;
2. the fact it is under-funded;
3. the frequent absence of available data and corpora, and its numerous consequences, among which we find the under-representation of lexical polysemy;
4. the shortcomings of the various existing techniques of lexical elicitation;
5. its reliance on mostly individual work;
6. the consequent lack of expertise of the lexicographer in many fields of knowledge, and specifically in environment, culture or area-specific aspects of life.

Nevertheless, some of the shortcomings are purely lexicographical (even when they are ultimately related to global ones), such as:

a. the fact that the description of words for many entries may be considered as insufficient or suboptimal;
b. the fact that the phonetic form is often absent in existing paper dictionaries and frequently unpredictable from the orthographic form and that few recordings and electronic dictionaries (other than mere electronic versions of the paper dictionaries) are available for these languages;
c. the fact that the coverage of certain lexical domains is often minimal;
d. the fact that aspects of a lexical entry which could be considered as especially relevant for a given language are not included.[7]

The direct consequence of such shortcomings is, as mentioned earlier, that field lexicography in its current state is often unable to really fulfill the goal of preserving a lexicon and fighting lexical erosion.

To a large extent though, this limited capacity to avoid lexical loss and to stop lexical erosion is also due to the fact that what lexicography ought to achieve in an EL context has not been clearly spelled out and transformed into a genuine constraint on lexicographical work, neither in terms of methods, nor priorities or content.

This is why our aim here will be to try to formulate areas in which changes in goals, methods and sometimes structures could significantly improve the capacity of field lexicography to truly contribute to language preservation and lexicon transmission.

7. Such as the specific numeral classifier used for a given noun in Palikur, even though it is especially important to ensure the effective use of such a word in context. This concerns a quite unique feature of Palikur in terms of linguistic typology and appears to be suffering strong generational loss; documenting it item by item should be considered compulsory.

3.2 Principles and priorities in EL lexicography

Bearing in mind the fact that producing a dictionary is a long and undervalued task, the result of which is as easy to criticize as it is difficult to achieve, and that our remarks or illustrations of the issues at stake should not be considered as a form of bashing the considerable work that they represent, we shall now try to consider what a dictionary should be if we want it to play an effective and significant role in the preservation of the lexicon.

We can state some basic principles concerning the priorities of EL lexicography, both at the level of individual languages and individual lexicographers, and at the level of countries or linguistic areas.

The first principle that must be implemented in order to fight lexical erosion is, as mentioned above, to consider the frailest part of the lexicon as a priority in lexicographical work. This may be called *the emergency principle*.

The second principle is that all should be done as if the dictionary was to become the only available source of information on the words or on the language at stake. This may be called *the exhaustivity principle*.

The third principle is that whatever can be done to reduce the time required to meet the requisites which follow from the first two principles should be done. This may be called *the economy principle*.

The fourth principle is that whatever can be done to enhance the quality of what is produced should be done, even if it deviates from the way lexicographical work on non-endangered languages is usually done. This may be metaphorically called *the ISO principle*. The idea behind this "ISO" principle (the norm referring here to classical ISO standards) is that we should have specific quality standards for this type of lexicography, standards that would be different from those of classical lexicography.

The last principle is that whatever can be done to ensure transmission – and not only scientific documentation per se – to the next generations should be done. This may be called *the transmission principle*.

Adopting and reconciling all these principles as both specific to EL lexicography and central to it has huge consequences for the methods, standards and formatting of the lexicographical work and the dictionaries themselves.

3.3 Methods in EL lexicography

The first consequence of the priorities and principles mentioned above is to concretely ask much more from the lexicographers of an EL than is required for a non-endangered language – in terms of exhaustivity of the lexicographical coverage of a language, the inclusion of portions of the lexicon whose treatment is

very demanding, and in terms of the precision of the information provided on each item.

The second consequence is that because raising the standards that have to be met when writing a dictionary to ensure actual preservation of the lexicon is a necessity that cannot be satisfied by individual lexicographers most of the time (with an individual lexicographer being either a trained researcher, linguist, ethnographer or a missionary or a trained native speaker), it is therefore necessary to considerably help them in doing so, but also to move EL lexicography away from some of the limitations of authorship. It is necessary to adopt a much more collective and collaborative framework based on multi-disciplinary, areal approaches and on lexicographical equivalents of freeware.

3.3.1 *Facing the lexicographer's limits*

The most demanding aspects of the work (from elicitation tools to actual lexicographical descriptions) can indeed be transformed by collective and collaborative work. If we consider, for instance, the biolexicon, based on our experience and the study of dozens of existing dictionaries of the same linguistic family (Maipurean/Arawakan) or area (Amazonia), the reality is that existing standards or practices are well below what would be necessary to allow preservation, for a number of reasons which are easy to understand, notably:

1. the lexicographer's own limits when it comes to dealing with realities he/she has no knowledge of;
2. the time needed to acquire such knowledge;
3. problems in elicitation techniques (using pre-determinate word lists and translations, and single stimuli like drawings or pictures, for instance, which lead to numerous errors especially as far as small animals and plants are concerned).

All of this leads to the fact that most of the time the authors of EL dictionaries cannot reasonably provide what is asked of them.

3.3.2 *Sharing resources and lexicographical glosses*

It follows from this that what has to be done cannot be done autonomously for each individual language, due to a lack of financial and human resources, and thus the only way for EL lexicography to meet its goals and to limit lexical erosion is to deeply transform the way field lexicography is carried out and conceived, and specifically to stop asking each individual lexicographer to deal with specialized domains such as the biolexicon in which he/she has little competence.

Fortunately, regarding this last point, it is clear that for large portions of the lexicon what has to be done could be done for all the languages of a given area or country if resources were fully shared.

Even though it is solitary work, EL lexicography would gain considerably if researchers made available, for each linguistic area, a kit of tools and knowledge which could be used by each lexicographer to work more efficiently and therefore limit the amount of time required to obtain satisfactory results in the description of the biolexicon. For example, the creation of a (animal) sound database for Amazonian lexicography would allow for work on dozens of languages to be far more efficient. Similarly, given the areal nature of many names, a collaborative compilation of existing knowledge would considerably simplify etymological issues.

3.3.3 *Sharing resources: Expertise*

To stick to our study of lexicography of the biolexicon among the Palikur, even though we had a non-basic knowledge of some parts of it, collaboration with biologists, ethno-linguists, ethno-pharmacologists and anthropologists indeed proved to a necessity, allowing both sides to obtain:

a. much more data;
b. much more precise data (in both biological and orthographic/linguistic dimensions of the issue), in a much shorter period of time.

An example of collaborative work of this kind is a two-week field trip conducted in 2011 with archeologists, botanists, anthropologists and Palikur informants, which consisted of:

– the identification and tagging (with number and scientific name) by botanists of 800 trees of an archeological site in the forest;
– the identification and naming by Palikur informants of each one of those 800 trees;
– the recording of short (2–3 minutes) ethno-linguistic comments (in Palikur) on each tree-type and its uses (including stories and information on traditional architecture, boat building and traditional object making according to the specific uses of the trees), thus allowing us to create a large corpus which could eventually be made available in online versions of the dictionary, and thus made available to the next generation;
– the use of this corpus to extract other unknown lexical items or unknown meanings of known ones.

Another joint research project was conducted on insects and ethno-pharmacology, leading to results which are crucial for the illustration of the dictionary but which could also be used for areal purposes.[8]

3.3.4 *Sharing resources: Elicitation tools*

When working on the description of lexical fields which are not familiar to him/her, the lexicographer cannot but rely on accessible tools (pictures, drawings, sounds, videos, etc.) and is nevertheless confronted with the difficulty of producing reliable information.

As far as the biolexicon is concerned, any linguist who has tried to obtain reliable denominations for specific animals, birds and plants knows how important and troublesome elicitation techniques are, even though the multiplication of field guides or websites about almost any kind of animal and plant are making it easier nowadays than it used to be.

Anyone wanting to use sound files to elicit bird names, for example, can use a website such as *http://www.xeno-canto.org* where he/she may easily obtain geographically relevant files of bird songs to be tested. This allows one to obtain names which otherwise would remain unknown or would be wrongly attributed.

Using drawings or pictures to elicit 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 is the fact that form is often not the main criterion of animal identification in the real life experience of the speakers, the fact that size is virtually impossible to represent through images, and more importantly, 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 is also tricky because of the fact that if A may be called N, this should never mean that N actually applies only to A, or the fact that a Western hyperonym H (general term indicating the class of objects the noun belongs to) is called N by an informant should never mean that N has indeed an hyperonymic status in his/her language, and also and most importantly, because presenting in a L2/L1 order what has been obtained by a L1/L2 question can be misleading. Indeed, linguistic inquiry based on images is always associated with verbal explanations from the inquirer which can deeply affect the nature of the answers obtained and can, for instance, produce artifacts and/or mere translations. Similarly, among

8. Because bees, for instance, are to a large extent named in Palikur according to the type of nest which they build, having a picture (or a verbal description) of the different kinds of nests is important for the dictionary but could also serve as a basis for the definition of an areal bee Swadesh-like list.

polyglot informants, the interference between naming systems may be so strong that extreme caution is required.

This is why direct field experience may prove to be crucial: after two complete "non-field" studies of Palikur birds by two distinct researchers (including sound use in one of them), a simple fifteen-minute walk provided no less than four unknown names of familiar birds, a one-hour trip to distinct biotopes can provide dozens more, with the only, but sometimes demanding, necessity being to identify the birds at stake.

While all this does not make things easy for the lexicographer, it is nevertheless true that simply knowing where to find the tools for the task allows one to considerably reduce the time needed to document a group of animals or plants and to increase the number of names that will be included in the dictionary.

But once again, it is obvious that if all the necessary preparation of the task was not left to the individual lexicographer but mutualized at country, or preferably area, level, it would become easier to document parts of the lexicon which until now have appeared too difficult to document. Having at one's disposal on a website all the shapes of bee/wasp nests in Amazonia, for instance, would allow extensive coverage and precision for dozens, if not hundreds, of minority languages. This proves, once again, that efficiency at the level of an individual language would be greatly enhanced by sharing resources at area level, at no extra cost.

3.3.5 *Sharing resources: Lexicographical glosses*

When the lexicographer of an individual language has managed to obtain all the data needed to fully include the biolexicon in his/her dictionary, he/she is left with the purely lexicographical task of describing each lexical item, in other words, producing specific lexicographical glosses and definitions for each word.

Without going into further detail, it appears from observing the way this task is performed in existing dictionaries that it is not an easy one for two main reasons:

- good definitions are not only true statements but should also provide sufficient information (see next section);
- good definitions need to be adapted to the user/s of the dictionary.

Regarding this last point, a difficulty of the task is that dictionaries of minority or endangered languages are most of the time bilingual dictionaries, with potentially three types of users: speakers of the majority language (school teachers, doctors or ordinary citizens), speakers of the minority language itself, and finally, scientists (Cristinoi & Nemo 2013). The definitions and glosses found in existing dictionaries thus oscillate (sometimes within a single dictionary) between the

three perspectives, often adopting only one of them, which makes them opaque or irrelevant for the other users.

Our claim here is that the main objective of a dictionary should be language (and culture) preservation and in order to secure it one should combine in each individual word definition/description all the information which would ensure lexical transmission and avoid lexical fuzziness.

As for the biolexicon, the following is an example of what may happen for each individual lexical item in a contemporary dictionary. It is well known that for certain Australian languages, only a small list of nouns could be documented before the language stopped being used. For instance, in a case where a name for "heron" was documented, it is actually the case that given the existence of no less than seventeen species of herons in the area, the contemporary descendants of the community who are making language revitalization efforts will never know if the term applied to a specific heron species, or whether it was a hyperonym. Quite too often, a similar pattern may be found in existing dictionaries: what the definition/gloss provides is half of the information needed.

But since providing optimal glosses or definitions is a difficult and time-consuming task, it is natural that most dictionaries have to adopt lower standards, and in a situation of lexical erosion face the loss of non-provided information.

As for the biolexicon, the issue appears once again unsolvable at the level of an individual language, and not only because of the workload, but because of the fact that if "good" definitions/descriptions/glosses of let's say all the Amazonian birds (of a given area) were ever created, they would remain (as far as publication is concerned) the property of the lexicographer who has provided them and would not be usable by other lexicographers/dictionary compilers lest they should be accused of plagiarism.

This is a strange paradox, since in many countries dozens of languages face the same problem, with little means to deal with it. They could benefit considerably from the availability of good lexicographical descriptions of animals or plants but cannot do so because of a conception of lexicographical work centered on single authorship and the risk of copyright-based accusations.

In a continent like South America, enhancing the capacity of dictionaries to reach a level of description of lexical items which allows the effective and full preservation of the words at stake would greatly benefit from making available ready-to-use lexicographical glosses[9] for sections of the lexicon such as the biolexicon.

9. For the Amazonian region, such glosses should become available in Spanish, Portuguese, French, English and Dutch. Because the way the gloss is written corresponds to a communicative strategy, its pattern allows the use of translation to make the glosses available in the five languages.

This does not mean that the individual lexicographer may not make up his/her own glosses or adapt or improve these freely available glosses, but only that given the scarcity of resources available for the creation of dictionaries of so-called small languages, sharing such lexicographical resources could be the only way to produce them.

3.4 Reconsidering standards and structures in EL lexicography

EL or minority language lexicography is too often considered in terms of negative comparisons with Western lexicographical standards that have emerged as the result of centuries of constant and renewed work and with the financial backing of powerful publishers. Such a perspective of course is partially grounded in huge and undisputable contrasts, in terms of the possibility to dedicate financial or human resources for lexicographical work, but has its shortcomings.

The main shortcoming of approaching small language lexicography as a substandard lexicography whose ultimate goal would be to resemble as much as possible the lexicography of the main languages of the world is that it does not question the assumption that small language lexicography may require a standard of its own, adapted to the sociolinguistic and cultural situation of the language and to its users and aimed at specific goals, such as preservation, which are mostly irrelevant in the lexicography of major languages.

3.4.1 *Beyond orthographic nomenclatures*

A very simple illustration of this subject in the Palikur situation is the issue of deciding whether the addressing system of the dictionary should or should not be orthographic in a context in which nearly half of the Palikur live in French Guyana where:

- the writing and spelling of Palikur is not taught in the school system (whereas the other half of the population is educated in Brazil in a bilingual primary school system);
- the existing orthographic standard, due to various circumstances which will not be detailed here, is to some degree opaque and unpredictable;
- many of the words that will be included in the dictionary have never been written so they have no standard spelling.

It thus becomes impossible to *presuppose* knowledge of word spelling and to rely on it for the ordinary user to be able to simply find any word in the dictionary. This leaves the lexicographer with the choice of sticking to standard alphabetical ordering of spelled words (used in Brazil) with the consequence that most

potential Palikur users will be unable to use the dictionary as it is[10] or will have to rely on the French-Palikur section to find the right word spelling (supposing they already know the spelling of the French word), or with the choice of permuting phonetic transcription and word spelling (using the phonetic and not the orthographic form as an entry in the dictionary), thus allowing access to the latter through a simplified version of the former, which can be easily guessed and can thus ensure the usability of the dictionary.

Leaving aside this illustration, it is obvious that once one starts asking oneself what could be changed in classical lexicographical standards in order to allow the dictionary to provide as much relevant information as possible – and moreover if this question is raised according to the various principles we have defined to allow language transmission and preservation – it becomes clear that it will only take a few seconds for the lexicographer to record his/her informant's pronunciation of a given word, and it will take him/her only a few minutes to record his/her informant speaking in Palikur (and possibly also Creole/French) about a given bird or plant. This will create thousands of small corpora which can be archived and thus remain accessible for the next generations and allow not only the documentation of a word but of the cultural representation and experience associated with it.

3.4.2 *Written examples*

Going back to the issue of whether or not EL lexicography should have its own standards, it is obvious that recording the pronunciation of a word and the knowledge associated with it (including the use of each word in context) may be far more important than creating examples for each word (the practice of standard dictionaries) in a context of lexical erosion, even if the latter is somewhat useful. Ultimately, having small recordings associated with thousands of individual words (or meanings) is not difficult to achieve – given that the lexicographer is documenting each word anyway with his/her informants – and may prove to be a new and efficient way to ensure the transmission of the knowledge at stake.[11]

3.4.3 *General constraints on lexicographical content*

Discussing standards, however, goes well beyond discussing the exact structuring of an entry, as lexicographical content and work are ultimately ruled by the same

10. Orthographic bidirectional tests have been conducted in order to measure this capacity, showing that it was actually lower than what could have been expected.

11. This of course implies moving away from the paper dictionary perspective, but once again, with mobile phones spreading in the Palikur community as in anywhere else in the world, transmission could well take the electronic path.

pragmatic constraints that apply to any contribution in language, namely the Gricean constraints of exhaustivity, clarity, truthfulness and relevance (Grice 1975).

Applying Grice's maxims (and Ducrot's law of exhaustivity) to defining what a lexicographical entry *has to say about something* requires understanding a dictionary or a dictionary entry as a contributional object which has to respect the following maxims:

1. *quantity*: "do not provide less information than required, do not provide more information than required";
2. *quality*: "do not say things for which you haven't got enough evidence";
3. *manner*: "do not say anything obscure";
4. *relation*: " be relevant".

It is indeed these general constraints that make the lack of exhaustivity of a dictionary somewhat unsatisfactory in terms of coverage of a lexical domain, as is the lack of exhaustivity of a lexicographical entry in terms of the information it provides in contrast with the information that is (or could be) expected – this second issue having strong consequences on the first one.

An important challenge in this respect and in the preservation perspective that we have adopted here is to combine such general constraints with the principles that we have introduced earlier. In other words, in the context of EL lexicography, defining the actual meaning of *do not provide less information than required* equals defining clearly what information is required, and defining the actual meaning of *do not say anything obscure* involves defining obscurity in relation to the different types of users, and dealing with the difficulties of satisfying various constraints simultaneously. This can be illustrated, for instance, by the situation where the name of an animal is described only by providing its scientific Latin name, this information being required for precise identification, even though it will be obscure for most users, because it is not common ground. In contrast, providing a common-ground description (such as *heron*) allows the maxim of clarity to be respected but not the maxim of quantity, as there is a lack of precision concerning which kind of heron we are referring to.

Reconsidering the standards in EL lexicography thus means clarifying what information should be provided, or to put it the other way round, clarifying which information is missing or is inadequate, if we want dictionaries to fully serve as tools for lexicon preservation and transmission.

3.4.4 *Specific constraints on biolexicon lexicography*

This can be illustrated by genuine examples of partial satisfaction of the constraint of completeness in some lexical entries of various dictionaries, all of which may be considered as providing good or rather good information about the biolexicon.

If we consider the dictionary of Yukuna (Arawakan):

Yukuna Dictionary (Schauer et al. 2005)

(1) a. *aroʾojómaji*, *ajoʾojómaji* s. piconcito: picón.[12] V.apén. aves.[13] (Schauer et al. 2005: 31)

b. *iñapimí* s. piconcito: picón. V.apén. aves. (Schauer et al. 2005: 46)

c. *choro* s. golondrina.[14] V.apén. aves. (Schauer et al. 2005: 35)

d. *juripíchaʾa* s. golondrina. V.apén. aves. (Schauer et al. 2005: 67)

e. *jutá*, *utá* s. golondrina negra.[15] V.apén. aves. (Schauer et al. 2005: 67)

f. *juwiche* s. azulejo[16] (especie de tángara). V.apén. aves. (Schauer et al. 2005: 67)

g. *kaijmeru* s. mariposa.[17] (Schauer et al. 2005: 70)
kaijméruna (término genérico) mariposas.
lachamaru mariposa nocturna.[18]
paʾajrú mariposa nocturna.
phichí panami mariposa grande, color café.[19]
pina mariposa grande, morfa azul.[20] V.apén. insectos.

h. *kalapichi perí* s. gavilán especie.[21]V. perí. V.apén. aves. (Schauer et al. 2005: 73)

i. *kapana* s. gavilán, especie más grande.[22] V. perí. V.apén. aves. (Schauer et al. 2005: 77)

---

12. Small woodpecker, woodpecker. The standard Spanish for woodpecker is nevertheless *pájaro carpintero*.

13. Bird.

14. Swallow.

15. Black swallow.

16. From azul/blue.

17. Butterfly.

18. Nocturnal butterfly.

19. Great butterfly with a coffee colour.

20. Great butterfly, blue morpho.

21. Kite species.

22. Kite, bigger species.

Moving now to a *Yine/Castellano Dictionary*:

(2) a. *pushchopu* s. victor díaz (especie de pájaro[23])
(Urquía Sebastián & Urquía Sebastián 2008: 21)

b. *pushropushro* s. pichihuichi (especie de pájaro)
(Urquía Sebastián & Urquía Sebastián 2008: 21)

c. *shaniyaka* s. huanayo (especie de garza[24])
(Urquía Sebastián & Urquía Sebastián 2008: 33)

d. *shawashka* s. paucarcillo (especie de pájaro) *Serolnikta nikanata shawashka*. El paucarcillo estaba comiendo maduro.[25]
(Urquía Sebastián & Urquía Sebastián 2008: 33)

e. *taki* s. tibe (especie de gaviota[26]) Ksatu gajerni taki. El tibe vive en la playa.[27] (Urquía Sebastián & Urquía Sebastián 2008: 24)

we may observe that:

- no literal translation of the noun is provided;
- no Latin name is provided;
- local Spanish names are provided, some of them being loans from other Amerindian languages (e.g. *pichihuichi*);
- class membership is provided but at various levels, ranging from direct hyperonyms (e.g. *gaviota*/*gull*) to the equivalent of what scientific classification calls class (e.g. *bird*);
- class membership is inserted as an explanation of the locally accepted name, *a victor díaz is a species of bird* or a *paucarcillo is a species of bird*, thus providing a piece of information which cannot be considered as common ground and may be said to provide *a piece of information that is required*, but may nevertheless not provide *all the information that is required*.

As for this last point, it may be the case that the ignorant reader does not know what kind of bird a *victor díaz* or *paucarcillo* actually is, and may need another dictionary to find out, and that every time the concerned species will not be a familiar one, which may be the case for most living beings. Nevertheless, all the

23. Bird species.

24. Heron/Ardeide species.

25. The paucarcillo is eating wood.

26. Gull species.

27. The tibe lives on the shore/beach.

information provided remains fully relevant, and could ultimately allow the exact referent to be found.

As a contrast, we may now consider the dictionary of another Arawakan language from Colombia, namely Piapoco.

Piapoco Dictionary (Klimpp 1995: 3)

(3) a. *àapi* s. culebra[28] (en general).
b. *àapi àleu minali* s. macabrel[29] (lit. culebra arbol morador[30]); *Corallus enydris enydris.*
c. *àapi íinu wiriichu* s. saltòn del monte (lit. culebra su esposa saltamontes[31]); *Tettigniidae.*
d. *àapi íiwitami* s. vinegarone (lit. culebra cabeza muerte[32]); *Thelyphonidae.*
e. *àapi wéetéeriwa iyú* s. culebra venenosa[33] (lit. culebra por medio de lo cual morimos[34]).

Leaving aside the last item,[35] we may notice this time that each entry associates three types of information, namely *denominative equivalent*, *meaning in the language itself* (Piapoco) and ultimately the *Latin/scientific identification* of the species or family. In some respect, it is more complete and satisfactory than the previous descriptions, but with the risk that whenever the denominative equivalent is not a common ground item (*culebra*) like here, providing only a denominative equivalent will not give sufficient information to the reader, who will know how something is named in three languages, Piapoco, Spanish and Latin, and what it means in Piapoco, but will ultimately not know what it is. This could have been the case in Yine for *taki*, whose denominative equivalent *tibe* is not common ground in Spanish and would have remained obscure if the information *a type of gaviota* had not been immediately provided.

---

28. Snake.

29. Tree boa, macabrel common.

30. Lit. Snake tree morador

31. Lit. Snake its spouse grasshopper.

32. Lit. Snake head dead.

33. Venomous snake.

34. Lit. Snake by which we die.

35. Which may be a usual name or might prove to be only a translation from Spanish if obtained through a L2/L1 question.

Such a structure is the consequence of the fact that when it comes to the biolexicon:

- all nouns are also names (e.g. *bluethroat* or *awiyaybu* in Palikur is a noun and the name of a bird);
- they have a meaning as nouns, independent of their meaning as names (*bluethroat* does not mean *bird*, it means *blue throat; awiyaybu* does not mean *bird*, it means *heavy tail*);
- what they name must be identified (a *bluethroat* may be identified as *Luscinia svecica*, an *awiyaybu* may be identified as *Colonia colonus*);
- what the noun stands for must be clarified, which often means decomposing its morphological structure and always means describing its motivation (*cuckoo* cannot be decomposed but refers onomatopoeically to the sound produced by the bird. Since *victor díaz* in Spanish refers to the bird's song, knowing what *pushchopu* means is a necessity);
- it is a necessity, beyond names and nouns, to describe what something *is* by providing the best possible definition of the denominative referent (a *small robin-like passerine bird* for *bluethroat*, a type of *gaviota*/*gull* living on shorelines for *tibe*).

What is important is to realize not only that all three types of information are required, but that the denominative equivalent is often provided in two forms (here Spanish and Latin).

Hence, three compulsory types of information should always be provided and one should be distinguished from the other in order to avoid any ambiguity of their nature:

- denominative equivalent(s) for the identification of the referent;
- lexical or morphemic decomposition for understanding the expression itself (in Frege's sense) and beyond that for understanding its motivation;
- encyclopedic description/definition for a characterization of the referent (as in classical monolingual dictionaries), providing basic synthetic information about the nature of what is named and which makes it accessible to the reader.

This information should be provided even if it is not always easy:

- to identify what exactly a name refers to when it comes to non-spectacular or familiar species;
- to find denominative equivalents in the main language;
- to avoid mismatching between different classificatory levels (i.e. presenting as the name of a class what is in fact only the name of a member of a class);

– to ensure morphemic or lexical decomposition, whenever word-formation is not transparent;
– to be sure of the motivated character of the word.[36]

Moreover, because the three types of information are relevant, it is necessary to achieve clarity to allow them to be discriminated from one another. Otherwise, as mentioned for the case of *golondrina negra*, the reader will be left having to guess whether *golondrina negra* is a translation, the actual name of the bird in Spanish or a short encyclopedic description.

Since a whole article would be necessary to address the issues related to each of these difficulties, we shall limit ourselves here to highlighting some specific issues.

First and most importantly, because of the almost impossible task for the ordinary lexicographer to solely produce thousands of encyclopedic characterizations/definitions which can combine correctness and accessibility to all readers, it would be a major boost for field lexicography if these definitions could be shared and made available for everyone within a country or area.

Such availability would leave each lexicographer with the possibility:

– to use them or not;
– to improve them;

but also:

– to sometimes avoid, through these definitions, imposing one's worldview on another.

It may, for instance, be the case that all birds classified as *avakni* in Palikur can be described as *rapaces* (raptors) in French, but it is also the case that some birds, such as the *kumak* (swallow-tailed kite, also a raptor), since they are not considered as raptors by the Palikur, should be described using a broader term such as *oiseaux* (birds) and/or that the description should mention the discrepancy as such (*Not considered to be a raptor*).

Similarly, it may often be useful and important to combine classificatory systems in the definition in order to overcome the mismatch between categories.

---

36. The term "motivated" refers to the fact that the name given to an animal, for example, is not arbitrary. For instance, the name "heavy-tail" (*awiyaybu*) is motivated by a physical/ethological feature of the bird. Motivation may be transparent, as in this example, or less so, for instance, when hummingbirds are named by the name of a plant because of the similarity of shape between their nest and the plant at stake.

For instance:

- by using monolingual descriptors, for example, describing in Palikur a *wiki* as "a small *titup*, flycatcher-like bird";
- by acknowledging discrepancies, either indirectly (e.g. *tuumwi*: name of a group of birds which includes all woodpeckers, treecreepers and some treecreeper-like ovenbirds/*Furnaridae*) or directly (e.g. *swiswi*: all waders/*Scolopacidae* except the *sakaska* giant snipe; *inuiri*: identifiable as harpy eagle juvenile but considered a distinct bird).

As for this last point, it must be stressed that if ultimately the constraint of relevance implies that all definitions be adjusted to the targeted audience (community members, local and non-local outsiders, scientists) – and thus finding compromises – prioritizing transmission of cultural knowledge and representations from generation to generation implies including (and favoring) them as much as possible in the encyclopedic definition provided (e.g. *mamatki*: vocal bird which can be heard only during the night and is widely feared and considered as a bad omen).

Ultimately, however, ensuring the transmission of such knowledge will clearly be more effective if it can be recorded in the first place during the lexicographical fieldwork and made available later on. Saving such knowledge has become quite easy and making it available can be achieved in the Palikur context in any school library or on any cultural association's computer. All conditions are thus met to allow the preservation of such knowledge for future generations.

3.4.5 *Sometimes adopting a non-Gricean perspective*

A last and paradoxical issue at this stage concerns the need in EL lexicography to also adopt a radically non-Gricean approach to endangered languages.

Very often, when working with informants, the lexicographer is exposed to what we can call *the partial knowledge syndrome* (PKS), in other words, to having less than complete information about a word.

For obvious reasons, one is then tempted, because of the missing part, to not mention the information at all, in order to avoid criticism or to look incompetent or ridiculous, with the result that too often, a large amount knowledge is finally lost.

This implies that in EL lexicography there is a strong necessity not only to refuse an all-or-nothing approach to language documentation according to which lexical descriptions should not be published if incomplete, but to create a section dedicated to documenting all such forms of partial knowledge.

It may, for instance, happen that we as lexicographers have only heard the name of a (yet unidentified) night bird but cannot provide its identification nor

precise description, and yet as EL lexicographers, we should not exclude this name from the dictionary, nor the little we know about it: storing it in a partial knowledge section could allow the entry to be improved at a later stage, provide the impetus to look for the missing information and eventually save the word.

Moreover, for obvious reasons, vanishing words and word uses being much more exposed to PKS than other words, the paradox is indeed that in many cases, for such "on the edge" words, the little we know will finally be forgotten and lost due to considerations of professional reputation. In Palikur, this concerns a large part of the *kyaptunka* vocabulary (ceremonial and vehicular variety of the language), especially the meaning of certain words found in songs recorded in the 1970s that appear to be lost.

In some cases, incompleteness may be limited or associated to residual uncertainty, as for example when a bird or bug name is known but its exact identification is not yet available, as was the case, for instance, in our Yukuna examples for various nocturnal butterflies (mariposas) and non-specified *picón* woodpeckers (*aroʾojómaji/iñapimí*).

In such cases, meta-communication would be satisfactory, a comment like "not yet identified" replacing the unavailable Latin or local identification. But providing identification clues when they are available would be better (e.g. *A hummingbird with atypical suspended nest*), for it could lead to later identification.

The same thing could be said regarding uncertainty (and the maxim of quality) or residual uncertainty, which should be mentioned as such. For instance, in Palikur the name *pakapka* should be noted as "most probably Pompadour Cotinga, *Xipholena punicea*", so long as the identification is not fully confirmed.

But in cases where incompleteness is more radical, we suggest that the solution of adopting explicitly an "emergency" approach to EL lexicography be implemented – according to 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.

In other words, dictionaries should accept presenting state-of-the-art lexical knowledge and not exclude nor mask partial knowledge.

4. Conclusion

What we have tried to show here is that because lexical erosion is a crucial issue for the survival of languages and an aggravating factor of endangerment and because it is often related to lifestyle changes which threaten word learning and

lexical knowledge, new modes of transmission of the lexicon must sometimes be considered in order to allow its preservation.

Furthermore, we asserted that because dictionaries are the usual form of representation of the lexicon of a language, they ought to play a major role in avoiding or countering lexical erosion, but also that in order to ensure actual preservation and transmission of lexical knowledge, field lexicography and EL lexicography have to set their own standards and reconsider their approach to issues such as lexicographical coverage of all lexical fields, completeness of the lexicographical descriptions proposed, usability and readability of the dictionary by community users, enlargement and oralizing of the former "example" section, phonetic form preservation and hearability, recording of the cultural knowledge associated with each item, avoidance of ethnocentric presentation of the items, and creation of a specific section for the documentation of partial lexical knowledge.

What we have finally advocated is a conception of lexicographical work in which the individual lexicographer is not asked to do everything. Rather, instead of hundreds of isolated and sometimes hopeless efforts to reach a certain level of completeness or to define methods and tools to address the most demanding section of the lexicon, individual lexicographers, in terms of tools, expertise but also in terms of availability of thousands of legally/freely usable lexicographical descriptions, could work more cooperatively to both enhance the capacity of dictionaries to play a significant role in the preservation of a language and minimize the cost of each dictionary.

This of course supposes to some extent a different organization of lexicographical work, but it is clear that either at a national or international level, much could be done in a rather short time by sharing resources and content that could benefit hundreds of languages at once.

References

Cristinoi, A. & Nemo, F. 2013. Challenges in endangered language lexicography. *Lexicography and Dictionaries in the Information Age*, 126–132. Denpassar: Airlangga University Press.

Green, D. & Green, H. 2010. *Yuwit Kawihka Dicionário Palikúr – Português*. Belem : SIL.

Grenand, F. 1989. *Dictionnaire Wayãpi-Français, Lexique Français-Wayãpi*. Paris: Peeters/Selaf.

Grenand, F. (ed.). 2009. *Encyclopédies Palikur, Wayana, Wayãpi: Langue, milieu et histoire, fascicule encyclopédie des Amérindiens de Guyane*. Paris: PUO-CTHS.

Grice, P. 1975. Logic and conversation. In *Syntax and Semantics, 3: Speech Acts*. P. Cole & J. Morgan (eds), 41–58. New York NY: Academic Press.

Hartmann, R.R.K. 2003. *Lexicography: Critical Concepts*. New York NY: Routledge.

Klimpp, D.A. 1995. *Vocabulario Piapoco-Español*. Bogotá: Asociación Instituto Lingüístico de Verano.

Launey, M. 2003. *Awna Parikwaki. Introduction à la Langue Palikur de Guyane et de l'Amapa*. Paris: IRD Éditions.

Mosel, U. 2011. Lexicography in endangered language communities. In *The Handbook of Endangered Languages*, P.K. Austin & J. Sallabank (eds), 337–353. Cambridge: CUP. doi:10.1017/CBO9780511975981.017

Schauer, J.G., Schauer, S., Yukuna, E. & Yukuna, W. 2005. *Meke Kemakánaka Puráka²aloji: wapura²akó chu, eyá Karíwana chu (Diccionario Bilingüe: Yukuna – Español; Español – Yukuna)*. Bogota: Editorial Fundación para el Desarollo de los Pueblos Marginados.

Urquía Sebastián, R. & Urquía Sebastián, W. 2008. *Diccionario yine–castellano (parcial)*. <http://www.lengamer.org/publicaciones/diccionarios/Dic_Prelim_Yine.pdf>