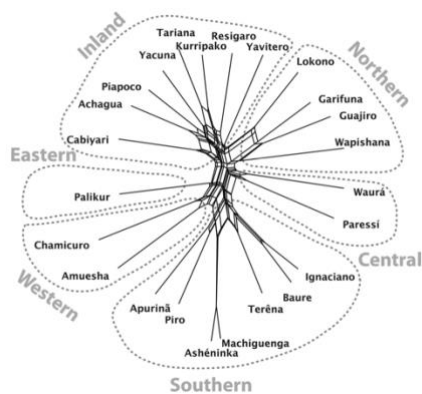
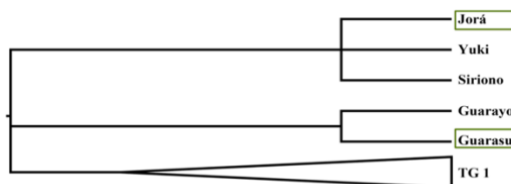


How language change in the context of minority issues can be described and what it holds for the field of linguistics

A collected volume for the Habilitation at the Europa-Universität Flensburg



How language change in the context of minority issues can be described and
what it holds for the field of linguistics

A collected volume for the Habilitation at the Europa-Universität Flensburg

Kumulative Habilitationsschrift
zur Erlangung der *Venia Legendi*
für das Fach Allgemeine Linguistik und Minderheitenforschung
vorgelegt von

Dr. Swintha Danielsen



eingereicht am 24.10.2024

angenommen von der Habitationskommission am 22.07.2025

erfolgreich bestanden mit einer mündlichen Prüfung am 24.09.2025

Anmerkungen

Die Habilitationsordnung der Europa-Universität Flensburg (2021) sieht vor, dass als Habilitationsschrift „mehrere allein oder gemeinschaftlich veröffentlichte [...] wissenschaftliche Arbeiten, die zusammengekommen einen wesentlichen Fortschritt der wissenschaftlichen Erkenntnis für das Fach oder Fachgebiet darstellen und insgesamt einer Habilitationsschrift gleichwertig sind (kumulative Habilitationsschrift)“ (§ 7 Schriftliche Habilitationsleistung, Absatz (1), Punkt 2.) eingereicht werden können.

Die vorliegende Arbeit fasst in der Rahmenschrift fünf publizierte Werke von Dr. Swintha Danielsen zusammen und ordnet sie in den wissenschaftlichen Kontext der Linguistik sowie ihrer jeweiligen Teilgebiete ein. Auf die publizierten Arbeiten wird verwiesen; sie sind der Rahmenschrift jedoch nicht erneut beigelegt.

Darüber hinaus sieht § 7 Absatz (3) der Habilitationsordnung (2021) vor, dass „[d]ie schriftliche Habilitationsleistung (...) im Regelfall in deutscher oder englischer Sprache verfasst sein [soll]. Sie kann aus besonderem sachlichen Grund auch in einer anderen Sprache verfasst sein, insbesondere wenn das Thema sich auf diese Sprache bezieht oder wenn die wissenschaftliche Diskussion ganz überwiegend in dieser Sprache stattfindet; dabei muss eine adäquate Beurteilung durch die Mitglieder der Habilitationskommission sichergestellt und der schriftlichen Habilitationsleistung eine Zusammenfassung in deutscher oder englischer Sprache beigelegt sein.“

Vier der fünf ausgewählten Publikationen sind auf Englisch erschienen. Entsprechend wurde auch die Rahmenschrift in englischer Sprache verfasst. Für den dritten ausgewählten Beitrag aus der Liste, der auf Spanisch publiziert wurde, ist aus diesem Grund eine englische Zusammenfassung beigelegt. Der Artikel wurde zunächst auf einer anerkannten argentinischen Linguistikkonferenz vorgestellt und anschließend in spanischer Sprache veröffentlicht. Spanisch ist die maßgebliche Wissenschaftssprache im Forschungsgebiet (Bolivien) von Dr. Swintha Danielsen und zugleich die weltweit am weitesten verbreitete panamerikanische Sprache.

Literatur

Europa-Universität Flensburg. 2021. *Habilitationsordnung (Satzung) der Europa-Universität Flensburg* (Habilitationsordnung 2021). Online: <https://www.uni-flensburg.de/fileadmin/content/portal/die-universitaet/dokumente/satzungen/amtliche-satzungen/2021/20210823-satzung-habilordnung-2021.pdf> (05.07.2026)

Table of contents

Introduction	3
1 Exchange of words and thoughts	5
1.1 Historical linguistics	6
1.2 Comparative linguistics	7
1.3 Linguistic typology	11
1.4 Language description, documentation and language endangerment	16
1.5 Ethics in linguistics	18
2 Collect and connect	20
2.1 Historical linguistics with limited data resources	20
Evaluating historical data (wordlists) in the case of Bolivian extinct languages (Danielsen 2013)	
2.2 Comparative Arawakan linguistics	21
The Spread of the Arawakan Languages: A View from Structural Phylogenetics (Danielsen, Dunn & Muysken 2011)	
2.3 The typology of polysynthetic languages	22
Afijos compuestos y su lexicalización en el baure (arawak) Compound affixes and their lexicalization in Baure (Arawak), summary	40
(Danielsen 2012)	
2.4 The description and documentation of endangered Bolivian Lowland languages ..	24
Guarayu. A revised dictionary by Alfred Hoeller (Danielsen, Sell & Terhart 2019)	
2.5 The discussion of ethics in linguistics – difficult cases	27
Discovering two almost extinct languages in Bolivia: Jorá and Guarasu (Danielsen & Gasparini 2017)	
3 Conclusions	29
References	30
Danksagung	42

List of maps

Map 1: South American language families and linguistic areas	10
--	----

List of figures

Figure 1: Language families of South America	8
Figure 2: Greenberg's Universals No. 1 and 3 in the original edition of 1963, 61	11
Figure 3: An excerpt from the Sahul questionnaire applied to Arawakan languages by Danielsen	13
Figure 4: Languages of the world in a Splitsgraph	14
Figure 5: Tupian and Tupi-Guarani languages	15
Figure 6: Lingua francas in Jesuit missions, so-called reductions	17
Figure 7: Krauss' (2007, 1) categorization of degrees of language endangerment	18

Bibliography, selected

Danielsen, Swintha. 2013. "Evaluating historical data (wordlists) in the case of Bolivian extinct languages". In *Historical reconstruction of American languages*, edited by Astrid Alexander-Bakkerus & Otto Zwaartjes. STUF – Language Typology and Universals/ Sprachtypologie und Universalienforschung, 66(3), 272-298. Bremen: Akademie Verlag.

Danielsen, Swintha, Michael Dunn, & Pieter Muysken. 2011. "The Spread of the Arawakan Languages: A View from Structural Phylogenetics". In *Ethnicity in Ancient Amazonia: Reconstructing past identities from archaeology, linguistics, and ethnohistory*, edited by Alf Hornborg & Jonathan Hill, 173-195. Boulder, CO: University Press of Colorado.

Danielsen, Swintha. 2012. "Afijos compuestos y su lexicalización en el baure (arawak)". In *Lenguas Indígenas de América del Sur I: Fonología y Léxico*, edited by Marisa Censabella, Cristina Messineo, Hebe A. González, & Beatriz Gualdieri, 151-171. Mendoza: Editorial FFyL-UNCuyo and SAL. Retrieved at: [https://www.academia.edu/5479523/Lenguas indígenas de América del Sur I Fonología y Léxico](https://www.academia.edu/5479523/Lenguas_indígenas_de_América_del_Sur_I_Fonología_y_Léxico) (29/09/2024).

Danielsen, Swintha, Lena Sell, & Terhart, Lena. 2019. "Guarayu. A revised dictionary by Alfred Hoeller". *Dictionaria* 6, 1-3590. Retrieved at: <https://dictionaria.cld.org/contributions/guarayu> (29/09/2024).

Danielsen, Swintha & Noé Gasparini. 2017. "Discovering two almost extinct languages in Bolivia: Jorá and Guarasu". *18e Rencontres Jeunes Chercheurs en Sciences du Langage*, ED 268 Paris 3, Jun 2015, Paris, France. Retrieved at: <https://hal-univ-paris3.archives-ouvertes.fr/hal-01495156v1> (29/09/2024).

How language change in the context of minority issues can be described and what it holds for the field of linguistics

Introduction to the collected volume for the Habilitation at the Europa-Universität
Flensburg

Introduction

This work illustrates what minority languages contribute to general linguistics and how they do so. The collection presents five different articles by the author¹ – Danielsen 2013*, Danielsen, Dunn & Muysken 2011*, Danielsen 2012*, Danielsen, Sell & Terhart 2019*, and Danielsen & Gasparini 2017* – with different perspectives and approaches to the major South American languages of the Arawakan and Tupian language families. The articles and their impact are described in the respective sections of this text (§§2.1–2.5). They follow two principal questions:

- Which points have been resolved and which are under debate in linguistics and linguistic communities when documenting languages, and who owns language as part of the culture?
- Which linguistic factors does the analysis of minority languages contribute to linguistics?

In the realm of language documentation, languages become the matter of interest, but they are also turned into objects. The ownership of these objects of study is connected to the question about the ownership of culture and this also brings ethics of research into play. The documentation of minority languages adds important variables to the discussion of objectives and the way of documentation, while language and knowledge become elements that have special value in very small language groups. This is the documentation level of the present work. On the analytic level, there is the second research question that refers to linguistics as a discipline. It focuses on the internally relevant factors for linguistics that are the outcomes of the analysis of the documented data. Both the documentation level and the analytic level share a relation to historical information, and the approach from these two angles helps understand the complexity of language. History is the temporal dimension relevant for language and culture, for the documentation of a language and its existence in time and the analysis of language change and the evolution of structures.

The two main research questions are again subdivided into specific questions, summarized in this introduction. The general thoughts in §1 are leading to the specific discussion of the studies of this

¹ The specific studies have been published elsewhere before, and their references are marked by “*” after the year throughout this text.

volume in §2. The issue of minority languages is exemplified with the work on critically endangered indigenous languages in Lowland Bolivia (South America). The specific research questions are related to the fields of historical and comparative linguistics, to linguistic typology, descriptive linguistics, and ethics in linguistics. It is shown that many small languages add to the historical reconstruction of language families (§2.1), and the available data do not only count for lexical comparison, but also grammatical information can be inferred from the documented data, which is important for the understanding of internal classifications. Apart from these internal classifications, the comparison of Arawakan languages, e.g., also provides information about the migration of peoples and languages in South America (§2.2). In linguistic typology, languages are classified into the different kinds of synthetic behaviour, where polysynthetic languages stand out as a particularly complex phenomenon. The Arawak language family is one of the largest and most wide-spread language families in South America, and Arawakan languages are mainly found in Amazonia and the lowlands of the continent, however, some have also migrated as far as Central America and to the Caribbean. Arawakan languages belong to the polysynthetic type of languages, and the study of the morphosyntax of these languages shows that there are templates within which morphemes are attached in different ways: Some morpheme combinations are rather fixed or co-occur more often, whereas others remain more flexible, and yet others maintain a clitic-like behaviour (§2.3). The historical dimension is also a factor here, since some morpheme combinations are older and develop into fixed units, while others just arise. Tupian is the other large language family in South America and has therefore also had great influence on other indigenous languages as well as the colonisers' languages. Arawakan and Tupian languages were the first languages the colonisers had contact with. This is also why the historical dimension has relatively early written notes and the study of these languages has a centuries-long tradition. The description of the Tupian language Guarayu (§2.4) shows that this language is a relatively old member of the language family, maintaining, e.g., final consonants and forms of lexemes that are most closely related to the reconstructed proto-forms of Tupian. This characteristic makes Guarayu a language of very much interest, comparable to long-extinct Tupinambá and other languages that were still spoken in South America in the 16th century. Finally, the case studies of the two very small and almost extinct Tupian languages Jorá and Guarasu (§2.5) address the questions: Can languages always be documented or described? How can the last speakers be approached, and how should their attitudes and wishes for (in)visibility be taken into account?

1 Exchange of words and thoughts

Language is (ex)change: Language is i.a. the exchange of information and the attitude, feeling about this information, but also the expression of social relations between human individuals. Just as humans grow, develop, adapt, evolve, and change, the languages they use change. This means, taking the somewhat overused metaphor of *LANGUAGES ARE CREATURES*,² a language is only “alive” (or “safe” in Krauss’ terminology, cf. §1.5), if it can adapt and change according to the environment and the people who use it. The linguist can then observe the patterns of change, the reasons or motivations of change, the outcomes and results of change, the attitudes and feelings about these changes, etc. The positive perception or interpretation of change is opposed to the purist view of language, which aims to conserve the character of a language at a certain point in time and considers change as a threat to an individual language. This is again related to a simplified and misleading view of what “language” is, called “the *Einzelsprachenideologie* (single-language ideology) [that] pretends that languages exist in isolation” (Langer & Havinga 2015, 31).

The different types of language change can be studied and offer a great field of investigation for **historical linguistics**. Furthermore, the knowledge of language change and the possible outcomes of changes is the basis for **comparative linguistics** and helps interpret the different constructions we find on a synchronic level within a language family. Languages in their different forms around the world are then taken by **linguistic typology** and patterns of synchronic systems are classified and categorized, some combinations of patterns and co-occurrences are statistically measured and arrive at the conclusion of certain predictable patterns and changes of the same. Linguistic typology is only possible with the specific and detailed **documentation** and **descriptions** of typologically diverse languages of the world. Thus, any accurate description of formerly unknown language data enriches the linguists’ understanding and makes generalizability not only possible, but also more realistic. In this respect, linguists have long discovered that a Eurocentric bias blurs the insights and gives misleading numbers when certain language types are always over-represented (cf. Dahl 2015). Last but not least, the phenomenon of this bias is also related to current scientific **ethics**, where we acknowledge that the world is diverse, and not only sources, but also authors and scientists should be freed from these Eurocentric biases. However, not all data can be collected anymore, because **language endangerment** often leads to language death, and in the history of humankind, change has also meant the loss of

² Even though, in fact, many a linguist argues quite right that there are limits to the analogy of language and living creatures (cf. discussion online by Mario Saraceni at <https://theconversation.com/language-isnt-alive-why-this-metaphor-can-be-misleading-205793>; June 6, 2023), the metaphor is being used in everyday’s language as well as in linguistics – cf. Dorian (2014: 13) on a critical consequence of using this metaphor and Thomason (2015: 3) on the decision of applying the metaphor to some degree. This book, however, does neither include the field of cognitive linguistics nor ecolinguistics, so the discussion is not extended here.

languages – always in different but generally relatively fast-progressing situations, in which a combination of factors causes an imbalance in the contact between two or more languages. While the colonization period spurred a rapid loss of many languages in South America, ongoing globalization can have the same effect. Ethics and language endangerment are also considered in sociolinguistics, where all the different factors connected to language and linguistic behaviour are investigated and compared. Indigenous languages have long been called “dialectos” in Latin America, which was a way of distinguishing them from what was considered a ‘real’ language. The word *dialecto* has the new meaning of “indigenous language” now in colloquial language, even though in the past 20 years, this issue has been discussed, and language policies have attempted to correct this naming practice that implies that indigenous languages were inferior or less independent.

In the following, we go into more detail about each of these fields in linguistics. The presentation develops from general observations to the more specific.

1.1 Historical linguistics

Even long before the modern field of linguistics was defined, scholars applied a historical lens, and determined that while specific languages changed over time, some of these languages were related to the same ancestor languages. These related languages were classified as a language family, and one of the well-known language families in the world is Indo-European, established by Sir William Jones in the year 1786 (cf. Campbell 1997, 406). Of course, Jones’ findings were based on a century-long (or longer) tradition of comparing languages, focussed mainly on the vocabulary of European languages, including some ancient languages, such as Old Greek, Latin, and Sanskrit. The historical observation of language change and similarities was especially motivated by travellers in the Americas, where completely distinct languages are spoken. We can now state that the existence of the large Arawakan language family in South America was discovered earlier than Jones’ Indo-European, namely by the Jesuit Father Gili in the early 1780s. In that time, the priest compared the languages of the Upper Orinoco and found similarities that could not be due to chance. He named the language family “Maipure” after one of the Amazonian representatives; later the language family was renamed “Arawak” or “Arawakan” (cf. Campbell 1997, 31-33). Father Gili introduced the concept of a *lingua matriz*³ (ibid., 33) and lay the foundations for the recognized “language family”. In his perspective, the parent language was not conceived as something preceding the existing languages, but he had rather a synchronic perspective and singled out one of the existing languages as the *lingua matriz*. The described similarities between languages were due to some historical relation, and motivated travellers, ethnographer and others to collect data on languages in different parts of the world.

³ “mother language”

The grammar of the Arawakan language Baure (ISO 639-3 brg, Glottocode baur1256), spoken in the Bolivian Amazonia, in Danielsen (2007) contributed to the field of Arawakan linguistics by one more specific language description, and its analysis benefits from the comparison to historical data on the same language, collected by Jesuits in the 18th century. Throughout the grammar, there is reference to changes of paradigms, constructions, and vocabulary, as could be observed in comparison to the first collected data on the language. This comparison was immensely relevant for the understanding of regular sound changes and relations of the Baure language to the proposed Proto-Arawakan (cf. Payne 1991). Historical linguistics once mainly worked with word forms, their stems, and particularly their phonemes. Lexicostatistics (cf. Swadesh 1955) is one of the early methods taken to prove the age of the split between two related languages in history. The modern classification of languages is generally not based on lexicostatistics anymore. However, some people still apply this method, others have elaborated it (Wichman and others, see in Hammarström 2014), always trying to make use of word lists: ASJP (Automated Similarity Judgement Program)⁴. The comparison of wordlists still yields good results when using basic vocabulary (Hammarström 2014, 69), in spite of the problem of chance similarities.

In order to understand the internal subdivisions in the Arawakan language family, e.g., some works have shown that not only lexical and phonetic comparison needs to be taken into consideration, but also the comparison of morphology and construction types (cf. Danielsen, Dunn, & Muysken 2011*, Eriksen & Danielsen 2014). This is addressed in §1.3.

Historical data have also taken a different approach, as shown by the recently-evolved field of missionary linguistics, a specific branch of linguistics that works with data in the missionary context. In Danielsen (2013*), e.g., Arawakan linguistics works with this approach, integrating data of already extinct Arawakan languages. Additionally, the studies by Danielsen, Sell, & Terhart (2019*) and Dietrich & Danielsen (2024) are based on the historical data of the Tupian language Guayayu (ISO 639-3 gyr; Glottocode guar1292). In all these examples, the manuscripts of Jesuit or Franciscan missionaries are analysed applying modern linguistic tools.

1.2 *Comparative linguistics*

Comparative linguistics is often done in historical linguistics, i.e., it is the comparison of languages and their “material”: lexicon, morphology, phonology, syntax, construction types. Everything can be compared. The outcome can be – but does not have to be – proof of the genetic relation between languages. Figure 1 gives an overview of the genetic classification of South American languages, as proposed by Muysken & O’Connor (2014).

⁴ <http://asjp.cld.org>

Table 1.1 *Schematic brief overview of some of the current families*

Type	Name	Location
<i>Larger families</i>	Arawakan	Widespread, from Belize and Honduras to Bolivia and Paraguay
	Tupian	Central Amazon, from Paraguay to French Guyana and Peru
	Cariban	Northern Amazon
	Macro-Jêan	Central and southern Amazon
	Chibchan	Central America, from Honduras through northwestern Colombia
	Tucanoan	Western Amazon and Pacific coast
	Pano-Takanan	Western Amazon
	Chon	Southern Cone
<i>Smaller families</i>	Barbacoan, Arawan, Chapacuran, Makúan, Nambikwaran, Witotoan	In the Andean foothills, central and southern Amazon basin
<i>Clusters</i>	Aymaran, Quechuan	Andes
<i>Isolates and small groupings</i>		Many in the Amazonian fringe

Figure 1: Language families of South America (source: Muysken & O'Connor 2014, 5)

After the ‘discovered’ genetic relation between some Arawakan languages by Father Gilli (cf. §1.1), many specialists concentrated on the language families of South America (e.g. Mason 1950, Loukotka 1968, Rodrigues 2013 [orig. 1984/1985]). The volumes that are often cited when it comes to South American languages are nowadays – in chronological order – the four volumes of Derbyshire & Pullum (1986-1998), Campbell 1997, Dixon & Aikhenvald 1999, and Adelaar & Muysken 2004. Greenberg (1987), who compared languages globally, proposed a very broad classification into only three phyla in the Americas. He claimed the majority of all American languages belonging to the Amerind phylum, and a few languages in North America, were classified as belonging to two phyla: Na-Dene and Eskimo-Aleut. This “lumping”⁵ proposal has never found substantial support in the scholarship (cf. also Bolnick, Shook, Campbell & Goddard 2004). In particular, the lumping of large Macro-clusters is problematic, e.g., Carib-Jê-Pano or Andean-Equatorial, including Andean, Jivaroan, Macro-Tucanoan, and what Greenberg calls Equatorial (with Macro-Arawakan and Tupian). Furthermore, there is no proof for an

⁵ In the sense of lumpers and splitters.

Amerind language family, and this large phylum is rather an outcome of unjustified generalization and exaggerated lumping.

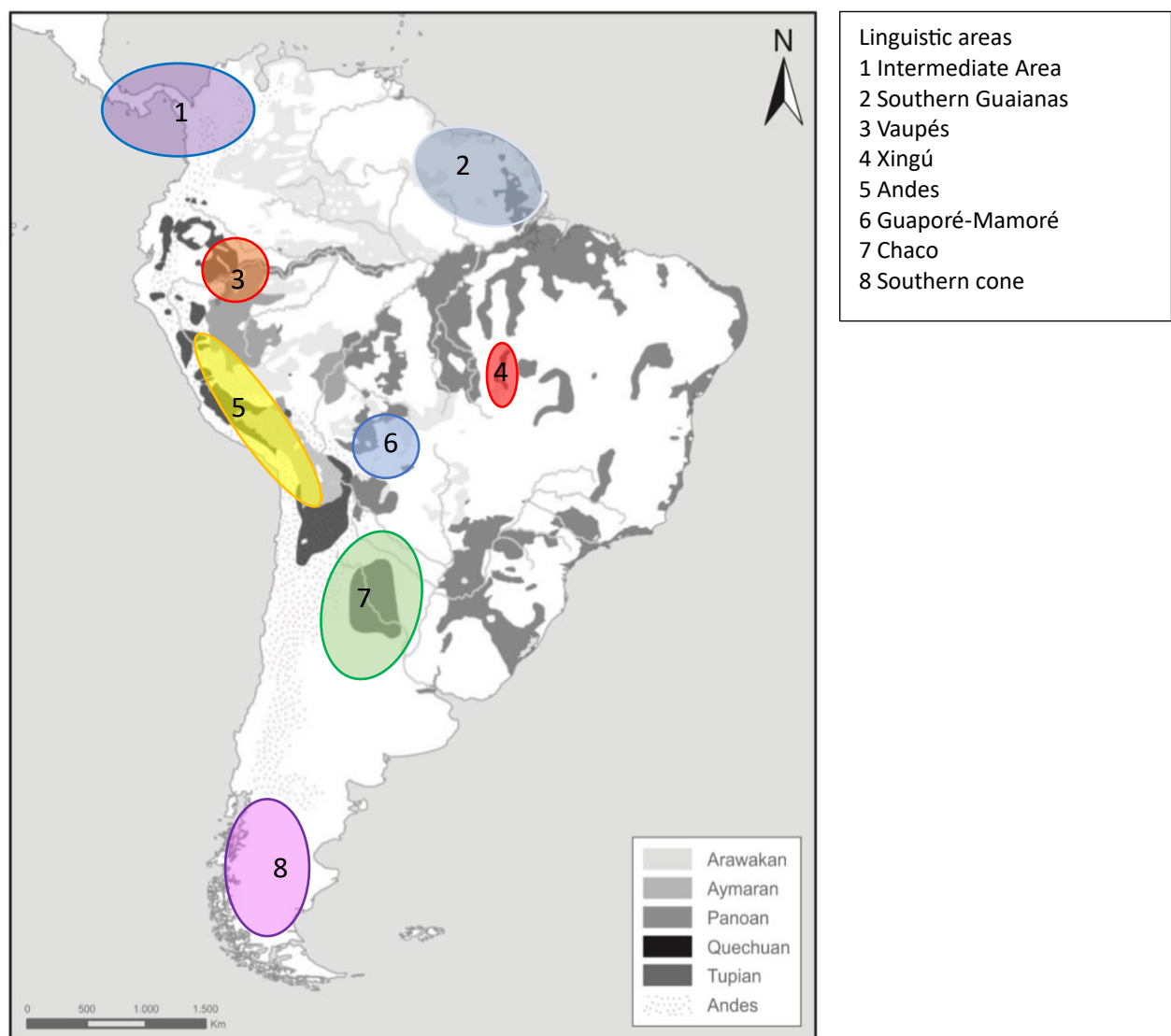
The outcome of comparative studies can also lead to the finding that there is a certain degree of areal diffusion between languages that are geographically close, but not necessarily genetically related. The concept of a *Sprachbund* was first applied to Balkan languages, which conveyed similarities across genetic boundaries. Areal linguistics have also become increasingly important for language comparison and description in South America. Constenla (1991) indicates areal diffusion in Central and northern South American languages (contact with Chibchan), and Aikhenvald's studies of the Vaupés region (2001) in the Amazon basin show intense linguistic diffusion of Arawakan and Tucanoan languages. Generally, more or less established linguistic areas in South America are (cf. Muysken & O'Connor 2014, 13 and others), from North to South (cf. Map 1):

1. Intermediate Area (*Area intermedia*) (Constenla 1991)
2. Southern Guaianas (Epps & Michael 2015)
3. Vaupés (Aikhenvald 2001)
4. Xingú (Seki 1999)
5. Andes (Adelaar & Muysken 2004)
6. Guaporé-Mamoré (Crevels & van der Voort 2008)
7. Chaco (Comrie & Golluscio 2015)
8. Southern cone (Klein 1992)

Linguistic areas are, of course, the consequence of language contact, linguistic and cultural exchange, and subsequent linguistic diffusion and convergence of certain characteristics. Linguistic diffusion is difficult to measure, and intermingles with genetic relation, and there are areas of more linguistic exchange and areas with less so. This and the different motivations of language diffusion, lead to Aikhenvald's (2022) claim to best distinguish between linguistic areas and linguistic regions. This specification is not followed up in this book.

Some linguists have claimed the whole of Amazonia to be a linguistic area and they list "a number of widely encountered features" (cf. Epps & Salanova 2013, 17), such as early attempts in Derbyshire & Pullum (1986-1998) or Dixon & Aikhenvald (1999). However, as Epps & Salanova (2003, 17) claim, "a great deal more work is needed to determine whether these or other features are in fact indicative of some kind of relationship among these languages (contact or otherwise), and whether their distribution is contiguous with the Amazon basin". Within Amazonia, some relatively contiguous areas have solidified (cf. also Muysken 2008, Epps & Michael 2015): the Vaupés basin, the Upper Xingú

region, the Chaco, and the Mamoré-Guaporé area (Crevel & van der Voort 2008). The latter is particularly relevant for the languages in the Bolivian Lowlands. Baure is an Arawakan language and belongs to this multilingual area at the border between Bolivia and Brazil. Paunaka (ISO 639-3 pnk; Glottocode paun1241) is an Arawakan language, but it is strictly speaking not located in this Amazonian area. Guarayu, a Tupian language, is spoken at the edge of the Mamoré-Guaporé linguistic area. As Danielsen (2011) showed, Arawakan features are very dominant in this declared linguistic area, and it seems difficult to tear them apart from areal linguistic features here, an argument that may be true for great parts of Amazonia.



Map 1: South American language families and linguistic areas (Map with language families taken from van Gijn 2014, 105)

It is probably impossible to prove or disprove the existence of a linguistic area; certain features can be found in more concentrated areas, so this needs to be mentioned and described. It is

certain that the major language families – Arawakan and Tupian in Amazonia – had influence in all areas in and around Amazonia. In the end, it is important to include features and characteristics that are due to contact rather than to genetic inheritance when describing a language fully; this belongs to the individual character of these languages, after all. The character, though, of language contact and the outcomes of it, is worthwhile investigating. Language contact is hardly ever balanced, as dominant languages generally spread their features more than less dominant ones. Language contact can also be observed in the linguistic exchange zone between Bolivian Amazonia and the Chiquitania, one within and the other outside the aforementioned linguistic area (Mamoré-Guaporé). Language contact in this exchange zone has been examined in the GIZAC project⁶, and examples can be found in Paunaka (Arawakan) with likewise Chiquitano, Guarayu, and Spanish loanwords, including functional words (cf. Danielsen & Terhart 2015, Terhart 2024).

1.3 *Linguistic typology*

In linguistic typology, languages are compared with the goal of generalizing patterns across languages into universal patterns. This is usually a functionalist approach to linguistics, and it observes the occurrence of certain structures in correlation with other certain structures, so that we may conclude, following Greenberg (1963, 2005), that the dominant word order in the world's languages is SO (subject-object), at least if the argument is nominal (see Figure 2).

Universal 1. In declarative sentences with nominal subject and object, the dominant order is almost always one in which the subject precedes the object.⁵

Universal 3. Languages with dominant VSO order are always prepositional.

Figure 2: Greenberg's Universals No. 1 and 3 in the original edition of 1963, 61

Greenberg's universals are generally still taken as the point of departure. But today, we are in a much better situation with respect to linguistic data in various ways. Through the extensive global collection of linguistic data, there are many more, and more precisely annotated, data, which are available in digital formats. This allows automatized analyses, such as frequency or co-occurrence analyses through scripts running over the corpora (cf. Seifart et al. 2018, Strunk et al. 2020). The information

⁶ <http://hdl.handle.net/2196/98dc4979-5835-4066-8067-a94a244b7459>

taken from the data can be encoded in different ways in matrices, and trends can be visualized in so-called “Neighbor-Nets” with the help of the programme SplitsTree (see also Figure 4). The algorithm applied in phylogenetic programmes was first used by Atkinson & Gray (2006) to show the branching within the Indo-European language family. For the categorization of Meganesian (Sahul), Oceanic, and Polynesian languages, Gray and colleagues designed and used the “Sahul questionnaire” with 201 specific linguistic features, ranging from phonology, via morphology, morphosyntax, word order, to certain lexical relations (cf. Reesink, Singer, & Dunn 2009). Gray’s methods that originally come from biology gave many different views on well-known and less well-known language families (see Figure 4). Gray’s methods have then also been applied to other languages for comparative and classificatory purposes, such as the Arawakan languages (cf. Danielsen et al. 2011*). Danielsen gathered all data in the Sahul matrices for Arawakan languages (cf. Figure 3), and since some distinctive features were missing, she added the Constenla (1991) linguistic features (81 features). In addition, since the Arawakan languages are structurally fairly similar, but can be distinguished for certain specificities, Danielsen created 73 Arawakan distinctive features in the matrix. With this information, we could arrive at a new internal classification proposal of the Arawakan language family and visualize the North-South split that is based on some morphosyntactic characteristics (cf. Danielsen et al. 2011*). Parts of these collected data for Arawakan languages were uploaded in 2014 to a large database of South American languages, called SAILS, together with various other specific projects and data collections. The follow-up project on a global scale is the Glottobank, based at the MPI-EVA at Leipzig (cf. Skirgård et al. 2023), where these and many more linguistic features of the world’s languages are being gathered.

1	Language ID: xxx		aca	ame	qap	apu	cni	bvv	bae	brg
2	Language name: Xxx		Achagua	Amuesha, Yanesha	Apolista	Apurinã	Ashaninka	Baniva	Bare	Baure
3	Questionnaire name: Sahul									
4	Description: The Pimdb extended and revised for the Sahul project									
5	Spreadsheet generated: 2006-08-22 16:50									
6	q_id	q_text	q_note	a_text	a_text	a_text	a_text	a_text	Mansyoni&Mansyoni	a_text
7	1.1.01	Give the phoneme inventory.		consonants: p, t, k, /, b, d, s, S, h, ts, m, n, l, r, w, j vowels: i, e, u, o, a, i:, e:, u:, o:, a:	consonants: p, py, b, by, t, ty, k, ky, s, S, Å, h, ts, ts, yS (retroflex), w, y, m, my, n, ny (ñ), ly, r,}	consonants: p, t, k, (/), s, S, h, ts, m, n, r, j, å vowels: i, ®, e, o, a, long and nasal same	consonants: p, t, k, s, S, x, h, ts, m, n, r, B/w, è (alveolar-palatal affricate) (Romani);	consonants: p, t, k, b, d, g, (h), ts, dz, sr, zr, m, n, l, r, rr, w, y vowels: a, e, i, u	consonants: p, t, k, ph, th, kh, b, d, (g), ts, s, h, m, n, hm, hn, Ó, w, y, hw vowels: i, u, e, a	Vowels: a, e, i, o; Consonants: p, t, k, (/), ts, v, s, S, m, n, r, w, j
8	1.1.02	How many places of articulation are there for stops?		4	3	3	3	3	3	3 or 4 (including glottal stop)
9	1.1.03	How many places of articulation are there for nasals?		2	4	2	3	2	2	2
10	1.1.03a	Are there as many points of articulation for nasals as there are for stops?	Only consider points of articulation where a nasal is phonetically possible	0	0	0	1	0	0	0
11	1.1.03b	Is there contrast between heterorganic and homorganic sequence of nasal and velar stop?	For example, does the language permit a phonetic contrast between -nk- and -ŋk- clusters	0	0	?	0	0	0	0
12	1.1.05	Are there fricative phonemes?		1	1	1	1	0	1	1
13	1.1.06	Are there phonemic prenasalised stops?		0	0	?	0	0	0	0
14	1.1.07	Is there a phonemic distinction between l/r?		1	1	1	0	0	0	0
15	1.1.08	Is there a phonemic velar fricative or glide?		0	1	?	1	1	0	0
16	1.1.09	Is there a voicing contrast between oral (i.e. non-prenasal) stops?		1	1	0	0	1	0	0
17	1.1.09a	Is there a laminal/apical contrast?		0	0	?	0	0	0	0
18	1.1.09b	Are there retroflexed consonants?		0	1	?	0	?	0	1
19	1.1.10	Is there phonemic consonant length?		0	0	?	0	0	0	0

Figure 3: An excerpt from the Sahul questionnaire applied to Arawakan languages by Danielsen (published in SAILS online, Muysken et al. 2016)

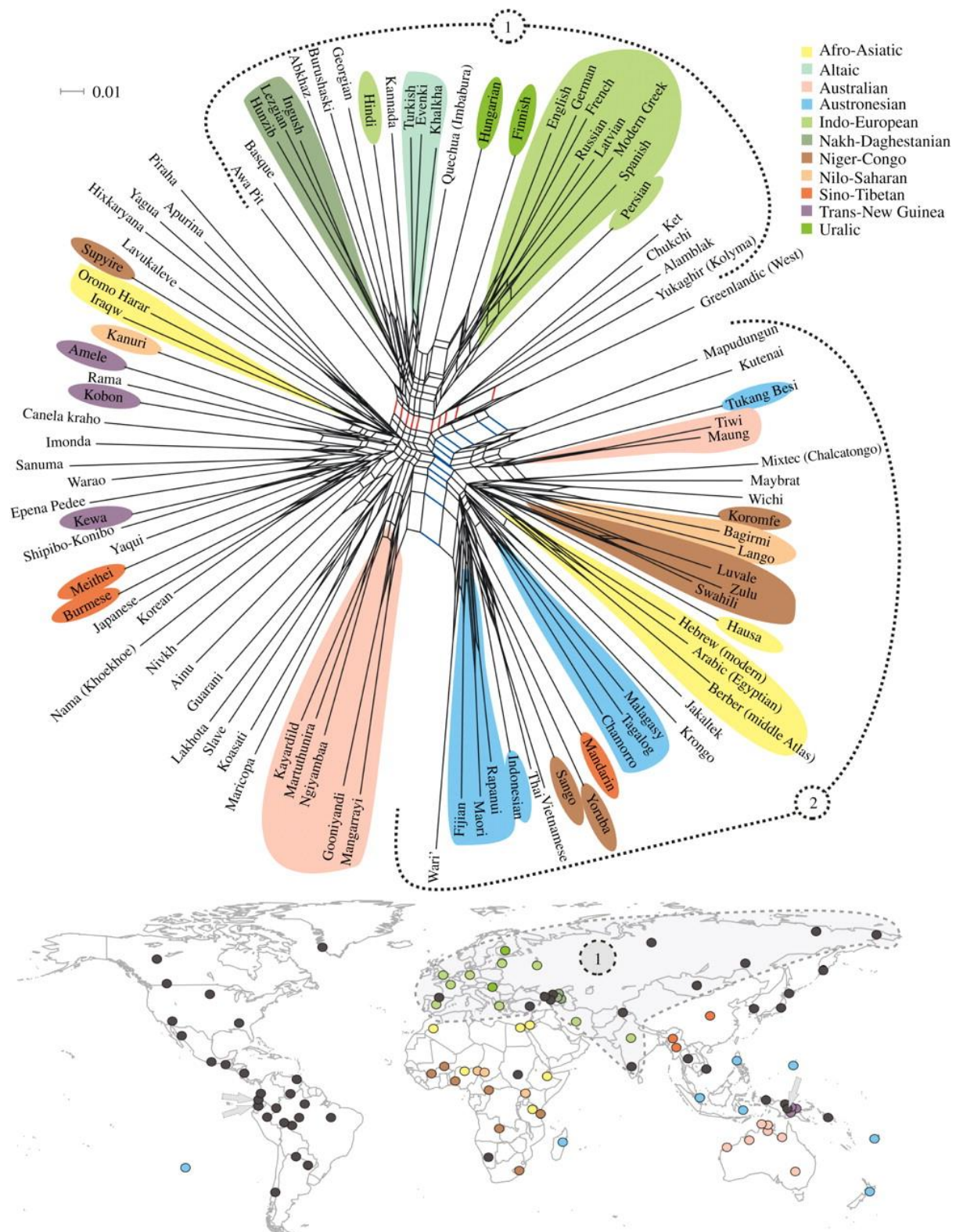


Figure 4: Languages of the world in a Splitsgraph (Greenhill et al. 2010)

The feature matrices have since served in many different projects to: gain a better understanding of internal relationships of related languages, pinpoint the time of the split of certain languages, and provide evidence of language contact in linguistic areas, etc.. In Gasparini et al. (2015), cf. Figure 5, the

data of Tupian and Tupi-Guarani languages have shown that certain subgroups underwent more regional contact and language change than others.

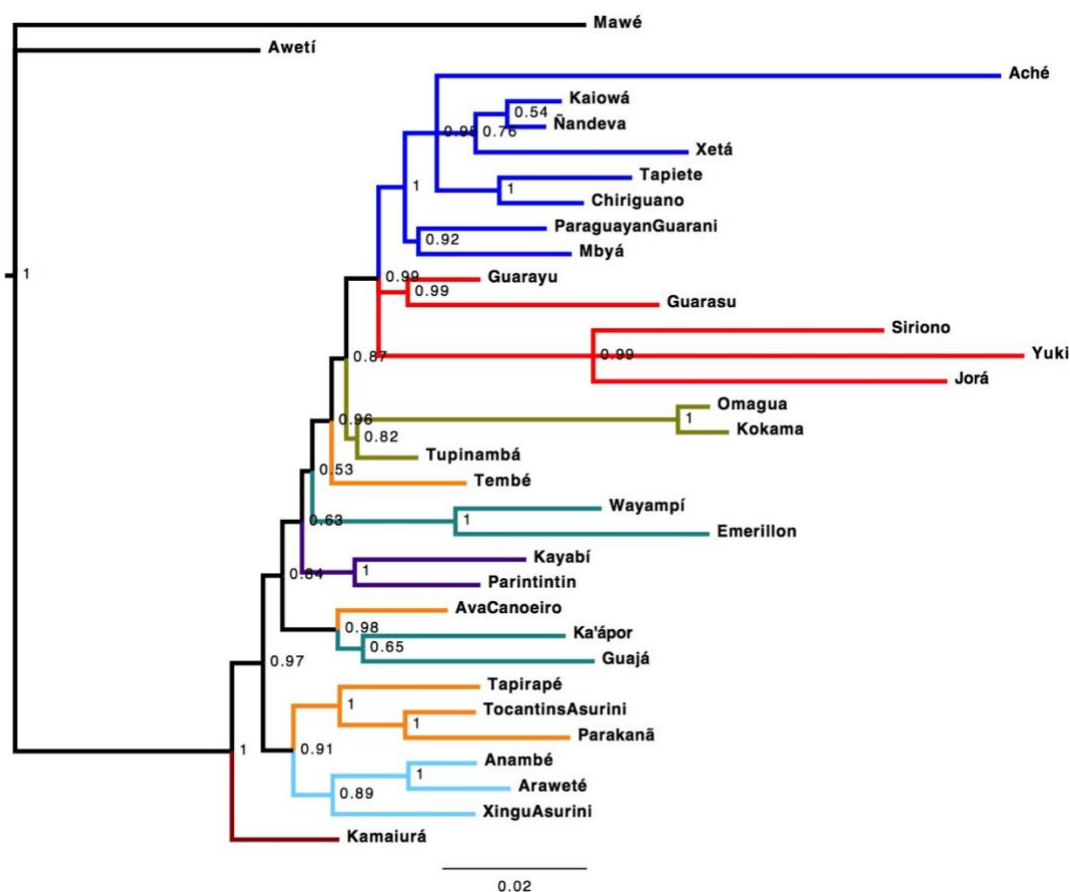


Figure 5: Tupian and Tupi-Guarani languages (Gasparini et al. 2015)⁷

Studies that make use of the collected language data in the archives are growing in number. There is now the possibility to give numbers to tendencies, and undertake frequency counts with specific interests, such as the study on noun incorporation into verb bases, with a sample as large as 50 languages world-wide (cf. Olthof et al. 2020).

Language typologists in the 21st century work differently than before: they no longer need to collect data themselves, they do not work as isolated specialists; and their work is not only relevant within linguistics. These shared databases are easily available to typological analysts, and the certain analytical trends do not stem only from within linguistics – they are also motivated and inspired by

⁷ Tupian includes Mawé and Awetí, the other great branch is Tupi-Guarani.

other fields of research, such as anthropology, archaeology, geography, biology. Linguistic typology can also contribute its findings to these fields.

Such open-source treatment of information requires clear ethical rules. These are referred to in §1.5.

1.4 *Language description, documentation and language endangerment*

All the fields of linguistics mentioned in §1.1–§1.3 are nourished by data collected and described in sensitive detail. Therefore, the work of field linguists who go out to collect the data of not-yet-described languages remains important. In the field, linguists can train their skills working with languages typologically distinct from the globally dominant languages; they can get to an understanding of the complexity of language, together with the speakers of these languages.

Also, the context of language endangerment is something that field linguists are confronted with, as the speakers often place great hopes upon the researchers, with the idea that their studies can assist with revitalization of their language. The Baure language still counted some 80 elderly speakers, when Danielsen's grammar (Danielsen 2007) was published. Today, however, there are probably fewer than 10 speakers. Fortunately, we can note some success in teaching the language to the youngest generation in the past 10 years (cf. Danielsen & Terhart forthcoming). The Paunaka language had very few speakers when the documentation project started in 2011, and there are currently no more than eight speakers, as the language suffered from less national support and as not yet been listed in the Bolivian Constitution as an official language (Terhart 2024). The Tupi-Guarani language Guarayu counts more than 10,000 speakers, constituting the fifth biggest language of Bolivia. It is still being transmitted intergenerationally, and with this number of speakers, it was also possible to integrate the speakers into the research team (cf. Danielsen 2024).

In South America there are manifold languages and language families, and Amazonia is a region with extreme linguistic diversity, for example, there are many small groups, with originally around 300 languages/groups and various linguistic isolates, and many people are originally multilingual. In Bolivian Amazonia alone, we find eight unclassified languages (or isolates) (Crevels 2002, 21). The majority of Amazonian languages is in danger of extinction (UNESCO 2023). Among the various reasons for the endangerment are (cf. also Thomason 2015):

- size of language group
- confrontation with a dominant language/s (mainly European) since and after colonization
- missionization and forced joint settlements with many other indigenous groups
- consequence of language contact, language shift

- (forced) movement of people (fleeing from colonizers, losing land = hunting grounds, rubber boom, etc.)
- legal status of the language
- language teaching

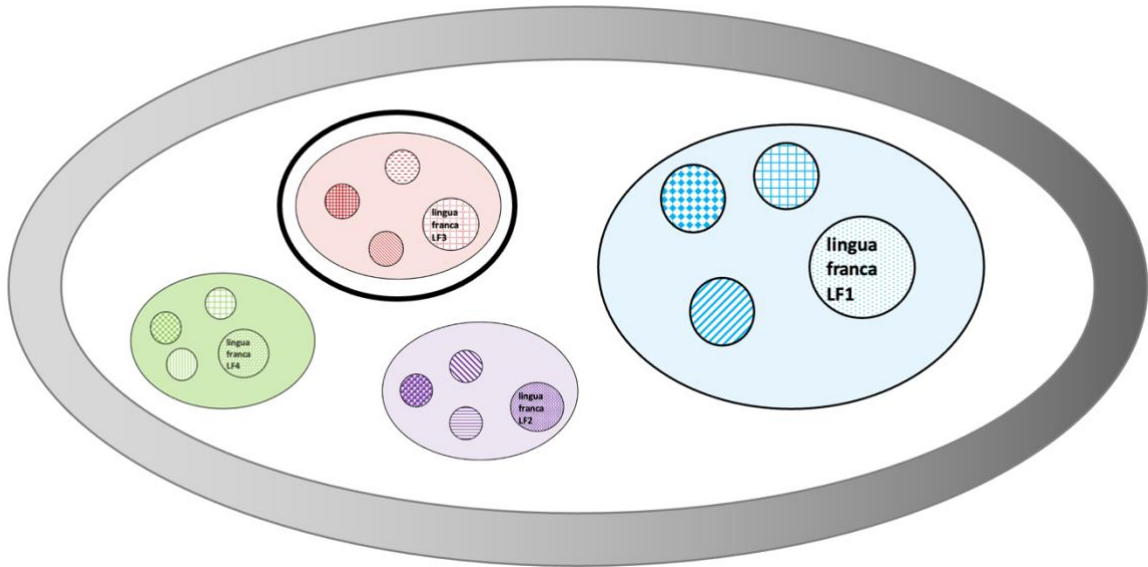


Figure 6: Lingua francas (LF = lingua franca) in Jesuit missions, so-called reductions

The situation of endangered languages is individual for each language, and there are usually multiple factors that play a role. A strong factor in Bolivia has been the language policies of the Jesuits throughout the 17th and 18th centuries (cf. Figure 6, cf. also Nikulin 2020), followed by the Franciscans through the 19th century until the middle of the 20th century, followed by the Evangelist missionaries since early 20th century until well into the 21st century. The manner in which the Jesuits' language policies influenced the linguistic landscape is exemplified in Figure 6. Jesuits firstly usually chose one lingua franca from a mission town that consisted of multiple language speakers. Then, the lingua francas with most speakers became more dominant over the others and functioned as lingua franca over a larger area. This kind of selective language support led to the rapid loss of many minority languages.

Language endangerment can be measured on different levels, and many terminologies have been used, especially because the borrowing from the field of biology does not always allow direct mapping between living species and language, and the terms “language death” and “language extinction” have a pathetic and emotional flavour. The categories by Krauss 2007 have been used quite widely, also by UNESCO (cf. Figure 7).

safe			a+	
endangered	stable		a–	all speak, children & up
	in decline	unstable; eroded	a	some children speak; all children speak in some places
		definitely endangered	b	spoken only by parental generation and up
		severely endangered	c	spoken only by grandparental generation and up
		critically endangered	d	spoken only by very few, of great-grandparental generation
extinct			e	no speakers

Figure 7: Krauss' (2007, 1) categorization of degrees of language endangerment

The languages discussed in this book all but one can be categorized as critically endangered or extinct, while one language (Guarayu) is a stable language in the plurilingual Bolivian context.

Language endangerment is only one reason why linguists have started documenting languages. While it is a matter of debate which reasons are the strongest, by and large, the scientific curiosity for the new, the different, the exotic, the awkward, the not-yet-discovered, are the romantic ideas that many researchers are stimulated by. Other questions include whether this is always good for the language community, and what is the position of the language community itself. Therefore, it is important to include some ethical considerations, not only in relation to this research and its context, but also more globally.

1.5 Ethics in linguistics

When we investigate something, this something becomes an object, and there may be the question of possession, such as in the case of archaeological sites and material culture. In such cases, the cultural heritage is in question and permissions must be given by the respective country or culture where the investigation takes place (cf. projects by Carla Jaimes Betancourt at the University of Bonn, cf. also a recent case of returning ancient feather cloaks of Tupinambá people to Brazil).⁸ When investigating

⁸ Carala Jaimes Betancourt, e.g. in the current project at: <https://www.iak.uni-bonn.de/de/institut/abteilungen/altamerikanistik/forschung-publikation/amazonas-anden/patrimonio-y-territorialidad-percepciones-pasadas-presentes-y-futuras-entre-los-pueblos-tacana-tsimane-moseten-y-waiwai/heritage-and-territoriality> (10/08/2024); the returning of the feather cloak to the Tupinambá: <https://www.kaea.eu/insights/a-tapestry-of-triumph-tupinamba-mantels-return-weaves-a-new-chapter-in-cultural-reclamation> and <https://agenciabrasil.ebc.com.br/en/geral/noticia/2024-07/indigenous-people-welcome-return-16th-17th-century-cloak-brazil> (28/09/2024).

language in the form of interviews or texts, the question of intellectual property cannot be ignored before information is published. Language communities should not simply be turned into objects of investigation. In the 21st century, researchers are more and more concerned about the participation of language communities in the course of investigation. This includes the contents of the culture that is being investigated, as well as the active participation of the communities in collecting data and analysing them. Therefore, a researcher working today in the field with speakers of endangered languages also seeks an active role for the speakers in this investigation. These discussions arise everywhere where humans are involved in research, and a number of publications have pointed out that the researcher should be sensitive to the needs of the people under study, and the imbalances in the global context. Danielsen (2024) shows the kind of ethical conflicts that may arise, and proposes solutions, such as the inclusion of the people under study in the design of research questions, and the production and publication of elements that can be useful for the communities, such as keyboard installations for different scripts and fonts, or publication in the local languages to enable more open access for them. The integration of the speech community in linguistic studies can also lead to much broader perspectives, such as the inclusion of description and perception of language change from the point of view of the community members, as well as the researcher.

Open access is also a key issue. Research outcomes are often published in expensive specialized volumes that only libraries can afford to purchase. Few libraries in the poorer countries, where the endangered languages are often located, can afford to buy them. In the past decade, the movement towards open access in the linguistics discipline (cf. link to the Open Access Network)⁹ has produced online journals such as Language Science Press.¹⁰ Additionally, publication in English language-only tends to exclude many people in Latin America, who need, at the least, a publication in Spanish in order to make use of the text. There is still much need for translation and publication in Spanish, for scholarship on South American languages to reach a wider readership, including the language communities under study. Language documentation programmes like DobeS (Dokumentation bedrohter Sprachen)¹¹ have developed ethical guidelines for work in the field, the collection of data, and the disposal and publication of data in open-access archives (cf. Code of conduct of DobeS, e.g. in Wittenburg, 2005). Researchers must ensure that the speakers/interviewees agree to a recording (otherwise, no recording can be collected), and after the interview, the speakers must approve the publication or otherwise no publication. But what if speakers do not understand what they consent to, because they are (digitally) 'illiterate', i.e. they may not have had the privilege of access to formal education, especially including digital world and its spheres? It is often up to the researcher to decide

⁹ <https://open-access.network/informieren/open-access-in-fachdisziplinen/linguistik>

¹⁰ <https://langsci-press.org>

¹¹ <https://dobes.mpi.nl>

once more if the data are not too sensitive for public exposure, and then the data must be carefully stored. The relevant metadata can be made available to future linguists, therefore the protection of the data must be ensured, as well as the accuracy of future data analysis. The problem of global data security has not yet been completely resolved; ongoing discussions will likely yield new solutions in future. It is therefore especially important that future researchers should know who to ask for further permissions, and that the data archives and metadata are regularly updated in formats by the curators. These minority issues are also widely discussed and worked on at the Institute of Frisian and Minority Issues at the Europa University of Flensburg.

2 Collect and connect

This section describes the texts that are part of this volume. The texts are related to the different fields and angles mentioned under §1. Each article contributes in part to the research questions: Which points have been resolved, and which are under debate when documenting languages, and who owns language as part of the culture? Which linguistic factors does the analysis of minority languages contribute to linguistics? The focus in this section lies on the description and analysis of small languages that are already extinct or critically endangered with very few remaining speakers. Part of the study addresses the area of Amazonian, and more specifically, Arawakan linguistics. The following summary briefly sets out the articles of the collection in separate subsections.

2.1 Historical linguistics with limited data resources

Many Amazonian languages are extinct today, but their names are mentioned in descriptions of South America in the 19th century. Missionaries and traveller-explorers, like Natterer (1820-1833; cf. Adelaar & Brijnen 2014, 176) and d’Orbigny (1835-1847), collected cultural, linguistic, biological, and geographical information about the peoples met. They produced short word lists or included single pieces of information in their significant works. In the 20th century, more data were collected in South America, and there are numerous manuscripts. Nordenskiöld (1911 and other works), e.g., travelled through Bolivia and took notes on languages and cultural information of various peoples, and was one of the first to publish his critical view on the position of indigenous peoples in national society. Hanke followed in his footsteps and is the only ethnographer who witnessed and described the genocide of the Jorá people in the 1940s in Bolivia (cf. Danielsen & Gasparini 2017*). All these historical data have not yet been fully digitized, nor have they all been analysed.

While wordlists would previously only be used for lexical comparison and the reconstruction of proto-lexemes, the historical lists of “words” often contain more than just lexical information. With some

knowledge of the language family, the data collection can reveal phonological and minimal morphological, morphosyntactic and syntactic information that is useful for the classification of the languages. Historical data are sometimes the only data we have. The already extinct Bolivian Arawakan languages Saraveka, Apolista, Paikoneka, and Joaquiniano can be classified on the basis of historical data, as illustrated in Danielsen (2013*). The available data are certainly limited, but they consist of wordlists, including a few short phrases. These data are compared and analysed, and with the knowledge of Arawakan languages, the information can be categorized with respect to, e.g., the types of morphemes marking the third person – one feature that distinguishes the languages within Arawakan. The data have been codified in feature matrices to be analysed in the SplitsTree programme as phylogenetic nets. This representation of the data allows inferences on how closely the languages are related. The visualization shows that the extinct Arawakan languages in Bolivia belong to different branches within the language family and are thus presumably the results of different migration routes and types of contact of Arawakan-speaking groups into Bolivia.

2.2 *Comparative Arawakan linguistics*

As described in §1.2, §1.3, and the preceding section §2.1, comparative linguistics and typology started with the comparison of words but moved on to the comparison of word structures and grammatical components, as well as word order. The comparison of linguistic features gives us insights into the structural profiles of languages, so we observe which structural characteristics occur and which other characteristics they possibly co-occur with. This leads to structural language types of the global languages (cf. §1.3). When comparing languages within one identified language family, we expect great structural similarities, and here the discrepancies are of relevance. The diverging linguistic features are those features that single out individual languages with internal innovations, and classify the languages into groups due to shared innovations, or they show the results of areal linguistic contact with languages outside the language family. The Arawakan language family has been investigated for more than two centuries, and various attempts have been made for its internal classification. There may be some tendencies as to Arawakan languages in the North showing less complexity in the verbal morphology than Southern Arawakan languages (Danielsen & Durand 2020, 142). However, many linguistic features are shared by all Arawakan languages. This can make us conclude that the spread of this large South American language family happened rather recently (cf. Danielsen et al. 2011*).

In the past decade, phylogenetic trees have been produced from linguistic feature matrices in order to visualize the relations between languages (cf. Figure 4 above). As shown in Danielsen et al. (2011*), the analysis of Arawakan data on a large scale allows implications on the movement of people and their languages, language spread, and language contact (cf. Danielsen et al. 2011* and Eriksen &

Danielsen 2014).¹² The comparison of the linguistic features of Arawakan languages indicates “that the Arawakan diaspora was in part a relatively recent phenomenon and that languages did not spread successively in one big migration phase, but rather in waves (cf. Hornborg 2005, 603)” (Danielsen et al. 2011*, 174). Language contact must have played an important role in the spread of Arawakan languages, in a similar manner to pidgin & creole languages. For example, Arawakan languages served as trade languages via the Amazonian riverine system, yet in the locations where we find Arawakan languages today, not all are the results of the movement of people as such, but they may also result from long-term language contact and the movement of an Arawakan language and local language shift (cf. Danielsen et al. 2011*, 186-189). These findings are not only relevant for the linguistics discipline, but they are picked up by other disciplines working on the Arawakan culture (cf. Hill & Santos-Granero 2002, Hornborg & Hill 2011).

2.3 *The typology of polysynthetic languages*

In linguistic typology, as shown in §1.3, the individual linguistic features play a role, and they are categorized and then compared between languages. There are linguistic features that are more common or frequent, such as pluralization of nouns, verb conjugation with relation to time (tense), or compounding of nominal roots. Other features may be considered “rare” and are not frequently found or only once in the languages of the world (cf. Wohlgemuth & Cysouw 2011). Here, it is especially interesting for the field of linguistics to find descriptions of languages that show such “rare” features, and often, these occur in small languages. We do not know much about the small languages – in contrast to the dominant languages, like English or Spanish – and their characteristics therefore ‘naturally’¹³ have not been integrated much in the generalizations of language in linguistics. A high degree of synthesis, called polysynthesis, is often considered an interesting feature to be compared. Haspelmath elucidates:

the notion of polysynthesis is primarily attractive because the languages concerned seem so exotic to linguists with a background in European languages. (...) In many areas of social science, it has taken scholars a long time before they recognized that a world-wide (or cross-species) perspective forces (or allows) us to abandon entrenched earlier views and notions that were strongly influenced by European or Western stereotypes. (Haspelmath 2018, 323-324)

¹² cf. also Campbell (1997: 95-100) on the “Implications of Linguistic Classification for Understanding the Original Population of the New World”.

¹³ The word ‘naturally’ is given in inverted commas, since the Eurocentric bias is not there by nature, but it is a consequence of the dynamics on the globe.

Some languages always appear in the examples, like Inuit or Yupik languages for the presentation of polysynthesis, possibly because they were first presented to the linguistic public. However, there are many more polysynthetic languages in the world. It seems that especially the Americas provide us with many examples of languages that are synthetic or polysynthetic (cf. Fortescue, Mithun & Evans 2017). Amazonian languages are described to provide a relatively high rate of synthesis (cf. Aikhenvald 2017):

Most Amazonian languages are agglutinating with some fusion. There are no isolating languages.(...) The majority of Amazonian languages are highly synthetic.(...) A major feature of highly synthetic languages in Amazonia is extreme internal complexity of grammatical words with numerous templatic slots for bound morphemes, and several morphemes in each slot. The order of bound morphemes—affixes or clitics—may be variable. (Aikhenvald 2017, 285-286)

Thus, the composition of morphemes, the way they co-occur, how the slots are organized where morphemes occur, and which categories are where to find within a word template, are features that are worth investigating with polysynthetic languages. Arawakan languages are largely polysynthetic. Such is the case of Baure, an Arawakan language from the Bolivian Amazonia. Danielsen (2012*) is an example of a typological description of the marginalized language Baure (Arawak), taking a look at affix composition and the role of these lexicalized units in a polysynthetic language. While Baure only shows limited synthesis in nominal words – with up to four morphemes combinable, the morphology of predicates tends to be highly complex, with up to 10 morphemes in one word. Baure has an average morpheme ratio of 2.74 morphemes per word (Danielsen 2007, 83) and is thus relatively synthetic. Considering only the most complex part of speech, i.e. the predicate, the ratio is 4.38 morphemes per word.

The verbal word in Baure can be very complex, with several morphological levels: reference to the arguments of the clause, aspect marking, mood and evidentiality, incorporation of nouns, classifiers, directionals, etc. It is generally difficult to determine the limits of derivation, because almost all marked categories have a tendency to function derivatively and result in lexicalization in many cases. Apart from the most recent research on this language, there are historical data from the Jesuit era. Comparing these old data with those of today, we note that some affix complexes were lexicalized and are presented today as units. These complexes could be called “composite affixes” that are processed as a single unit. Compound verbal suffixes do not have a nucleus, but rather a scope relationship that has been lexicalized. (Danielsen 2012*, 151, [translated into English here by the author]).

Danielsen (2012*), therefore, gives insights into the nature of polysynthesis and the behaviour of morphemes.

This line of investigation is also followed in Danielsen (*forthcoming*), where the focus lies on the change of the semantics of categories and the conversion in meaning of morphemes in this history of changing forms and converging patterns. That polysynthesis is not the end of a language development is illustrated in Danielsen (2015), where another strategy is illustrated that may, in the long run, lead to the loss of morphological complexity in this polysynthetic language. The article describes the TMA (tense, mood, aspect) categories in Baure, which are generally all suffixes in the complex verb template. The investigation of contemporary Baure, with comparison to historical descriptions of the Old Baure language, discovered alternative analytic strategies that started to replace the bound morphemes for TMA categories. A typological profile of a language is therefore never fix and always apt to change and adaption.

2.4 *The description and documentation of endangered Bolivian Lowland languages*

In the past 20 years, many South American, Amazonian, and Bolivian languages have been documented and described, last but not least due to the support of large documentation programmes, such as DobeS (Dokumentation bedrohter Sprachen)¹⁴ and ELDP (Endangered Languages Programme) with ELAR (Endangered Languages Archive)¹⁵. In these large and globally connected digital archives, linguists have uploaded the data of endangered languages they collected in the field. After these initial programmes for general data storage, other projects arose for the presentation of more specific data, such as grammatical features concerning certain verbal categories (TAME – tense, aspect, mood, evidentiality), the composition of the noun phrase, areal features, or certain language families and their features (cf. O’Connor & Muysken 2014 i.a.). Many of these data and the comparative Arawakan data mentioned in 2.2 and 2.3 have been uploaded to the database named SAILS online¹⁶, and additional Tupian data can be found in TuLaR¹⁷. More recently, the global scientific enterprise of Glottobank has gathered all available grammatical information of the world’s languages in one database, in the Grambank section¹⁸ (cf. Skirgård et al. 2023). Other databases present semantic relations between lexemes, such as in the colexification database CLICS^{3, 19} where lexemes in the world’s languages are connected in graphs according to their lexical overlap, i.e., questions such as: Does the language have separate words for the concepts of HAND and ARM, or is it one word? All these databases, these large collections of information and knowledge, allow much more integration

¹⁴ Dokumentation bedrohter Sprachen, at <https://dobes.mpi.nl>

¹⁵ Endangered Languages Archive, at <https://www.elararchive.org>

¹⁶ South American Indigenous Language Structures, at <https://sails.clld.org>

¹⁷ Tupian Language Resources, at <https://tular.clld.org>

¹⁸ <https://glottobank.org>

¹⁹ <https://clics.clld.org>

of the linguistic diversity in the generalization of linguistic features, typologies of languages. The databases offer statistical proof for statements that hold for many languages of the world.

Among the current digitalized data are also manuscripts of endangered languages, such as the Guarayu dictionaries by Hoeller (1929, 1932a), made available as an online dictionary in Danielsen et al. (2019*). This online publication in the *Dictionaria* series includes a grammatical description of the Guarayu language, which had not been described in detail since the early manuscripts of Franciscan priests. The *Dictionaria* page supports the idea of free availability of data:

Dictionaria is an open-access journal that publishes high-quality dictionaries of languages from around the world, especially languages that do not have a large number of speakers. The dictionaries are published not in the traditional linear form, but as electronic databases that can be easily searched, linked and exported. (<https://dictionaria.clld.org>)

The Guarayu language description, detailed in Danielsen et al. (2019*), with the linked and searchable dictionary, contributes to the field of language descriptions, being the first modern grammar of the small language that is public and open access. This availability of information and public access also gives a good example for future projects of this and other small languages.

Guarayu is particularly interesting for the study of Tupian languages and the great branch Tupi-Guarani within Tupian for two reasons other than individual interest in the language and the diversity argument: firstly, Guarayu is a comparatively old language of this family and provides evidence of a very close relationship to one of the oldest languages of the family, Tupinambá (Dietrich 2018); and secondly, Guarayu is a very good example of language perception and invisibility (in terms of Langer, Davies & Vanderbussche 2012, Langer & Havinga 2015) in the documents of 19th century missionaries. The documentation of Guarayu by Franciscan missionaries in the area stems back almost 200 years (Viudes 1841 in Dietrich & Danielsen 2024; Hoeller 1932a, 1932b), with the most recent documentation in the 21st century by the GIZAC project and ongoing research.²⁰

The claim that Guarayu is a rather old representative of the Tupian family is suggested by the remaining final consonant /r/, whereas the majority of Tupian languages have lost this consonant. Final consonants have also been documented for now-extinct Tupinambá (Ferraz Gerardi 2023), often taken as a basis for the reconstruction of proto-Tupian. The final consonant /r/ can be found in Guarayu *suindar* ‘owl’ and Tupinambá *suindara* (cf. Symoes & da Costa Almeida Almeida 2020[1771]), but in comparison, Guarani *suinda* has lost it. Some reconstructed examples in proto-Tupian indicate an original final syllable *-ir or *-ira, which demonstrates different developments in the subgroups of Tupian and Tupi-Guarani. An example is the ‘flower’: The word is either *potira* in Tupinambá of Belém

²⁰ GIZAC project at ELDP/ELAR: <http://hdl.handle.net/2196/98dc4979-5835-4066-8067-a94a244b7459>

(cf. Symoes & da Costa Almeida Almeida 2020[1771]) or *potir* in other Tupinambá variants (Ferraz Gerardi et al. 2022)²¹, just like old Guarayu *potir* (cf. Danielsen et al. 2019*)²². The presence or absence of the final vowel *-a* seems to be a variation, but in some languages, it is lost. Then the final consonant *-r* got lost in Guarani *potĩ* and many more languages. In Guarayu, on the other hand, we observe that this final consonant is not lost, but it has lately been subject to metathesis, and the modern Guarayu form is *potri* (already at the beginning of the 20th century, in *Dictionaria*, cf. Danielsen et al. 2019*). The retention of this final consonant in Guarayu – either in the final position or changed in metathesis – has further effects on the word structure and language change. The publications of old dictionaries, such as Danielsen et al. (2019*), and Symoes & da Costa Almeida Almeida (2020[1771]), allow for the observation of such language changes.

When the first Franciscan missionaries arrived in Bolivia in the area where Guarayu is spoken at the beginning of the 18th century, they were equipped with the early Guarani grammars written by Anchieta (1595) and Montoya (1993[1640]). As in other places, they presumably also used the Guarani versions of the Catechism and other texts for conducting the mass. The Guarayu speakers probably even understood half of the words, unlike the Chiquitanos in the closest proximity, who spoke a Macro-Jê language and could not understand any of the sermons – even though we can presume a certain degree of bilingualism of the Chiquitanos with Guarayu at the time of early colonization. The earliest Guarayu grammar was first written by the Franciscan Lacueva and then copied by the Franciscan Viudes in 1841 (published in Dietrich & Danielsen 2024). This grammatical description was mainly composed on the basis of the Guarani grammars the missionaries had at hand. This is not only obvious with whole copied paragraphs. On many pages, there are Guarani words in the description of what they argue to be Guarayu. Thus, in many examples, we cannot say with certainty if the priests referred to one or the other language, or were aware of the differences (cf. Dietrich & Danielsen 2024, 2, 22-23). Guarani and Guarayu may have been more similar 200 years ago than they are today. Some words used in Guarani were already obsolete in Guarayu by the beginning of the 20th century (Hoeller 1932a, b, and Danielsen et al. 2019*); other words have never been documented for Guarayu at all. The two languages are indeed similar, belonging to the Tupi-Guarani language family, but they took different paths of development, such as the aforementioned final consonant loss in Guarani. This also concerns grammatical changes. Some grammatical constructions may have existed in Guarayu at the beginning of the 19th century, because the morphological or lexical material can be evidenced. However, due to the scarce documentation, it is difficult to establish whether Guarayu actually used these constructions, or only understood them due to mutual similarity. The missionaries' view on the

²¹ Compare cognate set 6486 at: <https://tular.clld.org/contributions/tuled>

²² See specific entry at <https://dictionaria.clld.org/units/guarayu-LX001197>

Guarayu language – like many other Tupi-Guarani languages – was likely to be blinded or distorted by the Guarani descriptions. The recognition of the relation of the languages was often confused with the idea of there being two dialects of the same language. That indigenous languages belonged to related old language families was probably not well understood yet. This subject needs to be studied in the future: How visible were the particularities of the specific Tupi-Guarani languages to the language describers, and how much effect might their descriptions have had on the specific languages (as a kind of prescriptive grammar model of Guarani)? One basis for this comparison is the grammar sketch of early 20th century Guarayu in Danielsen et al. 2019* (based on Hoeller 1932b), and the related dictionary with 3,590 entries (based on Hoeller 1932a, 1929). Today, we can compare these data to the earliest grammar sketch from 1841 (Dietrich & Danielsen 2024) and to contemporary Guarayu, as documented in the GIZAC Archive (Danielsen 2016). The open access grammatical description of Guarayu on *Dictionaria* (Danielsen et al. 2019*) is presented in English and Spanish, so that Bolivian linguists and non-English speakers can also benefit from it.

2.5 *The discussion of ethics in linguistics – difficult cases*

When working with small endangered or almost extinct languages, we cannot investigate independently of the people who speak these languages, or who are the descendants of the last speakers. For the linguistic researcher, it may be a discovery that there is yet another undescribed language in the world. However, the speakers or descendants of the ethnic or linguistic group will have their own personal relation to the language. It must be acknowledged that the loss of the language is related to the marginalization of the people, and this is generally entwined with negative treatment towards them from outsiders. This context has been discussed in Danielsen & Gasparini (2017*), highlighting two case studies – the Tupi-Guarani languages Jorá and Guarasu Ñe'ë. The text points out that the Jorá people had suffered a genocide in the 1940s, and asks: “How to collect data from the surviving Jorá?” (Danielsen & Gasparini 2017*, 4). The text proposes a response to this dilemma:

The Jorá descendants do not talk about their memories, even on request. We do not know if they forgot or suppressed them due to the traumatic experiences, or because they were so young. Another Jorá descendent told us that many people have already come to ask the woman about her past, insisting repeatedly on Jorá vocabulary. [...] We decided to break away from this lurid 'hunt for the last speaker' and did not insist on trying to collect data. We should accept these limitations of the work from an ethical point of view. Instead, we relied on secondary information provided by Baure people who met Jorá captives in the

1940s and 1950s and remembered some of the vocabulary. (Danielsen & Gasparini 2017*, 4)

Danielsen & Gasparini (2017*) shows that linguists can only go as far as permitted by the people whose intellectual and cultural property they may be interested in. If the people under study feel uncomfortable, fear negative consequences of visibility, or have not come to an agreement with the outsider-researchers, researchers may need to reconsider alternative modes of linguistic investigation, other than the collection of data. Also, political tensions can mean that fieldwork is not possible, and alternative ways forward should be investigated. There are alternative solutions, such as working with the available secondary data.

Science and field research have evolved to include better integration of people who are the subjects of investigation. In recent decades, there has been a shift towards (support for) researchers conducting research in their own country, rather than following the colonizers' tradition of European or Western outsiders collecting data from 'exotic' people in less developed countries. The speakers of minority languages can themselves manage linguistic projects (cf. Danielsen, Cuñapiri & Guayarabey 2024), and they have been accepted as project managers by the ELDP, a major funding programme for the documentation of endangered languages.²³ National researchers and speakers of small languages can collaborate to promote a language in this manner. The publication of materials for and with the language communities has proven to be very well received, and this is also a supportive factor for the general perception and judgments about the investigation of a language (cf. Danielsen 2024). Nowadays, we find digital and paper dictionaries, transliterated historical data, story books, teaching materials, and many more text types for minority languages.

Danielsen & Gasparini (2017*) shows one mode of data treatment that respects the people whose language is being investigated. The investigation by the GIZAC project, mainly based on published vocabulary and personal notes of other researchers, produced a dictionary of Guarasu Ñe'ë (GIZAC 2015), which the speakers' descendants appreciate. The GIZAC researchers' studies supports today's work of the national language institutes of the Guarasu Ñe'ë language.

²³ ELDP: <https://www.eldp.net>

3 Conclusions

The two sections of this introduction set out this volume's contributions to the scholarship of minority languages within the discipline of linguistics, as well as the remaining unresolved questions. In §1, the exchange of words and thoughts in linguistics presented the different fields of the discipline we cross in this volume, ranging from historical linguistics via typology to the ethics of linguistic investigation. Section §2 outlined the ways in which the collection in this volume connects to the research in these different fields, and emphasized the specific gain of knowledge each article contains. The findings present the original work and individual contributions to the field of linguistics by the author of this volume.

The studies in (Danielsen 2013*), (Danielsen, Dunn & Muysken 2011*), (Danielsen 2012*), (Danielsen, Sell & Terhart 2019*), and (Danielsen & Gasparini 2017*) contribute to the fields of historical linguistics, comparative linguistics, linguistic typology, documentary and descriptive linguistics, and ethics of linguistics. They resolve some questions about the internal classification of Arawakan and Tupian languages, and contribute to our knowledge of the migration of the peoples in South America. The specific small languages studied in more detail in the articles gives us more access to certain branches of Arawakan and Tupian, including an understanding of language change over at least 200 years of documentation. Every new language contributes to the global view on language and makes the typological profile of human language more precise. As more small and previously undocumented languages are described and included in generalizations, the more precise our picture of language families and areal diffusion becomes. However, the role or the weight of the dominant languages is less mentioned and remains unresolved in the investigation of language contact and diffusion in South America, as well as the issue of how to measure their impact. Dominant languages have stronger influence on more languages, but we still lack a scientific measurement of the values, size, and strength of linguistic groups that are impacted.

While minority languages have suffered from a long tradition of ignorance and discrimination, their inclusion in the investigation of contemporary linguistics has certainly improved the diversity of information and representatives. Many issues have been discussed with relation to archiving data, collection of data of highly endangered languages, and the rights of the language communities and the individual speakers. What has not been presently resolved, is an ethically correct treatment of data in archives, and how to give language communities access and control over these archives and data. Thus, while our knowledge of linguistics advances as we change perspectives and search for solutions, science and linguistics do not act independently; many more perspectives, interests, and targets must

be considered. The vivid nature of science is the continuous search for new solutions. Just as language is not static and is apt to change, science is dynamic and constantly changing.

References

Adelaar, Willem F.H. & Pieter C. Muysken. 2004. *The languages of the Andes*. Cambridge: Cambridge University Press.

Adelaar, Willem F.H. & Hélène B. Brijnen. 2014. "Johann Natterer's Linguistic Heritage". *Archiv für Völkerkunde*, 162-183.

Aikhenvald, Alexandra Y. 2001. "Areal Diffusion, Genetic Inheritance, and Problems of Subgrouping: A North Arawak Case Study". In *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*, edited by R.M.W. Dixon & Alexandra Y. Aikhenvald, 167-194. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. 2017. "Chapter 15: Polysynthetic structures of Lowland Amazonia". In *The Oxford Handbook of Polysynthesis*, edited by Fortescue, Michael, Marianne Mithun, & Nicholas Evans, 284-311. Oxford: Oxford University Press.

Aikhenvald Alexandra Y. 2022. "The Amazon Basin: Linguistic Areas and Language Contact". In *The Cambridge Handbook of Language Contact, Volume 1: Population Movement and Language Change*, edited by Salikoko S. Mufwene, Anna María Escobar, 232-260. Cambridge Handbooks in Language and Linguistics. Cambridge University Press.

Anchieta, Joseph de. 1595. *Arte de grammatica da lingoa mais vsada na costa do Brasil, Feyta pelo P. Joseph de Anchieta da Cõpanhia de IESV*. Coimbra: Antonio de Mariz. Retrieved at: https://pt.wikisource.org/wiki/Arte_de_Grammatica_da_Lingoa_mais_Vsada_na_Costa_do_Brasil (23/01/2023).

Atkinson, Quentin D., & Russell D. Gray. 2006. "How Old Is the Indo-European Language Family? Illumination or More Moths to the Flame?". In *Phylogenetic Methods and the Prehistory of Languages*, edited by Peter Forster & Colin Renfrew, 91-109. Cambridge: McDonald Institute for Archaeological Research.

Bolnick, Deborah A., Beth A. Shook, Lyle Campbell, & Ives Goddard. 2004. "Problematic use of Greenberg's linguistic classification of the Americas in studies of Native American genetic variation". *American Journal of Human Genetics* 75(3), 519-522.

Campbell, Lyle. 1997. *American Indian Languages: The Historical Linguistics of Native America*. New York: Oxford University Press.

Campbell, Lyle, Thiago Chacon, & John Elliott. 2020. "Chapter 32: Contact and South American Languages". In *The Handbook of Language Contact*, edited by Raymond Hickey, 624-648. Hoboken, NJ: John Wiley & Sons Ltd.

Comrie, Bernard & Lucía Golluscio, eds. 2015. *Language Contact and Documentation*. Berlin, München, Boston: DeGruyter Mouton.

Constenla Umaña, Adolfo. 1991. *Las lenguas del área intermedia: Introducción a su estudio areal*. San José: Universidad de Costa Rica.

Crevels, Mily. 2002. "Why speakers shift and languages die: An account of language death in Amazonian Bolivia". In *Current studies on South American languages*, edited by Mily Crevels, Simon van de Kerke, Sérgio Meira, & Hein van der Voort, 9-30. Leiden: Research School of Asian, African, and Amerindian Studies (CNWS).

Crevels, Mily & Hein van der Voort. 2008. "The Guaporé-Mamoré region as a linguistic area". In *From linguistic areas to areal linguistics*, edited by Pieter C. Muysken, 151-179. Amsterdam, Philadelphia: John Benjamins.

d'Orbigny, Alcides. 1835-1847. *Voyage dans l'amérique méridionale. L'Homme Américain*. (various volumes). Paris: Pitois-Levrault; Strasbourg: F.G. Levraut.

Dahl, Östen. 2015. "How WEIRD are WALS languages?", paper at the MPI EVA Linguistics closing conference, May 3, 2015. Retrieved at:

https://www.eva.mpg.de/fileadmin/content_files/linguistics/conferences/2015-diversity-linguistics/Dahl_slides.pdf (04/11/2023).

Danielsen, Swintha. 2007. *Baure: An Arawak Language of Bolivia*. Indigenous Languages of Latin America (ILLA) 6. Leiden: CNWS.

Danielsen, Swintha. 2011. "Evaluating Arawakan features versus Guaporé-Mamoré features". Presentation at the workshop 'Historical perspectives on Sprachbund phenomena: the case of the Guaporé-Mamoré', Radboud University Nijmegen, November 25, 2011. Retrieved at: <https://de.scribd.com/document/469118366/Evaluating-Arawakan-features-versus-Guapore-Mamore-features> (29/09/2024).

Danielsen, Swintha. 2012. "Afijos compuestos y su lexicalización en el baure (arawak)". In *Lenguas Indígenas de América del Sur I: Fonología y Léxico*, edited by Marisa Censabella, Cristina Messineo, Hebe A. González, & Beatriz Gualdieri, 151-171. Mendoza: Editorial FFyL-UNCuyo and SAL. Retrieved at: https://www.academia.edu/5479523/Lenguas_indígenas_de_América_del_Sur_I_Fonología_y_Léxico (29/09/2024).

Danielsen, Swintha. 2013. "Evaluating historical data (wordlists) in the case of Bolivian extinct languages". In *Historical reconstruction of American languages*, edited by Astrid Alexander-Bakkerus & Otto Zwartjes. STUF – Language Typology and Universals/ Sprachtypologie und Universalienforschung, 66(3), 272-298. Bremen: Akademie Verlag.

Danielsen, Swintha. 2015. "Observing the evolution of TMA particles – replacing agglutinating strategies". In *Taming the TAME systems*, edited by Emmanuelle Labeau & Qiaochao Zhang, 1-23. Amsterdam, Philadelphia: Rodopi.

Danielsen, Swintha. 2016. "GIZAC – Gwarayu and the Intermediate Zone (Amazonia – Chaco, Bolivia) documentation project". *Endangered Languages Archive*. Handle: <http://hdl.handle.net/2196/00-0000-0000-000F-BF45-3> (07/09/2024).

Danielsen, Swintha. 2024. "Academic indigenous lands: Connecting academics with the indigenous community, a Bolivian case study". *Zenodo*. Retrieved at: <https://doi.org/10.5281/zenodo.10844212> (29/09/2024).

Danielsen, Swintha. *forthcoming*. “The history of morpheme convergence with an eye on nominalization in Baure”. In *Nominalization across Arawakan Languages*, edited by Magdalena Lemus & Zachary O’Hagan. Special issue of *Folia Linguistica*.

Danielsen, Swintha, Michael Dunn, & Pieter Muysken. 2011. “The Spread of the Arawakan Languages: A View from Structural Phylogenetics”. In *Ethnicity in Ancient Amazonia: Reconstructing past identities from archaeology, linguistics, and ethnohistory*, edited by Alf Hornborg & Jonathan Hill, 173-195. Boulder, CO: University Press of Colorado.

Danielsen, Swintha & Tom Durand. 2020. “Differential subject marking in Arawakan languages: Distribution and origins”. *Journal of Language Contact*, 13(1), 141-176.

Danielsen, Swintha & Noé Gasparini. 2017. “Discovering two almost extinct languages in Bolivia: Jorá and Guarasu”. *18e Rencontres Jeunes Chercheurs en Sciences du Langage*, ED 268 Paris 3, Jun 2015, Paris, France. Retrieved at: <https://hal-univ-paris3.archives-ouvertes.fr/hal-01495156v1> (29/09/2024).

Danielsen, Swintha, Lena Sell, & Terhart, Lena. 2019. “Guarayu. A revised dictionary by Alfred Hoeller”. *Dictionaria* 6, 1-3590. Retrieved at: <https://dictionaria.cld.org/contributions/guarayu> (29/09/2024).

Danielsen, Swintha & Lena Terhart. 2015. “Borrowed clause combining patterns in two Arawakan languages – Baure and Paunaka”. In *Perspectives on Hispanic Contact Linguistics in the Americas*, edited by Gonzalez-Rivera, Melvin & Sandro Sessarego, 121–152. Series *Lengua y Sociedad en el Mundo Hispánico*. Vervuert: Iberoamericana.

Danielsen, Swintha & Lena Terhart. *forthcoming*. “El estatus de las lenguas de las tierras bajas bolivianos”. In *Proceedings of the Simposio Lenguas regionales y minoritarias, Mayo 2022*, edited by Eva Gugenberger & Philine Hilbig.

Danielsen, Swintha, José Cuñapiri, & Miguel Guarayabey. 2024. *Vira virai: ¿Ves Aves?*, subtitle *¿Mba’e yaikwa virai eta resendar indo uma virai yasepia vitemi yuvireko? ¿Qué sabemos sobre las aves – qué aves vemos siempre? ... sobre las aves en el idioma guarayu*. Santa Cruz: Kipus.

Derbyshire, Desmond C. & Geoffrey K. Pullum, eds. 1986-1998. *Handbook of Amazonian Languages*, Volumes 1-4. Berlin: Mouton de Gruyter.

Dietrich, Wolf. 2018. "Tupian Languages". Oxford: Oxford Research Encyclopedia. Retrieved at: <https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-362> (08/09/2024).

Dietrich, Wolf & Swintha Danielsen. 2024. *La Primera Gramática del Guarayu: Manuel Viudes, 1841 – copia de Francisco Lacueva: Arte de la Lengua Guaraya. Transcripción del manuscrito de 1841, comentarios, introducción, estudio y análisis*. Wissenschaftliche Schriften der Universität Münster, Reihe XII, Band 4. Münster: ULB Münster. Retrieved at: <urn:nbn:de:hbz:6-65978523243> (29/09/2024).

Dixon, R.M.W., & Alexandra Y. Aikhenvald, eds. 1999. *The Amazonian languages*. Cambridge: Cambridge University Press.

Dixon, R.M.W. & Alexandra Y. Aikhenvald. 1999. "Introduction". In *The Amazonian languages*, edited by R.M.W. Dixon & Alexandra Y. Aikhenvald, 1-21. Cambridge: Cambridge University Press.

Dorian, Nancy C. 2014. *Small-Language Fates and Prospects: Lessons of Persistence and Change from Endangered Languages: Collected Essays*. Brill's Studies in Language, Cognition and Culture, 6. Leiden: Brill.

Epps, Patience & Andrés Pablo Salanova. 2013. "The Languages of Amazonia". *Tipiti: Journal of the Society for the Anthropology of Lowland South America*, 11(1), 1-28. Retrieved at: <http://digitalcommons.trinity.edu/tipiti/vol11/iss1/1> (29/09/2024).

Epps, Patience & Lev Michael. 2015. "The Areal Linguistics of Amazonia". In *The Cambridge Handbook of Areal Linguistics*, edited by Raymond Hickey, 934-63. Cambridge: Cambridge University Press.

Eriksen, Love & Swintha Danielsen. 2014. "The Arawakan Matrix". In *The Native Languages of South America: Origins, development, typology*, edited by Loretta O'Connor & Pieter Muysken, 152-176. Cambridge: Cambridge University Press.

Ferraz Gerardi, Fabrício, Stanislav Reichert, Carolina Aragon, Tim Wientzek, Johann-Mattis List, & Robert Forkel. 2022. "TuLeD. Tupían Lexical Database (v0.12) [Data set]". *Zenodo*. Retrieved at: <https://doi.org/10.5281/zenodo.6572576> (08/09/2024).

Online database at: <https://tular.cld.org/contributions/tuled> (08/09/2024).

Ferraz Gerardi, Fabrício. 2023. *A Role and Reference Grammar Description of Tupinambá*. PhD thesis, Universität Tübingen. Handle: <http://hdl.handle.net/10900/142263> (29/09/2024).

Fortescue, Michael, Marianne Mithun, & Nicholas Evans, eds. 2017. *The Oxford Handbook of polysynthesis*. Oxford: Oxford University Press.

Gasparini, Noé, Swintha Danielsen, Natalia Chousou-Polydouri, Zachary O'Hagan, Keith Bartolomei, & Lev Michael. 2015. "Southern Tupí-Guaraní Languages Combining classification with areal linguistics". Presentation at *The 18th Annual Workshop on American Indigenous Languages, May 8, 2015*. Retrieved at: http://www.ddl.cnrs.fr/fulltext/Gasparini/Gasparini_2015_SouthernTupiGuarani.pdf (29/09/2024).

GIZAC. 2015. *Diccionario flora y fauna Guarasu*. Edited by Swintha Danielsen & Lena Sell. Santa Cruz de la Sierra: Kipus.

Greenberg, Joseph H. 2005. *Language Universals: With Special Reference to Feature Hierarchies (with a preface by Martin Haspelmath)*. Berlin, New York: De Gruyter Mouton.

Greenberg, Joseph H. 1963. "Chapter 5: Some universals of grammar with particular reference to the order of meaningful elements". In *Universals of Language*, edited by Joseph H. Greenberg, 58-90. Cambridge: The MIT Press.

Greenberg, Joseph H. 1987. *Language in the Americas*. Stanford: Stanford University Press.

Greenhill, Simon, Quentin D. Atkinson, Andrew Meade, Russell D. Gray. 2010. "The shape and tempo of language evolution". *Proceedings of the Royal Society, B*, 277, 2443-2450. Retrieved at: <http://dx.doi.org/10.1098/rspb.2010.0051> (29/09/2024).

Hammarström, Harald. 2014. "Basic vocabulary comparison in South American languages". In *The Native Languages of South America: Origins, Development, Typology*, edited by Loretta O'Connor & Pieter Muysken, 56-70. Cambridge: Cambridge University Press.

Haspelmath, Martin. 2018. "The last word on polysynthesis: A review article (on The Oxford handbook of polysynthesis (2017), ed. by N. Evans, M. Fortescue, M. Mithun)". *Linguistic Typology* 22(2), 307-326.

Hill, Jonathan David & Fernando Santos-Granero, eds. 2002. *Comparative Arawakan Histories: Rethinking Language Family and Culture Area in Amazonia*. Urbana: University of Illinois Press.

Hoeller, P. Alfredo. 1929. *Diccionario Guarayo-Castellano*. Ms.

Hoeller, P. Alfredo. 1932a. *Guarayo-Deutsches Wörterbuch*. Hall in Tirol: Verlag der Missionprokura der P.P. Franziskaner.

Hoeller, P. Alfredo. 1932b. *Grammatik der Guarayu-Sprache*. Hall in Tirol: Verlag der Missionprokura der P.P. Franziskaner.

Hornborg, Alf. 2005. "Ethnogenesis, Regional Integration, and Ecology in Prehistoric Amazonia". *Current Anthropology* 46 (4), 589-620.

Hornborg, Alf & Jonathan Hill, eds. 2011. *Ethnicity in Ancient Amazonia: Reconstructing past identities from archaeology, linguistics, and ethnohistory*. Boulder, CO: University Press of Colorado.

Klein, Harriet Manelis. 1992. "South American languages". *International Encyclopedia of Linguistics* 4, 31-35.

Krauss, Michael. 2007. "Classification and terminology for degrees of language endangerment". In *Language Diversity Endangered*, edited by Matthias Brenzinger, 1-8. Berlin: Mouton de Gruyter.

Langer, Nils, Steffan Davies, & Wim Vandenbussche. 2012. *Language and History, Linguistics and Historiography*. Studies in Historical Linguistics 7. Oxford: Peter Lang.

Langer, Nils & Anna Havinga. 2015. *Invisible Languages in the Nineteenth Century*. Historical Sociolinguistics 2. Oxford: Peter Lang.

Loukotka, Čestmír. 1968. *Classification of South American Indian Languages*. Los Angeles: Latin American Studies Center, University of California.

Mason, J. Alden. 1950. "The Languages of South America." In *Handbook of South American Indians*, Vol. 6, edited by Julian Steward, 157-317. Washington, DC: Government Printing Office.

Montoya, P. Antonio Ruiz de. 1993[1640]. *Arte de la lengua guaraní*. Iuan Sanchez, Madrid. Edited by Bartomeu Melià. CEPAG, Asunción.

Muysken, Pieter. 2008. *From linguistic areas to areal linguistics*. Amsterdam, Philadelphia: Benjamins.

Muysken, Pieter & O'Connor, Loretta. 2014. "Introduction: South American indigenous languages; genealogy, typology, contact". In: *The Native Languages of South America: Origins, Development, Typology*, edited by Loretta O'Connor & Pieter Muysken, 1-28. Cambridge: Cambridge University Press.

Nikulín, Andrey. 2020. "Contacto de lenguas en la Chiquitanía". *Revista Brasileira de Línguas Indígenas* 2(2): 05, 5-30.

Nordenskiöld, Erland. 1911. *Indianer och hvita i nordöstra Bolivia*. Stockholm: Bonnier.

O'Connor, Loretta & Pieter Muysken, eds. 2014. *The Native Languages of South America: Origins, Development, Typology*. Cambridge: Cambridge University Press.

Olthof, Marieke, Eva van Lier, Tjeu Claessen, Swintha Danielsen, Katharina Haude, Nico Lehmann, Maarten Mous, Elisabeth Verhoeven, Eline Visser, Marine Vuillermet, & Arok Wolvengrey. 2020. "Verb-based restrictions on noun incorporation across languages, Linguistic Typology". *Linguistic Typology* 25(2), 211-256.

Payne, David L. 1991. "A classification of Maipuran (Arawakan) languages based on shared lexical retentions". In *Handbook of Amazonian languages*, Vol. 3, edited by Desmond C. Derbyshire & Geoffrey K. Pullum, 355–499. Berlin: Mouton de Gruyter.

Reesink Ger, Ruth Singer, & Michael Dunn. 2009. "Explaining the linguistic diversity of Sahul using population models". *PLoS Biology* 7(11), e1000241.

Rodrigues, Aryon Dall'Igna. 2013. "Relações internas na família linguística Tupí-Guaraní". *Revista Brasileira De Linguística Antropológica* 3(2), 233-252.

Muysken, Pieter, Harald Hammarström, Olga Krasnoukhova, Neele Müller, Joshua Birchall, Simon van de Kerke, Loretta O'Connor, Swintha Danielsen, Rik van Gijn & George Saad. 2016. *South American Indigenous Language Structures (SAILS) Online*. Jena: Max Planck Institute for the Science of Human History. Retrieved at: <https://sails.clld.org> (29/09/2024).

Saraceni, Mario. 2023. "Language isn't 'alive' – why this metaphor can be misleading". *The Conversation: Academic rigour, journalistic flair*, June 6, 2023. Online publication at: <https://theconversation.com/language-isnt-alive-why-this-metaphor-can-be-misleading-205793> (09/09/2024).

Seifart, Frank, Jan Strunk, Swintha Danielsen, Iren Hartmann, Brigitte Pakendorf, Søren Wichmann, Alena Witzlack-Makarevich, Nivja de Jonge, & Balthasar Bickel. 2018. "Nouns slow down speech across structurally and culturally diverse languages". *Proceedings of the National Academy of Sciences (PNAS)* 115 (22), 5720-5725.

Seki, Lucy. 1999. "The Upper Xingu as an incipient linguistic area". In *The Amazonian languages*, edited by R.M.W. Dixon & Alexandra Y. Aikhenvald, 417-428. Cambridge: Cambridge University Press

Skirgård, Hedwig, Hannah Haynie, Harald Hammarström, et al. (incl. Swintha Danielsen). 2023. "Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss". *Science Advances* 9(16), eadg6175.

Strunk, Jan, Frank Seifart, Swintha Danielsen, Irene Hartmann, Brigitte Pakendorf, Søren Wichmann, Alena Witzlack-Makarevich, Balthasar Bickel. 2020. "Determinants of phonetic word duration in ten language documentation corpora: Word frequency, complexity, position, and part of speech". *Language Documentation & Conservation* 14, 423-461.

Swadesh, Morris. 1955. "Towards greater accuracy in lexicostatistical dating". *International Journal of American Linguistics* 21(2), 121-137.

Symoes Pereira, Carlos & Arthur da Costa Almeida Almeida. 2020[1771]. *Tupi de Belém: Dicionário Tupinambá-Português 1771 (Portuguese Edition)*. Kindle E-Book.

Terhart, Lena. 2024. *A grammar of Paunaka*. Comprehensive Grammar Library. Berlin: Language Science Press.

Thomason, Sarah G. 2015. *Endangered Languages: An introduction*. Cambridge: Cambridge University Press.

UNESCO. 2023. *World Atlas of Languages*. Retrieved at: <http://www.unesco.org/languages-atlas/> (12/10/2023).

van Gijn, Rik. 2014. "Subordination strategies in South America: nominalization". In *The Native Languages of South America: Origins, Development, Typology*, edited by Loretta O'Connor & Pieter Muysken, 274-296. Cambridge: Cambridge University Press.

Wittenburg, Peter. 2005. DOBES: Code of Conduct. DOBES-CoC-V2, 19.7.2005. Retrieved at: https://dobes.mpi.nl/ethical_legal_aspects/DOBES-coc-v2.pdf (19/08/2023).

Wohlgemuth, Jan & Michael Cysouw. 2011. *Rara & Rarissima: Documenting the Fringes of Linguistic Diversity*. Berlin, New York: De Gruyter Mouton.

Summary in English of “Afijos compuestos y su lexicalización en el baure (arawak)”

Compound affixes and their lexicalization in Baure (Arawak)

Baure is an Arawak language of Bolivia that is critically endangered. Its documentation includes the early notes of Jesuit missionaries from 1749 and 1767, tagmemic grammars by SIL linguists from 1967 and 1968, and the documentation of the DobeS project run by Danielsen since 2008.

Like other Southern Arawakan languages, Baure is polysynthetic, especially in the verbal morphology. There are mainly suffixes in the language, there is a nominal classifier system, and incorporation of nouns and classifiers into verbs. Formerly, as the Jesuits’ notes show, the language had a reality status system, where realis and irrealis were obligatorily marked; however, this system was lost in contemporary Baure.

Apart from being a polysynthetic language with particularly complex verb morphology, the Baure language shows numerous tendencies of compounding. There is compounding of nominal roots with the head being the right-most lexical root. Generally, nominal compounds consist of two nominal roots (1), but in some rarer cases, this compound can again have another lexical root attached to the right.

(1) [N1+N2]N

Nominal compounding also includes classifiers. Classifiers are relatively grammaticalized elements in Baure, but they originate in lexical nominal roots, at least in the proto-language. Classifiers attach to the right of the nominal root, which questions the headedness of these compounds. The construction (2) parallels nominal compounding in (1).

(2) [N+CLF]N

Among the many classifiers in Baure, there are simple forms, usually monosyllabic, and disyllabic forms that are often transparent as consisting of a more specific classifier root and a more generic one -ki ‘boundaries’, generally referring to the contents of the element referred to by the first classifier root. These complex classifiers are lexicalized as units, but the at least eight composite classifiers show that also classifier compounding can be observed in Baure, at least diachronically, cf. (3).

(3) [CLF1+CLF2]CLF

Turning to verbal morphology, the article describes how verbal suffixes attach to the root, as in (4).

(4) [[[[[V-x1]-x2]-x3]-x4]-x5]

Even though some categories are traditionally known as inflection, the character of all the suffixes is rather derivational. They occur in a particular order, but not all the suffixes can co-occur. Often, one of the verbal suffixes has lexicalized with the verbal root as a unit, as in schematized in (5).

(5) [[[[[V+x1]V -x2]-x3]-x4]-x5]

Out of the 13 verbal suffixes, at least four are not only disyllabic, but also morphologically complex. Comparing to the earliest documented Baure data by the Jesuits from the 18th century, the diachrony of two of these morphemes can be detected: the complex suffixes -wapa 'COS' and -wana 'DEP'. Both go back to the composition of the irrealis form of the copula suffix -wo/-wa 'COP/COP.IRR' and another suffix. The second part of -wapa is the irrealis form of the punctual/perfective suffix -po/-pa 'PRFV/PRFV.IRR'. The second morpheme in -wana is the irrealis form of the nominalizer -no/-na 'NMLZ/NMLZ.IRR'. Thus, we observe composition of grammatical morphemes here, abstractly represented in (6).

(6) [[[[V -[y1 + y2]x1-x2]-x3]-x4]-x5]

Old Baure had an obligatory distinction of realis and irrealis on predicates. All affixes had a realis and an irrealis form. Contemporary Baure lost this reality status system. Forms from the old system are maintained with different specification, including the compound suffixes that are fixed units that cannot be decomposed anymore today.

In conclusion, the article demonstrates an important strategy of a polysynthetic language to reduce complexity of information to be processed, namely by forming units in the greater complex of morpheme chains. Nominal compounding and the more specific compounding of classifiers have the tendency to regard the right-most element as their head. Complex verbal suffixes are not headed; they rather reflect frequently co-occurring morpheme combinations that diachronically froze. Both types of compounding, nominal (incl. classifiers roots) and verbal share the proportioning of information in identifiable chunks rather than allowing one long row of equal (monosyllabic) morphemes.

Danksagung

Mein besonderer Dank gilt den Sprecherinnen und Sprechern der indigenen Sprachen Boliviens sowie den Gemeinschaften, die mich über viele Jahre hinweg mit großer Offenheit aufgenommen, ihr Wissen mit mir geteilt und meine Forschung in vielfältiger Weise unterstützt haben. Dazu gehören im Besonderen die Baure, Paunaka und Guarayu, aber auch die kleineren Gruppen seien hier genannt: Carmelito, Joaquiniano und Guarasuwe. Ohne ihr Vertrauen, ihre Geduld und ihre Bereitschaft zur Zusammenarbeit wäre diese meine Forschung, die in dieser Habilitation mündet, nicht möglich gewesen.

Ich danke meinen Kolleginnen und Kollegen sowie meinen wissenschaftlichen Kooperationspartnerinnen und -partnern für den fachlichen Austausch, die kritischen Diskussionen und die konstruktiven Anregungen, die meine Forschung über viele Jahre hinweg begleitet und bereichert haben. Hier sind insbesondere die Kolleginnen an den Universitäten und Max-Planck-Institute von Leipzig und Nijmegen genannt, die Universitäten in Addis Abeba, Sonora, Buenos Aires, Santa Cruz de la Sierra, La Paz, Cochabamba und ganz besonders die Europa-Universität Flensburg.

Mein Dank gilt außerdem den Institutionen und Förderorganisationen, die meine wissenschaftliche Arbeit, insbesondere die Feldforschung und die Durchführung der Forschungsprojekte, ermöglicht und unterstützt haben. Dazu gehören die Volkswagenstiftung und das ELDP-Programm an der SOAS-Universität in London.

Schließlich danke ich meiner Familie und meinen Freundinnen und Freunden für ihre Geduld, ihre Ermutigung und ihre stetige Unterstützung bei der Einreichung dieser Habilitation und im Laufe meiner gesamten Forschungslaufbahn.

Asoropaiy, chapie, aviye, Danke, thank you, gracias!
Baure, Paunaka, Guarayu, Deutsch, Englisch, Spanisch