LUND
UNIVERSITY

A phonological reconstruction of Proto-Omagua–Kokama–Tupinambá

Olof Lundgren

Supervisor:
Assoc. Prof. Gerd Carling
Lund University

Co-supervisor:
Assoc. Prof. Lev Michael
UC Berkeley

Centre for Language and Literature, Lund University

MA in Language and Linguistics, General Linguistics

SPVR01 Language and Linguistics: Degree Project – Master's (Two Years) Thesis, 30 credits

March 2020

Abstract in English

The Tupí–Guaraní languages Omagua [omg] and Kokama [cod] constitute interesting examples of heavy language contact in Amazonia. This is evident from their lexicon, which is mostly Tupí–Guaraní, but with a high percentage of non-Tupí–Guaraní forms, and the grammar, which is very distinct from other Tupí–Guaraní languages. The lexifying Tupí–Guaraní language in this contact situation is believed to be a language similar to Tupinambá [tpn], now extinct, but well-known from 16th century Jesuit grammars and texts. The circumstances which yielded the contact situation between the ancestral language of Omagua and Kokama and the non-Tupí–Guaraní language(s) are not widely known. Nor have the non-Tupí–Guaraní language(s) so far been identified. This thesis compares the phonology of Omagua and Kokama with their closest relative Tupinambá, and reconstructs the phonology of their most recent common ancestor, Proto-Omagua–Kokama–Tupinambá. In doing this, the thesis identifies which phonological changes were involved in the genesis of Omagua and Kokama, and what we can infer about the phonologies of the non-Tupí–Guaraní languages involved in the contact situation. This is of interest to the field of contact linguistics, as examples of contact languages of pre-Columbian origin in the Americas are rare.

Keywords: *Omagua, Kokama, Tupinambá, Tupí–Guaraní, historical linguistics, comparative linguistics, contact linguistics, phonology, creole languages*

Sammanfattning på svenska

Tupí-guaranígspråken omagua [omg] och kokama [cod] utgör intressanta exempel på omfattande språkkontakt i Amazonas. Detta framgår av språkens lexikon, som till största delen består av ord från tupí-guaranígspråk, men med en hög andel ord från obesläktade språk. Det framgår också av grammatiken, som skiljer sig mycket från andra tupí-guaranígspråk. Det språk som var huvudsaklig lexifierare i kontaktsituationen anses vara ett språk mycket likt tupinambá [tpn], som nu är utdött, men som finns väldokumenterat genom jesuitiska grammatikor och texter från 1500-talet. De omständigheter som gav upphov till kontaktsituationen mellan urspråket till omagua och kokama och icke-tupí-guaranígspråken är inte helt kända. Inte heller har man hittills lyckats identifiera vilka obesläktade språk det rör sig om. Denna uppsats jämför fonologin i omagua och kokama med deras närmaste

släkting tupinambá, och rekonstruerar fonologin till deras senast gemensamma förfader, proto-omagua-kokama-tupinambá. Genom att göra detta identifierar också studien vilka fonologiska förändringar som var inblandade i uppkomsten av omagua och kokama, och vad vi kan dra för slutsatser om fonologin i de obesläktade språk som var inblandade i kontaktsituationen. Detta är av intresse för forskningen inom kontaktlingvistik, eftersom kontaktspråk av förkolumbianskt ursprung i Amerika är ovanliga.

Nyckelord: *omagua, kokama, tupinambá, tupí-guaraní, historisk lingvistik, komparativ lingvistik, kontaktlingvistik, fonologi, kreolspråk*

Resumen en español

Las lenguas omagua [omg] y kokama [cod] de la familia tupí-guaraní constituyen ejemplos interesantes de contacto lingüístico en Amazonia. Esto es evidente de su léxico que es mayormente tupí-guaraní, con un gran porcentaje de formas no-tupí-guaraní, y la gramática, que es muy distinta de otras lenguas tupí-guaraní. Se cree que la lengua tupí-guaraní lexificadora en esta situación de contacto es parecida a la lengua tupinambá [tpn] que ya se ha extinguido pero que es bien documentada en textos jesuitas. Las circunstancias que cedieron a contacto entre el idioma ancestral de omagua y kokama y los idiomas no-tupí-guaraní no son bien conocidas. Tampoco se han identificado el/los idioma(s) no-tupí-guaraní. Esta tesis compara la fonología de omagua y kokama con su lengua más estrechamente relacionada, el tupinambá, y reconstruye la fonología de su ancestro común más reciente, el proto-omagua-kokama-tupinambá. Al hacer esto, la tesis identifica cuáles cambios fonológicos fueron involucrados en la creación de omagua y kokama y qué se puede inferir sobre los sistemas fonológicos de la lenguas no-tupí-guaraní que fueron involucradas en el contacto. Esto es de interés para el campo de la lingüística de contacto, ya que no existen muchos ejemplos de lenguas de contacto de origen precolombino en las Américas.

Palabras claves: *omagua, kokama, tupinambá, tupí–guaraní, lingüística histórica, lingüística contrastiva, lingüística de contacto, fonología, lenguas criollas*

Acknowledgements

First of all, I would like to thank my supervisors Gerd Carling and Lev Michael for guiding me through this thesis. I am very grateful that Lev eagerly introduced me to Amazonian linguistics during my time at UC Berkeley, and for his time and enormous patience throughout the project. I would also like to thank Eric Chen at UC Berkeley for greatly helping me with LingPy and EDICTOR, and for kindly explaining the technical bits to me.

I also wish to thank everyone who provided the original data used for this thesis, without whom this work would not be possible. These include the Omagua consultants Amelia Huanaquiri, Arnaldo Huanaquiri, Alicia Huanío and Lino Huanío, as well as UC Berkeley graduate students Zachary O'Hagan, Clare Sandy, Tammy Stark, and Vivian Wauters.

Writing an MA thesis turned out to be very heavy work, so I would also like to thank Maja Svärd at Psykoterapimottagningen and Anna Kettisen at Studieverkstaden for listening to my problems and for helping me overcome them and plan my work.

I am also grateful for having parents who are so very supportive of my studies. Not only have they never questioned my academic choices, they also always take a great interest in whatever I do, regardless of how niche it is.

Lastly, I am forever grateful to everyone who listened to my complaints and supported me when my morale was at its lowest, and who provided me with valuable advice. I could mention dozens of people, but thanks especially to Vendela Källmark, Therese Do, Verónica Giraldo, Linea Vergod Almgren, Sandra Cronhamn, Victor Bogren Svensson, Simone Morehed, Claudia Valdivia, Anna Stålmark, Daniel Kraft, and many others! Thank you Claudia for translating my abstract into Spanish as well!

Till min mormor Birgitta Annertz
(1931–2018)

Contents

List of Tables

List of Figures

List of Abbreviations

1	first person
2	second person
3	third person
A	agent
ABS	absolutive
AGT	agent
COR	coronal segment
COREF	coreferential
CPL	completive DER
derivative	
ERG	ergative
EXCL	exclusive
F.	female
GER	gerundive
INCL	inclusive
INTR.	intransitive
INSTR	Instrumental
IPA	International Phonetic Alphabet
KK	Kokama(–Kukamiria)
LOC	locative
M.	male
NC	nuclear case
NOMZ	nominalizer
OBTOP	oblique-topicalized
OMG	Omagua
P	patient PTG
Proto-Tupí–Guaraní	
PMAG	Proto-Mawéti–Guaraní
POK	Proto-Omagua–Kokama
POKT	Proto-Omagua–Kokama–Tupinambá
REL	relativizer
SG	singular
sp.	species
TPN	Tupinambá
TR.	intransitive

List of Symbols

For definitions of phonetic symbols, the reader is referred to the official full chart of the International Phonetic Alphabet (Kiel, 2015) `[link]`.

*	Reconstructed form
†	Non-occurring form
∅	Null, nothing
σ	Syllable
√	Root word
#	Word boundary
_	Location of a given sound in a phonological context (e.g. #_= word-initially)
/	In the phonological context of (in phonological rule notation)
~	Is in variation with
>	Turned into
<	Came from
/…/	Phonemic representation
[…]	Phonetic representation
⟨…⟩	Orthographic representation
C	Consonant
G	Glide, semivowel
N	Nasal consonant
P	Plosive (stop) consonant
^{N}P	Pre-nasalized stop consonant
V	Vowel

Chapter 1

Introduction and research questions

The languages Omagua and Kokama of Peruvian and Brazilian Amazonia have traditionally been considered members of the Tupí–Guaraní language family, a major branch of the Tupian language family of South America (e.g. Rodrigues, 1958a; Lemle, 1971). However, later work have shown that these languages have significant non-Tupí–Guaraní influence, and arose in an intense contact situation with a non-Tupí–Guaraní language or languages (Rodrigues, 1985; Cabral, 1995; Michael, 2014). This is evident from the lexicon, which is mostly Tupí–Guaraní, but with a high percentage of non-Tupí–Guaraní forms, and the grammar, which is very distinct from other Tupí–Guaraní languages. The lexifying Tupí–Guaraní language in this contact situation is believed to be a language similar to Tupinambá, a language now extinct, but well-known from 16th century Jesuit grammars and texts (Cabral, 1995). Indeed, a lexical phylogenetic study showed Tupinambá to be the language closest to Omagua and Kokama (Michael et al., 2015).

The circumstances which yielded the contact situation between Proto-Omagua–Kokama (POK), the ancestral language of Omagua and Kokama, and the non-Tupí–Guaraní language(s) are not widely known. Cabral & Rodrigues (2003) suggest that this contact situation arose in Jesuit mission settlements in the late 17th century or early 18th century, whereas Michael (2014) argues that it must have taken place much earlier in pre-Columbian times.

This Master's thesis is an exploratory study which aims to provide greater insight into the genesis of Proto-Omagua–Kokama by comparing the phonologies of Omagua and Kokama

with Tupinambá, using the Comparative Method (e.g. Weiss, 2014) in order to reconstruct the phonology of Proto-Omagua–Kokama–Tupinambá (POKT). This is of great interest as languages with heavy contact-induced restructuring in the Americas where the principal languages are indigenous, and whose origin can be dated to the pre-Columbian period are rare (Michael, 2014:311).

By identifying the phonological changes involved in the transition from POKT to Proto-Omagua–Kokama, future work will be able to use this phonological profile in order to further identify which substrate languages were involved in the contact situation.

In doing this, the study aims to answer the following research questions:

1. What are the phonological differences between Omagua, Kokama, and Tupinambá?
2. What phonological features can be reconstructed to the ancestral proto-language of these languages?
3. What phonological changes were involved in the genesis of Proto-Omagua–Kokama?
4. What can we infer about the phonologies of languages in contact with Pre-Proto-Omagua–Kokama in the contact situation?

1.1 Outline

Chapter 2 provides the reader with background information on Omagua, Kokama, and Tupinambá, their linguistic background and what previous research has been done on them. This is followed by an outline of their respective phonologies — their phonemic inventories, allophonic variation, stress, and phonological processes, which will be referred to throughout this thesis. This is followed by a section on what common morphology is found in the dataset, in order for the reader to be able to identify differences between cognate sets which are not due to phonological developments.

Chapter 3 describes the data used for this thesis, the origin of the data and how the data have been managed and transcribed. It also describes the process of constructing cognate and correspondence sets, and includes a section on the Comparative Method as well as related concepts that were used for reconstructing the phonology and identifying sound changes.

Chapter 4 lays out the results of the construction of correspondence sets, describing the distribution of each phonological segment and the segmental correspondences between the languages, as well as the reconstructed segment for each set.

Chapter 5 is a discussion of the correspondences of chapter 4, arguing for certain reconstructions and describing certain sound changes and their phonological context in greater detail.

Finally, Chapter 6 is a summary of the thesis and a conclusion of the findings. Following Chapter 6 and the references are two Appendices. Appendix A contains all the cognate sets used in this study, and Appendix B the LingPy code used for building the cognate and correspondence sets, as well as the input file that this algorithm used, and the unedited output that it gave.

Chapter 2

Background

2.1 Language background

2.1.1 Omagua & Kokama

Omagua and Kokama are two closely related languages, spoken in the Department of Loreto in northeastern Peru. Omagua is spoken by 2 elderly speakers as of February 2019 (Sandy & O'Hagan, 2020:1) in the community of San Joaquín de Omaguas, located south of Iquitos on the Amazon River, and is thus highly endangered. Kokama has two main dialects: Kokama, which is spoken along the Marañon, Samiria, Ucayali, and Amazon Rivers, and Kukamiria which is spoken in the upper Huallaga River. Linguistically, the differences between the dialects are phonological and lexical (Vallejos, 2010:28). Kokama is also endangered, with 1000 speakers out of an ethnic population of approximately 20 000. It is no longer transmitted to children, and the remaining speakers are elderly speakers spread across small villages who use the language only in restricted situations. The remaning population has shifted to Spanish (Vallejos, 2010:31-32).

2.1.2 Tupinambá

Tupinambá was the main language spoken along the coast of Brazil by the time of the European arrival in the late 16th century, located mainly around the area of modern Río de Janeiro and northwards. Because of this, the language is well known from sources written by Jesuit missionaries, such as de Anchieta (1595). The Tupinambá covered such a large

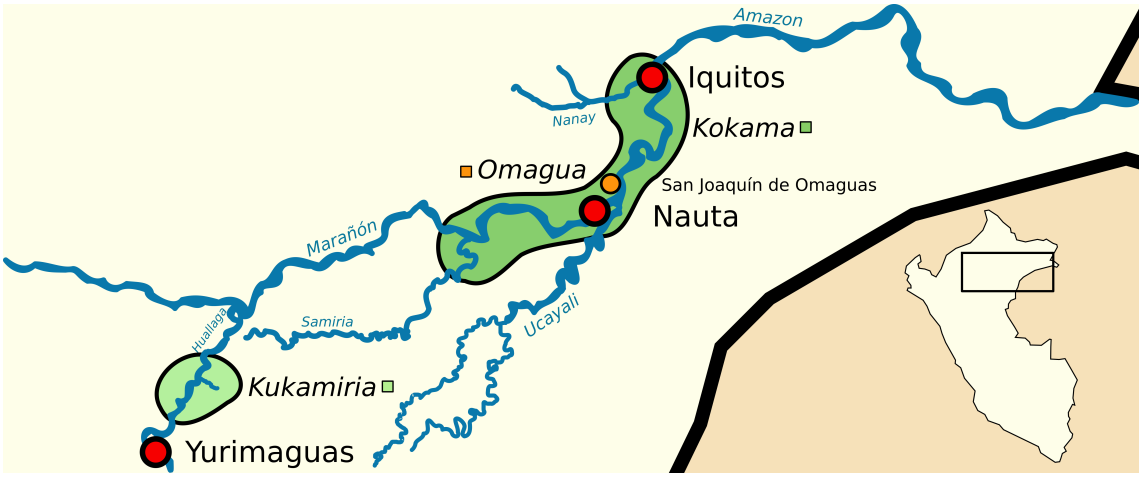

Figure 2.1: Approximate modern range of Kokama–Kukamiria and Omagua in Peru, based on the location of villages visited by Vallejos (2010).

area that the Tupinambá language was referred to as 'Brasílica' or 'Brasiliano' before the 18th century (Jensen, 1999:125).

Tupinambá as such can be considered extinct since the 18th century (Jensen, 1999:125), but in the 16th and 17th centuries, when colonists took Tupinambá-speaking wives, a contact version of Tupinambá called Nheengatú developed out of it, spoken by some 18 300 speakers today (Eberhard et al., 2019).

This thesis focuses only on Tupinambá as described in the older Jesuit sources, and work derived from them, e.g. Barbosa (1956) and Rodrigues (1958b), and does not take modern Nheengatú into consideration.

2.1.3 Earlier history

The Tupí–Guaraní family is spread over large parts of South America. By the time of contact with Europeans, it stretched from Tupinambá, spoken along the eastern coast of Brazil to the Guaranian languages, spoken in southern Amazonia, to Kokama spoken in modern day Peru (Michael et al., 2015). For this reason, the Proto-Tupí–Guaraní (PTG) homeland has been a topic of debate, some arguing for a homeland in southwestern Amazonia, on the Panará River basin, and some for a northeastern origin (O'Hagan et al., 2019:18). Rodrigues (2000) for instance argues for a southwestern origin largely based on the subgrouping of Tupí–Guaraní languages in Rodrigues (1985). According to this subgrouping, two of the three major Tupí–Guaraní branches can be found in southwestern Amazonia, and

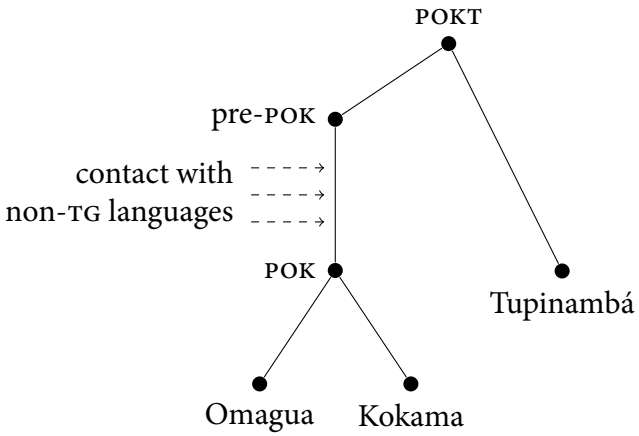

Figure 2.2: Relationship between Omagua, Kokama, Tupinambá, Proto-Omagua–Kokama (POK), and Proto-Omagua–Kokama–Tupinambá (POKT). Adapted from Michael (2017).

the dispersal from the homeland is therefore explained by positing one migration from southwestern to northeastern Amazonia.

O'Hagan et al. (2019), following the subgrouping of Michael et al. (2015), instead places the Proto-Tupí–Guaraní homeland on the lower Xingu River, positing a southward migration of a subgroup ('Southern') from the mouth of the Amazon River, after having split up from (Pre)-Proto-Omagua–Kokama–Tupinambá. According to this view, Proto-Omagua–Kokama–Tupinambá was spoken in the vicinity of the lower Amazon region, and then split into Pre-Proto-Omagua–Kokama and Pre-Tupinambá as Tupinambá spread southwards along the Atlantic coast, and Pre-Proto-Omagua–Kokama spread northwest upriver, arriving in the upper Amazon in approximately 1100 CE (O'Hagan et al., 2019:16). As our sources of Tupinambá stem from the 16th century, this means that there are only 400 years between the genesis of Proto-Omagua–Kokama and Tupinambá. Therefore we do not expect the differences between Tupinambá and Proto-Omagua–Kokama–Tupinambá to be very large (Michael, 2017).

2.2 Previous research

Given the presence of the Tupinambá along the coast of Brazil, there is a somewhat large amount of early documentation in the form of grammars (e.g. de Anchieta, 1595; Figueira, 1687), wordlists, ecclesiastical texts (e.g. de Araujo, 1618), and accounts of the life of the colonists. These sources are mainly Portuguese and French, as well as Dutch and German (Rodrigues, 1958b:7ff).

In the 19th century, important work on Tupinambá includes work by Barbosa (e.g. 1951, 1956, 1970) and Rodrigues (e.g. 1958b, 2010). Being an early documented Tupí–Guaraní language, it has had a central role in historical and comparative work on the Tupí–Guaraní language family, e.g. Rodrigues (1958a, 1985, 2010); Lemle (1971); Jensen (1998); Rodrigues & Cabral (2012).

In older sources such as Rodrigues (1958a) and Lemle (1971), Kokama and Omagua were classified as Tupí–Guaraní languages based on their lexical similarities with known Tupí–Guaraní languages such as Tupinambá or Guaraní. As research on the grammar of Kokama progressed, it was suggested by Rodrigues (1985) that, given its significant grammatical differences, Kokama must have been under influence by a non-Tupí–Guaraní language.[1]

In her doctoral thesis, Cabral (1995) compares Kokama with Tupinambá, the language with which Kokama shows the most lexical similarity, and the most conservative and earliest documented member of the Tupí–Guaraní language family. She compares Kokama and Tupinambá vocabulary, phonology, morphology, and syntax, and concludes that the similarities are restricted to vocabulary and phonology alone, and that Kokama has too few structural features to be classified as Tupí–Guaraní. Whereas Tupinambá is a polysynthetic language with much inflectional and derivational morphology, Kokama is described as an isolating language with no inflectional morphology and very limited derivational morphology. In addition, much of the derivational morphology in Tupinambá is prefixing, whereas in Kokama it is exclusively suffixing (Cabral, 1995:118).

She further concludes that the transmission of a Tupí–Guaraní language was interrupted in the history of Kokama speakers, and that Kokama because of this reason cannot be classified genetically at all (Cabral, 1995:5). She proposes that Kokama arose due to "imperfect learning in a process of language shift" in a contact situation between a Tupinambá-like language and other indigenous languages (Cabral, 1995:308). According to Cabral, this contact situation arose in *reducciones*, missionary settlements, in the 17th and 18th centuries, when speakers of indigenous languages from several language families rapidly had to learn a common language as a medium for communication. According to this hypothesis, this language was the Tupí–Guaraní language spoken by the Omagua and the Kokama, having

[1] "Como o Kokáma apresenta certas propriedades importantes não Tupí, dá a impressão de tratar-se de mais um casco de língua Tupi-Guarani adotada por um povo não Tupi." [The fact that Kokáma has certain important properties which are non-Tupí gives the impression that we are dealing with one more case of a Tupí–Guaraní language adopted by a non-Tupí people.] (translation by Jensen (1998:496))

the status of an official language within the *reducciones*. This language was then passed onto the children born within the *reducciones* as a native language (Cabral, 1995:309f).

Based on 17th and 18th missionary chronicles, older Omagua and Kokama texts, and modern data, Michael (2014) argues that the contact situation which gave rise to Proto-Omagua–Kokama must have occurred much earlier than the 17th century, before the Jesuit arrival, making Proto-Omagua–Kokama a pre-Columbian contact language. For instance, he shows that the Jesuit *reducciones* among the Omagua were single-ethnicity *reducciones* until the mid 1720s (Michael, 2014:331). He further writes that according to the hypothesis that Omagua–Kokama emerged in *reducciones*, Omagua has to have served as a *lingua franca*, which was learned by speakers of other languages. However, historical documents show that the Jesuits tried to establish Quechua as a lingua franca, rather than Omagua (Michael, 2014:326ff). In addition to this, the hypothesis that creoles emerge as a rapid imperfect learning of a foreign language due to restricted contact with native speakers is not supported by evidence presented by Michael, which shows that the Omagua were not outnumbered in the *reducciones*. Hence, he believes that there is no reason to believe that non-Omagua had limited access to Omagua (Michael, 2014:333f).

On the linguistic side, there is an early attestation of Kokama found in a letter by Lucero in one of the *reducciones* dated 3 June, 1681, an utterance which is more or less identical to modern Kokama. Michael writes that this poses difficulties for the *reducción* hypothesis, as the first multiethnic *reducción* had only existed for ten years when the utterance was produced. Hence, if one argues that Kokama emerged in the *reducciones*, it must have emerged in ten years at most (Michael, 2014:334f). There was also a continuous production of texts and descriptions of Omagua during the Jesuit contact with the Omagua, yet there are no remarks that the language changed rapidly, which the hypothesis suggests. In addition, according to this hypothesis, Omagua and Kokama emerged in different *reducciones*, with a week travel time from each other, yet the languages are very similar, so it is unlikely that two different social and multilingual contexts would produce such similar languages (Michael, 2014:337).

Michael (2017) proposes an alternative hypothesis that members of the Pre-Proto-Omagua-Kokama society incorporated a large numbers of captives from neighboring groups, resulting in creolization of Pre-Proto-Omagua–Kokama as non-Tupí–Guaraní speaking captives

eventually outnumbered the speakers of Pre-Proto-Omagua–Kokama, giving rise to Proto-Omagua–Kokama.

2.2.1 The creole origin of Omagua & Kokama

The conclusion reached by Cabral (1995) is that Kokama (and Omagua) are unrelated to other Tupí–Guaraní languages, and cannot be genetically classified, because of the interruption in transmission, and because of its phonology, morphology, and syntax, which differ greatly from those of Tupí–Guaraní languages. Michael (2017) however, describes Proto-Omagua–Kokama as a creole language, and considers it genetically related to Tupí–Guaraní. The question of whether Omagua and Kokama are Tupí–Guaraní languages therefore boils down to whether one considers creole languages to be genetically related to their lexifiers or not.

Creoles have traditionally been considered a distinct class of languages different from non-creoles, a view sometimes called *creole exceptionalism*. According to this view, creoles are exceptional in their structure, they develop in an exceptional acquisitional environment, and they differ from languages that emerged long ago (Ansaldo & Matthews, 2007:4). For instance, McWhorter (1998, 2001) identifies three features which will identify a language as a creole. These are (1) minimal use of inflection, (2) lack of tone, (3) semantically regular derivation.

With regards to creole genesis, creoles are often said to result from a 'break in transmission' of their lexifiers, i.e. the language from which the majority of the vocabulary in the creole is derived. Thomason & Kaufman (1988) distinguish between 'normal transmission' and 'abnormal transmission'. In the former, language is passed on from older members to younger members in the community with little change over time, but the entire language is transmitted. This may over time lead to a completely mutually unintelligible language, but these two languages can be said to have a 'genetic relationship' as transmission was normal.

In some cases of heavy contact, an entire population can acquire a new language within a very short amount of time, by other means than from parents or peer-group members, which causes the acquired language to have interference from the original language of the community, that is to say, the language was not perfectly transmitted (Thomason &

Kaufman, 1988:10). There is not a regular and systematic correspondence between the subsystem of the new language and the old language, and they can therefore not be said to have a genetic relationship, a process called 'abrupt creolization' (Thomason & Kaufman, 1988:11).

Other researchers, e.g. Mufwene (2001) and DeGraff (2003) instead believe that creoles developed gradually without a break in transmission, and that the development seen in creoles can be explained through usual processes of language evolution, except that second-language acquisition plays a significant role (Siegel, 2007:174). In this view, creoles can be said to lie "towards one end of a spectrum of languages exhibiting strong language contact effects" (Michael, 2017).

2.3 Morphology

Before describing the method used in this thesis, some background information on the morphology and phonology of the languages of study is needed. This section will describe the morphology of Omagua, Kokama, and Tupinambá with focus on the relevant morphology found in the data used for this thesis.

2.3.1 Omagua & Kokama morphology

Omagua and Kokama roots are characterized by a high amount of frozen morphology, i.e. historically morphologically complex words which have been reanalyzed as monomorphemic roots, a process quite common in creole languages (Crowley, 2008:90). That is, many roots in Omagua and Kokama contain segments which were historically affixes in Tupí–Guaraní (O'Hagan, 2011). This frozen morphology consists of a number of former verbal agreement prefixes, specifically a subset of the ergative and absolutive paradigms, namely the Proto-Omagua–Kokama–Tupinambá reflexes of the Proto-Tupí–Guaraní prefixes 1SG.ERG **a-*, 1PL.INCL.ERG **ja-*, 3.ERG **o-*, and 3.ABS **i-*, **ts-*, **t-* (O'Hagan, 2011:27).

According to O'Hagan (2011), the particular prefix frozen depends on the transitivity and semantic status of the prefix, and whether the prefix is bound vs. non-bound. It also depends on "the discourse frequency of particular referents" and "the event semantics of the verb" (O'Hagan, 2011:18).

Examples of these frozen prefixes can be seen in examples (1-2) from Proto-Omagua–Kokama and their origin in Proto-Tupí–Guaraní:

(1) apuka < *a-puka
laugh < 1SG.ERG-laugh
'to laugh' (O'Hagan, 2011:28)

(2) saku < *ts-akuβ
be.hot < 3.ABS-be.hot
'be hot' (O'Hagan, 2011:34)

Furthermore, there are also roots in Omagua–Kokama ending in *-a*, which were consonant-final in Proto-Tupí–Guaraní. These are considered to have frozen suffixes such as the nuclear case *-a* or the gerundive *-a* (Michael, 2017). The origin of these suffixes is discussed in subsection 2.3.2, but its origin is exemplified in (3).

(3) jatɨma < *ja-tɨm-a
plant < 1PL.INCL.ERG-plant-GER
'to plant' (Michael, 2017; O'Hagan, 2011:28f)

2.3.2 Tupinambá morphology

The Tupinambá lemmas that appear in the data and in dictionaries are typically not the bare root, but usually appear with a suffix. Presumably, when words were written down, they appeared in certain grammatical contexts, which often caused a root to appear together with a common suffix. In Proto-Tupí–Guaraní and in Tupinambá, roots can function both as nouns and verbs, for example in Proto-Tupí–Guaraní[2]:

(4) (a) *i-memɨ́ɾ-a 'her child'
(b) *i-memɨ́ɾ 'she gave birth'
(Jensen, 1999:149)

When a root is used syntactically as a noun, i.e. has the function of e.g. a subject or an object in a sentence, the root receives the case suffix *-a* as in (4a), often called *nuclear case* (NC) (alternatively called *nominal case*). This suffix only appears on roots that are consonant-final, and vowel-final roots are instead analyzed as having a zero-allomorph -∅ (Jensen, 1998:505).

[2]For consistency, the Americanist notation in the original sources has been converted into IPA in this section.

This is further illustrated in (5) from Tupinambá, where 'child' and 'woman' function as nouns, receiving the nuclear case, realized in both its allomorphs.

(5) kujã-∅ o-s-aɾõ o-memɨ́ɾ-a s-eɾekó-βo
woman-NC 3A-3P-care.for 3COREF-child-NC 3P-keep.with-GER
'The woman cares for her child, keeping it with her.' (Jensen, 1998:506)

A root does not receive the nuclear case, if it appears in a context where it does not function syntactically like a noun. Examples of such contexts are for instance the vocative or in circumstances where a noun can receive verbal morphology and function as a verb.

The lemmas in Tupinambá which are verbs, often appear with a final *-a* suffix as well. Since the nuclear case does not appear on roots when used predicatively, this suffix is what is named *gerundive*[3] by Rodrigues (2010:13). This suffix is used when multiple verbs share the same subject, and is common in constructions of position, movement, and direction (Jensen, 1998:529-531). Because of this, it is called a 'serial verb suffix' by Jensen (1998), shown in (6) from Tupinambá.[4]

(6) o-úɾ kunumí kuáp-a
3-come boy know-GER
'He came to meet the boy.' (Jensen, 1998:530)

The nuclear case and gerundive suffixes do not carry stress. This means that while Tupinambá has ultimate stress, lemmas with an *-a* suffix will have penultimate stress. This creates a stress contrast between words with a nuclear case suffix and a root-final *-a*, as in Tupinambá /ɨβ-a/ [ˈɨβa] 'stalk' vs. /ɨβa/ [ɨˈβa] 'fruit' (O'Hagan, 2013:3).

2.3.3 'Relational prefixes'

A feature of all Tupí–Guaraní languages is the so called relational prefix or r-prefix which in Proto-Tupí–Guaraní is described by Jensen (1998:501) as a "linking morpheme" occurring on 1) nouns preceded by a possessor, 2) a postposition preceded by its object, or 3) a verb preceded by a noun, see Table 2.1.

In Proto-Tupí–Guaraní, stems can be divided into three broad classes, stems which do not

[3] *Gerúndio* in Portuguese.

[4] Note that serial verb constructions per definition do not contain any markers of coordination or subordination (Aikhenvald, 2018:51), which is why *gerundive* is used here instead.

take a relational prefix (Class I), stems which do (Class II) and stems which are not inflected (Class III).

	Class I (without ɾ-)	**Class II (with ɾ-)**
Nouns	*tʃé tʃɨ ‘my mother’	*tʃé ɾ-úβ ‘my father’
Tr. verbs	*tʃé pɨtsɨk ‘grab me’	*tʃé ɾ-ekár ‘seek me’
Intr. verbs	*tʃé katú ‘I am good’	*tʃé ɾ-atsɨ ‘I hurt’
Postposition	*tʃé tsupé ‘for me’	*tʃé ɾ-etsé ‘with respect to me’

Table 2.1: The relational prefix in two PTG word classes (Jensen, 1998:499).

Jensen (1998:502) writes that there is a lack of uniformity in how the relational prefix is described, including its allomorphy. In her view, certain authors conflate the relational prefix **ɾ-* with the third person markers (**t-*, **ts-*, **i-* and **∅-*) treating them as allomorphs of the relational prefix.

	Class I	**Class II**
1st p.	*tʃé kó ‘my garden’	*tʃé ɾ-etʃá ‘my eye’
3rd p.	*i-ko ‘his garden’	*ts-etʃá ‘his eye’

Table 2.2: PTG relational prefix stems in 1st vs. 3rd person.

In Jensen’s view, shown in Table 2.2, the relational prefix does not occur in the third person since it is marked by a prefix, unlike the first or second person (Jensen, 1998:503).

In the view of Meira & Drude (2013), Jensen’s third person prefixes are actually also relational prefixes which occur when the possessor (as in this case) is not present, or present outside of the phrase. According to this view, Tupí–Guaraní languages do not have third person grammatical markers (Meira & Drude, 2013:2).

This is the view of Rodrigues & Cabral (2012) as well, who propose four relational prefixes for Proto-Tupí–Guaraní, the latter two of which correspond to the coreferential possessive markers and the marker for a human possessor of Jensen (1998).

Regarding the function of the relational prefix, Jensen (1998:557-559) gives several proposals, namely that it is (1) an epenthetic consonant, (2) a grammatical element, or (3) a phonologically conditioned morpheme. She discards hypothesis (1) and notes that “a rule of epenthesis would actually conflict with other phonological rules which operate at this level.” (Jensen, 1998:558) According to hypothesis (2), the relational prefix would be a grammatical element which shows “a grammatical relationship between the stem to which it attaches and the preceding morpheme.” She quotes Rodrigues (p.c.) saying that the

relational prefix only occurs "when the preceding morpheme is a separate word", meaning that it is a "phrase-level phenomenon." According to hypothesis (3), the relational prefix was originally phonologically conditioned, occurring at word boundaries when the following word was vowel-initial. In Proto-Tupí–Guaraní however, certain vowel-initial stems, such as **akáŋ* 'head' belonged to Class I, i.e. did not take the relational prefix. As an explanation to this she suggests that in Pre-Proto-Tupí–Guaraní, all vowel-initial stems belonging to Class I were originally consonant-initial (**i-C-akáŋ* > **i-akáŋ* 'his head').

Meira & Drude (2013) argue for a fourth hypothesis, namely that the relational prefix was originally part of the root, but underwent sound changes and reanalysis prior to the genesis of Proto-Tupí–Guaraní. By comparing paradigms in Proto-Tupí–Guaraní with the languages Awetí and Mawé outside the Tupí–Guaraní branch, they reconstruct contexts showing relational prefixes to Proto-Mawéti–Guaraní (PMAG), the ancestral language of Mawé, Awetí, and Proto-Tupí–Guaraní. As shown in Table 2.3, they reconstruct this original consonant as *T, symbolizing an alveopalatal consonant, perhaps /tʲ/, which went to Proto-Tupí–Guaraní *t by dissimilation and to *ɾ and *ts by lenition (Meira & Drude, 2013:14).

Rodrigues & Cabral (2012) reconstruct three relational prefixes to Proto-Tupían, the ancestral language of Proto-Tupí–Guaraní. In the view of Meira & Drude (2013) however, there would be no relational prefixes in PMAG, and obviously then neither in Proto-Tupían.

In the view of Meira & Drude (2013), the relational prefix in Tupí–Guaraní is more like consonant mutations in Celtic, a series of conditioned sound changes affecting initial consonants which through subsequent changes became grammaticalized (Meira & Drude, 2013:25). In the section for future research, Meira & Drude (2013:26) note that the analysis of these consonants as prefixes have led researchers to segment off these consonants from the root, comparing roots without their original initial consonant, and thereby obfuscating the patterns of alternation. They finally call upon the field to collect more accurate and comprehensive data, including paradigms and irregularities in order to facilitate comparative work within Tupí languages.

In the data used for this thesis, this means that the original relational prefix will be variably present in the data. Certain sets with an initial consonant found frozen to the root in Omagua– Kokama, have a vowel-initial cognate in Tupinambá, e.g. the set VILLAGE∼LAND, Omagua–Kokama *ɾitama*, Tupinambá *etama*, with an initial original relational prefix *ɾ-* in

PMAG		Mawé	Awetí	PTG	Additionally
iT	>	h	t	*ts	
jT	>	h	t	*t	*βT > PTG *p
[*T…]	>	s	t	*t	*rT > PTG *∅
[*NP T…]	>	s	∅	*r	*(p, t, k)T > Awetí p, t, k (lenition)
otherwise: *T	>	s	∅	*∅	*(m, n, ŋ)T > Awetí mp, nt, ŋk (fortition)

Table 2.3: Reflexes of PMAG *T (Meira & Drude, 2013:20).

Omagua–Kokama, but with the relational prefix not present in the Tupinambá word.

In these cases, the relational prefix has been added to the Tupinambá cognate, i.e. *etama* > *tetama*. This was done using Barbosa (1970) as a source, who lists the relational prefix in parenthesis after the lemma, e.g.: **terra** *etama (t)*. In other cases where the relational prefix was not frozen in Omagua and Kokama, this addition was not done for Tupinambá in order to facilitate segment comparison across the languages.

Lastly, this means that certain words have what looks like somewhat irregular correspondence patterns for initial consonants, e.g. *ɾ* : *ɾ* : *t* in the set GRANDCHILD (F. EGO): Omagua–Kokama *rimiariru*, Tupinambá *tembiarirõ*, but which are only different relational prefixes.

This addition was not done for e.g. *ts-* as it was synchronically a third person marker in Tupinambá, regardless of its original status as a initial consonant vis-à-vis morpheme in Pre-Proto-Tupí–Guaraní.

2.4 Phonology

The following section is a brief exposition of the segmental inventory of Omagua, Kokama, and Tupinambá, in order to serve as background for the chapter on phonological reconstruction.

2.4.1 Omagua & Kokama phonology

2.4.1.1 Vowels

The vowel inventory of Omagua and Kokama are shown in Figure 2.3 and consists of five vowel phonemes. The vowel phoneme /e/ only exists in Kokama, and the corresponding phoneme in Omagua is /ɪ/. The vowel phoneme /e/ in Kokama, is described by Vallejos

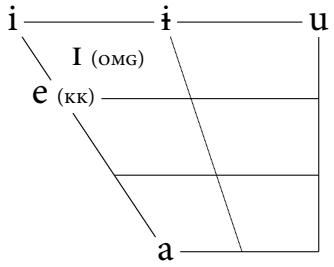

Figure 2.3: Omagua (OMG) and Kokama (KK) vowel phonemes. Unless shown in parentheses, a vowel is present in both languages.

(2010:51) as "higher and more centralized than a Spanish /e/." There is also an allophonic overlap between the phonemes /i/, /e/ and /ɪ/ in Kokama. For example, high vowels may be opened slightly word-finally, so that /nami/ 'ear' and /itimu/ 'die' may be realized as [namɪ] and [itimʊ], or even to a mid-high realization: [itimo]. For /i/, this may only occur after glides, so that /tsuwi/ 'tail' is opened to [tsuwe] (Vallejos, 2010:54).

Vallejos (2010) also writes that Kokama displays further allophonic variation in final position, and /a/ may be realized as [e∼ə], /e/ as [ɪ∼ə], and [ɨ] as [i]. Furthermore, Kokama displays further reduction of vowels medially, where some vowels may be deleted altogether in antepenultimate pretonal position, so that /japukɨ́ta/ 'paddle' is realized as [zap.kɨ.ta] (Vallejos, 2010:54).[5]

Omagua also show the mid-high vowels [e] and [o] allophonically. In Omagua's case, they surface "as the result of optional processes of assimilation or coalescence of adjacent vowels" (Sandy & O'Hagan, 2020:109), whereby [o] can surface as an allophone of /u/ in the sequences /ua/ or /au/, and [e] can surface as an the result of coalescence of the sequences /ai/ and /aɪ/ (Sandy & O'Hagan, 2020:109).

As described in Sandy & O'Hagan (2020), Omagua shows vowel nasality in six words, which according to their analysis is due to an unspecified nasal consonant /N/ which is deleted intervocalically or between a vowel and a glide, and instead surfaces as heavy nasalization on the preceding vowel and optionally on the following glide and vowel, e.g. /aNjá/ [ˈãjã] 'thus', /sɨNi/ [sɨ̃j] 'sweat', /suNi/ [sũj] 'tail' (Sandy & O'Hagan, 2020:106-109).

[5] Vallejos (2010) writes /z/ on p. 54 in the phonemic representation of this word (among others). This is presumably a mistake, as she lists the phoneme as /j/ in the consonant inventory, and writes "The alveolopalatal approximant /j/ is realized as alveolar fricative [z] word initially and in intervocalic position. Note that this process is optional but highly frequent, and for the majority of speakers the sounds [z] and [j] are in free variation." (Vallejos, 2010:45)

In both languages, the diphthong /aɨ/ is syllabified together, which could be described as the only phonemic diphthong (Vallejos, 2010:59-60; Sandy & O'Hagan, 2020:112).

2.4.1.2 Consonants

The consonant inventories of Omagua and Kokama are shown in Table 2.4 and consist of eleven phonemes with some differences between them.

	Bilabial	**Alveolar**	**Post-alv.**	**Palatal**	**Velar**
Stop	p	t			k, k^w (OMG)
Nasal	m	n			
Fricative		s (OMG)	ʃ (OMG)		x (KK)
Affricate		ts (KK)	tʃ (KK)		
Glides	w			j	
Tap		ɾ			

Table 2.4: Omagua (OMG) and Kokama (KK) consonant phonemes. Unless shown in parentheses, a consonant is present in both languages.

Stops Kokama has three stop phonemes: /p, t, k/. Omagua has the same consonants, and in addition to these, a labialized velar stop /k^w/ (Sandy & O'Hagan, 2020:99). In both Omagua and Kokama, these become voiced when following nasal consonant as part of regressive assimilation: Kokama /kunpetsa/ [kumbetsa] 'turtle sp.' (Vallejos, 2010:43); Omagua /iNdata/ [indata] 'bother' (Sandy & O'Hagan, 2020:99-101).

Sandy & O'Hagan (2020) choose to posit a /k^w/ phoneme. This is based on a number of observations. The first one is that both [kwa] and [ku.a] are possible Omagua sequences, and [kwa] can be a surface form of /kua/ in certain environments. There are however a number of /Cw/ words, where this alternation does not occur. Most of these words are instances of [kw], and very few of other consonant combinations and [w]. Based on this asymmetry, they choose to posit /k^w/ as its own phoneme, as well as on the fact that /k^w/ is a common areal feature (Sandy & O'Hagan, 2020:114).

In the analysis of Vallejos (2010), the corresponding phoneme to Omagua /k^w/ is a sequence of /k/ + /w/. She posits the syllable CCV where the second consonant is a glide, and the first consonant is /p/, /k/, /n/, or /ɾ/, but not /t/. As opposed to Sandy & O'Hagan (2020), she does not mention a difference in frequency between the syllable types.

This analysis difference means that e.g. the item 'sun' is phonologically analyzed in Kokama as

/kwaratʃi/ (Vallejos, 2010:50), but in Omagua as /kʷaraʃi/ (Sandy & O'Hagan, 2020:113f).

Nasals Both Kokama and Omagua have two phonemic nasals, /m/ and /n/. In both languages, these nasals undergo place assimilation before stops, so that the phonemic distinction is neutralized in this position. For this, both Vallejos (2010) and Sandy & O'Hagan (2020) use a capital unspecified nasal /ɴ/ in their phonemic analysis, e.g. Kokama /inaɴpika/ [inambika] 'careful not to' and Omagua /tʃuɴka/ [ˈtʃuŋga] 'ten'.

Moreover, in both languages there is an allophonic velar nasal [ŋ] which occurs in coda position. Sandy & O'Hagan (2020) find no evidence which phonemic nasal this would be derived from and use the unspecified nasal /ɴ/ in their analysis, e.g. /paɴ/ [paŋ] 'be rotten'. Vallejos (2010) analyzes the underlying nasal as the alveolar nasal /n/ based on the surface form in related words in non-coda position, e.g. /miʃan/ [miʃaŋ] 'small', related to /miʃananin/ 'the small one' (Vallejos, 2010:48).

In addition to this, /n/ is optionally palatalized to [ɲ] before /j/ in both Kokama and Omagua (Vallejos, 2010:47; Sandy & O'Hagan, 2020:103).

Affricates and fricatives Kokama has two affricates /ts/ and /tʃ/, where Omagua has corresponding fricatives /s/ and /ʃ/. The affricates are also found in Omagua, but are marginal phonemes resulting in borrowings from Kokama or Quechua (Sandy & O'Hagan, 2020:102). Moreover, the Kokama alveolar affricate is frequently lenited to [s] preceding non-high vowels, e.g. /tsetsa/ [tsetsa]∼[səsa] 'flower'. Similarly, the sound [ʃ] only shows up in a handful words, which is postulated as diachronic lenition as well (Vallejos, 2010:46,49). The sound [s] also shows up in borrowings from Quechua or Spanish (Vallejos, 2010:46).

In Kokama, /ts/ is also frequently palatalized before /i/: /tsitsa/ [tsitsa]∼[tʃitsa] 'face' (Vallejos, 2010:46-47). However, the corresponding process is not reported for Omagua (Sandy & O'Hagan, 2020).

Finally, there is also an infrequent velar fricative [x] in Kokama which only shows up in one word, the demonstrative pronoun /axan/ 'this' and its derivatives. It is a characteristic of female speech and has been hypothesized to be a borrowing from an unknown source language (Vallejos, 2010:49).

Glides and liquids Both Kokama and Omagua have two glides, a bilabial /w/ and a palatal /j/. In Kokama, /w/ may undergo fortition to [β] intervocalically before /i, e/: /tewe/ [teβe] 'salt'. Similarly, the palatal glide /j/ frequently undergoes fortition to [z] initially and intervocalically: /juwa/ [zuwa] 'thorn' (Vallejos, 2010:45). Such processes are not reported for Omagua in Sandy & O'Hagan (2020).

It should be noted that [w] and [j] also occur in Omagua as allophones of /u/ and /i/ in certain VV sequences, specifically falling sequences, e.g. /au/ [aw] and /ai/ [aj], and in level sequences, e.g. /ɪw/ [ɪw] and /ui/ [uj]. In medial position, the same process also affects rising vowel sequences, so that /ikua/ 'know' is [i.ˈku.a], but /ikua-pa/ 'know-CPL' is [i.ˈkwa.pa]. In medial position, /ui/ is also variably realized as [wi] or [uj], e.g. /amui/ [amuj] 'grandfather', but /amui=na/ [a.ˈmwi.na]∼[a.ˈmuj.na] (Sandy & O'Hagan, 2020:112-113).

Kokama follows similar patterns, but VV sequences remain two different syllables if the stress is associated with a high vowel, e.g. [ta.na.u.ka] 'our house', cf. Omagua /wau-pa/ [ˈwaw.pa] 'have scabies-CPL' (Vallejos, 2010:58). As opposed to Sandy & O'Hagan (2020), Vallejos (2010) considers coda-final glides to be phonemic (Vallejos, 2010:57-58).

The only phonemic liquid in both languages is the alveolar tap /ɾ/. However, in Kokama, a common allophone of the tap is the alveolar lateral [l], which occurs, especially in the Kukamiria dialect and especially among women. According to Vallejos (2010), this might formerly have been a characteristic of the Kukamiria dialect, which has shifted towards [ɾ] because of influence of the Kokama dialect (Vallejos, 2010:48).

2.4.1.3 Stress

Generally, the stress in both Omagua and Kokama falls on the penultimate syllable, e.g. Kokama [pa.ˈna.ɾa] 'banana', Omagua [ta.ˈpa.ka] 'piranha sp.'. The exception to this in Kokama is when a stress-bearing morpheme is attached, in which case the stress becomes final: /panara-pan/ [pa.na.ɾa.ˈpan] 'banana-DER' (Vallejos, 2010:62-66). The exception to penultimate stress in Omagua is if the word has a nasal coda, in which case it attracts stress: /saipura=ɴ/ [saj.pu.ˈraŋ] 'be.drunk=REL' (Sandy & O'Hagan, 2020:121).

2.4.1.4 Phonotactics

Syllables in Omagua–Kokama follow a (C)(C)V(C) pattern where the second consonant is typically a glide /j, w/, or in Omagua sometimes a tap /ɾ/ as a result of vowel syncope (Sandy & O'Hagan, 2020:109). The coda consonant in Kokama can only be a glide or the tap /ɾ/ or a nasal [ŋ] (phonemically /n/) (Vallejos, 2010:57). In Omagua, codas are either glides or [ŋ] phonemically /ɴ/ (Sandy & O'Hagan, 2020:109ff).

Syllables in Omagua–Kokama generally cannot contain two vowels, and two subsequent vowels are either assigned to separate syllables, or one of the vowel is glided if possible. This gliding process is covered in greater detail in section 2.4.1.2. Heteromorphemically in Omagua, vowel deletion or coalescence also occurs as a strategy, shown in (7) (Sandy & O'Hagan, 2020:127). Similar elision and merging also occurs in Kokama in similar contexts (Vallejos, 2010:67ff).

(7) (a) ta=ikua > [tekua] 'I know' (male speech)
(b) tana=ikua > [tanikua] 'we know'
(Sandy & O'Hagan, 2020:127)

Roots in Omagua are required to be bimoraic (CVV), which is an exception to the vowel hiatus constraint described above. When suffixed with e.g. the female speech plural form *=na*, the second vowel in the syllable receives primary stress: /jɨɨ/ 'axe' > [jɨ.ˈɨ.na] 'axes' (Sandy & O'Hagan, 2020:126).

Roots in Omagua–Kokama are required to be vowel-final (O'Hagan, 2011:25). This means that cognates consonant-final roots in Tupinambá end on typically either /a/ or /i/ in Omagua–Kokama, which are probably originally derived from the POKT reflexes of the Proto-Tupí–Guaraní suffixes such as nuclear case **-a*, gerundive **-a* and the oblique-topicalized construction suffix which is **-i* after consonant-final roots, shown in (8) (O'Hagan, 2011:25-26; Jensen, 1998:526). These suffixes became frozen to the root in Proto-Omagua–Kokama to satisfy the vowel-final word constraint (O'Hagan, 2011:25).

(8) *kwetsé i-ʔáɾ-i
yesterday 3-fall-OBTOP
'Yesterday he fell.' (Jensen, 1998:526)

2.4.2 Tupinambá phonology

2.4.2.1 Vowels

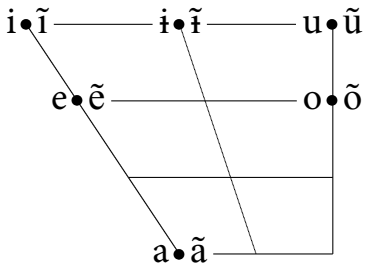

Figure 2.4: Tupinambá vowel phonemes.

The vowel phonemes of Tupinambá are shown in Figure 2.4, six oral phonemes and six nasal counterparts. The oral and the nasal do not have equal distribution; the oral vowel phonemes can occur in any position, whereas the nasal phonemes only occur in stressed (i.e. final) syllables, as in [nati'ʔũ] 'mosquito'. However, allophonic nasal vowels occur adjacent to nasal segments (Rodrigues, 1958a:101).

2.4.2.2 Consonants

	Bilabial	**Alveolar**	**Post-alv.**	**Palatal**	**Velar**	**Glottal**
Stop	p p^w (p^j)	t			k g k^w g^w	
Pre-nasalized stop	^{m}b	^{n}d			^{ŋ}g	
Fricative	β	s	ʃ			
Glides	(w)			j		
Tap		ɾ				

Table 2.5: Tupinambá consonant phonemes.

Stops Tupinambá has three stop phonemes: /p/, /t/, and /k/, as in /piɾa/ 'fish', /tĩ/ 'nose', and /oka/ 'house'. There is also a glottal stop [ʔ] which occurs allophonically to break up vowel sequences in polysyllabic words, e.g. /kai/ [ka'ʔi] 'monkey sp.'.

Tupinambá also has several labialized stops, /k^w/, /p^w/ and /g^w/, as in /k^waɾasɨ/ 'sun', /p^wã/ 'finger', and /g^wɨɾa/ 'bird'.

Fricatives Tupinambá has three fricative phonemes: /s/, /β/, and /ʃ/, as in /sɨ/ 'mother', /βaka/ 'to turn', and /poʃɨ/ 'ugly'. The phoneme /ʃ/ is considerably more infrequent than the other fricatives. Where it occurs, it is usually in the context of a high vowel /i/ or palatal

glide /j/, suggesting it most likely originally occurred as a palatalized allophone of /s/. Its phonological status is supported however, by few minimal pairs such as /ʃe/ 'my' and /se/ 'it tastes good' (Rodrigues, 1958b:115).

In addition to this Rodrigues (1958b:114f) also mentions [sʃ] as a possible allophone of /ʃ/ occurring in some words, but not in others.

Nasals On the surface level, Tupinambá has three plain nasals, [m], [n], and [ŋ], as in [mũ] 'friend', [naˈna] 'pineapple', and [tiˈŋa] 'white'. These alternate with their pre-nasalized plosive counterparts [ᵐb], [ⁿd], and [ᵑg] according to complex patterns which will be described further below. The palatal nasal [ɲ] is a common allophone of /j/ due to nasal harmony. At least O'Hagan (2011:6) treats [ɲ] as a phoneme in its own right, since it also occurs without the presence of any other nasal segment, as in /ɲota/ 'only', whereas Rodrigues (1958b:119) treats them as allophones of the same phoneme.

Relationship between plain nasals and pre-nasalized stops According to Rodrigues (1958b:107), the distribution between [m] and [ᵐb] (and likewise for [n] and [ⁿd]) is partially in complementary, partially free variation. Rodrigues' distribution is as follows (using [m]∼[ᵐb] as an example):

1. Before a stressed nasal vowel, or if a nasal sound occurs in the next syllable, the realization is [m]: [mwã] 'human finger', [ˈmina] 'spear'.
2. Before a stressed oral vowel and if no nasal sound follows it, the realization is [ᵐb]: [ᵐbɨ] 'human foot', [ˈᵐboja] 'snake'.
3. In an unstressed pretonic syllable both [m] and [ᵐb] occur in free variation, so long as the next syllable contains no nasal sound: [maˈʔej]∼[ᵐbaˈʔej] 'thing', [eɾimaˈe]∼[eɾiᵐbaˈe] 'earlier'.
4. In an unstressed posttonic syllable, the realization is always [m]: [ˈkãma] 'breast', [koˈẽma] 'morning'.
5. In final position, the realization is always [m]: [aˈam] 'I stand', [aˈsem] 'I go'.

As for the relationship between [ŋ] and [ŋg], one can assume that it is governed by the same principles as for [m]∼[ᵐb] and [n]∼[ⁿd], but the orthographic representation is typically

⟨ng⟩ for both, so the same analysis cannot be made for words with ⟨ng⟩.

Glides and liquids The sole liquid in Tupinambá is the voiced alveolar tap [ɾ] as in /wɨɾa/ 'bird', which is described as such in contemporary Portuguese sources by authors who knew both the tap [ɾ] and the trill [r] from their native language, and reported that the trill did not occur (Rodrigues, 1958b:82).

As for glides, Tupinambá has a palatal glide /j/ as in /ja^{n}de/ 'we.INCL', which has two allophones [ʒ] and [ɲ], the first of which occurs in free variation with [j] initially and medially, but not finally, as in [ʒakaɾe]∼[jakaɾe] 'caiman'. The nasal allophone [ɲ] occurs partially in complementary distribution partially in free variation with [j]∼[ʒ] in certain nasal contexts (cf. subsubsection 2.4.2.3) (Rodrigues, 1958b:116-119).

1. Before a stressed oral vowel, which is not followed by a stem-internal nasal consonant, the realization is [ʒ]∼[j] in free variation: [ʒu]∼[ju] 'thorn', [aˈʒaɾ]∼[aˈjaɾ] 'I take it'.
2. Before a stressed nasal vowel, or followed by a stem-internal nasal consonant, the realization is [ɲ]: [ɲũ] 'field', [kuɲã] 'woman'.
3. Before an unstressed vowel, which is not preceded by a nasal sound, the realization is [ʒ]∼[j] in free variation: [ʒaˈwaɾa]∼[jaˈwaɾa] 'dog', [aʒeɾuˈɾe]∼[ajeɾuˈɾe] 'I bite'.
4. Before an unstressed vowel, which immediately precedes a stressed syllable of the type [V], [VN], or [NVN], the realization is [ɲ]: [aɲeˈeŋ] 'I speak', [ɲaˈẽ] 'pot'.
5. Before an unstressed vowel, which immediately precedes a stressed syllable of the type [CV], [CVN], or [NV], or before an unstressed vowel which either immediately precedes an unstressed syllable, or indirectly precedes a stressed syllable containing a nasal sound, the realization is either [ɲ] or [ʒ]∼[j]: [ɲɨˈɾõ]∼[ʒɨˈɾõ] 'forgiveness', [ɲakuˈⁿda]∼[ʒakuˈⁿda] 'fish sp.'.
6. Medially after a stressed nasal vowel belonging to the same stem, [j]∼[ɲ] are in free variation when a vowel follows: [piˈɾãja]∼[piˈɾãɲa] 'piranha', [taˈkʷãja]∼[taˈkʷãɲa] 'penis'.
7. Medially after a stressed oral vowel belonging to the same stem, the realization is [j]: [ˈsaja] 'sour thing'.

8. Medially before a consonant, and finally, the realization is [j]: [saˈkʷãjᵐbae] ‘male’, [ⁿdaˈsoj] ‘I do not walk’.

There is also a bilabial glide [w], which will be further discussed below.

Labialization and phonological status of /w/ Labialization is generally marked by ⟨u⟩ as in ⟨cuarassy⟩ /kʷaɾasɨ/ ‘sun’. This writing is ambiguous as it can also represent the vowel /u/, which we will return to. Sometimes after ⟨g⟩, a velar glide or labialization is explicitly marked by a diaeresis, as in ⟨güyrá⟩ ‘bird’. It is clear from Barbosa’s work that this represents /gʷ/, and that the diaeresis is used to explicitly show the velar glide, as ⟨guy⟩ could be interpreted as /gɨ/ in Portuguese orthography, compare Portuguese ⟨guitarra⟩ /gi.ˈta.ʁa/ to the pre-reform spelling ⟨agüentar⟩ /a.gwẽ.taʁ/. Other authors, such as Rodrigues (1958b), uses a plain labio-velar approximant /w/ in his transcription of such words, e.g. /wɨˈɾa/ ‘bird’. Primary missionary sources point in different directions, with various transcriptions such as ⟨guirà⟩, ⟨guirâ⟩, ⟨guirá⟩, ⟨guyrá⟩ in Portuguese sources, but ⟨ouyra⟩ in French sources (Rodrigues, 1958b). It is possible that there was a variation between [gʷ]∼[w] in Tupinambá at the time when original transcriptions were made.

If one chooses to interpret this sound as /gʷ/, the phonological status of /w/ becomes less clear. Both Rodrigues (1958b) and Barbosa (1956) has a phonotactic analysis in which /CGV/ is a possible syllable, but if [w] is only found after certain consonants such as /k/ and /p/, a simpler analysis might be to posit only a /CV/ syllable, and treat /pw/ and /kw/ sequences as labialized /kʷ/ and /pʷ/, in which case Tupinambá would have a labialized set of stop phonemes including /gʷ/ as discussed above.

However, there are also words such as /pjaɾa/ ‘path’ which are written by Barbosa (1956) with a glide. These /CGV/ sequences are quite rare however, so if one chooses to treat /w/ as a phoneme, then its distribution would be quite restricted. Given the lack of other initial stop+glide sequences such as /kj/ or /tj/, a logical option would be to posit a phoneme /pʲ/ with quite restricted distribution. The glide /w/ is still found in certain suffixes such as the nominalizer *-swaɾa*, so under this analysis /w/ must still have phonemic status, but occurring mainly in this nominalizing suffix, whereas the other apparent occurrences can be treated as labialized stops.

2.4.2.3 Nasal harmony

Nasal processes play an important role in Tupinambá phonology and several phonemes have an oral or a nasal realization depending on the phonological context. The nasal harmony can be both regressive and progressive and is triggered by a phonemically nasal segment, such as a nasal consonant or a nasal vowel.

The progressive and regressive nasal harmony work in quite different ways. Progressive nasal harmony, as seen in (9) is triggered by a phonemic nasal in the final syllable and targets the nearest rightward consonant in a suffix, creating oral and nasal allomorphs for certain suffixes (O'Hagan, 2013:9ff).

(9) /pɨsɨɾõ-aβ/ 'rescue-NOMZ:INSTR' > [pɨsɨɾõˈʔam]
/pɨsɨɾõ-aɾ/ 'rescue-NOMZ:AGT' > [pɨsɨɾõˈʔan]
(O'Hagan, 2011:7)

Regressive harmony affects segments occurring leftward of the nasal segment. Regressive harmony targets pre-nasalized stop phonemes, turning them into plain nasals [m n ŋ]. Hence, pre-nasalized stops do not occur when followed by a nasal segment. Regressive harmony also targets /j/, turning it into [ɲ]. The nasalization of /j/ to [ɲ] seems to be optional, and does not always occur (O'Hagan, 2013:6). Phonemic nasal vowels only occur in stressed syllables, but regressive harmony also targets vowels, creating allophonic nasal vowels (Rodrigues, 1958b:101).

The nasal span of the harmony trigger seems to extend to all leftward segments, and no pre-nasalized stop will surface so long as there is a nasal trigger to the right of it. According to Rodrigues (1958b:100), there is variation regarding the representation of allophonic nasal vowels in the older European sources; one source might write ⟨amãna⟩, and another one ⟨amana⟩ 'rain'.

It is possible that nasal harmony affects all nasal vowels leftward of the trigger as well, but gets gradually weaker the further away it gets. Given the uncertainty in the representation of nasal allophony, it is difficult to make an accurate description of the nature of regressive nasal harmony.

2.4.2.4 Stress

Unsuffixed words in Tupinambá are always stressed on the final syllable, but roots which are followed by so-called unstressed suffixes can also receive penultimate or antepenultimate stress, which means that stress is always predictable from morphology, e.g. /ɨpa/ [ɨ.ˈpa] 'tree bark', but /ɨ-pa/ [ˈɨ.pa] 'water-LOC' (Rodrigues, 1958b:121-122).

Chapter 3

Method

This chapter begins with a description of the data used for this thesis, and a description of the data management and phonetic transcription. This is followed by a section on the construction of cognate and correspondence sets, followed by section on the Comparative Method and related concepts used for reconstructing the inventory of the ancestral proto-language and for characterizing sound changes.

3.1 Data

The comparative Omagua–Kokama–Tupinambá lexical dataset on which this thesis was based, as well as the comparative lexical data from other Tupí–Guaraní languages employed at a number of points in this thesis, was drawn from the Tupí–Guaraní Comparative Lexical Dataset (Chousou-Polydouri et al., 2019) developed by members of the Berkeley Tupí–Guaraní Comparative Project, led by Lev Michael. The Omagua, Kokama, and Tupinambá lexical data was principally harvested from published sources and unpublished lexical databases by Emily Leggitt in 2017.

The raw data used for this thesis consist of a spreadsheet with a word list in orthographic written form for each language, consisting of 720 words for Omagua, 890 words for Kokama, and 1666 words for Tupinambá.

The Tupinambá data ultimately come from dictionaries by Antônio Lemos Barbosa, especially his *Pequeno vocabulário Português-Tupi* (Barbosa, 1970). Barbosa's dictionary are in turn based on 16th century Jesuit texts and grammars, e.g. de Anchieta (1595).

The Omagua data come from fieldwork carried out mainly by UC Berkeley graduate students Zachary O'Hagan, Clare Sandy, Tammy Stark, and Vivian Wauters together with Omagua consultants Amelia Huanaquiri, Arnaldo Huanaquiri, Alicia Huanío and Lino Huanío in the community of San Joaquín de Omaguas in 2010–2013. It also stems from from 18th century Jesuit texts, see Michael & O'Hagan (2016).

The Kokama data stem from work by Rosa Vallejos, primarily Vallejos (2010) and Vallejos & Amías (2015).

3.1.1 Data transcription

In order to build correspondence sets, the data were converted from an orthographic transcription to the International Phonetic Alphabet (IPA).[1] Barbosa (1970) for instance, is written in a Portuguese-based orthography. In some cases, this was a fairly straight-forward automatic process, but since the original orthography does not reflect certain phonemic distinctions, much of the data had to be checked against other cognate data before being converted.

3.1.1.1 Tupinambá

For Tupinambá, certain graphs or digraphs could be automatically converted to IPA. No conversion has been carried out for vowels, except for ⟨y⟩ which represents /ɨ/, as in ⟨aty⟩ /atɨ/ 'wife'. For consonants, ⟨x⟩ was replaced by /ʃ/ as in ⟨pixé⟩ /piʃe/ 'burnt', ⟨nh⟩ by /ɲ/, as in ⟨nhũ⟩ /ɲũ/ 'field', ⟨b⟩ by /β/, as in ⟨uba⟩ /uβa/ 'father', ⟨ss⟩ by /s/ as in ⟨yssá⟩ /ɨsa/ 'stem, trunk'.

In the original orthography, /k/ is represented by several characters, ⟨qu⟩ before ⟨e, i⟩ and ⟨c⟩ elsewhere, including finally. These have both been replaced by /k/, as in ⟨yqué⟩ /ɨke/ 'side' and ⟨cama⟩ /kama/ 'breast'.

Labialization As discussed in subsubsection 2.4.2.2, there is an orthographic ambiguity in Barbosa (1970) between labialized stops + vowel sequences (/C^{w}V/) and consonant + vowel + vowel sequences (/CVV/). This could in most cases be resolved with double

[1] The IPA transcriptions are mostly phonemic, for the most part based on Sandy & O'Hagan (2020) for Omagua, Vallejos (2010) for Kokama, and Rodrigues (1958b) for Tupinambá. An exception is the result of nasal harmony, which is represented in the transcriptions, e.g. [m] vs. [^{m}b], or [j] vs. [ɲ]. For legibility, I have also chosen to write pre-nasalized stops without the superscript nasal in the data, e.g. *mboja* for *^{m}boja*.

checking with other work which explicitly marks labialization or glides, such as Barbosa (1956). In truly ambiguous cases, these have been left as /CVV/ rather than /C^{w}V/ in the data.

Glottal stop In Barbosa (1970), ⟨u⟩ and sometimes ⟨i⟩ can represent labialization and palatalization, respectively, e.g. ⟨puana⟩ /p^{w}ana/ 'to pass', but can also represent the full vowel /u/, as in ⟨puama⟩ /puʔama/ 'stand up'. Such orthographic vowel hiatuses often, but not always, imply the presence of an intervocalic glottal stop. The orthography in Barbosa (1970), usually distinguishes /j/ from /i/ initially, but not medially or finally, e.g. ⟨jaia⟩ /jaja/ 'mock', or ⟨pai⟩ /paj/ 'witch, shaman'. For this reason, many orthographic VV sequences had to be double-checked with cognates in other Tupí–Guaraní languages that preserve Proto-Tupí–Guaraní glottal stops, e.g. Kaiowá, Tembé, or Guajajara.

Velar nasals Orthographic ⟨ng⟩ is ambiguous in Barbosa (1970) as to whether it represents a velar nasal /ŋ/ or a pre-nasalized velar stop /^{ŋ}g/. For this reason, certain unsure ambiguous cases have been represented in IPA with a placeholder capital /N/ which is agnostic about which of the two segments the orthographic sequence corresponds to.

3.1.1.2 Omagua & Kokama

For Omagua and Kokama, the conversion process was more of an automatic process. For Omagua, it was just a matter of replacing e.g. ⟨y⟩ by /j/ and ⟨r⟩ by /ɾ/. For Kokama, the process was similar, except for certain cases where the orthography used by Vallejos & Amías (2015) was ambiguous in that both /pwa/ 'rotten' and /pu.(w)a/ are written ⟨pua⟩. Each of these ambiguous cases could be resolved manually however, either by double checking with Vallejos (2010) or inferring from how other words with different stresses were represented.

3.2 Methodology

In order to answer the research questions regarding what phonological features are reconstructable to the proto-language, as well as what phonological changes were involved in the genesis of Omagua and Kokama, this study employs the Comparative Method.

The Comparative Method is defined by Weiss (2014:127), as "the systematic process of reconstructing the segmental and suprasegmental inventory of an ancestral language from cognate reflexes in the genetically related daughter languages" and is described as the "key tool for investigating linguistic prehistory" since the mid 19th century. By using the Comparative Method, linguists can reconstruct ancestral languages which have existed in prehistory, but are unattested, i.e. never written down.

In order to apply the Comparative Method to a number of languages, one first needs a hypothesis that these languages are related, since languages that descend from a reconstructable ancestral language are per definition related. A data set is then assembled of words or roots that are thought to be related, i.e. *cognate*, and purged of borrowings to the extent possible (Weiss, 2014:128f). A presupposition when applying the Comparative Method is that sound change is regular and systematic, meaning that daughter languages will exhibit structural similarities or correspondences to one another, and that these correspondences will be confirmed by multiple instances (Weiss, 2014:133). These regular correspondences serve as the basis for reconstruction of proto-forms. For this reason, segments are grouped into *correspondence sets*. A simple example of three such correspondence sets is given by List (2019:141) for English, German, and Dutch, respectively: *d : t : d* as in *dead : tot : dood*; *θ : d : d* as in *thick : dick : dik*; *t : ts : t* as in *tongue : Zunge : tong*. These partially overlapping sets represent three different Proto-Germanic phonemes: *d, *θ, *t, respectively.

After creating the correspondence sets, one has to establish whether any of the sounds represent the original state of affairs, as in the case of English θ < *θ above, and in that case which one, or whether the original sound was a different sound all together. An important notion when establishing such proto-forms is *directionality* in sound change, meaning that certain changes is common in one direction but not in others. For instance, velar sounds quite often palatalize before front vowels, whereas the reverse change is impossible, or at best extremely rare (Weiss, 2014:135). This means that when confronted with a cognate set such as *kelu : tʃɛlo : tsjel* in Sardinian, Italian, and Old French, respectively, a strong contender for the proto-form of the initial segment would be *k and posit a sound change that Italian and Old French palatalized velar *k before a front vowel, which is in this case confirmed by the attested Classical Latin form *kaelum* (Weiss, 2014:130).

Another method employed in this thesis is that of *phonetic alignment*, meaning that when

creating correspondence sets, cognate sets are split up into segments and aligned in a matrix, where segments without correspondences is given a gap symbol. This is according to List (2019:138) a recently adapted concept in linguistic built on approaches within bioinformatics and computer science. It has always however implicitly been an integral part of the methodology within historical linguistics, he notes. An example of such alignment can be seen in Table 3.1.

		A		B	
Sanskrit	y	u	g	a	m
Greek	z	u	g	o	n
Latin	i	u	g	u	m
Gothic	j	u	k	–	–

Table 3.1: An example of phonetic alignment of some Indo-European cognates of 'yoke', where A and B are two examples of correspondence sets, and – is used to indicate the absence of corresponding sounds in Gothic. Adapted from List (2019:139).

3.3 Cognate sets

Once the data had been standardized, cognate sets were built between the three languages using computer-assisted tools and then manually double-checked, as well as cross-checked with existing Tupí–Guaraní cognate data from Chousou-Polydouri et al. (2019).

A computer-assisted framework, as opposed to a fully automated framework, is a method where the researcher lets an algorithm detect cognates automatically, and then manually edits the findings. In this way, the work is not entirely left to the computer, but is merely a way of speeding up the otherwise quite lengthy process of finding cognate sets manually in a dataset (List et al., 2017). The method used for cognate detection was LexStat, a method for cognate detection included in the Python linguistics library LingPy (List et al., 2018). The algorithms were run by undergraduate student Eric Chen at UC Berkeley, while the data preparation and the manual editing were performed by the author.

The algorithms used by LingPy use a clustering method where pairs of all words between all languages are compared and given a score. These pairs are then clustered based on their score. Pairs with a score under the given threshold are pruned, and words with the same meaning are divided into groups of cognate words. LexStat specifically uses a language-specific scoring scheme, by which all wordlists of all language pairs with different meanings are shuffled,

compared with the attested distribution and then clustered (List et al., 2017:5).

Out of the five automatic cognate detection methods tested by List et al. (2017), LexStat performed second-best, second only to Infomap, which builds on LexStat (List et al., 2017:13). List et al. (2017) note that the choice of method may depend on the task, and even such methods that performed worse than LexStat in their test, such as the Turchin method, may still be useful if the aim of the detection simply is to aid the researcher in creating cognate sets which are going to be manually checked anyway. They note that methods such as LexStat require at least 200 words for moderately close languages List et al. (2017:14). Since the dataset used in this study contains more than 200 words, and only includes three fairly closely related languages, LexStat should prove useful, especially as the cognate sets were manually checked afterwards.

The Python code for running the LexStat automatic cognate detection and alignment is included in Appendix B. This appendix also has instrutions on how to retrieve the input data file, and the output analysis used for manual editing in EDICTOR.

The cognate sets which LexStat yielded were then manually double-checked and edited using EDICTOR (`http://edictor.digling.org`; List, 2017), an online tool for managing cognate and correspondence sets. In this process, the cognate sets were cross-checked with existing Tupí–Guaraní cognate data. The sets were also morphologically segmented, so that a set can consist of words that are *partially* cognate, so that one part of e.g. a compound word is matched with that same part in another language. This means that a set such as FOUR: Omagua *ɪruaka*, Kokama *ɨrwaka*, and Tupinambá *ɨrundɨk* which were automatically identified as cognate, were manually segmented into Omagua *ɪru-aka*, Kokama *ɨrw-aka*, and Tupinambá *ɨru-ndɨk*, and the first parts was given the same cognate ID, whereas the the second parts were not, since *-aka* and *-ndɨk* do not appear to be cognate. In this step, the cognate sets were also properly aligned in EDICTOR. The total number of cognate sets finally used amounted to 275, shown in Appendix A.

Omagua	t	a	p	i	-	-	r	a
Kokama	t	a	p	i	-	-	r	a
Tupinamba	t	a	p	i	ʔ	í	r	a

Figure 3.1: Example of the cognate set TAPIR and the alignment of segments as viewed in the EDICTOR interface.

COGNATES	INDEX	PATTERN	CONCEPTS	Kok	Tup	Oma
517	3	u / 79	companion	i r w a	i r ũ	ɪ r u a
1058	3	u / 79	grandfather	a m w i	a m ũ j a	a m u i
1228	5	u / 79	island	ɨ p w a	ɨ p a ʔ ũ	ɨ p a u

Figure 3.2: Example of the correspondence set between Kokama /w/, Tupinambá /ũ/, and Omagua /u/ and the three sets where this correspondence is found, as viewed in the EDICTOR interface.

3.4 Correspondence sets

Once the cognate sets were built, sound correspondences across the languages needed to be identified in the data. This was done automatically using a package for LingPy called Lingrex (List, 2018, 2019). Lingrex takes as input a tab separated text file (.tsv) of the data which is coded for cognacy (as done in the previous step) and phonetically aligned. The algorithm then uses the alignments to build a network, and sorts corresponding segments into correspondence sets and outputs the text file with another column called "patterns", where the correspondence sets have been given an ID (List, 2019:147f). These correspondence sets can then be viewed and edited in EDICTOR. These sets were then manually inspected, and given preliminary names in order to be more easily managed, e.g. "general /a/" for the reflex of *a as /a/ in all languages. These correspondence sets are listed and described in the Results chapter.

Chapter 4

Results

The following chapter is an exposition of the correspondence sets between Omagua, Kokama, and Tupinambá. These correspondence sets have been generated by Lingrex as described in section 3.4, and given a number whose primary function is to serve as a name of the set in the description, but corresponds more or less to the size of the sets, i.e. set a_1 is a much larger correspondence set than a_2. For each set, I propose a reconstructed Proto-Omagua–Kokama–Tupinambá phoneme, and argue for the choice briefly for each set, and discuss the diachronic development and analyze the results in greater detail in the Discussion chapter.

Along with the correspondence sets a number of example cognate sets are given which can be found in Appendix A. For correspondence sets found in more than three cognate sets, three arbitrary examples are given. For correspondence sets found in three cognate sets or fewer, all of the examples are given. An example of the structure can be seen in Table 4.1 with the example cognate sets written out. For the full example cognate sets throughout the chapter the reader is referred to Appendix A however.

Set	OMG	KK	TPN	POKT	Example sets	OMG	KK	TPN
a_1	a	a	a	*a	AFTERNOON	karukatai	karuka	karuka

Table 4.1: Example of the Results chapter table structure.

The chapter starts with the proposed reconstructed vowel inventory for POKT followed by an exposition of all the correspondence sets for vowels. The proposed consonant inventory is then presented along with all the correspondence sets for consonants. At the end of

the chapter, a table is presented with correspondences that exhibit exceptions to the more general sound changes identified elsewhere in the chapter. At the end of the chapter, a list is given of all the major sound changes identified.

4.1 Reconstructed vowels

The phonological vowel inventory of Proto-Omagua–Kokama–Tupinambá is presented in Table 4.2. This inventory is identical to the reconstructed inventory for Proto-Tupí–Guaraní (PTG) by Jensen (1998:604), meaning that there were no major changes to the vowel phoneme inventory from PTG to Proto-Omagua–Kokama–Tupinambá.

*i *ĩ	*ɨ *ɨ̃	*u *ũ
*e *ẽ		*o *õ
	*a *ã	

Table 4.2: Reconstructed POKT vowels.

4.1.1 Oral vowels

4.1.1.1 *a

Set	OMG	KK	TPN	POKT	Example sets
a_1	a	a	a	*a	AFTERNOON, BREAST, GRANDFATHER
a_2	a	a	á	*a	ARM, EGG, FISH
a_3	∅	∅	a	*a	SWEAT (INTR.), TAIL

Table 4.3: *a correspondence sets.

The reconstruction of *a is a straightforward reconstruction, based on the sets in Table 4.3. In most cases *a has been preserved in all three languages, as seen in sets a_1 and a_2. These sets are the same, and the only difference is that stress is explicitly marked by an acute accent on certain Tupinambá words. In other cases, *a was lost in Omagua–Kokama, but preserved in Tupinambá. In certain cases, this is part of a general syllable loss in Omagua–Kokama, involving loss of *β, as in the words for BED, Omagua–Kokama *tupa*, Tupinambá *tupaβa*. In other cases, it is part of a vowel shortening following loss of *ʔ: VʔV > VV > V, as in the word for DEFECATE, Omagua *kapɪ*, Kokama *kape*, Tupinambá *kaʔapia*.

Set	OMG	KK	TPN	POKT	Example sets
e_1	ɪ	e	e	*e	FLY, WING, SHINE
e_2	i	i	é	*e	CAIMAN, KNIFE, SHAMAN
e_3	i	i	e	*e	ARRIVE (INTR.) LEAVE FROM (INTR.), NAME
e_4	i	j	é	*e	ALREADY∼ALSO, SPIRIT
e_5	ɪ	e	é	*e	FLY (INTR.), BACK
e_6	ɪ	i	e	*e	COCKROACH, RETURN
e_7	∅	∅	e	*e	PET, MONKEY, WIFE

Table 4.4: *e correspondence sets.

4.1.1.2 *e

The reconstruction of *e is based on the sets in Table 4.4 that involve various types of vowel raising. The correspondence e_1 is attested in approximately 20 cognate sets, and involves a vowel raising in Omagua from *e to /ɪ/. Set e_2 displays a raising in Omagua–Kokama from *e to /i/ when the *e is final and stressed. The same raising occurs in set e_3 when *e occurs adjacent to a coronal segment. In all of these cases Tupinambá preserves the original unraised vowel. This is evident from other Tupí–Guaraní languages, e.g. the set KNIFE, which is *kɨsé* in Tupinambá, *kɨʃi* in Omagua, and *kɨtʃi* in Kokama. Other Tupí–Guaraní languages generally preserve a final /e/, e.g. Paraguayan Guaraní *kɨse*, Araweté *tʃitʃe*. Similarly, the set MANY, which consists of Tupinambá *seta*, Omagua *ʃita*, and Kokama *tʃita*, generally have an unraised vowel in Tupí–Guaraní cognates, e.g. Guarayu *heta*, Guajajara *eta*.

In e_4, *e raised finally as in e_2 but preceded by a vowel, which caused *e to become an off-glide in Kokama.

The set e_5 are exceptions to e_2, i.e. where *e did not raise finally. Similarly, in e_6, *e raised in Kokama but not in Omagua. These are further discussed in subsection 5.1.4.

In e_6, *e was part of an objective nominalizer **emi-* and was lost in all of these cases. This is discussed further in subsubsection 5.1.7.3.

4.1.1.3 *i

Set	OMG	KK	TPN	POKT	Example sets
i_1	i	i	i	*i	AGOUTI, HAMMOCK, PIRANHA
i_2	i	i	í	*i	MANIOC FLOUR, SANDFLY, SISTER-IN-LAW (F. EGO)
i_3	∅	∅	í	*i	CATFISH, GRASS, TAPIR
i_4	i	j	i	*i	CHEST, TESTICLES, VULVA

Table 4.5: *i correspondence sets.

POKT *i as seen in Table 4.5 has generally been preserved in all three languages as in sets i_1 and i_2. In set i_3, it was lost following loss of *ʔ and vowel shortening as in CATFISH, Omagua–Kokama *mani*, Tupinambá *mandiʔí*. In set i_4, *i went to /j/ in Kokama following a consonant and preceding another vowel, as in MOSQUITO, Omagua *jatiú*, Tupinambá *ɲatiʔũ*, but Kokama *jatju*.

4.1.1.4 *ɨ

Set	OMG	KK	TPN	POKT	Example sets
ɨ$_1$	ɨ	ɨ	ɨ	*ɨ	ARM, BAT, WIND
ɨ$_2$	ɨ	ɨ	u	*ɨ	BELLY BUTTON, LAKE, REST (INTR.)
ɨ$_3$	u	u	ɨ	*ɨ	CHAMELEON, LUNG∼BREATHE (INTR.)
ɨ$_4$	∅	∅	ɨ	*ɨ	CUTBANK, DAUGHTER
ɨ$_5$	ɨ̃	ɨ	ɨ	*ɨ	SWEAT
ɨ$_6$	i	i	ɨ	*ɨ	GRAB, HURT, VINE

Table 4.6: *ɨ correspondence sets.

The reconstruction of *ɨ is based on the sets in Table 4.6. In most cases this vowel was preserved in all three languages, as in set ɨ$_1$. As seen in sets ɨ$_2$ and ɨ$_3$, Omagua–Kokama /ɨ/ sometimes alternates with Tupinambá /u/. The reconstruction for set ɨ$_2$ is *ɨ, as other Tupí–Guaraní languages generally has /ɨ/ in these words, as in LONG TIME AGO, Omagua *ɨmɨnua*, Kokama *ɨmɨnwa*, Tupinambá *umuã*, compared to e.g. Kamaiurá *ɨmawe*, Paraguyan Guaraní *ɨma*. Set ɨ$_2$ is generally next to labials, which might have caused a rounding assimilation from *ɨ > u in Tupinambá. The same process probably applies to set ɨ$_3$ which consists of two words, CHAMELEON, which has no other known cognates in other Tupí–Guaraní languages, and BREATHE, which generally has /ɨ/ in Tupí–Guaraní cognates. This is further discussed in subsection 5.1.8.

In ɨ$_5$, *1 was nasalized in Omagua, which is found in a handful of words (Sandy & O'Hagan, 2020:106-109).

In ɨ$_6$, these sets are caused by a Proto-Omagua–Kokama sound change *ɨ > i next to *ts. This is further discussed in subsection 5.1.1 and subsection 5.1.8.

4.1.1.5 *u

The reconstruction of *u, as seen in Table 4.7 is also straightforward, having been preserved in almost all cases in all three languages as in set u$_1$. In set u$_2$, it was lost in Kokama as part of

Set	OMG	KK	TPN	POKT	Example sets
u_1	u	u	u	*u	BLOOD, EGG, YELLOW
u_2	u	∅	u	*u	FALL (INTR.), SORUBIM SP., PADDLE (N.)
u_3	u	w	u	*u	BELLY BUTTON, FOUR, ROUND
u_4	ũ	u	u	*u	TAIL

Table 4.7: *u correspondence sets.

a general change of final syllable simplification in Kokama (see subsubsection 5.1.3.1) as in PEANUT, Omagua *munui*, Tupinambá *manduβi*, but Kokama *muni*. In u_3, *u turned to /w/ in Kokama following a consonant but before a vowel, e.g. BELLY BUTTON: Tupinambá *puru?ã*, Omagua *mɨrua*, Kokama *mɨrwa*. This process is also found in Omagua but is considered to be underlyingly /u/; see section 2.4.1.2 on Omagua–Kokama glides. In u_4, *u was nasalized in Omagua, which is found in a handful of words (Sandy & O'Hagan, 2020:106-109).

4.1.1.6 *o

Set	OMG	KK	TPN	POKT	Example sets
o_1	u	u	o	*o	BURST, HOUSE, TWO
o_2	u	u	ó	*o	FISH POISON, GO, ROPE∼THREAD
o_3	a	a	o	*o	BITTER, HAPPY, POOR
o_4	u	w	ó	*o	OTHER, ROOT, WIFE

Table 4.8: *o correspondence sets.

The reconstruction of *o is based on the sets in Table 4.8. Omagua–Kokama merged POKT *o and *u unconditionally, and the outcome was /u/, as in the set HOUSE: Tupinambá *oka*, Omagua–Kokama *uka*. In four cognate sets, as in set o_3, Tupinambá /o/ corresponds to Omagua–Kokama /a/, as in POOR: Tupinambá *poreausuβa*, Omagua *pariasu*, Kokama *parjatsu*. The reconstruction for this set is still *o, and Omagua–Kokama underwent a change o > a in these cases, as this vowel generally corresponds to /o/ in other Tupí–Guaraní languages, e.g. Paraguyan Guaraní *poriahu*, Kamaiurá *poryaup*. This is briefly discussed in subsection 5.1.8, but has an unknown cause. In o_4, POK *u (i.e. the outcome of the merger) turned to /w/ in Kokama following a consonant but before a vowel. This is the same process as in u_3 above.

4.1.2 Nasal vowels

In general, reconstructed nasal vowels follow the changes of their oral counterparts for the most part. Vowel nasality was generally lost unconditionally in Omagua–Kokama, but is

preserved in Tupinambá. Since nasal vowels only occurred phonemically on the final stressed syllables, and were much less frequent than their oral counterparts, the reconstruction for nasal vowels are based on much fewer cognate sets than the oral vowels, in some cases only a handful of sets.

4.1.2.1 *ã

Set	OMG	KK	TPN	POKT	Example sets
$ã_1$	a	a	ã	*ã	RIVER, ROUND, TOOTH
$ã_2$	a	∅	ã	*ã	SCRAPE (TR.), THING

Table 4.9: *ã correspondence sets.

Nasal *ã is reconstructed based on the sets in Table 4.9. Set ã corresponds to the oral counterpart a_1 in Table 4.3, and set $ã_2$ corresponds to a_5 in the same table.

4.1.2.2 *ẽ

Set	OMG	KK	TPN	POKT	Example sets
e_1	ɪ	e	ẽ	*ẽ	SPLIT (TR.)
e_2	ɪ	∅	ẽ	*ẽ	SWEET
e_3	i	i	ẽ	*ẽ	SMOKE FOOD

Table 4.10: *ẽ correspondence sets.

The reconstruction of *ẽ in Table 4.10 is based on three sets only, seen in Table 4.10. The development of *ẽ is parallel to that of its oral counterpart, seen in Table 4.4.

4.1.2.3 *ĩ

Set	OMG	KK	TPN	POKT	Example sets
$ĩ_1$	i	i	ĩ	*ĩ	HAMMOCK, NOSE, THIN

Table 4.11: *ĩ correspondence sets.

The reconstruction of *ĩ in Table 4.11 is a straightforward reconstruction based on one correspondence set and five cognate sets with parallel development to that of oral *i, seen in Table 4.5.

4.1.2.4 *ɨ̃

There are no cognate sets in the data that support the reconstruction of *ɨ̃, but as the nasal vowels generally behave the same as their oral counterparts in all other aspects, there is no

reason not to assume that Proto-Omagua–Kokama–Tupinambá had a low-frequent nasal *ɨ̃ phoneme as well.

4.1.2.5 *ũ

Set	OMG	KK	TPN	POKT	Example sets
$ũ_1$	u	w	ũ	*ũ	COMPANION, GRANDFATHER, ISLAND
$ũ_2$	u	u	ũ	*ũ	TONGUE
$ũ_3$	ú	u	ũ	*ũ	MOSQUITO

Table 4.12: *ũ correspondence sets.

The reconstruction of *ũ is based on five cognate sets, where Tupinambá has a preserved /ũ/ whereas in Omagua–Kokama they have merged with the reflex of POKT *u. In set $ũ_1$, *u has turned into a glide /w/ in Kokama following a consonant before a vowel (*u > w / C_V) as in e.g. GRANDFATHER: Tupinambá *amũja*, Omagua *amuj*, Kokama *amwi*. This process is also found in Omagua but is consdered to be underlyingly /u/; see section 2.4.1.2 on Omagua–Kokama glides.

4.1.2.6 *õ

Set	OMG	KK	TPN	POKT	Example sets
$õ_1$	u	o	õ	*õ	COPULATE, MOURN, TWO

Table 4.13: *õ correspondence sets.

The reconstruction of õ is based on four cognate sets, and behaves the same as set o_1 in Table 4.8, namely that both POKT *õ and *o merged with *u in Omagua–Kokama.

4.2 Reconstructed consonants

The phonological consonant inventory of Proto-Omagua–Kokama–Tupinambá is presented in Table 4.14. The inventory is similar to the reconstructed inventory for Proto-Tupí–Guaraní (PTG) by Jensen (1998:604), but not identical. First of all, PTG *tʃ and *ts merged to *ts in Proto-Omagua–Kokama–Tupinambá (Jensen, 1998:614). There is a marginal phoneme *ʃ in Tupinambá (see subsubsection 2.4.2.2), but it is rare in the Tupinambá data, and mostly seems to occur after /i/. According to Rodrigues (1958b:115), there are some minimal pairs, but as it does not occur in the data, it cannot be reconstructed for Proto-Omagua–Kokama–Tupinambá. Second of all, Jensen (1998) seemingly does not consider

the labiovelar stops to be phonemic in PTG, but rather consonant + glide sequences. For Proto-Omagua–Kokama–Tupinambá, I reconstruct *kʷ however, as the labiovelar is found in Tupinambá and Omagua, and is reconstructed for PTG by Rodrigues & Cabral (2012). Third of all, PTG *w corresponds to POKT *gʷ. Fourth of all, Tupinambá apparently had a labialized bilabial stop /pʷ/ (see subsubsection 2.4.2.2), but is only found in two cognate sets.

It is not entirely clear if the glottal stop *ʔ had phonemic status. It only survived in Tupinambá where its distribution is predictable, occurring between vowels to avoid hiatus, as in /ka.i/ [kaˈʔi] 'monkey sp.' vis-à-vis /kaj/ [kaj] 'to burn'. It has been included in this table as it is involved in sound changes related to vowel hiatus.

	Bilabial	**Alveolar**	**Palatal**	**Velar**	**Glottal**
Stop	*p (*pʷ)	*t		*k *g *kʷ *gʷ	(*ʔ)
Pre-nasalized stop	*ᵐb	*ⁿd		(*ᵑg)	
Fricative	*β				
Affricate		*ts			
Glides			*j		
Tap		*ɾ			

Table 4.14: Reconstructed POKT consonant phonemes.

4.2.1 Stops

Set	**OMG**	**KK**	**TPN**	**POKT**	**Example sets**
p_1	p	p	p	*p	BUTTERFLY, LATE, SOUND (V.)
$p^w{}_1$	pʷ	p	p	(*pʷ)	AFTER, BE ODOROUS
t_1	t	t	t	*t	ANTEATER, CORN, TAPIR
k_1	k	k	k	*k	AFTERNOON, CICADA, PIT VIPER SP.
$k^w{}_1$	kʷ	kw	kʷ	*kʷ	PAINT∼WRITE, SUN, YESTERDAY
$g^w{}_1$	w	w	gʷ	*gʷ	BIRD, JAGUAR, WALK (INTR.)
$g^w{}_2$	gʷ	w	u	*gʷ	ROCK∼SHAKE
$ʔ_1$	∅	∅	ʔ	(*ʔ)	AJÍ, CATFISH, TAPIR

Table 4.15: Correspondence sets for reconstructed stops.

The plain voiceless stops have been preserved in all daughter languages, represented in Table 4.15 by sets p_1, k_1, and t_1. I also reconstruct three labialized velar stops, *kʷ, *gʷ, and *pʷ, with *pʷ having very weak support from the data. It is only found in two sets: AFTER and BE ODOROUS, and is further discussed in Table 5.14. The reflexes of *kʷ remained unchanged in the daughter languages, whereas *gʷ turned to /w/ in Omagua–Kokama.

As lenition $g^w > w$ is typologically more common than fortition $w > g^w$, $*g^w$ is chosen for Proto-Omagua–Kokama–Tupinambá. For a further discussion of the nature of /g^w/ in Tupinambá, see section 2.4.2.2.

Tupinambá epenthensizes a glottal stop in V and VC syllables when these occur medially and finally (O'Hagan, 2013:3). The same [ʔ] occurs in other Tupí–Guaraní languages and was likely inherited from PTG *ʔ. This glottal stop was lost in Omagua–Kokama and many roots in Omagua–Kokama were shortened in the process, which is further discussed in subsection 5.1.3.

4.2.2 Nasals and prenasalized stops

Set	OMG	KK	TPN	POKT	Example sets
m_1	m	m	m	*m	ANTEATER, HORSEFLY, THING
$^{m}b_1$	m	m	^{m}b	*^{m}b	ASHES, FLY, SPIRIT
n_1	n	n	n	*n	CHAMELEON, PINEAPPLE, VOMIT (INTR.)
$^{n}d_1$	n	n	^{n}d	*^{n}d	BAT, KNEE, SPIDER
$ŋ_1$	n	n	ŋ	*ŋ	DRY, RED, WHITE

Table 4.16: Correspondence sets for reconstructed nasals and prenasalized stops.

In Tupinambá, the plain nasals alternate with prenasalized stops according to certain patterns which are further described in section 2.4.2.2. This allophony seems to have been eliminated in Omagua–Kokama which shows plain nasals in all instances where Tupinambá has a prenasalized stop, as seen in Table 4.16. Presumably the allophonic process is older (Jensen, 1998:579), so it is also reconstructed for Proto-Omagua–Kokama–Tupinambá. The exception is the prenasalized velar stop *^{ŋ}g, which presumably existed in POKT and Tupinambá based on symmetry, but cannot be said for certain due to orthographic ambiguity. For a further discussion of this, see section 2.4.2.2.

Omagua–Kokama also merged POKT *n and *ŋ to /n/ unconditionally, but the distinction is reconstructed to Proto-Omagua–Kokama–Tupinambá.

4.2.3 Affricates and fricatives

There are two sets for a Proto-Omagua–Kokama–Tupinambá affricate as see in Table 4.17. The first set, ts_1 has two reflexes, an affricate /ts/ in Kokama and /s/ in Tupinambá and Omagua. Since fortition of a fricative /s/ > /ts/ is very rare typologically, the reconstruction

Set	OMG	KK	TPN	POKT	Example sets
ts_1	s	ts	s	*ts	COVER (TR.), FISHING NET, SMOOTH
ts_2	ʃ	tʃ	s	*ts	HUNGRY, GRAB (TR.), KNIFE
β_1	w	w	β	*β	ARM, ARROW, WIND
β_2	∅	∅	β	*β	BED, HEAR (TR.), YELLOW
β_3	∅	∅	β	*β	DRUNK, SORUBIM SP., PEANUT
β_4	∅	w	β	*β	WRAP (TR.)

Table 4.17: Correspondence sets for affricates and fricatives.

*ts to Proto-Omagua–Kokama–Tupinambá is preferred for this set. The second set, ts_2, is the same segment but palatalized adjacent to /i/ in Omagua–Kokama. Since this palatalization does not occur in Tupinambá, it is not reconstructed to Proto-Omagua–Kokama–Tupinambá. A further discussion of this reconstruction is found in subsection 5.1.1.

The reconstruction of *β is based on four sets. Generally POKT *β was preserved in Tupinambá and had variable outcome in Omagua–Kokama. In certain roots, it was preserved as a glide /w/, shown in β_1, and in other roots it was elided, generally root-finally, shown in β_2 e.g. in BED: Tupinambá *tupaβ-*, Omagua–Kokama *tupa*, or medially, shown in β_3, as in SORUBIM SP.: Tupinambá *suruβi*, Omagua *surui*, Kokama *tsuri*. I reconstruct the *β as the Proto-Omagua–Kokama–Tupinambá phoneme for these sets, as a change from a fricative to an approximant is more likely as a type of lenition, than the opposite direction, which would be a type of fortition. In β_4, the cognate set is WRAP (TR.): Omagua *juana*, Kokama *juwana*, Tupinambá *uβana*. In this set, *β likely went to /w/ in Omagua as well, making this set identical to β_1, but has been interpreted synchronically as an allophonic glide in order to break up the vowel hiatus of /ua/, and therefore is not present in the phonemic representation.

The development of *β is further discussed in subsection 5.1.2, but appears to be part of a general constraint in Proto-Omagua–Kokama which disallows non-sonorants in coda position.

4.2.4 Glides and liquids

As for the glides, a *j is reconstructed in onset position (j_1) and coda position (j_3). In the case of j_3, the Omagua reflex is realized as [j] in the surface position but are considered to be underlyingly /i/, as described in section 2.4.1.2. Often, the /aj/ diphthong has been simplified in Kokama, but preserved in Omagua as underlying /ai/, represented in j_2. In

Set	OMG	KK	TPN	POKT	Example sets
j_1	j	j	j	*j	ARM, MOON, THORN
j_2	i	i	j	*j	DANCE, GRANDFATHER, PADDLE (N.)
j_3	i	j	j	*j	SNAKE, SPICY, TWO
j_4	j	j	ɲ	*j	LOSE ONESELF, MOSQUITO, SPIDER
j_5	∅	j	∅	–	SWEAT (INTR.)
j_6	j	j	∅	–	HEART
w_1	∅	w	∅	–	MANIOC FLOUR
$ɾ_1$	ɾ	ɾ	ɾ	*ɾ	ANCHIOTE, BASKET, ROAST (TR.)

Table 4.18: Correspondence sets for glides and liquids.

j_4, the reflex is virtually the same as in j_1, but the Tupinambá form has been allophonically nasalized due to nasal harmony. In j_5, a glide was inserted into the sequence /ɨi/ in Kokama, whereas the whole vowel sequence was simplified to /i/ in Omagua. In j_6, a glide was inserted into the sequence /ia/ < /ɨa/. Since these two are originally allophonic glides, they are not reconstructed to Proto-Omagua–Kokama–Tupinambá.

In w_1, a glide was inserted into the sequence /ui/ in Kokama. This is also the case in Omagua, but is considered an allophonic process there, whereas the form in Vallejos & Amías (2015) is listed as *uwi*. Nevertheless, this was originally an allophonic glide, and is therefore not reconstructed to Proto-Omagua–Kokama–Tupinambá.

The only liquid which is reconstructed is the alveolar tap *ɾ which is found in all daughter languages, alongside the Kukamiria dialect of Kokama where the allophone [l] is also found.

4.3 Preliminary 'sporadic' sound changes

A *sporadic sound change* is defined in Campbell (2013:23) as a sound change which does not affect all possible words, only one or a few, and it cannot be predicted which words the sound change will affect. As an example he gives the change from r > l in English *glamour* (< *grammar*), or the loss of /r/ in *speech* (< Old English *spræc*).

In addition to the sets given earlier in this chapter, there are thus a number of sets with correspondences which are more irregular than those listed above, and are only found in very few sets, generally 1-2. In very few cases, the change is found in more sets, but does not have a clear condition or motivation, and is therefore also included in this list. These

changes have been named 'sporadic sound changes', but it should be noted that several may appear sporadic, but would in fact be conditioned if more data were available, and may also be caused by other factors such as analogy.

The sets shown in Table 4.19 and Table 4.20 were manually extracted by going through Appendix A and extracting each cognate set which could not immediately be explained by the more 'regular' sound changes described elsewhere in this thesis. The sets in question have been split up into four categories: (a) epenthesis, i.e. where a segment was seemingly inserted into a word, (b) frozen nasal harmony, where an oral consonant was replaced by a nasal equivalent, (c) loss of segment, i.e. where a segment was sporadically lost in a word. The fourth category is shown in Table 4.20 which is a larger list of seemingly random correspondences, mostly between vowels. These correspondence sets are shown here, and further discussed in subsection 5.1.7.

Type	OMG	KK	TPN	Sets	OMG	KK	TPN
EPENTHESIS	ɪ	e	∅	BRING (TR.)	ɪɾuɾi	eɾuɾa	ɾuɾa
	ɨ	ɨ	∅	NEPHEW	mɪmɨɾɨa	memɨɾɨa	membɨɾa
				SON (M. EGO)	taɨɾɨa	taɨɾɨa	taʔɨɾa
	∅	∅	β	PEANUT	munui	muni	manduβi
	i	i	∅	THING	maɾai	maɾi	maɾã
NASAL HARM.	m	m	p	BELLY BUTTON	mɨɾua	mɨɾwa	puɾuʔã
	m	m	p	CURE	musana	mutsana	posaŋa
	j	nj	∅	FELLOW MAN	japɪsaɾa	njapitsaɾa	apiʃaɾa
LOSS OF SEGMENT	∅	∅	u	ACHIOTE	ɾuku	ɾuku	uɾuku
	u	a	ua	ANTEATER	tamanu	tamana	tamanduá
	∅	∅	ɨ	CUTBANK	ɨwama	ɨwama	ɨβɨʔama
				DAUGHTER	taiɾa	tajɾa	tajɨɾa
				BE ODOROUS	ʃapuni	tʃapuni	tɨap^{w}ana
	mɨ	mɨ	emi	PET	mɨma	mɨma	emimbaβa
				PREY	miaɾa	mjaɾa	embiaɾa
	n	n	∅	MOUSE∼RAT	sanuja	tsanuja	saujá
	∅	∅	a	ROCK∼SHAKE	wɨuta	wɨwɨta	moag^{w}ɨag^{w}ɨ
	∅	ɨ	ɨ	ROCK∼SHAKE	wɨuta	wɨwɨta	moag^{w}ɨag^{w}ɨ
	wɪ	∅	βe	FLY (INTR.)	uwɪwɪ	uwe	βeβé

Table 4.19: Three types of more sporadic changes: epenthesis, frozen nasal harmony, and loss of segment.

Table 4.20: Sets with irregular correspondences

OMG	KK	TPN	Sets	OMG	KK	TPN
p	p	p^{w}	BE ODOROUS	ʃapuni	tʃapuni	tɨap^{w}ana
a	a	ɨ	FLY (INTR.)	uwɪwɪ	uwe	βeβé
ɨ	ɨ	é	FLY (INTR.)			
ʃ	tʃ	t	BE ODOROUS	ʃapuni	tʃapuni	tɨap^{w}ana
u	u	a	BE ODOROUS			
i	i	a	BE ODOROUS			
i	e	j	BE SCARED			
ɨ	ɨ	o	BORROW	ɨpɨɾu	ɨpɨɾu	poɾu
u	∅	a	CALL (TR.)	sapukui	tsapuki	sapukaja
u	u	ɨ	CHAMELEON	sɪnɪmu	tsenemu	senembɨ
a	a	e	COLLARED PECCARY	taitatu	tajtatu	tajtetu
ai	e	oʔí	COVER (TR.)	jasai	jatse	asoʔí
ɪ	e	i	CULTIVATE (TR.)	kupɪ	kupe	kopiɾa
			DEFECATE			
ɨ	ɨ	u	DEER			
∅	∅	a	DEFECATE (INTR.)	kapɪ	kape	kaʔapia
u	i	ɨ	DIG	ʃukai	tʃiwiki	sɨβɨkoja
∅	w	β	DIG			
∅	i	ɨ	DIG			
a	∅	o	DIG			
i	i	j	DIG			

a	a	o	DOOR	jakina	jakina	okendaβa
u	ɨ	ɨ	DRIP (INTR.)	atukɨra	atɨkɨɾɨ	tɨkɨra
∅	∅	β	DRUNK	saipuɾa	tsajpuɾa	saβeɨpoɾa
∅	∅	e	DRUNK			
i	j	ɨ	DRUNK			
∅	∅	à	ENCOUNTER	sawiti	tsawiti	soβàitĩ
a	a	ei	ENTER	aki	aki	eiké
ɪ	e	ɨ	FLAME (V.)	sɪnɪ	tsene	sendɨ
e	i	ɨ	FORGET (TR.)	sɨsaɾai	tsitsaɾi	sesaɾaja
ɪ	i	i	FOUR	ɪɾuaka	iɾwaka	iɾu
i	i	ɨ	GRAB	japiʃika	japitʃika	pɨsɨka
ɨ	ɨ	a	HEAD	jakɨ	jakɨ	akaŋa
∅	∅	ŋ	HEAD			
ĩ	i	ɨ	HEART	ĩja	ija	ɲɨʔa
∅	∅	ɲ	HEART			
ɨ	ɨ	u	LAKE	ɨpasu	ɨpatsu	upaβa
n	n	∅	LONG TIME AGO	ɨmɨnua	ɨmɨnwa	umuã
a	a	au	POOR	paɾiasu	paɾjatsu	poɾeausuβa
ɨ	ɨ	i	RED	pɨtani	pɨtani	piɾaŋa
ɨ	ɨ	u	REST (INTR.)	japɨtu	japɨtu	putuʔú
p	p	β	SHINE (INTR.)	pɪɾa	peɾa	βeɾaβa
ɨ	ɨ	e	SHOUT	sasɨma	tsatsatsɨma	sasema
∅	∅	i	STAR	sɪsu	tsetsu	seiʃu
ɪ	u	e	THIGH	sɪtɨma	tsutɨma	etɨmã
a	i	i	THROW (V.)	atika	itika	itɨka
ɪ	e	ɨ	WIDE	ɪpɪwasu	epewatsu	pɨgʷasu
a	a	ɨ	WORM∼LARVA	sasuka	tsatsuka	sɨsoka
a	a	e	YESTERDAY	ikʷaʃi	ikwatʃi	kʷeisé

Chapter 5

Discussion

The following chapter will take the phonological correspondences presented in the Results chapter, compile the phonological developments and examine and discuss them in greater detail. Phonological changes that were only found in very few sets (named 'sporadic changes') will also be discussed on a case-to-case basis in order to account for all the data. Many discussions of this sort reference Tupí–Guaraní cognate data, which are drawn from the Tupí–Guaraní Comparative Lexical Dataset (Chousou-Polydouri et al., 2019).

5.1 Phonological developments

5.1.1 Development of *ts

In almost all cases, Tupinambá has the most conservative phonology of the three languages, quite understandably, being an older language and without the same substrate influence. There are exceptions to this however, one being the Proto-Omagua–Kokama–Tupinambá affricate *ts, where Tupinambá and Omagua show the reflex /s/, and Kokama preserves /ts/. The opposite scenario, i.e. an affricatization from s > ts in Kokama is less likely, since deaffricatization or lenition (ts > s) is more common than affricatization or fortition (s > ts). The affricate *ts is what is reconstructed for PTG as well (Schleicher, 1998:13-19).

As seen in Table 5.1, POKT *ts, palatalized in Omagua and Kokama before /i/.[1] In some cases

[1] There are very few words in Vallejos & Amías (2015) that begin with /tsi/, one example being 'eye', found in the set EYE∼FACE. According to Vallejos (2010:47) it is very frequently realized as [tʃitsa], so the motivation for not spelling it ⟨chitsa⟩ is unclear.

the palatalization trigger was lost as well in a vowel hiatus resolution process, resulting in the phonemes /tʃ/ in Kokama, and /ʃ/ in Omagua. In many cases the palatalization occurred after a change *e > i, shown in (a), meaning that the raising of *e must have occurred prior to the palatalization. A similar case is shown in (b), where palatalization occurred after the change ɨ > i.

In some cases, shown in (c), the sound change ɨ > i did not occur. It is difficult to come up with a clear-cut explanation for this, but one can notice that most (but not all) of the examples in (b) have a labial element, which possibly blocked ɨ > i. It is however hard to imagine why ɨ > i occurred, in e.g. the HURT set, but not in the MOON set, where the only distinguishing segment is an initial palatal glide.

Interestingly, there are no instances of the sequence *tsi in POKT, which means that all cases of palatalization in Omagua–Kokama come from either older *tse or *tsɨ, but not *tsi.

	Set	**Omagua**	**Kokama**	**Tupinambá**
(a)	ARRIVE (INTR.)	jauʃima	jawatʃima	gʷasema
	CRY (INTR.)	jaʃua	jatʃu	jaseʔó
	DANCE (INTR.)	japuɾaʃi	japuɾatʃi	poɾaseja
	KNIFE	kɨʃi	kɨtʃi	kɨsé
	LEAVE (TR.)	iʃaɾi	itʃaɾi	sejaɾa
	LEAVE FROM (INTR.)	uʃima	utʃima	sema
	LOOK FOR (TR.)	ʃikaɾi	tʃikaɾi	sekaɾa
	MANY∼MUCH	ʃita	tʃita	setá
	NECK∼THROAT	jaʃuka	jatʃuka	aseoka
	YESTERDAY	ikʷaʃi	ikwatʃi	kʷeisé
(b)	DIG	ʃukai	tʃiwiki	sɨβɨkoja
	GRAB	japiʃika	japitʃika	pɨsɨka
	HEAVY	ipuʃi	iputʃi	posɨja
	HUNGRY	jamaʃi	jamatʃi	ambɨasɨ
	HURT	saʃi	tsatʃi	asɨ
	SUN	kʷaɾaʃi	kwaɾatʃi	kʷaɾasɨ
	VINE	iʃipu	itʃipu	ɨsɨpó
(c)	ANT	sasɨwa	tsatsɨwa	tasɨβa
	BE SCARED	ɨsɨi	ɨtse	sɨja
	CUT	jasɨkata	jatsɨkataka	asɨka
	GO DOWNRIVER (INTR.)	asɨɾɨka	atsɨɾɨka	sɨɾɨka
	MOON	jasɨ	jatsɨ	jasɨ
	SMOOTH	ɨsɨma	ɨtsɨma	sɨma
	SWEAT (INTR.)	sɨ̃i	tsɨji	sɨʔaja

Table 5.1: Palatalization in Omagua & Kokama.

	Set	Omagua	Kokama	Tupinambá
(a)	BEARD	muta	muta	amotaβa
	BED	tupa	tupa	tupaβa
	DOOR	jakina	jakina	okendaβa
	DRY (V.)	tɨpa	tɨpa	tɨpaβa
	FRIEND∼LOVER	tɨwasa	tɨwatsa	atɨβasaβa
	HEAR (TR.)	sɪnu	tsenu	senduβa
	PEANUT	munui	muni	manduβi
	PET	mɨma	mɨma	emimbaβa
	POOR	paɾiasu	paɾjatsu	poɾeausuβa
	SORUBIM SP.	suɾui	tsuɾi	suɾuβi
	STRAIN (TR.)	jumukua	jumuka	mog^{w}aβa
	YELLOW	iju	iju	juβa
(b)	ARM	jɨwa	jɨwa	jɨβá
	BROTHER (F. EGO)	kɨwɨɾa	kɨwɨɾa	kɨβɨɾa
	CLOUD∼FOG	ɨwɨtini	ɨwɨtini	ɨβɨtiŋa
	COCKROACH	aɾawɪ	aɾawi	aɾaβé
	CORN	awati	awati	aβati
	CUTBANK	ɨwama	ɨwama	ɨβɨʔama
	FRIEND∼LOVER	tɨwasa	tɨwatsa	atɨβasaβa
	HIGH UP (ADV.)	ɨwati	ɨwati	ɨβaté
	WRAP	juana	juwana	uβana
(c)	ARROW	uwa	uwa	uʔúβa
	BITTER	ɪɾawa	ɪɾawa	ɾoβa
	FAT∼LARD	ikawa	ikawa	kaβa
	FLAT	pɪwa	pewa	peβa
	LOUSE	kɨwa	kɨwa	kɨβa
(d)	ANT	sasɨwa	tsatsɨwa	tasɨβa
	CROSS (V.)	sasawa	tsatsawa	sasaβa
(e)	ALSO∼ALREADY	awi	aj	aβé
	DRUNK	saipuɾa	tsajpuɾa	saβeɨpoɾa

Table 5.2: Development of POKT *β.

5.1.2 Development of *β

As mentioned in subsection 4.2.3, Tupinambá preserves POKT *β whereas in Omagua–Kokama, the development is either /w/ or ∅ , as shown in Table 5.2. The development seems to be *β > ∅ in final syllables, generally root-finally, shown in (a), e.g. POKT **tupaβ-* 'bed', Tupinambá *tupaβa*, Omagua–Kokama *tupa*, and *β > w in non-final syllables, shown in (b), e.g. POKT **aβati*, Tupinambá *aβati*, Omagua–Kokama *awati*. In most of the cases in (a) the Tupinambá vowel *-a* is a suffix and not part of the root, making most of the /β/ in (a) root-final, except for PEANUT and SORUBIM SP. In Tupinamá nouns for instance, the final *-a* is in many cases a nuclear case or a gerundive suffix, see subsection 2.3.2. Therefore,

it seems as if the deletion of *β is connected to the requirement of vowel-final roots in Omagua–Kokama, as described in subsubsection 2.4.1.4.

There are however a number of cases where *β > w occurred in Omagua–Kokama even when *β was root final, shown in (c), again in Table 5.2. This is apparently related to root size, as almost all of examples in (c) are cases where the resulting root would be monomoraic if *β were lost. Therefore, to satisfy the minimal root size of two morae, *β was retained. This can be seen in e.g. POKT **uʔuβ-* 'arrow', Tupinambá *uʔuβa*, Omagua–Kokama *uwa*. Since Omagua–Kokama underwent loss of *ʔ and subsequent vowel shortening, the resulting root would have been *†u* in Omagua–Kokama. In (a), however, the resulting root after the loss of *β is longer than two morae, the exception being YELLOW *iju*, which was instead augmented to fill the requirement by the addition of the initial /i/.

In two cases, *β did not go to /w/ even in longer roots, shown in (d). Looking at cognates in other Tupí–Guaraní languages, some languages seem to preserve root-final *β in some form. In the set LAKE for instance, the Tupinambá form is *upaβa*. In Tembé and Guajajara, the cognates are *ipaw* and *ypaw* respectively, but in Aché and Paraguyan Guaraní, they are *i̷pa* and *ypa*. Looking at the cases in (d), the cognates of the ANT and CROSS (v.) are *tahĭ̷w* and *ahaw* in Tembé, and *tahyw* and *wahaw* in Guajajara. However, in Aché they are *tabiu* and *tahyí*, and in Paraguayan Guaraní *watʃa* and *(a)hasa*, i.e. generally without a reflex of *β as expected. The exception is the Aché form *tabiu* where the final /u/ could possibly represent a reflex, but it is unclear whether the medial /b/ has any connection. Nevertheless, it does not seem like these two words in particular had a unique development in other Tupí–Guaraní languages. It is difficult to provide an explanation for why *β was preserved in Omagua–Kokama in these particular sets, but perhaps the final *-a* was treated as part of the root rather than a suffix in these particular words, so that the words in (d) really belong in (b).

The word for ANT is somewhat irregular for other reasons as well as the original initial POKT *t seems to have lenited to *ts in Proto-Omagua–Kokama.

In one case, shown in (e), *β > w as expected in Omagua, but was lost in Kokama where the final vowel was turned into a glide: *e > i > j, resulting in a monosyllabic root. Given the small size of this root, the expected outcome would be to preserve the consonant as in the examples in (b). In DRUNK, the expected outcome *β > w did not occur, and *β was deleted

instead. It is unclear why it was not preserved, but could be part of a general reduction of the /aβeɨ/ sequence.

5.1.3 Vowel hiatus resolution

As described in subsubsection 2.4.1.4, Omagua–Kokama disallows vowel hiatus within a syllable, which means that such VV contexts had to be resolved in Proto-Omagua–Kokama. A common source of vowel hiatus was the loss of POKT glottal stop *ʔ. POKT *ʔ was found in V and VC syllables when these occurred medially and finally, in e.g. [tapɨˈʔɨra] 'tapir'. It was then lost in Proto-Omagua–Kokama creating vowel hiatus, as glottal stop always occurred intervocalically. This glottal stop was most likely inherited from PTG given its presence in other Tupí–Guaraní languages, e.g. Tembé *tapi'ir*, Warazu *tapɨ́ʔɨ*, Kamaiurá *tapi'it*. Loss of glottal stop is the most common source of vowel hiatus, but there are a few other examples. There are generally no diphthongs in Omagua–Kokama, nor long vowels (see subsection 2.4.1), so vowel hiatus was resolved in a number of ways, shown in Table 5.3.

As shown in (a), two identical vowels were shortened to one, VV > V, e.g. POKT **kapiʔi* 'grass', Tupinambá *kapiʔi*, Omagua–Kokama *kapi* (< †*kapii*). At least in one case, Omagua seems to have preserved both vowels, in POKT **seʔẽ* 'sweet', Tupinambá *seʔẽ*, Omagua *sii*, Kokama *tse* (< †*tsee*). Perhaps this is a way of preserving root size, if †*sɪ* would be considered too small a root to Omagua speakers, see subsection 5.1.5. In TERMITE in (a), POKT */iʔi/ seemingly turned to /ia∼ea/. Presumably, the *-a* here is originally a frozen suffix, so the result of the reduction was rather /i/, hence the placement in group (a). The *i > e change in Kokama is most likely vowel opening before the final /a/.

If the first element of the sequence was *i or *u, they became their respective glides /j/ and /w/, as shown in (b). In Omagua, these are considered to be underlyingly /i/ and /u/ however and are therefore represented as such in the table, as explained in subsection 2.4.1. Some of these also come from the raising of *e next to coronals: *e > i > j, as in TIRED.

In (c), we see some exceptions to (b), i.e. where the first vowel of the sequence did not turn into a glide. In ANTEATER, the final sequence /ⁿdua/ was simplified in Proto-Omagua–Kokama to /nu/ in Omagua and /na/ in Kokama. Similarly in SPIDER, the final /ĩ/ was deleted. The *ⁿd > /n/ change is regular, but based on (b), we would assume the first *u of these sequences to

	Set	Omagua	Kokama	Tupinambá
(a)	AJÍ	ɨkɨi	ɨki	kɨʔɨɲa
	ARROW	uwa	uwa	uʔúβa
	BRAIN	apɪtuma	apetúma	aputuʔuma
	CATFISH	mani	mani	mandiʔí
	DEFECATE (INTR.)	kapɪ	kape	kaʔapia
	GRASS	kapi	kapi	kapiʔí
	INDIGENOUS PERSON	tapɨja	tapɨja	tapɨʔɨja
	REST (INTR.)	japɨtu	japɨtu	putuʔú
	SING (INTR.)	ikaɾa	ikaɾa	ɲeʔeŋaɾa
	SWEET	sɪɪ	tse	seʔẽ
	TAPIR	tapiɾa	tapiɾa	tapiʔíɾa
	TERMITE	kupia	kupea	kupiʔí
(b)	BELLY BUTTON	mɨɾua	mɨɾwa	puɾuʔã
	DAWN	kʷɪma	kwema	koʔema
	LONG TIME AGO	ɨmɨnua	ɨmɨna	umuã
	MOSQUITO	jatiú	jatju	ɲatiʔũ
	ROUND	japua	japwa	apuʔã
	TIRED	kaniú	kanju	kaneʔõ
	VULVA	tamatia	tamatja	tamatiʔá
(c)	ANTEATER	tamanu	tamana	tamanduá
	BE ODOROUS	ʃapuni	tʃapuni	tɨapʷana
	CRY (INTR.)	jaʃua	jatʃu	joseʔó
	DIG	ʃukai	tʃiwiki	sɨβɨkoja
	FIREFLY	mua	muwa	mamuã
	NECK∼THROAT	jaʃuka	jatʃuka	aseoka
	SPIDER	janu	janu	ɲanuĩ
(d)	ISLAND	ɨpau	ɨpwa	ɨpaʔũ
	SPIRIT	mai	maj	mbaʔé
(e)	LIVER∼HEART	pɨa	pɨa	pɨʔá
	SLOTH	aɨ	aɨ	aʔɨ
	SON (M. EGO)	taɨɾɨa	taɨɾɨa	taʔɨɾa
	SWEAT (INTR.)	sɨ̃ɨ	tsɨji	sɨʔaja
(f)	CUTBANK	ɨwama	ɨwama	ɨβɨʔama
	SISTER-IN-LAW (F. EGO)	uki	uki	ukeʔí

Table 5.3: Vowel hiatus resolution.

turn into a glide /w/. Possibly, this change is due to the infrequency of syllable types such as /nua/ [nwa] in Omagua and Kokama, and the change could therefore possibly be induced by a phonotactic constraint. For instance, Sandy & O'Hagan (2020) report only one instance of a /nwa/ syllable in Omagua. A similar correspondence can be seen in the Kokama cognate of Tupinambá *atuá* 'nape' namely *ata*∼*atu* (Vallejos & Amías, 2015:40). A similar alternation is found also internally in Tupinambá, e.g. *ɨmuã*∼*ɨmã* 'already' (Barbosa, 1970:124). However, we also see a preserved sequence alongside a simplified one in *ɨmɨnwa*∼*ɨmɨna* 'long time ago' (Vallejos & Amías, 2015:76), and a /ɾua/ sequence in the BELLY BUTTON set, Omagua *mɨrua*, Kokama *mirwa*, cognate to Tupinambá *puruã*, so no hard conclusions can be drawn.

The same phenomenon can be seen in BE ODOROUS, CRY (INTR.), DIG, and NECK∼THROAT in (c), in which the first element of the vowel sequence was deleted after causing palatalization on the preceding *ts. This means that at least this specific type of vowel hiatus resolution must have occurred after both the raising of *e > i and the subsequent deletion of /i/.

In FIREFLY, the *u did not turn into a glide. Instead *ua was broken up into two different syllables with an allophonic glide [w]. This is somewhat expected, since Omagua synchronically does not display this process when the vowel sequence is final, hence /ikua/ [i.ˈku.a], but /ikua-pa/ [i.ˈkwa.pa], cf. section 2.4.1.2.

Similarly, in one case, the POKT sequence *koʔe turned into /kʷe/ in Omagua–Kokama: POKT **koʔema* 'dawn', Tupinambá *koʔema*, Omagua *kʷɪma*, Kokama *kwema*. Presumably this went via an intermediary stage of **kuema*, after the change of *o > *u in Proto-Omagua–Kokama.

As shown in (d), the same glide transformation also seems to have occurred when *i and *u were the second element of the sequence, although the examples are scarce. In SPIRIT in (d), final *e raised regularly to /i/ and then turned into a glide: POKT **mbaʔe* 'spirit', Tupinambá *mbaʔe*, Omagua *mai*, Kokama *maj*. In ISLAND however, there was a sporadic metathesis in Kokama, leading to the form *ɨpwa* instead of expected *†ɨpaw*, which is what we find in Omagua.

In the case of POKT *aʔɨ and *ɨʔa sequences, they were more or less preserved, as in POKT **aʔɨ* 'sloth', Tupinambá *aʔɨ*, Omagua–Kokama *aɨ*, shown in (e). The sequence /aɨ/ could be described as the only phonemic diphthong in Omagua–Kokama (Vallejos, 2010:59-60; Sandy & O'Hagan, 2020:112). In LIVER∼HEART in (e), an epenthetic segment is inserted in Kokama

	Set	Omagua	Kokama	Tupinambá
(a)	AJÍ	ɨkɨi	ɨki	kɨʔɨɲa
	BURN	ukai	uki	kaja
	CALL (TR.)	sapukui	tsapuki	sapukaja
	DIG	ʃukai	tʃiwiki	sɨβɨkoja
	FALL (INTR.)	ukukui	ukuki	kuja
	FISH (TR.)	sɪkɨi	tsɨki	sekɨja
	FORGET (TR.)	sɨsaɾai	tsitsaɾi	sesaɾaja
	LOOK (INTR.)	umai	umi	maʔẽ
	PIRANHA	ipiɾai	ipiɾi	piɾaɲa
	SCRAPE (TR.)	kaɾai	kaɾi	kaɾãja
	THING	maɾai	maɾi	maɾã
	YOUNG WOMAN	kuniatai	kunjati	kuɲãtaĩ
(b)	ALL	upai	upi	opaβĩ
	PEANUT	munui	muni	manduβi
	SORUBIM SP.	suɾui	tsuɾi	suɾuβi
(c)	SNAKE	mui	muj	mboja
	SPICY	tai	taj	taja
	SPIRIT	mai	maj	mbaʔé
	DUST∼SAND	kui	kuj	kuj
	TOOTH	ai	aj	ãja

Table 5.4: Vowel-glide simplification in Kokama final syllables.

to break up the vowel sequence, either [ɣ] or [z] (Vallejos, 2010:59). Such an epenthetic sound is not reported for Omagua (Sandy & O'Hagan, 2020:112-113), but in both languages this sequence is syllabified in two syllables. In SWEAT (INTR.), the presumably root-final *-aj* was reduced to *-i* in both Omagua and Kokama. Normally, this development happened only in Kokama unless the root would be too short, as described in subsubsection 5.1.3.1. In this case, the reason is probably because the loss of *ʔ would yield a triple vowel sequence /ɨ̃ai/, which was reduced to /ɨ̃i/.

In (f) in Table 5.3, in the set CUTBANK, the first vowel element *ɨ was lost altogether instead of perhaps expected *†ɨwɨama*. In SISTER-IN-LAW (F. EGO), *e was apparently lost altogether as well. Presumably also in this case, *e raised to /i/ due to influence from the final vowel, and then disappeared in the same vowel shortening process that yielded *kapi* 'grass' in Omagua–Kokama above.

5.1.3.1 Development of final vowel-glide sequences

A similar sound change to those in Table 5.3 is the development of original POKT sequences of a vowel + a palatal glide /Vj/ in Kokama shown in Table 5.4. The difference is that these

are not vowel hiatus resolutions, as the second element in POKT was a glide and not a vowel, otherwise they would be broken up by a glottal stop. These /Vj/ sequences were either inherited from Proto-Omagua–Kokama–Tupinambá, or from Proto-Omagua–Kokama following a loss of some medial segment, typically *ʔ or *β. These sequences were reduced in Kokama to only a final /i/, after the split from POK. This can be seen in examples (a-b) in Table 5.4, for example POKT **kaɾãj-* 'to scrape' > POK **kaɾaj* > Kokama *kaɾi*, or POKT **tsuruβi* > POK **tsuruj* > Kokama *tsuri*. In these cases, Omagua preserved the sequence, which is analyzed by Sandy & O'Hagan (2020) to be underlyingly /Vi/ but [Vj] in the surface representation. I believe that this is a synchronic analysis, and that they were nevertheless derived from POK *Vj, as Omagua–Kokama do not allow diphthongs and because such final /Vi/ sequences are not found in Tupinambá.

This simplification did not occur in monosyllabic roots, as evident from the examples in (c) where the simplification has not occurred. This is most likely due to a minimum root length requirement, which does not allow roots to be too short.

5.1.4 Development of *e and *ẽ

Whereas *e was preserved in Tupinambá, POKT *e developed in different directions in Omagua–Kokama based on the position of the sound and the surrounding environment, as shown in Table 5.5. Shown in (a), *e was preserved in Tupinambá and Kokama in most cases, but raised to /ɪ/ in Omagua. In Kokama, /e/ has a variable realization and may be realized as [ə∼ɪ] alongside [e] (Vallejos, 2010:75). There are many examples of this correspondence, but for Table 5.5, only three have been chosen as representative.

When *e was stressed and final, it was raised to [i] in Omagua–Kokama. This process is apparently still productive in Kokama, as /e/ and /i/ are in free variation word-finally (Vallejos, 2010:76). Following a coronal consonant, commonly *ts, *e was also raised to /i/ in Omagua–Kokama, as shown in (c). Often, *ts was palatalized after this change to /tʃ∼/ʃ/, see subsection 5.1.1.

Although the evidence for this is somewhat scarce, it seems like an initial *je- sequence developed into /i/ in Proto-Omagua–Kokama, shown in (d). In the case of Tupinambá *ɲeʔeŋaɾa*, [ɲ] is a possible allophone of /j/ when occurring to the left of a nasal segment, in this case [ŋ]. In the set BAD, POKT **jeatsej-* did not yield *†iatsej(a)*, but apparently metathesized to

	Set	Omagua	Kokama	Tupinambá
(a)	COPULATE	mɪnuka	menuka	menõ
	FLY	mɪɾu	meɾu	mbeɾu
	OPEN (TR.)	ɪpɪka	epeka	peká
(b)	HIGH UP (ADV.)	ɨwati	ɨwati	ɨβaté
	KNIFE	kɨʃi	kɨtʃi	kɨsé
	SHAMAN	sumi	tsumi	sumé
(c)	INVITE (TR.)	paɾisaɾa	paɾitsaɾa	paɾesaɾa
	LOOK FOR (TR.)	ʃikaɾi	tʃikaɾi	sekaɾa
	MANY∼MUCH	ʃita	tʃita	setá
(d)	SING (INTR.)	ikaɾa	ikaɾa	ɲeʔeŋaɾa
	SWEET POTATO	itika	itika	jetɨka
	BAD	aisɪ	ajtse	jeaseja
	RETURN	ɪɾɪwa	iɾiwa	jeɾeβa
(e)	FLY (INTR.)	uwɪwɪ	uwe	βeβé
	BACK	jatukupɪ	jatukupe	atukupé
	COCKROACH	aɾawɪ	aɾawi	aɾaβé
(f)	RETURN	ɪɾɪwa	iɾiwa	jeɾeβa
	SMELL	sɪtuni	tsetuni	setuna
(g)	NAIL	pɨsapi	pɨtsape	pɨsapẽ
	SPLIT (TR.)	pɪsɪ	petse	pesẽ
	SWEET	sɪɪ	tse	seʔẽ

Table 5.5: Development of POKT *e and *ẽ.

Omagua *aisɪ* and Kokama ajtse. In RETURN *je seemingly went to /ɪ/ in Omagua rather than /i/. This is most likely a case of the same vowel harmony discussed in subsection 5.1.6, where it probably went through the regular form **iɾɪwa* before assimilating to the word-medial vowel. However, this does not explain why the medial vowel did not go to /i/ as in Kokama, as usual next to coronals.

Finally, as shown in (e), it seems like some final *e were not raised to /i/ but rather behave, or partially behave as the sets in (a). In the case of FLY (INTR.), it might be explained by the likeness of the two syllables or some kind of iconicity, but it does not explain BACK. In COCKROACH the regular change occurred in Kokama but not in Omagua.

Similarly in (f), the sound change conditioned by a coronal did not, or did only partially occur in the sets RETURN and SMELL. As for RETURN it occurred regularly in Kokama but not in Omagua. It cannot be said why not, but the RETURN set is also characteristic in that the *je > i change explained in (d) did not occur in this particular word either.

As for the sets (g), they show that nasal *ẽ seemingly developed differently. More data is

Set	Omagua	Kokama	Tupinambá
AXE	jɨɨ	jɨ	jɨ
CHACRA	kuu	ku	kó
NOSE	tii	ti	tĩ
PATH	pɪɪ	pe	pé
SWEET	sɪɪ	tse	seʔẽ

Table 5.6: Examples of the bimoraic minimal root size in Omagua.

needed to give a certain conclusion, but this suggests that while *e was raised to /i/ finally, nasal *ẽ was not. This would be an interesting finding if confirmed, since vowel nasalization generally did not leave any trace in Proto-Omagua–Kokama. It is possible that it is a type of vowel harmony as well, similar to the case of FLY (INTR.) discussed above. An exception is Omagua *pɨsapi* which shows regular development in its final vowel.

5.1.5 Minimal root size in Omagua

There is some evidence for that simple CV roots are not permitted in Omagua, and in many cases where a CV root is expected and found in Kokama, Omagua shows CVV instead. Presumably there is a phonological constraint in Omagua which relates to root size, disallowing small monomoraic roots. A bimoraic root is a quite common word minimality requirement across the languages of the world; many languages — such as English (Harris, 1994:261) — disallow words consisting of only one light CV syllable. In certain languages such as Lardil or Mohawk, such roots are augmented to fit the minimal requirement (Piggott, 2010:1). In the case of Omagua, this seemingly leads to the blocking of the sound change VV# > V# in certain cases. Compare e.g. the regular change *iʔi > *ii > i in POKT **mandiʔi* > Omagua–Kokama *mani*, to the counterexample POKT **tseʔẽ* > Kokama *tse*, but Omagua *sɪɪ* (expected †*sɪ*).

In other cases it leads to root augmentation by lengthening V > VV, shown in Table 5.6, as in AXE: POKT **jɨ* > Kokama *jɨ*, but Omagua *jɨɨ* (expected †*jɨ*), or NOSE: POKT **tĩ* > Kokama *ti*, but Omagua *tii* (expected †*ti*).

The presence of this lengthening rule leads us to believe that diachronically there was no blocking of VV# > V# in Omagua, and presumably this sound change occurred already in Proto-Omagua–Kokama. Omagua then innovated the lengthening rule and lengthened all CV roots to CVV, making it seem like the VV > V change never took place in Omagua. By

	Set	Omagua	Kokama	Tupinambá
(a)	FAT∼LARD	ikawa	ikawa	kaβa
	FISH	ipiɾa	ipiɾa	piɾá
(b)	AJÍ	ɨkɨi	ɨki	kɨʔɨɲa
	BE FAT	ɨkɨɾɨwasu	ɨkɨɾatsu	kɨɾá
	BE SCARED	ɨsɨi	ɨtse	sɨja
	BORROW	ɨpɨɾu	ɨpɨɾu	poɾu
	GREEN	ɨkɨɾa	ɨkɨɾa	kɨɾa
	NEW	ɨpɨsasu	ɨpɨtsatsu	pɨsasu
	NIGHT	ɨpɨsa	ɨpɨtsa	pɨsajé
	SMOOTH	ɨsɨma	ɨtsɨma	sɨma
(c)	OPEN (TR.)	ɪpɪka	epeka	peká
	WIDE	ɪpɪwasu	epewatsu	pɨgʷasu

Table 5.7: Vowel harmony in Omagua and Kokama.

posing this lengthening rule, we do not have to assume that vowel shortening was innovated separately in Omagua and Kokama.

5.1.6 Vowel harmony

There are certain traces of vowel harmony that can be found in Omagua and Kokama, shown in Table 5.7. Most often this is found in the initial vowel *i-* stemming from the POKT 3rd person prefix *i-*. In Omagua and Kokama, this initial vowel will be realized as *ɨ-* if the vowel of the following syllable is /ɨ/, shown in (b) in Table 5.7. In at least two cases the *i-* opened to *e-* in Kokama and *ɨ-* in Omagua, in the sets OPEN and WIDE in (c). In these cases the following syllable contained a /e/ in the case of OPEN. As for the case of WIDE, there must have been a sound change ɨ > e prior to the addition of the initial *i-* prefix, as the expected output would have been e.g. †*ɨpɨwatsu* in Kokama otherwise. The original root vowel /ɨ/ is confirmed by looking at cognate data for other Tupí–Guaraní languages, e.g. Paraguyan Guaraní *pɨgʷasu* and Chiriguano *pɨwasu*. In other cases the normal realization is /i/, shown with two selected examples in (a) in Table 5.7.

5.1.7 Minor sound changes

5.1.7.1 Epenthesized sounds

Table 5.8 shows a couple of sets with epenthesized segments. These sets appear to have little in common and need to be discussed on a case-by-case basis.

In BRING in Table 5.8, the Omagua and Kokama forms display an initial reflex of POKT *e whereas the Tupinambá root has no such vowel. Most likely, this cognate set contains the root 'come' PTG, POKT **ur-* (see the COME set), with a comitative causative prefix, PTG **ero-∼ro-* (Jensen, 1998:533f), where *ro-* occurred with first and second person agent morphemes, and *ero-* elsewhere. Given that the allomorphy existed already in PTG and was inherited in Tupinambá (Rodrigues, 2010:13), it seems safe to assume that the Omagua–Kokama forms go back to the allomorph with the initial **e-*, and the Tupinambá form to the other allomorph. It is not clear however, what conditioned the allomorphy and explains why different allomorphs appear in Omagua–Kokama vis-à-vis Tupinambá.

Change	Set	Omagua	Kokama	Tupinambá
∅ > e	BRING	ɪɾuɾi	eɾuɾa	ɾuɾa
∅ > ɨ	NEPHEW	mɪmɨɾɨa	memɨɾɨa	membɨɾa
	SON	taɨɾɨa	taɨɾɨa	taʔɨɾa
∅ > i	THING	maɾai	maɾi	maɾã
∅ > n	LONG TIME AGO	ɨmɨnua	ɨmɨnwa	umuã

Table 5.8: Cognate sets with epenthesized segments.

The sets NEPHEW and SON in Table 5.8 both contain a final sequence *-ɨɾa* in Tupinambá corresponding to *-ɨɾɨa* in Omagua–Kokama. Looking at cognates in other Tupí–Guaraní languages, there is no second /ɨ/, e.g. Guajajara *ta'yr* and *imemyr*, or Xingú Asuriní *ta'yra* and *membyra*, respectively. This suggests that this was innovated in Proto-Omagua–Kokama. It is likely that these two words are related, so it in fact concerns one change and not two separate ones, but an explanation for the epenthesized /ɨ/ cannot yet be offered.

The set THING contains a final *-i* in Omagua and Kokama. The Kokama form probably stems from a POK form **marai* with regular Kokama final VV simplification (see subsubsection 5.1.3.1). This final generally /i/ does not have correspondences in other Tupí–Guaraní languages, e.g. Paraguyan Guaraní *marã* or Xingú Asuriní *mara (pa)*, except Parintintín *marãi*. It is possible that this final *-i* is a piece of morphology added in Proto-Omagua–Kokama, or that it has undergone analogy with e.g. the POK nominalizer *-mai* from POKT **mbaʔe* 'thing'.

Finally, in LONG TIME AGO, there is a /nu/ sequence which is not present in Tupinambá, nor in other Tupí–Guaraní languages. Most likely, this is a piece of morphology added to the original root in Proto-Omagua–Kokama.

5.1.7.2 Frozen nasalization

Proto-Omagua–Kokama–Tupinambá had a productive nasal harmony process where certain segments changed to nasal counterparts in the presence of a rightward nasal element. This process was not inherited in Proto-Omagua–Kokama and similar processes are not mentioned in phonological descriptions of neither Omagua nor Kokama. In certain cases however, the reflex in Omagua–Kokama comes from the nasal form, as seen in Table 5.9. For all cases except for FELLOW MAN, this can be explained by a nasal segment (vowel or consonant) present in Tupinambá. As nasal vowels were lost in Omagua and Kokama, the (original) nasal allophone is the only trace of the original trigger. As for FELLOW MAN, the nasality is only present in Kokama and also has the reported variant form *napitsara* (Vallejos & Amías, 2015:145). As there is no nasal trigger for this word, it is likely that the initial nasal has a different origin than frozen nasal harmony, perhaps the second person pronominal clitic *n(a)*=, or the first person inclusive pronominal clitic *ni(a)*= (Vallejos, 2010:149).

Set	Omagua	Kokama	Tupinambá
BELLY BUTTON	mɨɾua	mɨɾwa	puruʔã
CURE	musana	mutsana	posaŋa
FELLOW MAN	japɪsaɾa	njapitsaɾa	apiʃaɾa
SISTER	kunia	kunja	kuɲã
YOUNG WOMAN	kuniatai	kunjati	kuɲãtaĩ

Table 5.9: Frozen nasal harmony in Omagua and Kokama.

The sets SISTER and YOUNG WOMAN in Table 5.9, which appear to be related, show a similar process. In Tupinambá, the strict underlying phonemic form for these words would have been /kujã/ with a nasal realization of /j/ as [ɲ] triggered by the final nasal. The Omagua–Kokama forms have a preserved /n/ as a remnant of the nasal harmony process not found in Omagua–Kokama. It is also possible that the underlying form was originally /kunjã/ with an identical surface form, which would explain the /n/ in Omagua–Kokama.

5.1.7.3 Loss of segments

Table 5.10 shows a number of cognate sets that display a loss of a segment in one way or another. In ACHIOTE, the initial vowel was lost in Proto-Omagua–Kokama (apheresis). This is evident from looking at cognates in other Tupí–Guaraní languages, most of which display such an initial vowel, e.g. Guarayu *urucu* and Araweté *irikũ*, but it was unexpectedly lost in Omagua and Kokama. Possibly, the initial /u/ was in Proto-Omagua–Kokama reanalyzed

as the PTG 3rd person coreferential possessive prefix **o-*, which shows us as a frozen *u-* in some Proto-Omagua–Kokama roots.

No.	Change	Set	Omagua	Kokama	Tupinambá
(a)	u > ∅	ACHIOTE	ɾuku	ɾuku	uɾuku
(b)	ɨ > ∅	CUTBANK	ɨwama	ɨwama	ɨβɨʔama
		DAUGHTER	taiɾa	tajɾa	tajɨɾa
		ROCK∼SHAKE	wɨuta	wɨwɨta	ag^{w}ɨag^{w}ɨ
(c)	e > ∅	PET	mɨma	mɨma	emimbaβa
		PREY	miaɾa	mjaɾa	embiaɾa
		WIFE	miɾikua	miɾikwa	embiɾekó
(d)	ɲ > ∅	HEART	ĩja	ija	ɲɨʔã
(e)	n > ∅	MOUSE∼RAT	sanuja	tsanuja	saujá
(f)	σσ > σ	FLY (INTR.)	uwiwi	uwe	βeβé

Table 5.10: Cognate sets with loss of segments.

The sets in (b) have in common that they involve the loss of *ɨ, but in different circumstances. In CUTBANK and DAUGHTER in Table 5.10, *ɨ was unexpectedly lost following a glide. In ROCK∼SHAKE, *ɨ was lost in addition to several other segments. The original root here is likely **ag^{w}ɨ-* with a reduplication reflecting the repeated action of shaking something, and prefixed in Tupinambá with the causative prefix *mo-* (Rodrigues, 2010:13). In Kokama, the initial /a/ was lost, possibly due to reanalysis of the PTG 1st person ergative prefix **a-* (cf. example 1 in subsection 2.3.1), and with addition of the causative suffix *-ta* (Vallejos, 2010:86). In Omagua, the *ɨ vowel of the reduplicated syllable was elided. While it might look like the glide was vocalized, it is pronounced as a glide on the surface level with the stress on the first syllable: /wɨuta/ [ˈwɨwta] 'be dizzy' (Sandy & O'Hagan, 2020:113).

The examples in (c) all contain the original PTG object nominalizer **emi-* (Jensen, 1998:541), which have in all three examples lost its initial *e. In addition, the original PTG *i > POK *ɨ in PET, but not in the other two. This sound change is described more in the following section. The alternation between /^{m}b∼m/ in Tupinambá is regular, see subsection 2.3.2.

In HEART, Tupinambá displays an initial palatal consonant corresponding to zero in Omagua and Kokama. Looking at cognates in other Tupí–Guaraní languages, some display this initial palatal, e.g. Xingu Asuriní *ɲi'ʔã*, Parintintin *nha'a'ĩ*, Paraguayan Guaraní *ɲeʔã*, whereas others show an initial fricative: Xetá *'hĩɲa*, Tapirapé *xỹhỹ*, Tocantins Asuriní *sy'ó*. Likely, the PTG form contained an initial consonant of some sort which in Tupinambá was /j/ underlyingly, but nasalized due to the final nasal vowel. Generally, the sequence /ji/ does

not exist initially, or is at best very rare, so when the form underwent the sound change *ɨ > i in Omagua–Kokama, it is possible that the initial *j was lost due to phonotactic constraints.

In MOUSE∼RAT in Table 5.10, there seems to have been a sporadic loss of the medial nasal in Tupinambá. Looking at Tupí–Guaraní cognates, most languages display a medial nasal of some sort, e.g. Paraguayan Guaraní *anguja*, Tapirapé *anoxã*, Wayampí *anuya*, Guarayu *zamyya*, or a velar stop likely corresponding to the velar nasal, e.g. Awetí *tãkuja'jýt*. It is unclear what caused the loss of this nasal.

Finally, in FLY in Table 5.10, the Omagua and Kokama words display an initial /u/ not found in Tupinambá. Possibly, this was originally a 3rd person ergative person marker (cf. subsection 2.3.1). However, this initial vowel is variably found across the Tupí–Guaraní family, so either it was seperately innovated in several languages, or the initial vowel goes further back than Proto-Omagua–Kokama, cf. e.g. Tocantins Asuriní *owéwe∼wewe* and Guajajara *uwewe*. This root contains a duplicated syllable *βe found across the Tupí–Guaraní language family. In Kokama however, the repeated syllable was lost, possibly due to haplology, yielding *uwe*.

5.1.8 'Sporadic' developments

Finally, Tables 5.11 to 5.14 show a number of cognate sets that display sound changes not covered elsewhere in this thesis. These tables have been grouped into rough categories based on the type of change, but for legibility, these categories have been split up in four tables within this section. These tables are named 'sporadic' developments, but it should be noted that several may appear sporadic, but would in fact be conditioned if more data were available.

The first two sets in (a) in Table 5.11 show a correspondence between /a/ in Tupinambá and /u/ in Omagua–Kokama. Other Tupí–Guaraní languages generally show an /a/, e.g. Paraguayan Guaraní *(a)sapukái*, Mbyá *-japukai*, Warazu *-hapúkai*, etc., so presumably there was a change in Proto-Omagua–Kokama from *a > u and with regular final vowel sequence simplification in Kokama. It is unclear what caused this change, but it may be influence from the preceding /u/ as a kind of vowel harmony. In BE ODOROUS, the change is most likely caused by the labialized /pʷ/ which turned the following vowel into a /u/, which is a labial

No.	Change	Set	Omagua	Kokama	Tupinambá
(a)	a > u	CALL	sapukui	tsapuki	sapukaja
		BE ODOROUS	ʃapuni	tʃapuni	tɨapʷana
	au > a	POOR	paɾiasu	paɾjatsu	poɾeausuβa
	aŋ > ɨ	HEAD	jakɨ	jakɨ	akaŋa
(b)	e > a	COLLARED PECCARY	taitatu	tajtatu	tajtetu
	e > ɨ, i	FORGET	sɨsaɾai	tsitsaɾi	sesaɾaja
	e > ɨ	SHOUT	sasɨma	tsatsatsɨma	sasema
		AFTER	sakapɨɾɨ	tsakapɨɾɨ	takɨpʷéri
	e > u	THIGH	sɪtɨma	tsutɨma	etɨmã

Table 5.11: 'Sporadic' sound developments, (a-b).

segment as well, and then disappeared as secondary articulation on the consonant.

In POOR in (a) the vowel sequence /eau/ corresponds to /ia/ [ja] in Omagua and Kokama. It is not completely clear what caused this simplification, but such three-vowel sequences are at best rare in Omagua and Kokama; there are no examples of /iau/ or /eau/ in the Omagua and Kokama data, and only one example each of /jau/ and /wau/ in Omagua, so it is possible that it was simplified due to phonotactic reasons.

In HEAD, the sequence /aŋ/ in Tupinambá corresponds to Omagua and Kokama /ɨ/. Looking at other Tupí–Guaraní languages, most of them have a corresponding /aŋ/ as well, e.g. Tembé *akãng*, Kamaiura *akaŋ*. Interestingly, Central Tupí–Guaraní also show an /ɨ/ however, e.g. Tapirapé *ãkyg*, Tocantins Asuriní *iákyŋa*, Parakanã *akyg*, Xingu Asuraní *a'kɨ̃ŋ* etc. The original vowel was probably still /a/, but it does not seem like this development is unique to Proto-Omagua–Kokama.

In COLLARED PECCARY in (b), Tupinambá shows an /e/ where Omagua and Kokama shows an /a/ instead. This /e/ is likely the original vowel, since cognates in other Tupí–Guaraní in general show an /e/ as well, e.g. Paraguayan Guaraní *taitete*, Tembé *taitetú*, Emerillon *taitetu* etc. It is unclear what caused the change from *a > e in Proto-Omagua–Kokama, but possibly it was influenced by the /a/ in the preceding syllable as a kind of vowel harmony. In FORGET, the original *e has been raised regularly in Kokama to /i/ next to a coronal. In Omagua however, the corresponding sound is an /ɨ/. The reason for the Omagua change from *e > ɨ, possibly via a raising to *i, is not entirely clear, but there is a certain correspondence between /ɨ/ and /i/ in this environment discussed e.g. in subsection 5.1.1 manifested in that all sequences of /tsɨ/ went to /(t)si/. A similar development is seen in SHOUT and AFTER where an /ɨ/ in Omagua–Kokama corresponds to an /e/ in Tupinambá. The relationship

No.	Change	Set	Omagua	Kokama	Tupinambá
(c)	ei > a	ENTER	aki	aki	eiké
		YESTERDAY	ikʷaʃi	ikwatʃi	kʷeisé
	ei > e	STAR	sɪsu	tsetsu	seiʃu
(d)	i > a	THROW	atika	itika	itɨka
		BE ODOROUS	ʃapuni	tʃapuni	tɨapʷana
	i > e	CULTIVATE	kupɪ	kupe	kopiɾa
		DEFECATE	kapɪ	kape	kaʔapia
	i > ɪ	FOUR	ɪruaka	iɾwaka	iɾundɨk

Table 5.12: 'Sporadic' sound developments, (c-d).

between these sounds might be further illuminated in future studies.

In THIGH, the original first vowel *e, which is the vowel found in virtually all Tupí–Guaraní cognates, was regularly raised to /ɪ/ in Omagua, but shows an irregular change to /u/ in Kokama. The reason for this is unknown.

In (c) in Table 5.12 in ENTER and YESTERDAY, there is a correspondence between /ei/ in Tupinambá and /a/ in Omagua–Kokama. It is possible that this correspondence would be clearer with more data, but at the moment the development is not clear. In STAR, Tupinambá /ei/ corresponds to /e/. In this case it is important to note that the diphthong /ei/ in Tupinambá often alternates with /e/ in the data, for example *akʷéme*∼*akʷéime* 'long time ago', *kʷesé*∼*kʷeisé* 'yesterday', or *pʷeraja*∼*pʷeiraja* 'tired'. Hence, it is quite possible that the Omagua and Kokama forms in the case of ENTER corresponds to a Tupinambá form with a monophthong.

In (d) in THROW and BE ODOROUS in Table 5.12, there is a correspondence between /i/ in Tupinambá and Kokama vis-à-vis /a/ in Omagua. According to Barbosa (1951), there is an infinitive form of *itɨka* which is *eitɨka*, so possibly the Omagua form could be derived from this diphthong and show the same correspondence as ENTER and YESTERDAY in (c).

In CULTIVATE and DEFECATE in (d), *i was unexpectedly lowered to *e in POK, as evident from Tupí–Guaraní cognates such as Parintintin *-kopir*, Kaʼápor *kupixa* for CULTIVATE, Aché *puči* and Kamaiurá *o-potsi-m* for DEFECATE, which show an /i/. Normally, we expect the reverse development, namely that *e is raised to /i/ in final position, which is not what we find in this case.

In FOUR in (d), the initial /i/ has unexpectedly been raised to /ɪ/ in Omagua.

No.	Change	Set	Omagua	Kokama	Tupinambá
(e)	ɨ > a	AFTER	sakapɨrɨ	tsakapɨrɨ	takɨpʷéri
	ɨ > e	FLAME	sɪnɪ	tsene	sendɨ
		WIDE	ɪpɪwasu	epewatsu	pɨgʷasu
		BE SCARED	ɨsɨi	ɨtse	sɨja
	ɨ > i	GRAB	japiʃika	japitʃika	pɨsɨka
		ROAST	mitʃɪra	mitʃɪra	miʃɨra
		VINE	iʃipu	itʃipu	ɨsɨpó
		DIG	ʃukai	tʃiwiki	sɨβɨkoja
		DRUNK	saipuɾa	tsajpuɾa	saβeɨpoɾa
		THROW	atika	itika	itɨka
	ɨ > i, ĩ	HEART	ĩja	ija	ɲɨʔã
	ɨ > u	CHAMELEON	sɪnɪmu	tsenemu	senembɨ
		LUNG∼BREATHE (INTR.)	putu	putu	pɨtu
		DRIP	atukɨra	atɨkɨrɨ	tɨkɨra

Table 5.13: 'Sporadic' sound developments, (e).

Category (e) in Table 5.13 shows different vocalic developments of *ɨ. For example, in AFTER, original *ɨ, as evident from cognates such as Mbyá *akykue*, Parakanã *akykwer* and Tembé *haykwêr* (⟨y⟩ = [ɨ]), changed to /a/ in Omagua and Kokama, possibly because of vowel harmony triggered by the /a/ in the initial syllable.

In FLAME in Table 5.13, there is again a correspondence between /ɨ/ in Tupinambá and /e/ in Omagua–Kokama. Possibly, this is due to influence from the /e/ in the preceding syllable, but it could also be a general case of an /ɨ∼e∼i/ correspondence. Similarly, in WIDE in Table 5.13, *ɨ seems to have shifted to *e in Proto-Omagua–Kokama for unknown reasons. Other Tupí–Guaraní languages generally agree with Tupinambá in this word, e.g. Paraguyan Guaraní *pyguasu*, Chiriguano *pɨwasu*. In BE SCARED in Table 5.13, /ɨ/ in Tupinambá and Omagua also corresponds to /e/ in Kokama.

There are a number of words in which /ɨ/ in Tupinambá corresponds to /i/ in Omagua–Kokama. There is a regular sound change from *tsɨ > /(t)si/ described above and in subsection 5.1.1, so the sets included here are sets which display this correspondence elsewhere than in this context.

In GRAB, ROAST, and VINE in Table 5.13, it appears that the original *ɨ changed to /i/ also before *ts. In DIG, the first *ɨ changed regularly after *ts, but was elided in Omagua due to vowel hiatus. The second *ɨ also went to /i/ however, possibly due to influence from the first, in a kind of vowel harmony. Similarly, in THROW, it is possible that the *ɨ went to /i/ due to

influence from the *i in the preceding syllable. If this is the case, we have to assume that this change occurred before the change to the initial /a/ in Omagua discussed above.

In DRUNK in Table 5.13, there apparently was spontaneous change from *ɨ > i following loss of the medial segments. It is possible that this change was a way of avoiding vowel hiatus between *a and *ɨ which is a rare sequence in Omagua-Kokama. Through the change to /ai/ the second element of the sequence can be regularly realized as a glide instead.

In HEART in Table 5.13, the original *ɨ changed to /i/ in both Omagua and Kokama, but seems to have been nasalized in Omagua to in the process, possibly due to the original final nasal vowel.

In CHAMELEON in Table 5.13, the final original *ɨ changed to /u/ in Omagua–Kokama, as evident from cognates such as Parakanã *enemy*. This is also the case in LUNG∼BREATHE which based on Tupí–Guaraní cognates such as Warazu *pɨtjú-ʔu*, Parakanã *pytoʾo*, Guajajara *upytuʾu*, *pytuhem*, and Kamaiurá *pytu* seemed to have /ɨ/ as the original vowel, which changed to /u/ in Omagua–Kokama. The reason for this is not entirely clear, but there seems to be a connection between /ɨ/ and /u/ around labial segments. This is discussed more in connection to (g) in Table 5.14.

In DRIP, the original *ɨ changed to /u/ in Omagua, but not in Kokama, for unknown reasons. Cognates in other Tupí–Guaraní languages generally show /ɨ/ or /i/ in this position, e.g. Mbyá *tyky* or Warazu *tɨki*, but not a back rounded vowel.

In (f) in Table 5.14, there are a number of words that changed from *o to *a in Proto-Omagua–Kokama. It is difficult to say what caused this sound change, and it is possible that its condition would become more obvious with more data available. Likewise, the change from *o > ɨ in BORROW has an unknown cause.

In (g) in Table 5.14, there are a number of sets that show a correspondence between /ɨ/ and /u/. As mentioned in set (e) (shown in Table 5.13), there is a correspondence between these sounds. A vowel /u/ is not generally found in any Tupí–Guaraní cognates of these words, which generally have /ɨ/ or /i/, e.g. for DEER, e.g. Emerillon *ɨuhu*, Awetí *tɨzɨwatu*; for LONG TIME AGO, e.g. Paraguyan Guaraní *yma*, Tocantins Asuriní *-ymáwe*, Kamaiurá *ymawe*, but Araweté *ima*, Warazu *ðímə*; and for REST e.g. Warazu *-pɨtjú-ʔu*, Guajajara *upytu'u*, and Tocantins Asuriní *pyto'ó*), but Emerillon *putuʔu*. Set LAKE does not have any cognates in

No.	Change	Set	Omagua	Kokama	Tupinambá
(f)	o > a	BITTER	ɨrawa	ɨrawa	ɾoβa
		DOOR	jakina	jakina	okendaβa
		ENCOUNTER	sawiti	tsawiti	soβáitiĩ
		POOR	paɾiasu	paɾjatsu	poɾeausuβa
		DIG	ʃukai	tʃiwiki	sɨβɨkoja
		COVER	jasai	jatse	asoʔí
	o > ɨ	BORROW	ɨpɨɾu	ɨpɨɾu	poɾu
(g)	ɨ > u	DEER	ɨsɨwasu	ɨtsɨwatsu	sugʷasu
		LAKE	ɨpasu	ɨpatsu	upaβa
		LONG TIME AGO	ɨmɨnua	ɨmɨnwa	umuã
		REST	japɨtu	japɨtu	putuʔú
(h)	p^w > p	AFTER	sakapɨɾɨ	tsakapɨɾɨ	takɨpʷéri
		BE ODOROUS	ʃapuni	tʃapuni	tɨapʷana
(i)	β > p	SHINE (INTR.)	pɪɾa	peɾa	βeɾaβa

Table 5.14: 'Sporadic' sound developments, (f-i).

its exact form, but most likely contains the word *ɨ 'water', which places it in this set as well. It is difficult to find a condition for this change, but it might be triggered by an adjacent labial element which caused rounding assimilation of *ɨ. More data is needed in order to be sure of the exact condition.

In (h), there has been a delabialization of earlier *p^w in AFTER and BE ODOROUS. In the latter, it has most likely affected the following vowel and caused the change from *a > u. In these words, most Tupí–Guaraní cognates have a corresponding labialized stop, e.g. Guarayu *zaquìcuer*, Tembé *haykweêr* for AFTER, and Guarayu *zìãcuã* and Guajajara *hyàkwen* for BE ODOROUS. It is possible that all earlier instances of /p^w/ delabialized, but more data is needed to be certain.

Finally, in SHINE (INTR.), there is an initial /p/ in Omagua–Kokama corresponding to an initial /β/ in Tupinambá. Root-initial *β is quite rare, and is not present in any of the sets discussed in subsection 5.1.2, therefore it is hard to say whether this is a regular development of initial *β. There are however many root-initial *p, none of which shows lenition to /β/ in Omagua and Kokama. For this reason, it is likely either a case of fortition or an unrelated change, for example rest morphology, or analogy with another /p/ initial word. Looking at cognates in other Tupí–Guaraní languages, these generally show a more expected reflex of *β such as /w/ or /v/, but two cognates also show an initial *p*-, Wayampí *ɔ-pɛla* and Emerillon *pelab*.

5.2 Summary of sound changes

The research questions asked in the introductory chapter of this theses were the following:

- What are the phonological differences between Omagua, Kokama, and Tupinambá?
- What phonological features can be reconstructed to the ancestral proto-language of these languages?
- What phonological changes were involved in the genesis of Proto-Omagua–Kokama?
- What can we infer about the phonologies of languages in contact with Pre-Proto-Omagua–Kokama in the contact situation?

The Results chapter of this thesis has answered research question 1 and 2, by building correspondence sets and reconstructing proto-segments. In this process several sound changes have been identified and further discussed in the Discussion chapter which answers research question 3. Figure 5.1 is a list of all the regular sound changes that have been discussed so far, not counting the ones labeled 'sporadic' above. The list consists of changes from POKT to Omagua and Kokama and apply to both, i.e. were present already in POK, unless stated otherwise.

Among the regular sound changes, there are only two from POKT, namely ts > s, and ɨ > u in the context of a labial consonant. This means that the Tupí–Guaraní language involved in the contact situation that gave rise to Proto-Omagua–Kokama was phonologically very similar to Tupinambá. This inference is also drawn by Michael (2017), who writes that (Pre-)Tupinambá and Pre-Proto-Omagua–Kokama split around 800 CE, and speakers of Pre-Proto-Omagua–Kokama arrived in the upper Amazon basin around 1100 CE, based on archaeological evidence. This means that by the time of the genesis of Proto-Omagua–Kokama, Pre-Proto-Omagua–Kokama and Pre-Tupinambá had only been separated for about 300 years. The sources for Tupinambá stem from the 16th century, which is only 400 years after the genesis of Proto-Omagua–Kokama (Michael, 2017). Therefore we do not expect the differences between Tupinambá and Proto-Omagua–Kokama–Tupinambá to be very large.

Vowel changes	**Consonant changes**
1. *$\tilde{V}$ > *V	1. *ʔ > ∅
2. *o > *u	2. *ts > s (OMG)
3. *Vj > i / _# (KK)	3. *ts > *tʃ / _i
4. *e > *i / _#	4. *β > ∅ / _#
5. *e > *i / _$C_{[+COR]}$, $C_{[+COR]}$_	5. *β > *w / elsewhere
6. *e > ɪ / elsewhere (OMG)	6. *g^w > *w
7. *je > *i / #_	7. *^{N}P > *P
8. *ɨ > *i / ts_(some cases)	8. *ŋ > *n
9. *VʔV > *V	
10. *u > *w / {ɾ, k, p}_a	
11. *u > ∅ / n_{a, i}	
12. *i > *j / {t, n}_u	
13. *i > ∅ / tʃ_{u, a}	
14. √CV > √CVV (OMG)	
15. *iCɨ > *ɨCɨ	
16. *iCe > *eCe	

Figure 5.1: Summary of sound changes from POKT to POK with some later individual sound changes shown in brackets.

5.3 Non-Tupí–Guaraní phonological influences on POK

So far, the discussion chapter has mostly covered the phonological developments from Proto-Omagua–Kokama–Tupinambá to its daughter languages. The fourth research question asks what can be inferred about the phonologies of languages in contact with Pre-Proto-Omagua–Kokama. According to Winford (2003:319), creoles generally preserve sounds that are similar between the languages involved in the contact situation. When sounds differ however, superstrate sounds are generally replaced by a substrate sound, which is characteristic of general L2 acquisition.

In this case, according to the hypothesis of Michael (2017), members of the Pre-Proto-Omagua–Kokama society incorporated a large numbers of captives from neighboring groups, resulting in creolization of Pre-Proto-Omagua–Kokama as non-Tupí–Guaraní-speaking captives eventually outnumbered the speakers of Pre-Proto-Omagua–Kokama, giving rise to Proto-Omagua–Kokama.

By looking at sound changes from POKT to POK, we can hypothesize that certain phonological substitutions are due to L2 interference as non-Tupí–Guaraní-speaking captives

were acquiring Pre-Proto-Omagua–Kokama, and thereby infer some characteristics of the phonologies of the non-Tupí–Guaraní languages which were influential in the contact situation.

For example, some of the major sound changes from Proto-Omagua–Kokama–Tupinambá to Proto-Omagua–Kokama include:

1. The loss of nasal vowels.
2. The loss of pre-nasalized stops in favor of their nasal counterparts.
3. The loss of nasal harmony as a whole.
4. The merger of POKT *o and *u into POK *u.
5. The merger of POKT *ŋ and *n into POK *n.
6. The loss of glottal stop *ʔ.
7. The constraint that roots must be vowel-final.
8. As a result of the above, the loss of root final *β.

It is therefore possible that the above sound changes were caused by the lack of a distinction in the phonology of the influential non-Tupí–Guaraní language(s), such as the the absence of vowel nasality and nasal processes, and a failure in producing /o/ and /ŋ/, substituting and merging them with /u/ and /n/. One could also imagine that the phonotactics of the non-Tupí–Guaraní-language(s) disallowed certain coda consonants, giving rise to the loss of *β and the vowel-final word constraint in Proto-Omagua–Kokama.

It should be noted that a sound change can also be the result of an internal development, so the absence of a feature in Proto-Omagua–Kokama is not necessarily proof of non-Tupí–Guaraní influence. For example, glottal sounds are less audible than consonants produced with a closure in the mouth and are easily lost as an internal development, so one needs to be wary of what one ascribes to contact influence. However, without arguing for the cause of these changes, a neighboring language exhibiting the features noted above would indeed be a good potential candidate in a hypothesis regarding non-Tupí–Guaraní languages involved in the contact situation.

Chapter 6

Conclusions

This chapter presents the concluding remarks of this thesis, with a summary of the findings and how they relate to the research questions asked in the introduction, followed by suggestions for future research.

6.1 Summary

This exploratory comparative thesis examined the phonologies of Omagua and Kokama, two closely related heavily contact-induced Tupí–Guaraní languages, and compared these with the phonology of Tupinambá, an older attested Tupí–Guaraní language without the same contact influence. As the details of the contact situation which gave rise to Omagua and Kokama are not widely known, this thesis also reconstructed the phonology of Proto-Omagua–Kokama–Tupinambá, the ancestral language from before the contact situation. By doing this, the thesis aimed to shed light on the phonological differences between these languages, in order to identify what phonological changes were involved in the genesis of Omagua and Kokama, and whether any of these changes could be attributed to non-Tupí–Guaraní contact-induced influence.

The research questions asked in the beginning of this thesis were as follows:

- What are the phonological differences between Omagua, Kokama, and Tupinambá?
- What phonological features can be reconstructed to the ancestral proto-language of these languages?

- What phonological changes were involved in the genesis of Proto-Omagua–Kokama?
- What can we infer about the phonologies of languages in contact with Pre-Proto-Omagua–Kokama in the contact situation?

The thesis used lexical data from the Tupí–Guaraní Comparative Lexical Dataset (Chousou-Polydouri et al., 2019) in order to build cognate sets and correspondence sets for segment comparison. This was done using the automatic cognate detection tool LexStat (List et al., 2017) and the cognates were then manually double-checked, segmented and aligned using EDICTOR (List, 2017). Correspondence sets consisting of sound correspondences between the three languages were then identified using List (2018, 2019), and then again manually inspected in EDICTOR. Once the correspondence sets had been built, a proto-segment was reconstructed for each correspondence set in the Results chapter using the Comparative Method, with arguments for the choice of reconstruction.

Through this process of reconstruction, phonological changes were identified from the POKT reconstruction to the reflexes in its daughter languages. These phonological changes were compiled and cognate sets exhibiting similar phenomena were grouped together and more closely examined. Phonological changes that were only found in very few sets (named 'sporadic changes') were also discussed on a case-to-case basis in order to account for all the data. Lastly, a summary of all the non-sporadic changes were given along with a section of possible causes to the phonological changes in the genesis of Proto-Omagua–Kokama, and what can be inferred about other languages involved in the contact situation.

The thesis found that the phonology of Proto-Omagua–Kokama–Tupinambá was generally very similar to that of Tupinambá, which was expected due to shallow time depth between the split of these languages and the contact situation which gave rise to Proto-Omagua–Kokama. A major difference is that Tupinambá underwent lenition of *ts > s in common with Omagua, whereas Kokama preserves *ts. The thesis identified a number of sound changes from Proto-Omagua–Kokama–Tupinambá to Proto-Omagua–Kokama, involving for instance the loss of nasal vowels, the merger of *o and *u, the merger of *ŋ and *n, raising of *e, the loss of *ʔ, the partial loss of root-final *β, palatalization of *ts, among others. The thesis also identified and discussed phonological developments such as the resolution of vowel hiatus in Proto-Omagua–Kokama, vowel harmony in Proto-Omagua–Kokama, monomoraic root augmentation in Omagua, among others. Furthermore, the thesis also identified and

discussed more irregular phonological developments, for example some cases of vowel epenthesis, vowel loss, frozen nasalization, and more sporadic changes without an obvious phonological condition.

6.2 Future research

The genesis of Proto-Omagua–Kokama is an interesting topic where there is much room for future studies. This thesis only constitutes a small piece in the bigger puzzle of this contact situation. For further studies, one could make a comparative study within other areas than phonology, such as syntax and morphology, and identify syntactic and morphological changes as well. Alongside this, it would be useful to compare the changes identified from Proto-Omagua–Kokama–Tupinambá to Proto-Omagua–Kokama, and the inferences made about possible non-Tupí–Guaraní traits with actual non-Tupí–Guaraní languages spoken in the vicinity of Omagua and Kokama. Cabral (1995) identifies certain lexical and grammatical features in Kokama that can be attributed to Arawakan influence, but as Michael (2014) notes, it would be useful to look at other neighboring families as well, such as Peba–Yaguan and Zaparoan languages as the typological profile of Proto-Omagua–Kokama is more similar to these families than to Tupí–Guaraní.

References

Aikhenvald, A. Y. (2018). *Serial verbs*. Oxford: Oxford University Press.

de Anchieta, J. (1595). *Arte de Grammatica da Lingoa mais usada na costa do Brasil*. Coimbra: Antonio de Mariz.

Ansaldo, U. & Matthews, S. (2007). Deconstructing creole: The rationale. In S. Matthews, L. Lim, & U. Ansaldo (Eds.), *Deconstructing creole*. John Benjamins Publishing Company.

de Araujo, A. (1618). *Catecismo na lingoa brasilica*. Lisboa: Pedro Crasbeeck.

Barbosa, P. A. L. (1951). *Pequeno Vocabulário Tupi-Português*. Rio de Janeiro: Livraria São José.

Barbosa, P. A. L. (1956). *Curso de Tupi Antigo: Gramática, Exercícios, Textos*. Rio: Livrario Sãn José.

Barbosa, P. A. L. (1970). *Pequeno vocabulário Português-Tupi*. Rio de Janeiro: Livraria São José.

Cabral, A. S. A. C. (1995). *Contact-induced language change in the Western Amazon: The non-genetic origin of the Kokama language*. PhD thesis, University of Pittsburgh.

Cabral, A. S. A. C. & Rodrigues, A. D. (2003). Evidências de criouslização abrupta en Kokáma? *Papia*, 13, 180–186.

Campbell, L. (2013). *Historical linguistics: an introduction*. Edinburgh: Edinburgh University Press, 3rd revised edition.

Chousou-Polydouri, N., Michael, L., & O'Hagan, Z. (2019). Tupí-Guaraní Comparative Lexical Dataset. Unpublished database.

Crowley, T. (2008). Pidgin and creole morphology. In S. Kouwenberg & J. V. Singler (Eds.), *The handbook of pidgin and creole studies* (pp. 74–98). Oxford: Wiley-Blackwell, 1st edition.

DeGraff, M. (2003). Against creole exceptionalism. *Language*, 79(2), 391–410.

Eberhard, D. M., Simons, G. F., & Fennig, C. D., Eds. (2019). *Ethnologue: Languages of the World*. Dallas, Texas, USA: SIL International, 22nd edition. `http://www.ethnologue.com`.

Figueira, L. (1687). *Arte de grammatica da lingua brasilica*. Lisboa: Officina de Miguel Deslandes.

Harris, J. (1994). *English sound structure*. Oxford: Blackwell.

Jensen, C. (1998). Comparative Tupí-Guaraní morphosyntax. In D. C. Derbyshire & G. K. Pullum (Eds.), *Handbook of Amazonian languages*, volume 4 (pp. 489–618). Berlin/New York: Mouton de Gruyter.

Jensen, C. (1999). Tupí-guaraní. In R. M. W. Dixon & A. Y. Aikhenvald (Eds.), *The Amazonian languages* (pp. 125–164). Cambridge: Cambridge University Press.

Lemle, M. (1971). Internal classification of the Tupi-Guarani linguistic family. In D. Bendor-Samuel (Ed.), *Tupi studies*, volume 1 (pp. 107–129). Norman: Summer Institute of Linguistics of the University of Oklahoma.

List, J.-M. (2017). A web-based interactive tool for creating, inspecting, editing, and publishing etymological datasets. In *Proceedings of the 15th conference of the European chapter of the association for computational linguistics. System demonstrations* (pp. 9–12).

List, J.-M. (2018). Lingrex: Linguistic reconstruction with LingPy. Version 0.1.1. Jena: Max Planck Institute for the Science of Human History.

List, J.-M. (2019). Automatic inference of sound correspondence patterns across multiple languages. *Computational Linguistics*, 45(1), 137–161.

List, J.-M., Greenhill, S., Tresoldi, T., & Forkel, R. (2018). Lingpy. a Python library for historical linguistics. `http://lingpy.org`. Version 2.6.4. Jena: Max Planck Institute for the Science of Human History.

List, J.-M., Greenhill, S. J., & Gray, R. D. (2017). The potential of automatic word comparison for historical linguistics. *PLOS ONE*, 12(1).

McWhorter, J. (1998). Identifying the creole prototype: Vindicating a typological class. *Language*, 74, 788–818.

McWhorter, J. (2001). The world's simplest grammars are creole grammars. *Linguistic Typology*, 5(2/3), 125–166.

Meira, S. & Drude, S. (2013). Sobre a origem histórica dos "prefixos relacionais" das línguas Tupí-Guaraní. *Cadernos de Etnolingüística*, 5(1).

Michael, L. (2014). On the pre-Columbian origin of Proto-Omagua-Kokama. *Journal of Language Contact*, 7(2), 309–344.

Michael, L. (2017). Proto-Omagua-Kukama, a pre-Columbian Amazonian creole language. Presentation slides. `[link]`.

Michael, L., Chousou-Polydouri, N., Bartolomei, K., Donnelly, E., Wauters, V., Meira, S., & O'Hagan, Z. (2015). A Bayesian phylogenetic classification of Tupí-Guaraní. *LIAMES*, 15(2), 1–36.

Michael, L. & O'Hagan, Z. (2016). *A Linguistic Analysis of Old Omagua Ecclesiastical Texts*. Cadernos de Etnolingüística: Série Monografias 4.

Mufwene, S. S. (2001). *The ecology of language evolution*. Cambridge: Cambridge University.

O'Hagan, Z. (2011). Proto-Omagua-Kokama: Grammatical sketch and prehistory. Undergraduate honors thesis, UC Berkeley.

O'Hagan, Z. (2013). Syllable structure, nasal harmony and phonemicization in Tupinambá. Course paper, UC Berkeley.

O'Hagan, Z., Chousou-Polydouri, N., & Michael, L. (2019). Phylogenetic classification supports a northeastern amazonian proto-tupí-guaraní homeland. *LIAMES*, 19:e019018.

Piggott, G. (2010). Cyclic spell-out and the typology of word minimality. Manuscript, McGill University.

Rodrigues, A. D. (1958a). Classification of Tupi-Guarani. *International Journal of American Linguistics*, 24(3), 231–234.

Rodrigues, A. D. (1958b). *Phonologie der Tupinambá-Sprache*. PhD thesis, Universität Hamburg, Hamburg.

Rodrigues, A. D. (1985). Relações internas na família lingüística Tupi-Guarani. *Revista de Antropologia*, 27/28, 33–53.

Rodrigues, A. D. (2000). Hipótese sobre as migrações dos três subconjuntos meridionais da família tupí-guaraní. In *II Congresso Nacional da ABRALIN e XIV Instituto Linguístico* (pp. 1596–1605). Florianópolis: Associação Brasileira de Linguística.

Rodrigues, A. D. (2010). Estrutura do Tupinambá. In A. S. A. C. Cabral, A. D. Rodrigues, & F. B. Duarte (Eds.), *Línguas e culturas tupí*, volume 2 (pp. 11–42). Campinas: Curt Nimuendajú.

Rodrigues, A. D. & Cabral, A. S. A. C. (2012). Tupían. In L. Campbell & V. Grondona (Eds.), *The Indigenous Languages of South America: A Comprehensive Guide* (pp. 59–166). Berlin: De Gruyter Mouton.

Sandy, C. & O'Hagan, Z. (2020). A phonological sketch of Omagua. *International Journal of American Linguistics*, 86(1), 1–37.

Schleicher, C. O. (1998). *Comparative and internal reconstruction of the Tupi-Guarani language family*. PhD thesis, University of Wisconsin-Madison.

Siegel, J. (2007). Transmission and transfer. In S. Matthews, L. Lim, & U. Ansaldo (Eds.), *Deconstructing creole*. John Benjamins Publishing Company.

Thomason, S. G. & Kaufman, T. (1988). *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.

Vallejos, R. (2010). *A grammar of Kokama-Kokamilla*. PhD thesis, University of Oregon.

Vallejos, R. & Amías, R. (2015). *Diccionario kukama-kukamiria / castellano*. Iquitos: FORMABIAP.

Weiss, M. (2014). The comparative method. In C. Bowern & B. Evans (Eds.), *The Routledge Handbook of Historical Linguistics*. Routledge Handbooks Online.

Winford, D. (2003). *An introduction to contact linguistics*. Oxford: Blackwell.

Appendices

Appendix A

List of cognate sets

The following appendix is the list of all the cognate sets on which the results and the discussion in this thesis are based. The original lexical data were drawn from the Tupí–Guaraní Comparative Lexical Dataset (see Chousou-Polydouri et al., 2019). Any work which wishes to cite the data in this appendix should cite Chousou-Polydouri et al. (2019), i.e.:

`Chousou-Polydouri, Natalia, Lev Michael, and Zachary O'Hagan. 2019. Tupí-Guaraní Comparative Lexical Dataset. Unpublished database.`

NO.	SET	OMAGUA	KOKAMA	TUPINAMBÁ
1	ACHIOTE	ɾuku	ɾuku	uɾuku
2	AFTER	sakapɨɾɨ	tsakapɨɾɨ	takɨpʷéɾi
3	AFTERNOON	kaɾukatai	kaɾuka	kaɾuka
4	AGOUTI	akuti	akuti	akuti
5	AJÍ	ɨkɨi	ɨki	kɨʔɨɲa
6	ALL	upai	upi	opaβĩ
7	ALSO∼ALREADY	awi	aj	aβé
8	ANGRY	jumɨɾa	jumɨɾa	jemoɨɾõ
9	ANT	sasɨwa	tsatsɨwa	tasɨβa
10	ANTEATER	tamanu	tamana	tamanduá
11	ARM	jɨwa	jɨwa	jɨβá
12	ARMADILLO	tatu	tatu	tatu

NO.	SET	OMAGUA	KOKAMA	TUPINAMBÁ
13	ARRIVE (INTR.)	jauʃima	jawatʃima	gʷasema
14	ARROW	uwa	uwa	uʔúβa
15	ASHES	tanimuka	tanimuka	tanimbuka
16	ATTACH	ujaɾi	ujaɾi	jaɾa
17	AUNT	jaiʃɪ	jajtʃi	ajʃé
18	AXE	jɨɨ	jɨ	jɨ
19	BACK	jatukupɪ	jatukupe	atukupé
20	BAD	aisɪ	ajtse	jeaseja
21	BAT	anɨɾa	anɨɾa	andɨɾá
22	BATHE (INTR.)	jasuka	jatsuka	jasuka
23	BE BORN (INTR.)	uwaɾi	uwaɾi	aɾa
24	BE DRY	ikana	ikana	kaŋa
25	BE FAT	ɨkɨɾɨwasu	ɨkɨɾatsu	kɨɾá
26	BE ODOROUS	ʃapuni	tʃapuni	tɨapʷana
27	BE SCARED	ɨsɨi	ɨtse	sɨja
28	BEARD	muta	muta	amotaβa
29	BEAT	inupa	inupa	nupã
30	BED	tupa	tupa	tupaβa
31	BELLY BUTTON	mɨɾua	mɨɾwa	puɾuʔã
32	BENT	japaɾa	japaɾa	apaɾa
33	BIG	wasu	watsu	gʷasu
34	BIRD	wɨɾa	wɨɾa	gʷɨɾá
35	BITTER	iɾawa	iɾawa	ɾoβa
36	BLACK	suni	tsuni	una
37	BLOOD	suɨ	tsuwɨ	ugʷɨ
38	BOIL (TR.)	apupuɾi	apapuɾi	pupuɾa
39	BONE	kanuaɾa	kanwaɾa	kaŋa
40	BORROW	ɨpɨɾu	ɨpɨɾu	poɾu
41	BRAIN	apɪtuma	apetúma	aputuʔuma
42	BRANCH	sakama	tsakamɨ	akã
43	BREAST	kama	kama	kama
44	BRING (TR.)	ɪɾuɾi	eɾuɾa	ɾuɾa
45	BROTHER (F. EGO)	kɨwɨɾa	kɨwɨɾa	kɨβɨɾa
46	BURN	ukai	uki	kaja
47	BURST	upuka	upuka	poka

NO.	SET	OMAGUA	KOKAMA	TUPINAMBÁ
48	BURY (TR.)	jatɨma	jatɨma	tɨma
49	BUTTERFLY	panama	panama	panama
50	CAIMAN	jakaɾi	jakaɾi	jakaɾé
51	CALL (TR.)	sapukui	tsapuki	sapukaja
52	CANDELA∼FIRE	tata	tata	tatá
53	CANOE	ɨaɾa	ɨɾaɾa	ɨgaɾa
54	CATFISH	mani	mani	mandiʔí
55	CHACRA	kuu	ku	kó
56	CHAMELEON	sɪnɪmu	tsenemu	senembɨ
57	CHEST	putia	putja	potiʔá
58	CICADA	jakɨɾana	jakɨɾana	jakɨɾana
59	CLAY	tawa	tawa	tag^{w}á
60	CLOUD∼FOG	ɨwɨtini	ɨwɨtini	ɨβɨtiŋa
61	COCKROACH	aɾawɪ	aɾawi	aɾaβé
62	COLLARED PECCARY	taitatu	tajtatu	tajtetu
63	COME (INTR.)	uɾi	uɾi	uɾa
64	COMPANION	ɪɾua	iɾwa	iɾũ
65	COOK	ɨjɨ	ɨjɨ	jɨβa
66	COPULATE	mɪnuka	menuka	menõ
67	CORN	awati	awati	aβati
68	COTTON	amaniú	amanju	amaniju
69	COVER (TR.)	jasai	jatse	asoʔí
70	CROSS	sasawa	tsatsawa	sasaβa
71	CRY (INTR.)	jaʃua	jatʃu	joseʔó
72	CULTIVATE (TR.)	kupɪ	kupe	kopiɾa
73	CURE	musana	mutsana	posaŋa
74	CUT	jasɨkata	jatsɨkataka	asɨka
75	CUT HAIR	japina	japina	apina
76	CUTBANK	ɨwama	ɨwama	ɨβɨʔama
77	DANCE (INTR.)	japuɾaʃi	japuɾatʃi	poɾaseja
78	DAUGHTER (M. EGO)	taiɾa	tajɾa	tajɨɾa
79	DAWN	k^{w}ɪma	kwema	koʔema
80	DEEP	tɨpɨ	tɨpɨ	tɨpɨ
81	DEER	ɨsɨwasu	ɨtsɨwatsu	sug^{w}asu
82	DEFECATE (INTR.)	kapɪ	kape	kaʔapia

NO.	SET	OMAGUA	KOKAMA	TUPINAMBÁ
83	DIG	ʃukai	tʃiwiki	sɨβɨkoja
84	DOOR	jakina	jakina	okendaβa
85	DRIP (INTR.)	atukɨra	atɨkɨrɨ	tɨkɨra
86	DRUNK	saipura	tsajpura	saβeɨpora
87	DRY (V.)	tɨpa	tɨpa	tɨpaβa
88	DUST∼SAND	kui	kuj	kuj
89	EAR	nami	nami	nambi
90	EGG	supia	tsupja	upiʔá
91	ENCOUNTER	sawiti	tsawiti	soβàitĩ
92	ENTER	aki	aki	eiké
93	EYE∼FACE	sisa	tsitsa	esá
94	FALL (INTR.)	ukukui	ukuki	kuja
95	FAR	amusɪ	amutse	amõ
96	FAT∼LARD	ikawa	ikawa	kaβa
97	FELLOW MAN	japɪsara	njapitsara	apiʃara
98	FIREFLY	mua	muwa	mamuã
99	FISH	ipira	ipira	pirá
100	FISH (TR.)	sɪkɨi	tsɨki	sekɨja
101	FISH POISON	timu	timu	timbó
102	FISHING NET	pɨsa	pɨtsa	pɨsá
103	FLAME (V.)	sɪnɪ	tsene	sendɨ
104	FLAT	pɪwa	pewa	peβa
105	FLY	mɪru	meru	mberu
106	FLY (INTR.)	uwɪwɪ	uwe	βeβé
107	FOOT	pɨta	pɨta	pɨtá
108	FOREST	kawa	kawa	kaʔá
109	FORGET (TR.)	sɨsarai	tsitsari	sesaraja
110	FOUR	ɪruaka	irwaka	iru
111	FRIEND∼LOVER	tɨwasa	tɨwatsa	atɨβasaβa
112	FROG	kurʊru	kurʊru	kurʊru
113	FRUIT	ia	ija	ɨʔá
114	GARBAGE	ɨtɨ	ɨtɨ	ɨtɨ
115	GIVE	mi	mi	meʔeŋa
116	GO DOWNRIVER (INTR.)	asɨrɨka	atsɨrɨka	sɨrɨka
117	GOURD	kuja	kuja	kuja

NO.	SET	OMAGUA	KOKAMA	TUPINAMBÁ
118	GRAB	japiʃika	japitʃika	pɨsɨka
119	GRANDCHILD (F. EGO)	ɾimiaɾiɾu	ɾimjaɾiɾu	tembiaɾiɾõ
120	GRANDFATHER	amui	amwi	amũja
121	GRASS	kapi	kapi	kapiʔí
122	GREEN	ɨkɨɾa	ɨkɨɾa	kɨɾa
123	GRILL	juɾa	juɾa	juɾá
124	HAMMOCK	ini	ini	inĩ
125	HAND	pua	puwa	pó
126	HAPPY	saɾɨwa	tsaɾɨwa	oɾɨβa
127	HARD	jumuata	tata	atã
128	HEAD	jakɨ	jakɨ	akaŋa
129	HEAR (TR.)	sɪnu	tsenu	senduβa
130	HEART	ĩja	ija	ɲɨʔa
131	HEAVY	ipuʃi	iputʃi	posɨja
132	HIGH UP (ADV.)	ɨwati	ɨwati	ɨβaté
133	HOLE	kʷaɾa	kwaɾa	kʷaɾa
134	HORN	ijaka	ijaka	aka
135	HORSEFLY	mutuka	mutuka	mutuka
136	HOUSE	uka	uka	oka
137	HOWLER MONKEY	akɨkɨ	akɨkɨ	akɨkɨ
138	HUNGRY	jamaʃi	jamatʃi	ambɨasɨ
139	HURT	saʃi	tsatʃi	asɨ
140	HUSBAND	mɪna	mena	mena
141	INDIGENOUS PERSON	tapɨja	tapɨja	tapɨʔɨja
142	INVITE (TR.)	paɾisaɾa	paɾitsaɾa	paɾesaɾa
143	ISLAND	ɨpau	ɨpwa	ɨpaʔũ
144	JOIN∼ATTACH	jatɨɾɨ	jatɨɾɨ	atɨɾa
145	KILL	ajuka	ajuka	juká
146	KNIFE	kɨʃi	kɨtʃi	kɨsé
147	LAKE	ɨpasu	ɨpatsu	upaβa
148	LAND	tujuka	tujuka	tujuka
149	LARVA	uɾa	uɾa	uɾa
150	LATE	ɨpɨsa	ɨpɨtsa	pɨsajé
151	LEAVE (TR.)	iʃaɾi	itʃaɾi	sejaɾa
152	LEAVE FROM (INTR.)	uʃima	utʃima	sema

NO.	SET	OMAGUA	KOKAMA	TUPINAMBÁ
153	LIVER∼HEART	pɨa	pɨa	pɨʔá
154	LONG	ipuku	ipuku	puku
155	LONG TIME AGO	ɨmɨnua	ɨmɨnwa	umuã
156	LOOK (INTR.)	umai	umi	maʔẽ
157	LOOK FOR (TR.)	ʃikaɾi	tʃikaɾi	sekaɾa
158	LOSE WAY	supaɾa	tsupaɾa	sopaɾa
159	LOSE ONESELF	kajma	kajɨma	kaɲema
160	LOUSE	kɨwa	kɨwa	kɨβa
161	LUNG∼BREATHE (INTR.)	putu	putu	pɨtu
162	MACAW	aɾaɾa	aɾaɾa	aɾaɾa
163	MANIOC FLOUR	ui	uwi	uʔí
164	MANY∼MUCH	ʃita	tʃita	setá
165	MIX (TR.)∼PRESS	kamiki	kamika	kambɨka
166	MONKEY	kai	kaj	kaʔí
167	MOON	jasɨ	jatsɨ	jasɨ
168	MOSQUITO	jatiú	jatju	ɲatiʔũ
169	MOURN	sapiɾu	tsapiɾu	sapiɾõ
170	MOUSE∼RAT	sanuja	tsanuja	saujá
171	MOUTH	juɾu	juɾu	juru
172	NAIL (BODY PART)	pɨsapi	pɨtsape	pɨsapẽ
173	NAME	iɾa	iɾa	eɾa
174	NECK∼THROAT	jaʃuka	jatʃuka	aseoka
175	NEPHEW	mɪmɨɾɨa	memɨɾɨa	membɨɾa
176	NEW	ɨpɨsasu	ɨpɨtsatsu	pɨsasu
177	NIGHT	ɨpɨsa	ɨpɨtsa	pɨsajé
178	NOSE	tii	ti	tĩ
179	OBEY (TR.)	sapiaɾi	tsapjaɾi	sapjaɾa
180	OPEN (TR.)	ɪpɪka	epeka	peká
181	OTHER	amua	amwa	amó
182	OWNER	jaɾa	jaɾa	jaɾa
183	PACA	paka	paka	paka
184	PADDLE (N.)	japukuita	japukita	pukujtaβa
185	PAINT∼WRITE	kʷatiaɾa	kwatjaɾa	kʷatiaɾa
186	PATH	pii	pe	pé
187	PATIO	ukaɾa	ukaɾa	okaɾa

NO.	SET	OMAGUA	KOKAMA	TUPINAMBÁ
188	PEANUT	munui	muni	manduβi
189	PERSON	awa	awa	aβá
190	PET	mɨma	mɨma	emimbaβa
191	PINEAPPLE	nana	nana	naná
192	PIRANHA	ipiɾai	ipiɾi	piɾaŋa
193	PIT VIPER SP.	jaɾaɾaka	jaɾaɾaka	jaɾaɾaka
194	PLANT (TR.)	jatɨma	jatɨma	tɨma
195	POOR	paɾiasu	paɾjatsu	poɾeausuβa
196	PREY	miaɾa	mjaɾa	embiaɾa
197	RAIN	amana	amana	amana
198	RED	pɨtani	pɨtani	piɾaŋa
199	REST (INTR.)	japɨtu	japɨtu	putuʔú
200	RETURN (INTR.)	ɪɾɪwa	iɾiwa	jeɾeβa
201	RIVER	paɾana	paɾana	paɾanã
202	ROAST (TR.)	mitʃiɾa	mitʃiɾa	miʃɨɾa
203	ROCK∼SHAKE	wɨuta	wɨwɨta	moag^{w}ɨag^{w}ɨ
204	ROPE∼THREAD	inimu	inimu	inimbó
205	ROUND	japua	japwa	apuʔã
206	SANDFLY	maɾiwi	maɾiwi	maɾig^{w}í
207	SAP∼CAUIM	kai	kaj	kawĩ
208	SCRAPE (TR.)	kaɾai	kaɾi	kaɾãja
209	SHAMAN	sumi	tsumi	sumé
210	SHINE (INTR.)	pɪɾa	peɾa	βeɾaβa
211	SHOUT	sasɨma	tsatsatsɨma	sasema
212	SING (INTR.)	ikaɾa	ikaɾa	ɲeʔeŋaɾa
213	SISTER	kunia	kunja	kuɲã
214	SISTER-IN-LAW (F. EGO)	uki	uki	ukeʔí
215	SIT (INTR.)	japɨka	japɨka	g^{w}apɨka
216	SKIN	piɾuaɾa	piɾwaɾa	piɾueɾa
217	SLOTH	aɨ	aɨ	aʔɨ
218	SMELL (INTR., TR.)	sɪtuni	tsetuni	setuna
219	SMOKE (N.)	tatatini	tatatini	tatatiŋa
220	SMOKE FOOD	mɪmukai	memuki	mokaʔẽ
221	SMOOTH	ɨsɨma	ɨtsɨma	sɨma
222	SNAKE	mui	muj	mboja

NO.	SET	OMAGUA	KOKAMA	TUPINAMBÁ
223	SON (M. EGO)	taɨɾɨa	taɨɾɨa	taʔɨɾa
224	SORUBIM SP.	suɾui	tsuɾi	suɾuβi
225	SPICY	tai	taj	taja
226	SPIDER	janu	janu	ɲanduĩ
227	SPILL∼POUR	iʃini	itʃini	ena
228	SPIRIT	mai	maj	mbaʔé
229	SPLIT (TR.)	pɪsɪ	petse	pesẽ
230	STAR	sɪsu	tsetsu	seiʃu
231	STEAL (TR.)	muna	muna	mondá
232	STEM	ɨwa	ɨwa	ɨβa
233	STONE	itakɨ	itaki	itakɨ
234	STRAIN (TR.)	jumukua	jumuka	mogʷaβa
235	SOUND (V.)	ipu	ipu	pũ
236	SUN	kʷaɾaʃi	kwaɾatʃi	kʷaɾasɨ
237	SWALLOW (V.)	jumukuni	jumukuni	mokona
238	SWEAT (INTR.)	sɨ̃i	tsɨji	sɨʔaja
239	SWEET	sɪɪ	tse	seʔẽ
240	SWEET POTATO	itika	itika	jetɨka
241	TAIL	sũi	tsuwi	uaja
242	TAPIR	tapiɾa	tapiɾa	tapiʔíɾa
243	TERMITE	kupia	kupea	kupiʔí
244	THIGH	sɪtɨma	tsutɨma	etɨmã
245	THIN	miɾi	miɾi	miɾĩ
246	THING	maɾai	maɾi	maɾã
247	THORN	jua	juwa	juatĩ
248	THREE	musapɨɾɨka	mutsapɨɾɨka	mosapɨt
249	THROW	atika	itika	itɨka
250	THUNDER	tupana	tupa	tupana
251	TIRED	kaniú	kanju	kaneʔõ
252	TOBACCO	pɪtɨma	petɨma	petɨmbu
253	TONGUE	kumɨɾa	kumɨɾa	apekũ
254	TREE	ɨwɨɾa	ɨwɨɾa	ɨβɨɾá
255	TOOTH	sai	tsaj	ãja
256	TWO	mukuika	mukujka	mokõj
257	URINATE (INTR.)	kʷaruka	kwaɾuka	kaɾuka

NO.	SET	OMAGUA	KOKAMA	TUPINAMBÁ
258	VILLAGE∼LAND	ɾitama	ɾitama	tetama
259	VINE	iʃipu	itʃipu	ɨsɨpó
260	VULTURE	uɾupu	uɾupu	uɾuβu
261	VULVA	tamatia	tamatja	tamatiʔá
262	WASP	kawa	kawa	kaβa
263	WATER	tɨa	tɨja	tɨ
264	WHITE	tini	tini	tiŋa
265	WHITE-LIPPED PECCARY	tajasu	tajatsu	tajasu
266	WIDE	ɪpɪwasu	epewatsu	pɨg^{w}asu
267	WIFE	miɾikua	miɾikwa	embiɾekó
268	WIND	ɨwɨtu	ɨwɨtu	ɨβɨtu
269	WING	pɪpu	pepu	pepó
270	WORM∼LARVA	sasuka	tsatsuka	sɨsoka
271	WRAP	juana	juwana	uβana
272	YELLOW	iju	iju	juβa
273	YESTERDAY	ik^{w}aʃi	ikwatʃi	k^{w}eisé
274	YOUNG MAN	kunumi	kunumi	kuɾumĩ
275	YOUNG WOMAN	kuniatai	kunjati	kuɲãtaĩ

Appendix B

Source code and raw datasets

The following appendix contains the Python code for automatic cognate detection and alignment for Omagua, Kokama, and Tupinambá. The spreadsheet files with the LingPy input and output are too large to be included here, but can be downloaded separately on `https://lup.lub.lu.se/student-papers` as an appendix to this thesis. These spreadsheets contain the original lexical data analyzed by LingPy, and the result of the cognate detection and alignment, which was then uploaded to EDICTOR and manually corrected there, as described in section 3.2. The algorithms were run by undergraduate student Eric Chen at UC Berkeley, while the data preparation and the manual editing were performed by the author.

```
# coding: utf-8

# In[1]:

import lingpy as lp
import pandas as pd
import re

from functools import reduce

# In[2]:

pd.set_option('display.max_rows', 2000)
pd.set_option('display.max_columns', 50)

# # Tokenize

# In[3]:

false_affricate = re.compile(r'([^t])([s])')

def tokenize(fun):
    replacement_dict = {' ': '_', '-': '+', 'N': ''}
    for original, replacement in replacement_dict.items():
        fun = fun.replace(original, replacement)
    subbed = false_affricate.sub(r'\1-\2', fun)
```

```
    return lp.ipa2tokens(subbed, semi_diacritics='s', merge_vowels=False, merge_geminates=False)

# ## Omagua

# In[4]:

df_oma = pd.read_csv('omagua.csv', encoding='utf8', keep_default_na=False)
df_oma = df_oma[1:]
print(f'Omagua average number of forms per meaning: {df_oma.TUE.count() / df_oma.TUE.nunique()}')

# In[5]:

df_oma['TOKENS'] = df_oma['FUN'].apply(tokenize)
df_oma

# In[6]:

oma_token_sets = df_oma['TOKENS'].apply(set)
reduce(set.union, oma_token_sets.tolist())

# ## Kokama

# In[7]:

df_kok = pd.read_csv('kokama.csv', encoding='utf8', keep_default_na=False)
df_kok = df_kok[1:]
print(f'Kokama average number of forms per meaning: {df_kok.TUE.count() / df_kok.TUE.nunique()}')

# In[8]:

df_kok['TOKENS'] = df_kok['FUN'].apply(tokenize)
df_kok

# In[9]:

kok_token_sets = df_kok['TOKENS'].apply(set)
reduce(set.union, kok_token_sets.tolist())

# ## Tupinamba

# In[10]:

df_tup = pd.read_csv('tupinamba.csv', encoding='utf8', keep_default_na=False)
df_tup = df_tup[1:]
print(f'Tupinamba average number of forms per meaning: {df_tup.TUE.count() / df_tup.TUE.nunique()}')

# In[11]:

df_tup['TOKENS'] = df_tup['FUN'].apply(tokenize)
df_tup

# In[12]:

tup_token_sets = df_tup['TOKENS'].apply(set)
reduce(set.union, tup_token_sets.tolist())

# # Combine data

# In[13]:

df_oma['DOCULECT'] = 'Omagua'
df_kok['DOCULECT'] = 'Kokama'
df_tup['DOCULECT'] = 'Tupinamba'

def fix_empty_concept(concept):
    return concept if concept else 'NONE'

df_wordlist = pd.concat([df_oma, df_kok, df_tup], ignore_index=True)
df_wordlist = df_wordlist.rename(index=str, columns={'FUN': 'IPA', 'TUE': 'CONCEPT'})
df_wordlist['TOKENS'] = df_wordlist['TOKENS'].apply(' '.join)
df_wordlist['CONCEPT'] = df_wordlist['CONCEPT'].apply(fix_empty_concept)
df_wordlist
```

```
# In[14]:

# convert to tsv format that lexstat wants as input

df_wordlist.to_csv('wordlist.csv', index=False, encoding='utf8')
wordlist = lp.basic.wordlist.get_wordlist('wordlist.csv')
wordlist.output('tsv', filename='wordlist')

# # Cognate detection

# In[15]:

lex = lp.LexStat('wordlist.tsv', check=True)
lex.cols

# In[16]:

lex.get_scorer(runs=10000)
lex.output('tsv', filename='tupi-guarani.bin')

# In[17]:

lex.cluster(method='lexstat', threshold=0.55, ref='cogid', cluster_method='infomap')

# # Alignment analysis

# In[18]:

alm = lp.Alignments(lex, ref='cogid', segments='tokens')
alm.align(method='library', scoredict=lex.cscorer)

# In[19]:

alm.output('html', filename='tupi-guarani_alignment')
alm.output('tsv', filename='tupi-guarani_alignment')
```