FROM FIELDWORK TO RECONSTRUCTION: HISTORICAL ISSUES IN HOTSPOTS OF LINGUISTIC DIVERSITY*

Luca Ciucci

1. The need for linguistic documentation

This special issue aims to explore the advances in historical linguistics that are made possible by the ongoing documentation of underdescribed and endangered languages. Indeed, apart from its contribution to typology, any linguistic description offers data to be analyzed from a diachronic perspective. It hardly needs to be noted that both documentation and historical linguistics significantly contribute to the culture of the communities involved.

The most urgent task of linguistics today is perhaps the description and preservation of the amazing linguistic diversity that is still present on the planet. UNESCO reaffirmed

this priority with the proclamation of 2019 as the *International Year of Indigenous Languages*. The awareness of the need for more efforts to protect minority languages led to the declaration of an entire decade, 2022-2032, as the *International Decade of Indigenous Languages*. Already since 1999, on February 21 of each year, UNESCO has been celebrating the *International Mother Language Day*, with particular attention to the promotion and protection of those linguistic varieties that are currently threatened.

Meanwhile, the COVID-19 pandemic has made the description of minority languages even more urgent, as coronavirus has often hit the hardest those vulnerable communities in which endangered languages are spoken, causing the death of many language keepers. At the same time, the rapid expansion of the epidemic has made it necessary to quickly and effectively transmit coronavirus-related information to these communities. Although communication during the pandemic has highlighted many linguistic barriers and a huge divide between hegemonic and minority languages (see López 2020 for Latin America), there has been an unprecedented worldwide production of health information materials in indigenous languages (see CEJIS 2020, among others, for some examples in Bolivian languages), which was also made possible by linguistic research.

In recent decades, efforts to describe minority languages have increased considerably. Hale et al. (1992) is often cited as a turning point in the general attitude towards language loss. However, Hale et al.'s urgent appeal for more investigation on endangered languages was not the first call for scholars and institutions to pay more attention to the issue of language extinction (see Seifart et al. 2018). Although fieldwork has been increasing our knowledge of linguistic diversity, a large number of languages are still poorly or not at all described, and many of them will be lost forever (for some

numerical estimates, see Austin & Sallabank 2011, Seifart et al. 2018, among others).

Fieldwork is "the backbone of an empirical science of linguistics" (Aikhenvald 2007:3),

but within linguistics itself it has sometimes been considered 'unfashionable' and has

been given low priority, in favor of more theoretical activities, which has discouraged

data collection. Such an attitude has partly changed, but there is nevertheless the need for

linguistic description to find greater scientific recognition (Seifart et al. 2018:333).

2. From language documentation to linguistic classification and reconstruction

For most languages of the world, there can be no historical linguistics owing to the lack

of adequate documentation, and this already emerged at the dawn of modern linguistics.

Although doing historical linguistics without a written tradition was sometimes

considered impossible, first-hand data have always contributed to diachronic analysis

(Campbell 2016). Three years before 1786, when Sir William Jones postulated the

common origin of Sanskrit, Greek and Latin (among others), the Jesuit father Filippo

Salvatore Gilij classified the languages of the Orinoco River basin (Gilij 1783),

understood the genetic relationship between Carib languages and realized that Maipure

(a now extinct Arawak language from the upper Orinoco region) was related to Mojeño

(Arawak), spoken in the Jesuit missions of Bolivia. In so doing, Gilij laid the foundations

for the classification of Arawak languages. Gilij relied on data collected by missionaries,

including his own. In the same period, another precursor of the comparative method,

Lorenzo Hervás y Panduro, collected data provided by exiled Jesuit missionaries to

attempt the first classification of all languages known at his time. The first volume of his

linguistic works, *Catalogo delle lingue conosciute* ('Catalog of known languages') was published in 1784 (Hervás y Panduro 1784). The writings and materials of Hervás y Panduro influenced Alexander von Humboldt (Zimmermann 2006), and still offer useful data on some little-known and currently extinct languages.

The extinction of a language implicates a loss of linguistic and cultural diversity, a loss of identity, the loss of a unique way of seeing the world and a reduction of the data available for enhancing our understanding of human language (Crystal 2000). Since language documentation is based on cooperation with the speakers and contributes to the preservation or revitalization of a language, it has stakeholders outside academia and can have very concrete repercussions for the speakers. Preservation of the traditional language improves well-being, mental health and social cohesion (cf. Hallet et al. 2007, Taff et al. 2018, Zuckermann 2020). The bilingualism between the traditional language and the country's majority language brings substantial cognitive advantages (Kovács & Mehler 2009, Keysar et al. 2012, Krizman et al. 2012, Garraffa et al. 2017). Furthermore, there is a correlation between the loss of biological and linguistic diversity (Maffi 2001, Loh & Harmon 2014).

The death of a language makes it impossible to study its historical development (unless it was documented at some stage), the linguistic history of its family or its classification. For example, in South America, within Amazonia, at least 60% of its languages were lost following the conquest (Aikhenvald 2012:21). As a consequence, the genetic affiliation of many South American languages (including documented ones) is now problematic because of the scarcity of data. Furthermore, the loss of a language also reduces the possibility of studying the typology of language change.

3. The contribution of historical linguistics to culture preservation

Although historical linguistics may seem exceedingly theoretical with respect to the needs of the speakers' communities, a diachronic perspective can also contribute to the preservation and revitalization of a language, and such contribution is hardly mentioned in the scientific literature.[1] Indeed, to revitalize a language, it is often necessary to use texts collected in the past, and historical linguistics can help interpret those data. The use of regular correspondences with other related languages can help fill the gaps in the documentation or create a neologism (Rosés Labrada & Spence 2019). The study of historical documents can reveal lost grammatical structures of vital importance for revitalization (Campbell 2016). Hence the need to teach the basics of historical linguistics to language teachers involved in revitalization programs.

Historical linguistics also contributes to the appreciation of the language and the cultural heritage of a given population, and strengthens its identity and self-esteem. In fact, the combination of historical linguistics with non-linguistic data can answer crucial questions for the native speakers, concerning, for instance, their origin, their history, contacts with other populations or the etymology of a culturally significant word. Historical linguistics helps give back indigenous people a part of their history, which has not been recorded or whose memory has often been canceled by traumatic events.

The speakers of hegemonic languages, on which most research has focused, have the privilege of getting answers to many questions about the history of their language. This

---

[1] This specific topic was addressed at the workshop *Historical-comparative linguistics for language revitalization*, organized by Jorge Emilio Rosés Labrada & Justin Spence at the *2019 LSA Linguistic Institute*, University of California, Davis, June 29-30, 2019 (see also Rosés Labrada & Spence 2019).

knowledge originated within circles of specialists, became part of university courses and gradually spread to the school system. The ongoing efforts to document minority languages allow historical linguistics to create, *mutatis mutandis*, a comparable wealth of knowledge for indigenous languages, so that in the medium to long term anyone interested in them can have access to this kind of information.

In the past, linguistic discrimination downplayed the complexity of minority languages, declaring them 'primitive' and thus worthless. In some cases, such a prejudice persists even today. By contrast, a detailed diachronic analysis of a specific linguistic feature (as in the studies that constitute the present volume) returns full (often challenging) complexity to a marginalized language, thus eliminating prejudices that lack any scientific basis. Although a specialized study of historical linguistics is initially aimed at a public of scholars, over time, its results will enter the educational system of the speakers' communities, contributing to raising the quality of education and to safeguarding their intangible heritage, both of which are goals of the upcoming *International Decades of Indigenous Languages* (2022-2032). The training of language teachers involved in the preservation or revitalization process needs more studies (Mihas et al. 2013) and historical linguistics can be a part of it.

4. About this special issue

The present special issue evolved out of the two-day workshop *From fieldwork to reconstruction: Language documentation and historical linguistics*, convened by the editor on 7-8 November 2018 at the Language and Culture Research Centre (James Cook

University), Cairns, Australia. The workshop was funded by the Australian Research Council Discovery Project *The integration of language and society*. The authors have undertaken original fieldwork, contributing significatively to the description of the languages involved. Every paper follows the established methodological pathways of historical and comparative linguistics. The new advances in historical linguistics here presented are made possible by the first-hand information the contributors have collected on these underdocumented and endangered languages. The volume focuses on two areas of the world whose remarkable linguistic diversity needs more investigation and is currently under threat: South America, and Mainland East and South East Asia. The first four papers focus on South America.

The Arawak family, the most widespread in South America, extends from Guatemala and the Caribbean islands to Argentina and Paraguay. It comprises about 40 living languages, plus several dozen of extinct ones (Aikhenvald 2012:32–33). In *Removing the owner: Non-specified possessor marking in Arawak languages*, Alexandra Y. Aikhenvald systematically examines the non-specified possessor suffix and its uses in all Arawak languages for which there is enough information available, with special attention to the North Arawak languages on which the author has undertaken original fieldwork: Baniwa of Içana, Bare, Tariana and Wakerena of Xié. The family-internal comparison shows that one can reconstruct the non-specified possessor suffixes *–ʧi* and *–hi* for Proto-Arawak, where they occurred on body parts and culturally relevant items, but not on kinship terms. Reflexes of *–ʧi*/*–hi* are also employed for lexical nominalizations across the family. Such a polysemy of unspecified possessor marker and deverbal nominalizer is an unusual typological feature, and the available data suggest that it also characterized Proto-Arawak.

Contiguous to the area where Arawak languages are spoken is the Gran Chaco lowland, which spans Argentina, western Paraguay and southeastern Bolivia. The Gran Chaco used to be a cultural and linguistic area (Comrie et al. 2010). Guaycuruan and Mataguayan languages were at the geographical center of the area, while other families, including Enlhet-Enenlhet, Lule-Vilela and Zamucoan, were located in its periphery.

The Mataguayan family consists of four languages, Chorote, Maka, Nivaĉle and Wichi, whose documentation has grown considerably in recent years. In *Determiners and the development of grammatical nominalizations in Nivaĉle*, Manuel Otero, Doris Payne and Alejandra Vidal focus on Nivaĉle, which is spoken by more than 17,000 people in northern Argentina and Paraguay. The authors did fieldwork in several communities close to the Argentina-Paraguay border. Their paper analyzes the origin of Nivaĉle complement clauses introduced by *ka=* and headless relatives. Both structures can be considered 'grammatical nominalizations' according to Shibatani's (2019) definition. Headless relative clauses were not discussed in previous studies on Nivaĉle. The authors describe the complex system of Nivaĉle determiners and show that they grammaticalized to markers of headless relative clauses via amalgam structures. A similar grammaticalization involves the complementizer *ka=*. Although the grammaticalization of a demonstrative to a complementizer is cross-linguistically well-known (Kuteva et al. 2019:134–135), the comparison with the other Mataguayan languages and the analysis of sound correspondences is necessary to show that the complementizer *ka=* stems from the grammaticalization of a homophonous determiner.

The Zamucoan family, also spoken in the Chaco, shows traces of past contact with Mataguayan (Ciucci 2014, 2020). Zamucoan consists of two endangered languages,

Ayoreo and Chamacoco, plus the extinct Old Zamuco. Two authors in the volume documented Zamucoan: Bertinetto (Ayoreo) and Ciucci (Ayoreo, Chamacoco).

Zamucoan languages are fusional, while most South American languages are agglutinating. In *Zamucoan person marking as a perturbed system*, Pier Marco Bertinetto starts from the available reconstruction of the Proto-Zamucoan person system (Ciucci & Bertinetto 2015, 2017) and shows that a change of morphological type from agglutinating to fusional took place in the very distant past, before the Proto-Zamucoan stage that we can reconstruct. The author analyzes the dynamics that determined the change of morphological type. The fusional nature of Zamucoan is evident in nouns and adjectives, which have a prefix expressing gender, number and a peculiar case-like category (called 'form'). Proto-Zamucoan verbs maintained some traces of agglutination, since the person was marked by a prefix and its number by a suffix. By contrast, in possessable nouns both the possessor's person and number had to be expressed by a prefix, since suffixation was already recruited to express the number of the possessed (along with other features). In addition, unusual instances of syncretism arose in possessive inflection. These factors triggered a structural conflict that challenged the one-to-one association between morpheme and meaning.

When a language is extinct, finding new historical documents helps to better understand the language and the historical linguistics of its family. This is the topic of Luca Ciucci's paper *How historical data complement fieldwork: New diachronic perspectives on Zamucoan verb inflection*. The only extant dictionary of Old Zamuco, spoken in the 18th century, was recently rediscovered by the author, who combined fieldwork with archival research. The Old Zamuco dictionary was written by the Jesuit missionary Ignace Chomé (1696-1768) and is the main source of the language. It provides

a wealth of morphological data on Old Zamuco verb inflection, which turns out to be more complex than previously known. The paper compares these data with Ayoreo and Chamacoco, which allows us to explain some changes involving inflectional classes and irregular verbs across Zamucoan. The existing reconstruction of Proto-Zamucoan verbs (Ciucci & Bertinetto 2015) is confirmed, and more details are added to it. Several questions raised in previous works on Old Zamuco and Proto-Zamucoan verb inflection are finally answered thanks to the new data.

The last two papers explore underdocumented languages from Mainland East and South East Asia, where the classification of languages and their variation can be problematic, owing to the scarcity of studies. For instance, documenting a little-known language often allows us to improve an existing dialectological classification that is based on incomplete data. This is the topic of Bai Junwei's paper *Northern and southern Munya dialects: Towards a historical perspective*. Munya (Qiangic, Tibeto-Burman) is an endangered language spoken by 16,000 people in the Sichuan Province (China). The first studies on Munya date back to the 1980s (Bai 2019:21). Munya is traditionally divided into two dialects, a western and an eastern one, but the definition of 'dialect' is questionable, since they are not mutually intelligible. The author provides evidence that western Munya should be further divided into a southern and a northern dialect (mutually intelligible), and analyzes the changes that led to such a dialectal distinction. The northern dialect is usually more innovative than the southern in phonology, grammar and lexicon. While Bai's fieldwork data permit new diachronic insights on the western Munya dialects, there is still a severe shortage of information on the eastern Munya dialect. A description of the latter is essential for the reconstruction of Proto-Munya.

Southern China is very likely the homeland of the Hmong-Mien family (Ratliff 2010:239–240), which has two main branches, Hmongic and Mienic. Most Hmong-Mien languages are underdocumented. Indeed, not even the exact number of languages is known. It is realistic to assume that Hmong-Mien comprises over a hundred languages (Gerner & Bisang 2010:620, Sposato 2015), scattered across southern China, Laos, Myanmar, Thailand and northern Vietnam.

After the Vietnam war, speakers of Hmong-Mien languages, such as Hmong (Hmongic), fled southeast Asia, and were resettled in several Western countries, such as the United States, France and Australia. Nathan M. White's *Prehistory of verbal markers in Hmong: What can we say?* focuses on the development of four verbal markers in Hmongic. The paper analyzes the grammaticalization pathways that took place at different stages of Hmongic and contributes to its internal classification, which is still debated. There are no written sources for Hmongic languages before 1870, so that all data come from recent documentation. To produce a reference grammar of Hmong (White 2021), the author did fieldwork in several diaspora communities and created the *Hmong Medical Corpus* (White 2019) to help Hmong refugee communities access medical information in their language. This is an example of the linkage between fieldwork and benefits to the community that I have outlined in this introduction, together with their connection to historical linguistics.

Both language and history are essential elements of people's identity. Historical linguistics permits us to better understand the history of a population in the absence of a written tradition, thus giving a significant cultural contribution to many indigenous peoples at risk of losing their language and identity. Exploring the connection between

fieldwork and historical linguistics also aims at encouraging further investigation of endangered languages, whose documentation and preservation constitute an urgent task.

References

AIKHENVALD, A. Y. 2007. Linguistic fieldwork: Setting the scene. *Linguistic fieldwork*, ed. A. Y. Aikhenvald, 3–11. Special issue of *Sprachtypologie und Universalienforschung* 60.1.

AIKHENVALD, A. Y. 2012. *The languages of the Amazon*. Oxford: Oxford University Press.

AUSTIN, P. K. & SALLABANK, J. (eds.) 2011. *The Cambridge handbook of endangered languages*. Cambridge: Cambridge University Press.

BAI, J. W. 2019. *A grammar of Munya*. James Cook University, Cairns: PhD diss.

CAMPBELL, L. 2016. Language documentation and historical linguistics. *Language contact and change in the Americas. Studies in honor of Marianne Mithun*, eds. A. L. Berez-Kroeker, D. M. Hintz & C. Jany, 249–271. Amsterdam / Philadelphia: John Benjamins.

CEJIS (Centro de Estudios Jurídicos e Investigación Social) 2020. *Cuida tu comunidad*. Available at: https://www.cejis.org/cuida-tu-comunidad/, accessed on 2021-03-06.

CIUCCI, L. 2014. Tracce di contatto tra la famiglia zamuco (ayoreo, chamacoco) e altre lingue del Chaco: prime prospezioni. *Quaderni del Laboratorio di Linguistica* 13 n.s.

CIUCCI, L. 2020. Matter borrowing, pattern borrowing and typological rarities in the Gran Chaco of South America. *Borrowing matter and pattern in morphology*, ed. F.

Gardani, 283–310. Special issue of *Morphology* 30.4.

CIUCCI, L. & BERTINETTO, P. M. 2015. A diachronic view of the Zamucoan verb inflection. *Folia Linguistica Historica* 36.1, 19–87.

CIUCCI, L. & BERTINETTO, P. M. 2017. Possessive inflection in Proto-Zamucoan: A reconstruction. *Diachronica* 34.3, 283–330.

COMRIE, B., GOLLUSCIO, L. A., GONZÁLEZ, H. A., & VIDAL, A. 2010. El Chaco como área lingüística. *Estudios de lenguas amerindias 2: contribuciones al estudio de las lenguas originarias de América*, eds. Z. Estrada Fernández & R. Arzápalo Marín, 85–131. Hermosillo, Sonora (Mexico): Editorial Unison.

CRYSTAL, D. 2000. *Language death*. Cambridge: Cambridge University Press.

GERNER, M. & BISANG, W. 2010. Classifier declinations in an isolating language: On a rarity in Weining Ahmao. *Language and Linguistics* 11.3, 579–623.

GILIJ, F. S. 1783. *Saggio di storia americana*, Vol. 3, *Della religione, e delle lingue degli Orinochesi, e di altri Americani*. Roma: Luigi Perego.

GARRAFFA, M., OBREGON, M. & SORACE, A. 2017. Linguistic and cognitive effects of bilingualism with regional minority languages: A study of Sardinian–Italian adult speakers. *Frontiers in Psychology* 8. Available at: https://www.frontiersin.org/article/10.3389/fpsyg.2017.01907, accessed on 2021-03-01.

HALE, K., KRAUSS, M., WATAHOMIGIE, L. J., YAMAMOTO, A. Y., CRAIG, C., JEANNE, L. M. & ENGLAND, N. C. 1992. Endangered languages. *Language* 68.1, 1–42.

HALLETT, D., CHANDLER, M. J. & LALONDE, C. E. 2007. Aboriginal language knowledge and youth suicide. *Cognitive Development* 22.3, 392–399.

HERVÁS Y PANDURO, L. 1784. *Catalogo delle lingue conosciute e notizia della loro*

affinità, e diversità. Cesena: Gregorio Biasini.

KEYSAR, B., HAYAKAWA, S. L. & AN, S. G. 2012. The foreign-language effect thinking in a foreign tongue reduces decision biases. *Psychological Science* 23.6, 661–8.

KOVÁCS, Á. M. & MEHLER, J. 2009. Flexible learning of multiple speech structures in bilingual infants. *Science* 325.5940, 611–612.

KRIZMAN, J., MARIAN, V., SHOOK, A., SKOE, E. & KRAUS, N. 2012. Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proceedings of the National Academy of Sciences* 109.20, 7877–7881.

KUTEVA, T., HEINE, B., HONG, B., LONG, H. NARROG, H. & RHEE, S. 2019. *World lexicon of grammaticalization*. Second edn. Cambridge: Cambridge University Press.

LOH, J. & HARMON, D. 2014. *Biocultural Diversity: Threatened species, endangered languages*. Zeist, The Netherlands: WWF Netherlands.

LÓPEZ, L. E. 2020. *Otra peste más: la subalternidad de las lenguas originarias en tiempos de pandemia*. Ciudad de Guatemala: Maya' Wuj.

MAFFI, L. (ed.) 2001. *On biocultural diversity: Linking language, knowledge, and the environment*. Washington / London: Smithsonian Institution Press.

MIHAS, E., PERLEY, B., REI-DOVAL, G. & WHEATLEY, K. (eds.) 2013. *Responses to language endangerment. In honor of Mickey Noonan. New directions in language documentation and language revitalization*. Amsterdam / Philadelphia: John Benjamins.

ROSÉS LABRADA, J. E. & SPENCE, J. 2019. Historical linguistics for language revitalization: Connecting communities past and present. Paper presented at the *6th International Conference on Language Documentation and Conversation*. March 3, 2019. Available at: http://hdl.handle.net/10125/44776, accessed on 2021-02-23.

RATLIFF, M. 2010. *Hmong-Mien language history*. Canberra: Pacific Linguistics.

SEIFART, F., EVANS, N., HAMMARSTRÖM, H. & LEVINSON, S. C. 2018. Language documentation twenty-five years on. *Language* 94.4, e324–e345.

SHIBATANI, M. 2019. What is nominalization? Towards the theoretical foundations of nominalization. *Nominalization in languages of the Americas*, eds. R. Zariquiey, M. Shibatani & D. W. Fleck, 15–167. Amsterdam / Philadelphia: John Benjamins.

SPOSATO, A. 2015. *A grammar of Xong*. State University of New York, Buffalo: PhD diss.

TAFF, A., CHEE, M., HALL, J., DULITSEEN HALL, M. Y., MARTIN, K. N. & JOHNSTON, A. 2018. Indigenous language use impacts wellness. *The Oxford handbook of endangered languages*, eds. K. L. Rehg & L. Campbell, 862–883. Oxford: Oxford University Press.

White, N. M. 2019. *Hmong Medical Corpus*. Available at: http://www.hmong-medical-corpus.org/, accessed on 2021-03-04.

White, N. M. 2021. *The Hmong language of North Queensland*. James Cook University, Cairns: PhD diss.

ZIMMERMANN, K. 2006. Lorenzo Hervás y Panduro: su contribución a la lingüística moderna y su influencia sobre Wilhelm von Humboldt. *Castilla-La Mancha: caminos de universalidad*, eds. R. Sevilla & A. Serrano, 266–283. Bad Honnef: Horlemann.

ZUCKERMANN, G. 2020. *Revivalistics: From the genesis of Israeli to language reclamation in Australia and beyond*. Oxford: Oxford University Press.

Luca Ciucci

Language and Culture Research Centre

James Cook University

PO Box 6811

4870 Cairns, QLD

Australia

luca.ciucci@jcu.edu.au