

*Lexica: the dictionary series of the Pangloss Collection*

# Na (Mosuo)-Chinese-French dictionary

Alexis Michaud

Version 1.2.2  
July 2018

## Table of contents

Introduction .....	iii
1. Presentation of the language and data .....	iii
1.1. The Yongning Na language .....	iii
1.2. Dialect and language consultants .....	iv
2. User guide .....	vi
2.1. Structure of dictionary entries .....	vi
2.2. Versions of the dictionary .....	viii
2.3. Perspectives for technological improvements .....	xi
3. A research programme: <i>lexicology</i> .....	xii
Acknowledgments .....	xiv
References .....	xv

# Introduction

This introduction is divided into (i) a presentation of the language and the consultants, (ii) a User’s guide (also containing some technical information on the software environment) and (iii) some remarks about the broader research agenda.

## 1. Presentation of the language and data

### 1.1. The Yongning Na language

This dictionary documents the lexicon of the Na language (ႬႬ-ᱦᱮᱰᱤ) as spoken in and around the plain of Yongning, located in Southwestern China, at the border between Yunnan and Sichuan, at a latitude of 27°50’ N and a longitude of 100°41’ E. This language is known in China as ‘Mosuo’ 摩梭话. The language’s *Ethnologue* code is nru (Lewis, Simons & Fennig 2016) and its *Glottolog* code is yong1270 (Nordhoff 2012).

A review of the literature about Na and the other languages of the Naish group is provided (in Chinese) by Lǐ (2015). For an English-language introduction, see Michaud, He Limin & Zhong Yaoping (2017). Important references include a reference grammar of a dialect close to the one described here: that of Luoshui 落水 (Lidz 2010), as well as research into the tone systems of previously undocumented dialects (Ā 2016; Dobbs & La 2016). New to 2017-2018 is the successful application of a tool for automatic phonemic transcription to Yongning Na data (Adams et al. 2017; Adams et al. 2018); the tool is available online as open-source software<sup>1</sup>.

---

<sup>1</sup> <https://github.com/oadams/persephone>

Concerning lexicographic work, *An anthology of everyday words and expressions in the Mosuo language* (Zhíbā & Xǔ 2013) presents vocabulary and expressions arranged by semantic field. The authors are a native speaker from the Lake area (泸沽湖) and a doctor in linguistics from Yunnan University. Their fieldwork is described as covering the Yongning plain and the Lake area, but with the Yongning plain as the main research area (p. 2). Approximations in phonetic notation are so numerous that they make the volume unreliable as a work of reference. Voicing contrasts were challenging for the linguist in the team, whose training was mainly focused on the theory and practice of teaching Chinese as a foreign language. Thus, the name of the mountain  $k\gamma\text{-}m\gamma\text{-}l$  is transcribed as  $g\alpha^{55}mu^{55}$ , with a voiced initial (p. 17 and elsewhere). The mountain’s name in Chinese, *Gemu* 格姆山, may have exerted an influence here. Conversely, the adjective  $dz\gamma\text{-}l$  ‘good’ is transcribed as  $t\phi\alpha^{13}$ , with an unvoiced initial. Some phonemes, such as uvulars, are absent from the notations.

## 1.2. Dialect and language consultants

Unless otherwise stated, all the data are from one language consultant, Mrs. Latami Dashilame ( $la\text{-}t^h\alpha\text{-}mi\text{-}l$   $t\alpha\text{-}s\gamma\text{-}l\text{-}la\text{-}m\gamma\text{-}l$ ; Chinese: 拉它米打史拉么). She was born in 1950 in the hamlet called  $\alpha\text{-}l\alpha\text{-}l\text{-}xw\gamma\text{-}\#l$ , close to the monastery of Yongning. The administrative coordinates of this hamlet are: Yúnnán province, Lǐjiāng municipality, Nínglàng Yí autonomous county, Yǒngníng district, Ālāwǎ village (云南省丽江市宁蒗彝族自治县永宁乡阿拉瓦村). The choice to work in one location only, and essentially with one consultant, is, again, based on the investigator’s focus on the tone system. There is considerable dialectal diversity within the Na area (much more so than in the Naxi-speaking area); the tone systems of different villages are conspicuously different, and this geographical diversity combines with dramatic differences across social groups, and across generations. The obvious thing to do seemed to be an in-depth description and analysis of the language as spoken by one person (simultaneously making a few forays into other idiolects and dialects). Data from other speakers are indicated using their codes in the author’s database of speakers of Naish languages. Table 1 provides the speaker codes.

language consultant code	name	year of birth
F4	$la\text{-}t^h\alpha\text{-}mi\text{-}l$ $t\alpha\text{-}s\gamma\text{-}l\text{-}la\text{-}m\gamma\text{-}l$	1950
F5	$ki\text{-}zo\text{-}l$	1973
F6	$t\phi^hi\text{-}d\gamma\text{-}\#l$	1987
M18	$la\text{-}t^h\alpha\text{-}mi\text{-}l$ $t\alpha\text{-}s\gamma\text{-}l\text{-}t\alpha\text{-}l\gamma\text{-}l$	1972
M21	$ho\text{-}dz\gamma\text{-}ts^he\text{-}l$	1942
M23	$d\gamma\text{-}l\text{-}d\gamma\text{-}l$	1974

Table 1. Language consultants.

My fieldwork on Yongning Na began in October 2006.<sup>2</sup> A list of words was begun through elicitation, and gradually expanded and corrected as narratives were recorded and transcribed; addition of new words was therefore a slow process. An advantage of placing the emphasis on text collection is that a context is available to help clarify the meaning of newly encountered words, also offering a basis for further discussion of their usage with language consultants. Systematic elicitation of large amounts of vocabulary was not carried out, hence the limited number of entries: currently on the order of 3,000.

In the classical tradition of linguistic fieldwork, a language description should include a dictionary, a grammar, and a collection of texts. Linguistic fieldwork consists in “going into a community where a language is spoken, collecting data from fluent native speakers, analysing the data, and providing a comprehensive description, consisting of grammar, texts and dictionary” (Dixon 2007:12). The grammar, texts and dictionary are referred to as the “Boasian trilogy” (Foley 1999) by reference to Franz Boas’s foundational work collecting North American languages (Boas 1902:190; Boas & Swanton 1911); the trilogy can now be said to have become a *tetralogy* as it integrates a multimedia component: audio and video recordings (Musgrave & Thieberger 2014). In addition to this dictionary, the following resources are available:

- A set of Na recordings with time-aligned transcriptions is available from the Pangloss Collection (Michailovsky et al. 2014; Michaud et al. 2016).<sup>3</sup>
- A book-length study of Na morphotonology (Michaud 2017) is available online.<sup>4</sup> It also contains detailed information on the phonemic analysis.

---

<sup>2</sup> Some notes about fieldwork are presented at [http://lacito.vjf.cnrs.fr/pangloss/fieldwork/Na\\_fw-Yongning\\_en.htm](http://lacito.vjf.cnrs.fr/pangloss/fieldwork/Na_fw-Yongning_en.htm) (consulted 15/03/ 2018).

<sup>3</sup> The current web address is [http://lacito.vjf.cnrs.fr/pangloss/corpus/list\\_rsc.php?lg=Na](http://lacito.vjf.cnrs.fr/pangloss/corpus/list_rsc.php?lg=Na)

<sup>4</sup> <http://langsci-press.org/catalog/book/109>

## 2. User guide

Entries and examples have translations into English, Chinese and French. Three language settings are offered for the PDF: English-based (this version), Chinese-based,<sup>5</sup> and French-based.<sup>6</sup> Chinese translations are often a useful complement to the translation in English, as there are often closer equivalents: for instance, 𐀀𐀮 translates straightforwardly as Chinese 扛 whereas the English translation is more roundabout: ‘to carry on the shoulder’.

### 2.1. Structure of dictionary entries

Each entry contains

- a *phonological transcription*: the form of the word in phonetic alphabet; tone is indicated in terms of phonological categories
- a *surface form*: the surface-phonological form of the word in phonetic alphabet. Tone is indicated in terms of tonal realizations. This form can be read by anyone with a knowledge of the International Phonetic Alphabet, without requiring an understanding of the mapping of underlying phonological tone categories to surface tone in Yongning Na
- an *orthographic representation*: notation in a Romanized script devised by Roselle Dobbs with Mosuo collaborators
- *part of speech*: an indication of the part of speech, using a simple set of labels
- *tone*: the tone category of the word. This information is already present in the phonological transcription; having it repeated on its own facilitates searches
- *definitions* in English, Chinese and French
- *examples* with translations
- *links* to related words, such as synonyms, or constituent parts of complex words
- *classifier*: for nouns, an indication on the more commonly associated classifiers

For surface-phonological forms, the aim was to achieve greatest simplicity. Special symbols used in the word’s underlying phonological form are removed from the surface form: dashes indicating junctures internal to the word, and tilde in reduplicated forms. For classifiers, the numeral ‘one’ is added before the classifier, because classifiers are not free forms: they cannot be said on their own. For other bound morphemes, such as affixes and clitics, no surface form is indicated.

---

<sup>5</sup> Available at <https://halshs.archives-ouvertes.fr/halshs-01744420>

<sup>6</sup> Available at <https://halshs.archives-ouvertes.fr/halshs-01204645>

The proposed orthography in Latin alphabet was added by Roselle Dobbs in 2017. The transcription was developed by R. Dobbs with Na consultants, with a view to use within the Na community. Importantly, this is not a transliteration of the phonology: phonological forms are all based on the dialect of consultant F4, whereas proposed orthographic representations are intended by R. Dobbs and her collaborators as a cross-dialect writing system. Given the high degree of dialectal diversity, proposing a transcription that is acceptable for speakers of several dialects implies some compromises. For instance, only the low tone (L), presumed to be more stable, is indicated in the orthography. Until a full description of this romanized writing system is published, requests for information about orthographic developments should be directed to Roselle Dobbs ([rosellemav@hotmail.com](mailto:rosellemav@hotmail.com)).

Among examples, those elicited to verify the output of certain combinations of tones are preceded by the mention “(Phonological elicitation)”. Proverbs and sayings are marked as “(Proverb)”.

Some pieces of information are not shown in the PDF and online versions. These are:

- An indication of *semantic domain*: ‘society’, ‘house’, ‘body’, ‘plant’, ‘animal’, and so on. No attempt was made to use a fine-grained classification of the sort found in the WordNet database of English, where nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonym (Fellbaum 2005). This is a rough division; the labelling relies partly on form, and partly on semantic contents. As for other aspects in the dictionary, choices made reflect the investigator’s research priorities: for instance, the entries for ‘day’, ‘night’, ‘month’, and ‘year’ were tagged as “classifiers”, along with all other nouns that can appear immediately after a numeral. This allowed easy extraction of all classifiers for the purpose of a study of the tone patterns of classifiers (Michaud 2013). These lexical items could just as well have been tagged as belonging to the semantic domain of ‘time’.
- *Notes on past notations*: information tracing the history of notations, from the first fieldwork to the current version. For instance, the entry  $\eta w \mathbf{w} \mathbf{r} \mathbf{p}^h \mathbf{\ae} 1$  ‘tile’ has a note that indicates that it was initially written with a M.H tone pattern, and with vowel  $\mathbf{\ae}$  in both syllables:  $\eta w \mathbf{\ae} \mathbf{p}^h \mathbf{\ae} 1$ . The note explains that the perception of  $\mathbf{\ae}$  in the first syllable is due to a phonetic tendency towards regressive vowel harmony. Verifications are also consigned in this field. About half the entries have information of this type.
- *glosses*: glosses in English, Chinese and French, intended for the glossing of texts. The dictionary adopts the abbreviations recommended in the Leipzig Glossing Rules (Comrie, Haspelmath & Bickel); all other terms are provided in full. Glosses mostly follow the choices made by Lidz (2010).

Some monosyllabic roots extracted from disyllables are indicated by the symbol †. No surface form is provided, as these monosyllabic forms are not currently in use in the language.

Borrowings from Chinese and Tibetan are indicated as such in cases where identification seems straightforward. No efforts at systematic elicitation of borrowings from either language were made, but all loanwords occurring in texts were added to the dictionary. The information provided includes: the donor language, the form in the donor language, and some explanations. When the number of syllables in the borrowed word is the same as in the donor language, the glosses in English (and French) start by the original word followed by two colons and a translation: e.g. ‘办法::solution’ for *pælhwaŋ*†.

## 2.2. Versions of the dictionary

### 2.2.1. Principles

The Lexica series aims to combine *readability* for users who browse through the dictionaries with *computer-readable encoding* suitable for Natural Language Processing. Dictionaries are offered (i) as online dictionaries in HTML format, (ii) as PDF documents, and (iii) as databases in Toolbox/MDF format and in XML format. This appears most suitable for dictionaries, which constitute work-in-progress. Successive versions are numbered, following the model used in software development. Major releases will probably occur every few years. The versions hosted in the HAL-SHS archive are presented in Table 2.

version number	date	caractéristiques
1.0	September 2015	first public release
1.1	November 2016	improvements in contents
1.2	March-April 2018	addition of orthography and surface forms, and changes to the typesetting

*Table 2. Versions of the Na dictionary released to date in the HAL-SHS archive.*

The list of words as of 2011 was deposited in the STEDT database.<sup>7</sup> The same year, under the impetus of Guillaume Jacques and Aimée Lahaussais, plans were made to bring the word list closer to the standards of a full-fledged dictionary. A project was deposited with the *Agence Nationale de la Recherche*, accepted in 2012, and begun in 2013: the HimaCo project (ANR-12-CORP-0006).<sup>8</sup>

<sup>7</sup> <http://stedt.berkeley.edu/>

<sup>8</sup> <https://himalco.hypotheses.org/>



### 2.2.2. *Versions 1.0 and 1.1: complying strictly with the LMF standard*

Céline Buret, a computer science engineer, worked with the project team for two years (Nov. 2014-Oct. 2015). She converted the data to the format of the Field Linguist's Toolbox (MDF), then produced scripts for conversion to a XML format complying with the Lexical Markup Framework standard, a pivotal format designed for machine-readable dictionaries (Francopoulo 2013; Romary 2013). The scripts constitute a Python 2 library called PYLMFLIB, for: Python LMF library.<sup>9</sup> In 2015, version 1.0 of the online and PDF versions of the dictionary were produced and published online, along with the source document in MDF (Toolbox) format. Version 1.1, released in 2016, used the same library of computer scripts.

### 2.2.3. *The Lexika software and the Lexika collection of dictionaries*

A limitation of the Lexical Markup Framework in its 2013 formulation gradually became apparent: it places a constraint on what an entry can contain. Subentries that belong to different grammatical categories need to have separate entries set up. For instance, 1a4-kʰɿ4 can mean both 'year of the Tiger' and 'born in the year of the Tiger' (in Chinese: 虎年 and 属虎). From the linguist's point of view, it is desirable to set up two subentries within the same entry. But the part-of-speech categories are different ('year of the Tiger' is a noun phrase, and 'born in the year of the Tiger' is a predicate, categorized as an adjective), which, in the Lexical Markup Framework, necessitates setting up two different entries.

In 2016, Benjamin Galliot, working at CNRS-LACITO under a fixed-term (six-month) contract funded by CNRS (Délégation Paris-Villejuif), wrote a new library, LEXIKA,<sup>10</sup> using Python 3, which allows for native management of the Unicode standard. The name of the software, LEXIKA, differs (in its written form) from that chosen for the series of dictionaries of the Pangloss Collection: LEXICA; this is intended to prevent confusion between the software package, on the one hand, and the dictionary series. Benjamin Galliot also wrote a XSL script for generating the PDF versions of the dictionary. The new version of the Na dictionary released in the year 2018 (version 1.2) now has

- a phonetic transcription of tone as it surfaces on the item pronounced in isolation: a surface-phonological transcription of tone, in addition to the indication of the underlying tone category
- a romanized representation: a proposed spelling, devised by Roselle Dobbs and her collaborators (see more below)
- more cross-references between entries, pointing to synonyms, for instance.

---

<sup>9</sup> <https://pypi.python.org/pypi/pylmlib/1.0>

<sup>10</sup> <https://bitbucket.org/BenjaminGalliot/lexika>

Intermediate versions (1.2.1, etc) are generated as the work progresses and made available through a GitHub repository that hosts the database.<sup>11</sup> The database is available from the GitHub repository in two formats:

- an XML file using the model designed by Benjamin Galliot (Bonnet et al. 2017)
- a MDF file, which can be opened with the TOOLBOX software package.

Planned improvements for future versions include the addition of

- *audio files for each head word*: this function has successfully been tested, but the editing of audio files still needs to be conducted
- *links to the entire set of online recordings*: listing all textual occurrences in the lexicon entry, with links to the audio file and its aligned transcription. Textual occurrences ultimately constitute the best resource to document a word's usage. The examples currently presented in the dictionary are few in number, compared to the occurrences in texts, and their context of use may not be clear, despite efforts at providing contextual information for examples jotted down during fieldwork
- *additional cross-references* between entries
- *more Chinese loans*: indicating the Na pronunciation of Chinese words that are now commonly used by speakers of Na.

I would gratefully receive any comments or notifications of errors that the reader may wish to bring to my attention: please send e-mail to [alexis.michaud@cnrs.fr](mailto:alexis.michaud@cnrs.fr) or create 'Issues' in the dictionary's GitHub repository.<sup>12</sup> Collaborations are especially welcome for the following improvements :

- *the vocabulary of religion*: the field of religion remains mostly unexplored; the main consultant and I both lack the command of Txbtxn that would be essential for this part of the investigation, and involvement of consultants from the Yongning monastery did not prove feasible in view of current restrictions on contacts with foreigners
- *plants and animals*: as a dweller of the plain, the main consultant does not have extensive knowledge of wild plants and animals; the number of entries recorded so far remains small, and some definitions are currently limited to general indications such as 'a type of pine'. To arrive at exact identification, and at more extensive lexicographic coverage, would require collaboration with other consultants, and with botanists.

---

<sup>11</sup> <https://github.com/alexis-michaud/na/tree/master/Dictionary>

<sup>12</sup> <https://github.com/alexis-michaud/na/issues/new>

It is also tempting to include in the linguistic description (either the dictionary or the grammar) expressive noises, such as clicks and the ingressive sound ɬ̤. The meaning of this expressive noise in Na can be characterized in the same way as that of words in the dictionary: the full definition would be that *it expresses enjoyment of food or drink ('Yummy!'), and is also used to express admiration of a beautiful object, scene, or prospect*. A reason for not listing it simply under ɬ is that, unlike interjections, ɬ̤ is not pronounced on expiratory airflow, but on ingressive airflow. The air flows through the sides of the mouth, which is where saliva flows when one's mouth waters. Increasing precision in the annotation in audio files, aiming at complete (exhaustive) transcription of all relevant events in the speech signal, has potential to enrich the linguistic description of such events.

## 2.3. Perspectives for technological improvements

An exciting prospect consists in creating dynamic links between resources, so that dictionaries and grammars can become interconnected, and also link to the texts that constitute the core of linguistic resources. The way forward has been described clearly (Nordhoff 2008; Maxwell 2012); time will tell whether we are able to bring this vision to the stage of practical achievements within a few years, or whether the project will be postponed year after year for want of the human resources required for its realization.

### 3. A research programme: *lexicology*<sup>13</sup>

Lexicography has been considered as somewhat peripheral among the language sciences: its task would consist in carrying out a simple inventory. Samuel Johnson, in the Preface to his 1755 *Dictionary of the English Language*, expresses this observation with much eloquence:

It is the fate of those who toil at the lower employments of life, to be rather driven by the fear of evil, than attracted by the prospect of good; to be exposed to censure, without hope of praise; to be disgraced by miscarriage, or punished for neglect, where success would have been without applause, and diligence without reward.

Among these unhappy mortals is the writer of dictionaries; whom mankind have considered, not as the pupil, but the slave of science (...). Every other authour (*sic.*) may aspire to praise; the lexicographer can only hope to escape reproach, and even this negative recompence (*sic.*) has been yet granted to very few.

A pithy summary of the same observation (two and a half centuries later) by a professor of linguistics is that “the lexicographer is the equivalent in linguistics of the guy stacking the shelves at Sainsbury’s” (anonymized personal communication, 1996). In contrast to the description of the morphosyntax of a language, which aims to highlight the cogency of the grammatical system as a whole, the description of the lexicon tends to be seen as piecemeal in nature. The alphabetical order in which dictionaries are presented looks like a confession that the lexicon lacks internal organization.

This somewhat disparaging view overlooks the many possibilities for developing lexicography into *lexicology*. To write a dictionary is to explore the structure of the lexicon of a language (François 2008), in connection with the study of social and cultural structures. Writing a dictionary requires delving into meaning, attempting to delineate connotations, polysemy, and relationships between words; many dictionary entries could easily be expanded into essays. Thus, the vocabulary of kinship provides valuable insights about family structures and their history. (In the case of the Na language, the usefulness of this source of information has long been recognized (Fù 1980), but it remains to be exploited systematically.) Beyond the semantic domain of kinship (a classic in the field of ethnolinguistic studies), various areas call for exploration, including emotions (Tersis & Boyeldieu 2017) and spirituality (François 2013). In-depth lexicographical descriptions

---

<sup>13</sup> This section, which aims to highlight the importance of lexicography for linguistic research, is based on insights from a research programme of the LACITO research centre.

can serve as a basis for various research approaches, which combine typology with a dynamic (diachronic) dimension (Gast, König & Moyse-Faurie 2014; Koptjevskaja-Tamm, Rakhilina & Vanhove 2015; Juvonen & Koptjevskaja-Tamm 2016; Schapper, San Roque & Hendery 2016). One promising strand of research is the study of contact effects on the lexicon.

“The tendency for bilingual individuals to align the semantic structures of the languages they speak leads to the diffusion of certain lexical categorizations across vast linguistic and cultural areas: thus, specific semantic distinctions, specific patterns of polysemy, specific set phrases become telltale signs of a given area. It is sometimes possible to explain these areal phenomena in terms of links between language use and social practices in the region: certain modes of family organization, for example, may be correlated with specific lexical structures in the field of kinship, or in the vocabulary of marriage and interpersonal relationships.” (Alexandre François & Lameen Souag, séminaire « Structures du lexique : typologie et dynamiques »<sup>14</sup>, LACITO, janvier 2018)

Specifically, comparative research into the lexicon of Na and Pumi (Daudey 2014) appears especially promising for an in-depth understanding of the languages and cultures of these two peoples which have coexisted in the Yongning Plain for centuries. This is one of many research issues which I hope the present dictionary will help investigate.

---

<sup>14</sup> Source: <http://lacito.vjf.cnrs.fr/themes/lexique.htm> (consulted on February 15th 2018). Original text: « La tendance, chez les individus bilingues, à aligner les structures sémantiques des langues qu'ils parlent, a permis la diffusion de certaines catégorisations lexicales à l'échelle de vastes aires linguistiques et culturelles : c'est ainsi que certains découpages sémantiques, certaines polysémies ou phraséologies, deviennent les symptômes d'une aire donnée. Parfois, il est possible d'expliquer ces phénomènes areaux par des liens entre pratiques langagières et pratiques sociales répandues dans la région : certains modes d'organisation familiale, par exemple, pourront être corrélés à des structures lexicales spécifiques dans le domaine de la parenté, ou dans le vocabulaire du mariage et des relations interpersonnelles. »

# Acknowledgments

Many thanks to Picus Ding for putting me in touch with the Mosuo scholar Latami Dashi 拉他咪·达石. Special thanks to Latami Dashi for supporting and encouraging my work with his mother Latami Dashilame 拉他咪·达石拉么 over the years. Many thanks to the main consultant, Latami Dashilame, and to all family members.

Many thanks to Céline Buret, Benjamin Galliot, Séverine Guillaume and Mathieu Mangeot-Nagata for their much-appreciated computational expertise, and to Guillaume Jacques for suggestions all along the way. Many thanks to connoisseurs of the Na culture and language for useful exchanges: Lamu Gatusa 拉木嘎吐萨 (Chinese pen-name: Shi Gaofeng 石高峰), Liberty Lidz, Christine Mathieu, and Pascale-Marie Milan. Special thanks to Roselle Dobbs for extensive discussions and vigorous proofreading over the years. Many thanks to Latami Dashi and A Hui 阿慧 (to my knowledge the first speaker of Mosuo to read a M.A. degree in language and linguistics) for suggesting corrections. Remaining errors and shortcomings are my own responsibility.

I am grateful for the opportunity allowed me by my home institution, Centre National de la Recherche Scientifique, of staying in China in 2011-2012 for extensive fieldwork, through a temporary affiliation with the CNRS's Humanities & social sciences research centre in China: CEFC (*Centre d'Etudes Français sur la Chine contemporaine*). From November 2012 to June 2016, I was based at the international research institute MICA, in Hanoi, in an exceptionally stimulating environment allowing for close collaboration with colleagues from Asia and elsewhere. Special thanks to the heads of the institute, Phạm Thị Ngọc Yến (succeeded in 2015 by Nguyễn Việt Sơn) and Eric Castelli, for their support and encouragement.

I am grateful to the Dongba Culture Research Institute (丽江市东巴文化研究院) in Lijiang for inviting me to become an Adjunct member (外聘研究员), and for facilitating administrative and practical matters; special thanks to Li Dejing 李德静 and to Mu Jihong 木霁弘. At Yunnan University, many thanks are due to Duan Bingchang 段炳昌, Wang Weidong 王卫东, Zhao Yanzhen 赵燕珍 and Yang Liquan 杨立权 for their careful and sensitive management of fieldwork-related administrative matters.

So many people have supported this project that I must apologize for those names that should be here but were inadvertently left off the list.

This work was supported financially by the ANR project HimalCo (ANR-12-CORP-0006), and constitutes a contribution to the LabEx “Empirical Foundations of Linguistics” project (ANR-10-LABX-0083).

## References

- Ā, Hui 阿慧. 2016. 永宁摩梭话“le+V+se”结构的声调实验分析——阿拉瓦和舍垮的对比 [An experimental analysis of the tones of “le+V+se” in the Yongning Mosuo language: a comparison between Alawa and Shekua dialects]. 昆明: 云南大学文学院 Yunnan University 中国少数民族语言文学专业硕士研究生学位论文 M.A. thesis.
- Adams, Oliver, Trevor Cohn, Graham Neubig, Hilaria Cruz, Steven Bird & Alexis Michaud. 2018. Evaluating phonemic transcription of low-resource tonal languages for language documentation. *Proceedings of LREC 2018 (Language Resources and Evaluation Conference)*. Miyazaki. <https://halshs.archives-ouvertes.fr/halshs-01709648>.
- Adams, Oliver, Trevor Cohn, Graham Neubig & Alexis Michaud. 2017. Phonemic transcription of low-resource tonal languages. *Proceedings of ALTA 2017 (Australasian Language Technology Association Workshop)*, 53–60. Brisbane. <https://halshs.archives-ouvertes.fr/halshs-01656683>.
- Boas, Franz. 1902. *Tsimshian texts*. (Bulletin of the Smithsonian Institution. Bureau of American Ethnology 27). Washington: Government Printing Office.
- Boas, Franz & James R. Swanton. 1911. Siouan (Dakota). *Handbook of American Indian Languages I*, 875–965. Washington: Government Printing Office, Bureau of American Ethnology, Bulletin 40.
- Bonnet, Rémy, Céline Buret, Alexandre François, Benjamin Galliot, Séverine Guillaume, Guillaume Jacques, Aimée Lahaussais, Boyd Michailovsky & Alexis Michaud. 2017. Vers des ressources électroniques interconnectées: Lexica, les dictionnaires de la collection Pangloss. *9èmes Journées Internationales de la Linguistique de corpus*, 48–51. Grenoble.
- Comrie, Bernard, Martin Haspelmath & Balthasar Bickel. Leipzig Glossing Rules. <http://www.eva.mpg.de/lingua/resources/glossing-rules.php> (12 July, 2011).
- Daudey, Henriëtte. 2014. A grammar of Wadu Pumi. Victoria: LaTrobe University Ph.D.
- Dixon, Robert M. 2007. Field linguistics: a minor manual. *Sprachtypologie und Universalienforschung* 60(1), 12–31.
- Dobbs, Roselle & Mingqing La. 2016. The two-level tonal system of Lataddi Narua. *Linguistics of the Tibeto-Burman Area* 39(1), 67–104.

- Fellbaum, Christiane. 2005. WordNet and wordnets. In Keith Brown (ed.), *Encyclopedia of Language and Linguistics*, 665–670. 2nd edition. Oxford: Elsevier.
- Foley, William A. 1999. Review of Gerrit van Enk & Lourens de Vries, *The Korowai of Irian Jaya*. (Oxford studies in anthropological linguistics, 9). New York: Oxford University Press, 1997. Pp. xiv, 321. *Language in Society* 28(3). 470–472.
- François, Alexandre. 2008. Semantic maps and the typology of colexification. In Martine Vanhove (ed.), *From polysemy to semantic change: Towards a typology of lexical semantic associations*, 163–215. (Studies in Language Companion Series 106). Amsterdam: John Benjamins.
- Francopoulo, Gil (ed.). 2013. *LMF: Lexical Markup Framework*. Wiley Online Library.
- Fù, Màoqi 傅懋勤. 1980. Yǒngníng Nàxīzú de mǔxì jiātíng hé qīnshǔ chéngwèi 永宁纳西族的母系家庭和亲属称谓 (The matriarchal family structure and kinship terms of Yongning Naxi people). *Ethno-National Studies* 民族研究 3.
- Gast, Volker, Ekkehard König & Claire Moyse-Faurie. 2014. Comparative lexicology and the typology of event descriptions: A programmatic study. In Doris Gerland, Christian Horn, Anja Latrouite & Albert Ortmann (eds.), *Meaning and grammar of nouns and verbs*, 145–183. (Studies in Language and Cognition 1). Düsseldorf University Press.
- Juvonen, Päivi & Maria Koptjevskaja-Tamm. 2016. *The lexical typology of semantic shifts*. (Cognitive Linguistics Research 58). Berlin & Boston: Walter de Gruyter.
- Koptjevskaja-Tamm, Maria, Ekaterina Rakhilina & Martine Vanhove. 2015. The semantics of lexical typology. In Nick Riemer (ed.), *The Routledge Handbook of Semantics*, 434–454. London: Routledge.
- Lewis, M. Paul, Gary F. Simons & Charles D. Fennig (eds.). 2016. *Languages of China: an Ethnologue country report*. (19th edition). Dallas: SIL International. <http://www.ethnologue.com/>.
- Lǐ, Zǐhè 李子鹤. 2015. Nàxī yǔyán yánjiū huígù - Jiānlùn yǔyán zài wénhuà yánjiū zhōng de jīchǔ dìwèi 纳西语言研究回顾——兼论语言在文化研究中的基础地位 (A review of Naxi language studies, with a discussion of the fundamental role of cultural studies for linguistic research). *Chama Gudao Yanjiu Jikan* 茶马古道研究集刊 4. 125–131.
- Lidz, Liberty. 2010. A descriptive grammar of Yongning Na (Mosuo). Austin: University of Texas, Department of linguistics Ph.D. <https://repositories.lib.utexas.edu/bitstream/handle/2152/ETD-UT-2010-12-2643/LIDZ-DISSERTATION.pdf>.
- Maxwell, Mike. 2012. Electronic grammars and reproducible research. In Sebastian Nordhoff (ed.), *Electronic Grammaticography*, 207–235. (LD&C Special Publication 04). Honolulu: University of Hawaii Press.



- Michailovsky, Boyd, Martine Mazaudon, Alexis Michaud, Séverine Guillaume, Alexandre François & Evangelia Adamou. 2014. Documenting and researching endangered languages: the Pangloss Collection. *Language Documentation and Conservation* 8. 119–135.
- Michaud, Alexis. 2013. The tone patterns of numeral-plus-classifier phrases in Yongning Na: a synchronic description and analysis. In Nathan Hill & Tom Owen-Smith (eds.), *Transhimalayan Linguistics. Historical and Descriptive Linguistics of the Himalayan Area*, 275–311. (Trends in Linguistics. Studies and Monographs [TiLSM] 266). Berlin: De Gruyter Mouton.
- Michaud, Alexis. 2017. *Tone in Yongning Na: lexical tones and morphotonology*. (Studies in Diversity Linguistics 13). Berlin: Language Science Press. <http://langsci-press.org/catalog/book/109>.
- Michaud, Alexis, Séverine Guillaume, Guillaume Jacques, Đăng-Khoa Mạc, Michel Jacobson, Thu Hà Phạm & Matthew Deo. 2016. Contribuer au progrès solidaire des recherches et de la documentation : la Collection Pangloss et la Collection AuCo. *Actes de la conférence conjointe JEP-TALN-RECITAL 2016, volume 1 : Journées d'Etude de la Parole*, vol. 1, 155–163. Paris: Association Francophone de la Communication Parlée.
- Michaud, Alexis, He Limin & Zhong Yaoping. 2017. Nàxī 納西 language / Naish languages. In Rint Sybesma, Wolfgang Behr, Zev Handel & C.T. James Huang (eds.), *Encyclopedia of Chinese Language and Linguistics*, vol. 3, 144–157. Leiden: Brill.
- Musgrave, Simon & Nicholas Thieberger. 2014. Rethinking grammatical description: from Heath to hypertext. Lecture. Research Unit for Indigenous Language, University of Melbourne. <https://indiglang.arts.unimelb.edu.au/events/rethinking-grammatical-description-from-heath-to-hypertext/>.
- Nordhoff, Sebastian. 2008. Electronic reference grammars for typology: challenges and solutions. *Language Documentation and Conservation* 2(2). 296–324.
- Nordhoff, Sebastian. 2012. Linked data for linguistic diversity research: Glottolog/langdoc and asjp online. In Christian Chiarcos, Sebastian Nordhoff & Sebastian Hellmann (eds.), *Linked Data in linguistics: representing and connecting language data and language metadata*, 191–200. Berlin: Springer.
- Romary, Laurent. 2013. TEI and LMF crosswalks. *arXiv preprint arXiv:1301.2444*.
- Schapper, Antoinette, Lila San Roque & Rachel Hendery. 2016. Tree, firewood and fire in the languages of Sahul. In Päivi Juvonen & Maria Koptjevskaja-Tamm (eds.), *The lexical typology of semantic shifts*, vol. 58, 355–422.
- Zhībā, Ěrchē 直巴·尔车 & Ruìjuān Xǔ 许瑞娟. 2013. *Mósuōyǔ chángyòng cíjù huìcui* 摩梭语常用词句荟萃 (*An anthology of everyday words and expressions in the Mosuo language*). Kunming 昆明: 云南人民出版社.