LUND UNIVERSITY

So close and yet so different:

Reconstructing the phonological history of three Southern New Caledonian languages

William Zetterberg

Supervisors: Arthur Holmer, Mechtild Tronnier

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

June 2021

Abstract

This thesis investigates the phonological history of three languages of New Caledonia, an overseas special collectivity of France, located in southwest Pacific Ocean. New Caledonia is home to remarkable linguistic diversity, with around 28 distinct indigenous languages varieties spoken today. These languages, known as Kanak languages, are members of the Oceanic subgroup of the Austronesian language family. Most of these languages are spoken on the main island, Grande Terre, and are commonly organized into two subgroups, a Northern and a Southern group. The Southern languages have previously been proposed to form two distinct subgroups alongside the Northern subgroup, a Mid-Southern and Far-Southern subgroup respectively. However, little research has so far been conducted on the phonological history of the languages of the Mid-Southern group, and it has not been possible to systematically evaluate the position of these languages in relation to the languages of the Northern and Far-Southern subgroups. This thesis therefore focuses on three previously described members of the Mid-Southern group, Ajië, Tîrî, and Xârâcùù, with the aim to reconstruct the phonological structure of the last common ancestor of these languages. The goal of this reconstruction was to clarify the position of the Mid-Southern languages within the New Caledonian group. In order to reconstruct the phonological system of this common ancestor, a large set of lexical and morphological items were secondarily collected from various published wordlists and dictionaries in the three languages. From these, 266 sets of cognate words were compiled between the three languages, from which sound correspondences were generated. The correspondences between the languages were systematically analyzed, based on which predictions were made about phonological properties in the common proto-language. As such, this study presents the first phonological reconstruction of the ancestral form of the Mid-Southern languages of mainland New Caledonia. The results of this study indicate that many of the characteristic traits found in the Mid-Southern languages evolved already in the common Proto-Mid-South language. By further comparing the results with previous comparative work on the Northern and Far-Southern subgroups in comparison with higher-order Austronesian reconstructions, the phonological reconstruction present strong evidence in favor of a distinct Mid-Southern subgroup of the New Caledonian mainland, which is characterized by a number of phonological innovations that can be credited to the common ancestor, of which several cannot be reconstructed to the neighboring Northern and Far-Southern subgroups.

Keywords: Historical linguistics, Comparative method, Reconstruction, Sound change, Phonology, Austronesian, Oceanic

Acknowledgments

First and foremost, I want to thank my two supervisors Arthur Holmer and Mechtild Tronnier for all their guidance and patience throughout this journey, and for all their helpful feedback on my many drafts over this last year. A huge thank you also goes to Arthur Holmer for first introducing me to the wonderful world of Austronesian languages, and for suggesting that I take a closer look at the languages of New Caledonia when I was looking for topics to explore in my thesis.

Lastly, I am grateful to all my friends and family who have supported me and cheered me on throughout this process. A special thanks goes to all of you who have offered helpful suggestions about my work, however small these may have been. You have all made a difference!

Table of contents

List of maps

List of images

List of figures

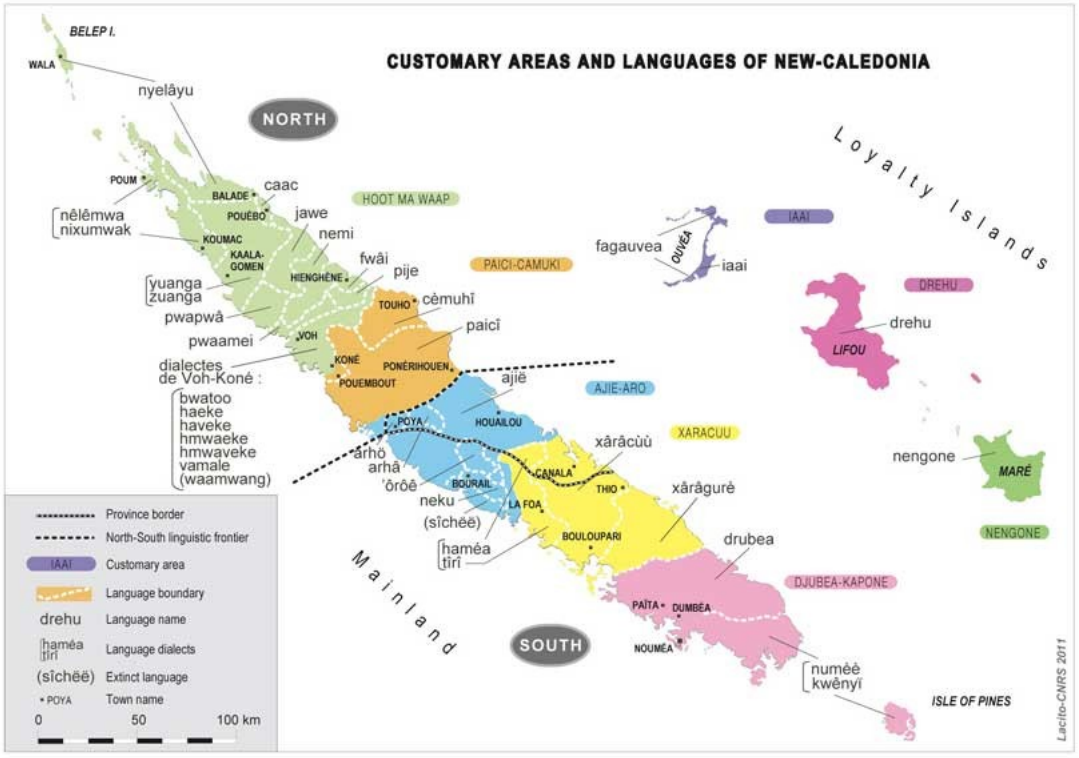
List of tables

Abbreviations

AJE	Ajië language (A'jië, Ajie, Houailou, Wailu)
DL	dual number
EXCL	exclusive pronominal form
FOC	focal or independent form
INCL	inclusive pronominal form
OBJ	object form (accusative case)
PL	plural number
PNC	Proto-New Caledonian
POc	Proto-Oceanic
SBJ	subject form (nominative case)
SG	singular number
TIR	Tîrî language (Tîrî-Mea, Tiri, Tinrin, Ciri, Cîrîî, Grand Couli)
XAC	Xârâcùù language (Canala, Aneyu)

1 Introduction

The New Caledonian or Kanak languages refers collectively to the traditional languages of the Kanak people, the indigenous Melanesian inhabitants of the archipelago of New Caledonia or Kanaky[1], an overseas special collectivity of France, located in southwest Pacific Ocean, roughly 1,500 km east of Australia. The archipelago consists of a main island, Grande Terre, the neighboring Loyalty Islands, Belep Islands, Île de Pins, as well as a number of smaller uninhabited islets. The region has a total population of 271,407, of which more than 90% live on the mainland (ISEE, 2019).[2] Kanaks are the largest ethnic group in the region, comprising 39% of the total population, followed by Europeans (27%), Wallisians and Futunians (8%), while other communities (Tahitians, Ni-Vanautu, etc.) constitute the remainder of the total population (ISEE, 2014).[3]



Map 1: Customary areas and languages of New Caledonia (Lacito-CNRS, 2011).

1 The name Kanaky has been introduced by indigenous and pro-independence groups (Sallabank, 2015, p. 33).

2 https://www.isee.nc/population/recensement/structure-de-la-population-et-evolutions (retrieved September 9, 2020).

3 https://www.isee.nc/population/recensement/communautes (retrieved September 9, 2020).

New Caledonia is home to remarkable linguistic diversity. The exact number, names, and spellings of all the indigenous languages varies between sources. Ethnologue lists 34 living languages (Eberhard et al., 2020), while Glottolog lists 33, of which one is reported to be extinct.[4] A commonly cited number is 28 (see Moyse-Faurie & Néchérö-Jorédié, 1989; Osumi, 1995). The majority of the languages are spoken on Grande Terre, an area around 50 km wide and 500 km long. As illustrated in Map 1, the languages are commonly illustrated according to the "customary areas" (French *aires coutumières*) from which they originate, but many are now reported to be spoken in the capital of Nouméa, a result of an increasing urbanization in the region (Sallabank, 2015, p. 36).

The languages of the mainland are classified according to a north-south linguistic border as illustrated in Map 1, where the languages of the mainland are more closely related to each other than to the neighboring languages of the Loyalty Islands (Lynch & Ozanne-Rivierre, 2001). The Northern varieties have since been demonstrated to together form a Northern branch of the Mainland group (Ozanne-Rivierre, 1995). However, the linguistic classification of the Southern languages is debated, which have previously been classified into two groups, a Far-Southern group, which comprises the two languages Numèè and Drubea, and a Mid-Southern group, which comprises the remaining languages of the south, further organized into three subgroups (Haudricourt, 1971). Later proposals have placed the two Southern groups together in a Southern branch (see Lynch & Ozanne-Rivierre, 2001), but this remains to be demonstrated.

In addition to the Kanak languages, there are also multiple non-Melanesian languages spoken in the archipelago.[5] These have all been introduced by later migration or colonization and are not of interest to this study. This includes first French, the official and most widely spoken language in the region. Other significant language communities listed by Ethnologue include the Polynesian languages Futuna and Wallisian, as well as Javanese, Tahitian, and Vietnamese, all introduced to the region within the last two centuries, as well as a French-based creole, Tayo, spoken in the South province (Eberhard et al., 2020). An additional language, Fagauvea (or West Uvean), is spoken on the island Ouvéa together with the Melanesian language Iaai, but the former is a Polynesian outlier language, brought to the region before French colonization (Haudricourt, 1971).

The indigenous people, culture and languages, have a history of being marginalized since French colonization in 1853 (Sallabank, 2015, p. 31). Today, the Kanak languages are all spoken by small

4 https://glottolog.org/resource/languoid/id/newc1243 (retrieved September 14, 2020).

5 I will refer to the Melanesian languages of New Caledonia as "Kanak languages" or "New Caledonian languages" throughout this thesis to avoid confusion with the term "languages of New Caledonia", which may refer to any language of the New Caledonian archipelago.

communities relative to the population, with speakers being typically bilingual in French, which functions as the lingua franca across languages communities throughout the archipelago (Sallabank, 2015, p. 37). Many of the language are severely threatened. Of 34 listed living languages, Ethnologue reports that one language has no native speakers left, six languages are used only by the older generation, seven languages are reported to not be actively transmitted to children, and another 13 languages are reported to be diminishing (Eberhard et al., 2020). From a comparative perspective, the languages exhibit a great deal of variation in both phonology (Haudricourt, 1971) and syntax (Moyse-Faurie & Ozanne-Rivierre, 1983), and present an extraordinary tale of linguistic diversification. With this thesis, I want to shed some further light on the phonological history of the Mid-Southern languages in particular.

1.1 History of research

Because the Kanaks had no prior written traditions, the earliest sources on the Kanak languages derive from the first contacts between the indigenous population and European sailors. These are wordlists collected by the Cook and d'Entrecasteaux expeditions in the late 18th century. More documentation of the languages appeared shortly thereafter, starting in the 19th century with works by Christian missionaries. Much of the data collected during these early expeditions and later missions is largely phonologically and grammatically deficient for comparative work, either being produced in English or French orthographies, and reflecting a low understanding of phonological aspects and grammatical properties of the indigenous languages (Haudricourt, 1971, pp. 359–360).

Starting from the first half of the 20th century, a French pastor stationed on the mainland, Maurice Leenhardt, produced important comparative works on Kanak languages and dialects, as well as a grammar and dictionary of Ajië. Research by trained linguists did not take off until after the second world war, however. Since then, research on Kanak languages has focused mainly on language documentation and description, but comparative historical linguistic work has also been undertaken, focusing initially on phonological reconstruction, but morphological and syntactic topics have also been investigated more recently. At the present date, most languages have been documented on some level, including bilingual dictionaries, and phonological and grammatical descriptions or sketches. The bulk of this research has been published in French, which may explain why many findings have not gained attention in English speaking research circles.

French linguist André-Georges Haudricourt conducted a comparative survey of Kanak languages in 1962 and 1966 and offered the first diachronic considerations and internal subgrouping of the languages, proposing that the languages form a part of the Oceanic group of the Austronesian language family. Haudricourt's classification has later been refined by other researchers. The most prominent expert on Kanak languages, French linguist Françoise Ozanne-Rivierre, published many descriptive and comparative works on the Kanak languages, focusing mainly on phonological reconstruction in the languages of the northern mainland and the Loyalty Islands (Ozanne-Rivierre, 1992; 1995). Other contributors include Australian linguist John Lynch, who has conducted comparative work on the phonological systems of the languages of the Loyalty Islands (Lynch, 2003), and French linguist Jean-Claude Rivierre, who has investigated tonogenesis on the mainland (Rivierre, 1993) and produced a comparative phonology of two languages of the far south in 1973.

The languages of the mid south have so far been subject to very little historical linguistics work. The late linguist George Grace investigated the phonological history of two Mid-Southern languages Tîrî and Xârâcùù in the late 20th century but made little progress in deciphering the phonological history of the two languages (see Grace, 1996). Grace (1992, p. 120) concluded that:

> "we're still left with the question of how these languages came to be the way they are. My direct attempt to reconstruct, by means of the comparative method, the precise changes they [Xârâcùù and Tîrî] had undergone has failed."

Grace's findings have since contributed to the discussion about what he first called the "aberrant" Austronesian languages of Melanesia and why these are so difficult to reconstruct (see Pawley, 2006). However, since then, no further research has been undertaken to bring us any closer to understanding the phonological history of the Mid-Southern languages. There is no extensive list of cognates or sound correspondences between languages of the Mid-Southern group, which means that it has not been possible to systematically evaluate the position of these languages in relation to the proposed Northern and Far-Southern groups of the New Caledonian mainland.

1.2 Aim and research question

The aim of this study is therefore twofold. The primary aim is to reconstruct as much as possible about the phonological distinctions of the last common ancestor to the Mid-Southern languages, by means of comparing lexical and morphological data from a selected set of descendant languages. The secondary aim and goal of the phonological reconstruction is to clarify the position of the Mid-

Southern languages within the New Caledonian group in relation to reconstructions of earlier ancestral forms. The study was driven by the following research question:

RQ: What phonological distinctions can be reconstructed to the last common ancestor of the Mid-Southern subgroup?

This research question is best addressed using the comparative method of linguistic reconstruction. This method relies on lexical and morphological data from a selected set of languages and builds on the two fundamental hypotheses about the material. First, that the selected set of languages are (typically) related by virtue of being descended from a common, hypothesized ancestor, a so-called "proto-language", and second, that this relationship can be demonstrated through regular correspondences between the descendant languages, from which hypotheses can be derived about phonological distinctions in the proto-language. I will return to principles of this method in more detail in section 3.2.

Three languages classified in the Mid-Southern subgroup were chosen to address this research question: Ajië, Tîrî, and Xârâcùù. This choice followed with a basic hypothesis about their linguistic classification. In this case, each of the three languages have previously been classified into three separate subgroups within the Mid-Southern group (Haudricourt, 1971), which initially minimized the risk of any of the three being more closely related to each other, and therefore provided a good starting point for comparative research on the Mid-Southern group.

1.3 Structure of the thesis

The remainder of the thesis is organized into four chapters. In chapter 2, I provide a background to the external and internal classification of the Kanak languages, as well as giving a linguistic background to the three languages in question, including a phonological sketch of the three respective languages. The background also includes the results of previous comparative work on Kanak languages, to be referenced in light of the phonological reconstruction. In chapter 3, I describe the method, including what type of data was used, where the data came from, and how it was managed and prepared for analysis. In the same vein, I introduce the theory and process behind the comparative method. In chapter 4, I present and discuss the results of the phonological reconstruction and address my two aims as formulated above. Chapter 4 concludes with a general discussion of the method and results. Finally in chapter 5, I summarize the results of the thesis, and consider prospects for future research.

2 Background

2.1 Subgrouping

2.1.1 Language change and types of subgroups

When considering how languages change over time, an analogy can be drawn with the two fundamental processes of biological evolution, that of divergent and convergent evolution. That is, much like in the process of divergent evolution in biology, where a single population splits into two or more populations by evolving an increasing number of independent traits over time, a linguistic ancestor gives rise to two or more distinct descendant varieties, each defined by the number of independent traits they evolved over time in relation to their ancestral form. Likewise with the process of convergent evolution in biology, where two or more genetically distinct populations evolve an increasing number of shared traits over time, giving rise to analogous structures in organisms, a set of unrelated languages may evolve an increasing number of analogous traits over time. But there are also limitations to this analogy. Unlike organisms, linguistic traits may freely converge across both related and unrelated language varieties by means of language contact, forming “hot spots” for convergence, often called “diffusion areas”. The issue with classifying languages in such areas have gained much attention in Oceanic linguistics, and researchers have since introduced two types of subgroups in linguistic classification, so-called “innovation-defined” subgroups, and “innovation-linked” subgroups (Pawley & Ross, 1995).

A so-called innovation-defined subgroup is a subgroup in which all members share a set of innovations that evolved in their immediate ancestor and are reflected in each of the daughter languages. This kind of subgrouping is what is represented in the traditional family-tree model, and what I will refer to using the term “subgroup” in this thesis.

A so-called innovation-linked subgroup, often called a “linkage” (Lynch, 1999), is a subgroup in which all the members form a network of languages with overlapping linguistic innovations, where no single innovation is shared throughout the entire network, and the languages as such cannot be defined by any single trait like with proper subgroups. A linkage may be a result of diffusion across languages that share more than one immediate ancestor, or it may be a result of

an earlier dialect continuum, in which case the languages share a single immediate ancestor, but still cannot be assigned any exclusively shared innovations because the parent did not exist long enough as a unit to evolve any defining innovations of its own (Ross et al., 2016, p. 12).

2.1.2 External and internal relations

The earliest evidence of human settlement on New Caledonian are remnants of stoneware associated with the Lapita cultural tradition. This archaeological tradition appeared first in the Bismark Archipelago around 1600 BCE, and rapidly spread outward, its bearers reaching New Caledonia, Vanuatu, Fiji, Tonga, and Samoa in Remote Oceania by boat around 1300–1000 BCE (Pawley & Ross, 1995, p. 64). The earliest archaeological sites of the Lapita tradition from the northern part of the New Caledonian mainland date back to around 1000 BCE. Judging from the archaeological record, the mainland appears to have been settled first, with the neighboring smaller islands being settled shortly thereafter (Sand, 1999, p. 142).

Around 600–800 BCE, Lapita associated stoneware is replaced by novel traditions throughout the archipelago. This divergence of archaeological traditions in the region is already observed within the first centuries after settlement, and by the common era, separate archaeological traditions had taken hold in the north and south of the mainland respectively (Sand, 1999, p. 155).

Haudricourt first explored the internal relationship of the Kanak languages, working under the hypothesis that these languages descended from a single Austronesian language that arrived to the archipelago before the common era (Haudricourt, 1971, p. 383). The closest relatives of the languages of New Caledonia are now believed to be the Austronesian languages of north, central, and south Vanuatu (Lynch, 1999; Lynch & Ozanne-Rivierre, 2001), which together are hypothesized to form an innovation-linked subgroup, reflecting an early dialect continuum that has been named the Southern Oceanic linkage (Lynch, 1999). This subgroup forms one of multiple higher order subgroups within the Oceanic group, as shown in Figure 1. Note that each node in figure 1 represents a split in the family tree which reflects a hypothesized linguistic ancestor[6], where proposed intermediary subgroups are marked in parenthesis. There is today a strong consensus that the spread and diversification of the last common ancestor of all Oceanic Austronesian languages, what is known as Proto-Oceanic, was directly associated with the spread of Lapita pottery into Remote Oceania (Pawley & Ross, 1995).[7]

6 Hypothesized ancestors are marked with the prefix *proto-*, e.g. Proto-Oceanic (POc), etc.

7 For an overview of this discussion, see Spriggs (1995).

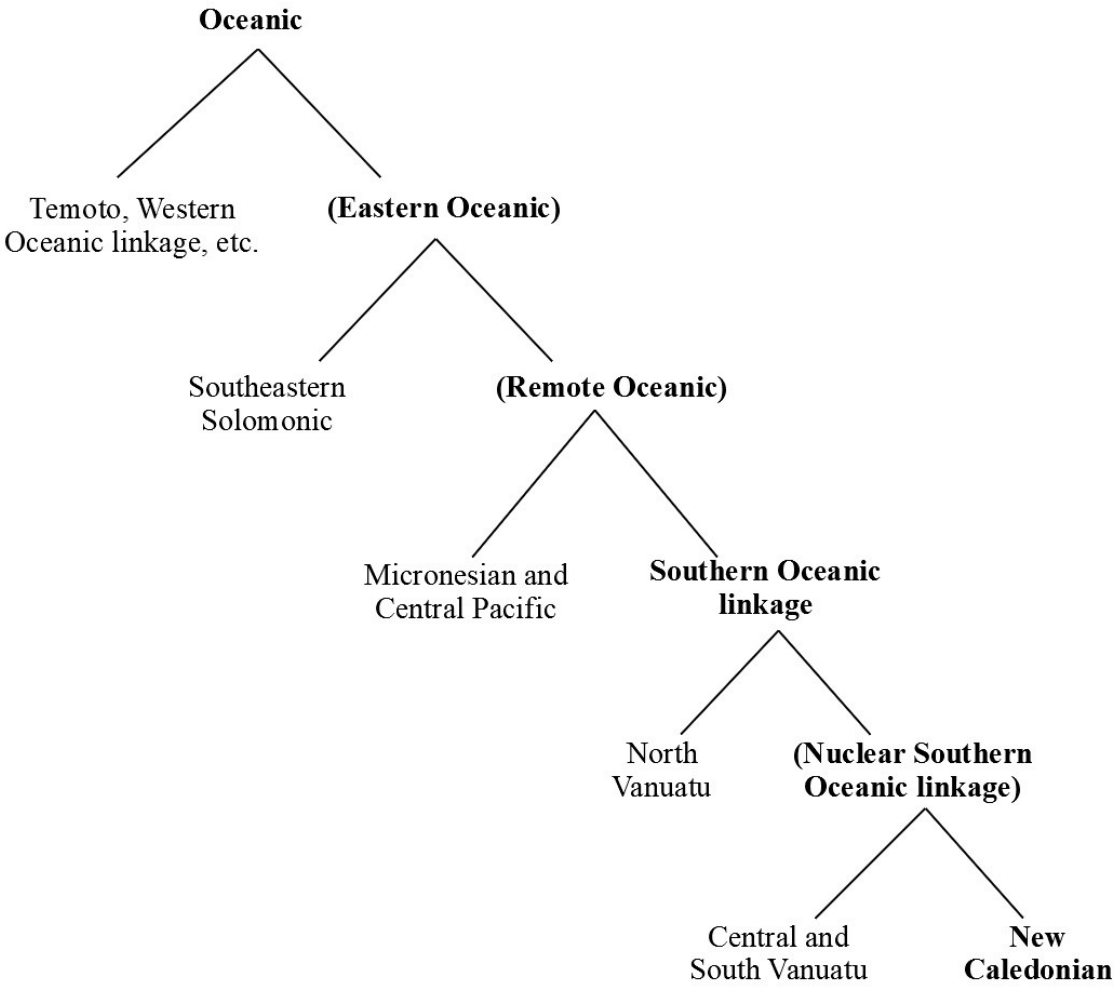


Figure 1: Major Oceanic subgroups (Ross et al., 2016).

There are currently two proposals for the internal classification of the languages of New Caledonia. Haudricourt's first suggestion for an internal subgrouping of the Kanak languages organized them into five groups on the mainland, Far north, North, Center, South, and Deep south, to which he added the three Kanak languages of the Loyalty Islands, without further specification about their internal relationships (Haudricourt, 1971, p. 359), as illustrated in Figure 2 below. The two subgroups of the southern mainland, which Haudricourt (1971) named the South and Deep south groups respectively, are I call the Mid and Far-Southern subgroups respectively in this thesis. Within the Mid-Southern subgroup, Haudricourt (1971) organized the language varieties into three additional subgroups, named after the primary settlements around which they are spoken (p. 372). As before, note that each node in the family tree also represents a hypothesized linguistic ancestor.

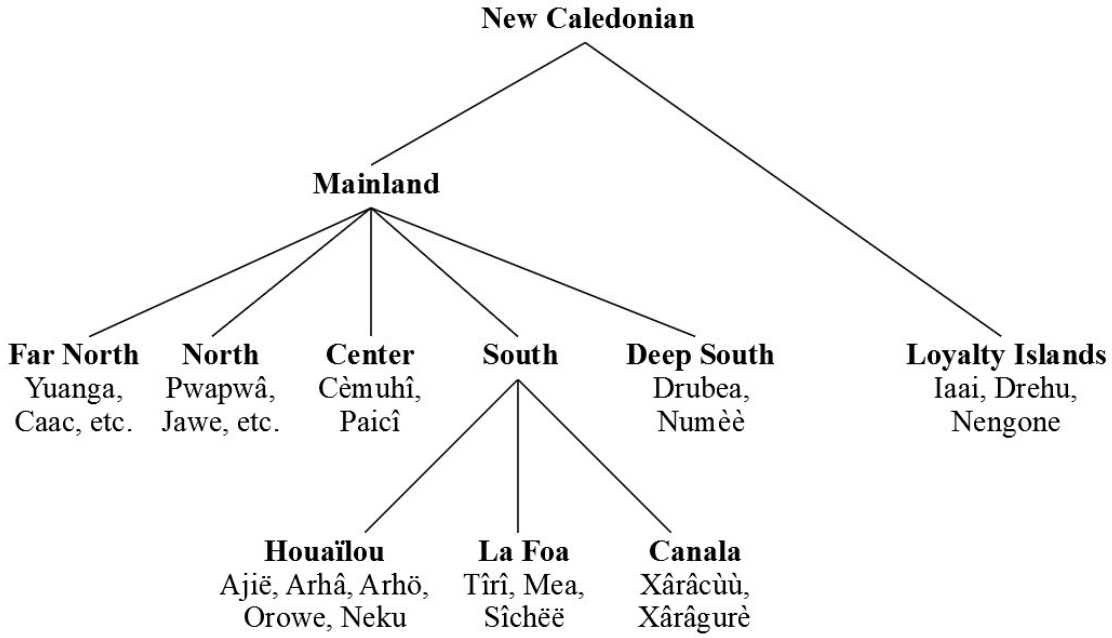
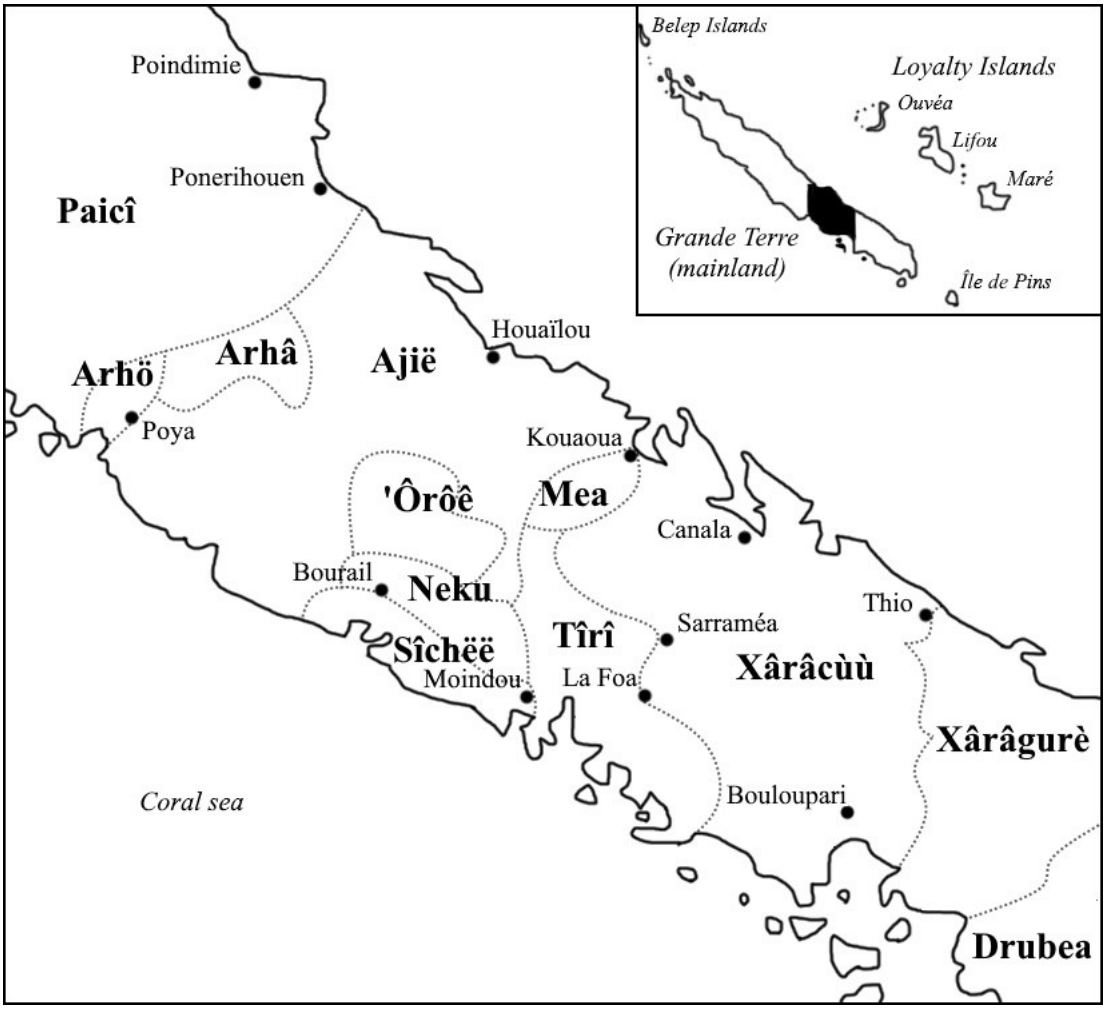


Figure 2: New Caledonian subgroups (Haudricourt, 1971).

Later research has concluded that the first split in the group must have been between the ancestor of the Kanak languages of the mainland and those of the Loyalty Islands (Lynch & Ozanne-Rivierre, 2001; Lynch, 2003). Haudricourt's classification (see Figure 2) was later elaborated by Ozanne-Rivierre, who demonstrated that the languages classified as Far north, North, and Center are linked by a number of phonological innovations, and together form a Northern branch of the Mainland group (Ozanne-Rivierre, 1995). This has led to the Mid and Far Southern languages to be later classified in a joint Southern branch (Lynch & Ozanne-Rivierre, 2001), but so far there is little evidence in support of this branch as a proper, innovation-defined subgroup. In this regard, Haudricourt's classification remains the most reliable model to date.

2.2 Language background

The Mid-Southern languages include at least ten distinct varieties, which are reported to be closely related within their three respective subgroups (see Figure 2), within which they have variously been described as either languages or dialects. The distribution of the respective language varieties is illustrated in Map 2. The two peripheral languages Paicî and Drubea belong to the Northern and Far Southern subgroups respectively. In this thesis, I will only focus on the three Mid-Southern languages, Ajië, Tîrî, and Xârâcùù, which each represent one of the three respective subgroups proposed by Haudricourt (1971), as discussed above.



Map 2: Mid-Southern languages.

2.2.1 Ajië

Ajië or Houailou (Wailu), known by native speakers as *mêrê a'jië* 'Ajië language' (IPA: [ˈmẽɽẽ ˈaʔⁿɟiʌ]), is spoken in the North and South Provinces of mainland New Caledonia, centered around Poya in the west, and Houaïlou and Kouaoua in the east, as illustrated in Map 2. The language borders the Northern language Paicî to the north, the closely related varieties Arhö and Arhâ in Poya, as well as 'Ôrôê (Orowe, Boewe), Neku, and Sichëë (Zire) at Bourail, and Mea and Xârâcùù to the east in Kouaoua. Ajië is among the most widely spoken Kanak languages today, for which Ethnologue reports more than 5000 speakers in a 2009 census (Eberhard et al., 2020).

Jacqueline de La Fontinelle (1976) distinguished between three areas of regional variation, primarily between the varieties of the inland valleys, known as *mêrê-a'kûrû-é* [ˌmẽɽẽˈaʔ ˈkũɽũe] 'language of the bottom of the valley', where there is a degree of variation between the varieties of the east side, centered around Houaïlou, and the west side, centered around Poya, and the varieties

of the east coast, known as *mêrê-a' parawiè* [ˌmẽɽẽˈaˀ ˈpaɽawiɛ] 'language of the shore'. In addition to regional variation, de La Fontinelle (1976) reported significant phonological differences between age groups at the time of her research, which overlapped to an extent with regional variation as well (p. 21). The neighboring languages Arhâ and Arhö are sometimes regarded as more divergent dialects of the same language (Haudricourt, 1971, p. 372).

In terms of modern research, Pastor Leenhardt published in 1935 a grammar sketch and dictionary of the language, *Vocabulaire et grammaire de la langue Houailou*. Jacqueline de La Fontinelle published in 1976 a detailed phonological and syntactic description, *La langue de Houaïlou*. In 1986, Fédération de l'Enseignement Libre Protestant developed an orthography for Ajië for use in schools and published a bilingual Ajië-French dictionary in 2001, *Dictionnaire a'jië-français* (Sylvain Aramiou, Jean Euritein, Georges Kaviviorio). For publications in English, a bilingual wordlist from Fédération de l'Enseignement Libre Protestant can be found in the *Comparative Austronesian Dictionary* (Aramiou & Euritein, 1995). Frantisek Lichtenberk published a grammar sketch of the language in 1978, *A sketch of houailou grammar*, based on previous documentation. An introduction to the language by Darrell Tryon can also be found in the *Comparative Austronesian Dictionary*, based on de La Fontinelle (1976).

2.2.2 Tîrî

Tîrî (IPA: [ˈʇ̪ĩɾ̃ĩ]), also known as Tinrin or the "language of Grand Couli", is spoken in reserves in the commune of Sarraméa, La Foa, and Moindou in the South Province, as illustrated in Map 2. A closely related variety called Mea is spoken in Kouaoua in the North Province, which is often regarded as a dialect of the same language (Osumi, 1994, p. 2). The two varieties are diminishing and were reported to only have around 600 speakers in a 2009 census (Eberhard et al., 2020).

Grace (1976) noted three distinct varieties of the language, one spoken in La Foa, called Tîrî, a second variety spoken in Kouaoua to the north, i.e. Mea, and third variety, which he interpreted as a leveled dialect, which was spoken primarily in the reserve of Grand Couli in Sarraméa. Midori Osumi (1995) identified dialectal differences between the varieties spoken in reserves of La Foa and the reserve Grand Couli in Sarraméa, as well as a now exist variety spoken by deportees on Île de Pins after the suppression of a revolt in 1878–1879 (pp. 3–10).

In terms of modern research, George Grace published a bilingual Tîrî-English dictionary in 1976, *Grand Couli Dictionary*, focused primarily on the variety spoken in the Grand Couli reservation in Sarraméa, with notes on regional variation. Midori Osumi later published a detailed phonology and

grammar of the language in 1995, *Tïnrin grammar*, based primarily on the variety spoken in La Foa, though she also made note of regional variation.

2.2.3 Xârâcùù

Xârâcùù, or the "language of Canala", known by native speakers as *nââ xârâcùù* 'Canala language' (IPA: [ˈnãː ˈxãɾãciː]), is spoken around Kouaoua and Canala in the North Province, and Sarraméa and Thio in the South Province. The language borders Ajië and Mea in Kouaoua to the northwest, Tîrî in Sarraméa to the south, and closely related Xârâgurè in Thio along the east coast, as illustrated in Map 2. The language is among the more widely spoken Kanak languages today, with more than 5000 speakers of the language in a 2009 census (Eberhard et al., 2020).

According to Claire Moyse-Faurie & Marie-Adèle Néchérö-Jorédié (1989), the language shows slight variation locally, with greater lexical and phonological differences observed primarily between the varieties spoken to around Canala and Sarraméa to the west, and the varieties spoken around Thio to the east. The eastern varieties are spoken alongside Xârâgurè, which is reflected by some shared properties between the two varieties. Xârâgurè is sometimes regarded as a more divergent dialect of the same language (Moyse-Faurie & Néchérö-Jorédié, 1989, pp. 16, 27).

Regarding modern research, George Grace published a bilingual Xârâcùù-English dictionary in 1975, *Canala Dictionary*. In 1989, Claire Moyse-Faurie and Marie-Adèle Néchérö-Jorédié published an extensive bilingual Xârâcùù-French dictionary, *Dictionnaire xârâcùù-français*. Claire Moyse-Faurie published a syntactic description of the language in 1995, *Le xârâcùù: langue de Thio-Canala (Nouvelle-Caledonie) – elements de syntaxe*. For sources in English, a grammar sketch by John Lynch can be found in *The Oceanic languages* (John Lynch, Malcolm Ross, & Terry Crowley), based on Moyse-Faurie's syntactic description. An introduction and bilingual wordlist contributed by Claire Moyse-Faurie can also be found in the *Comparative Austronesian Dictionary*.

2.3 Phonological background

In the following section, I will give a brief phonological sketch of each of the three languages. The purpose of this section is to aid the phonological reconstruction presented in chapter 4.

2.3.1 Ajië

The phonological description I present here is based on de La Fontinelle (1976).

2.3.1.1 Vowels

There are ten oral and six nasal vowels in Ajië, as shown in table 1, in the transcription proposed by de La Fontinelle (1976). Note that de La Fontinelle (1976) transcribed the "central" vowels using graphs for both central and back unrounded vowels. These vowels are importantly non-front and unrounded, where the degree of backness is debatable.

Table 1: Vowels of Ajië.

	Oral			Nasal		
	Front unrounded	Central unrounded	Back rounded	Front unrounded	Central unrounded	Back rounded
Close	i	ɯ	u	ĩ	ɯ̃	ũ
Mid-close	e	ə	o	ẽ		õ
Mid-open	ɛ	ʌ	ɔ			
Open		a			ã	

Vowel length is also phonemically distinct in Ajië, and each vowel may occur as short or long, and differ only in duration. This is analyzed as an aspect of the syllable structure, which is marked with double letters for long vowels. For this reason, only short vowels are exemplified in table 1 above. Minimal pairs include /ⁿgi/ 'horn' vs /ⁿgii/ 'cast a spell', /ⁿgɔ/ 'I (1SG.SBJ)' vs /ⁿgɔɔ/ 'to vomit', etc. (de La Fontinelle, 1976, p. 344).

According to de La Fontinelle (1976), the nasal vowels were unstable at the time of her study, with some older speakers differentiating between a mid-close back /õ/ and mid-open back /ɔ̃/, and mid-open central /ʌ̃/ and open central /ã/, which had merged as /õ/ and /ã/ respectively in the speech of younger generations. A mid-open front /ɛ̃/ was also marginally used by older speakers, which later generations merged with /ã/ (de La Fontinelle, 1976, p. 115). The reduced system is reflected in Aramiou & Euritein (1995), which I have also chosen to present in table 1 above.

The vowels are subject to some allophonic processes. The mid-open and mid-close oral vowels are neutralized in vowel sequences before unstressed /i, u, a, ã/, so that /ˈlə-a/ 'this yam (yam sp.)' and /ˈlʌ-a/ 'this landslide' are homophonous. In monomorphemic words containing such sequences, it is not possible to determine what the underlying vowel is, and in such cases de La Fontinelle (1976) marked the segment with a capital letter for an archiphoneme, e.g. /ˈpEi/ 'sick' is realized as

either [pej] or [pɛj] (pp. 135–138).

There is also a degree of regressive assimilation across syllable boundaries, particularly in the coastal variety, e.g. [ˈⁿɟakɛ] ~ [ˈⁿɟɛkɛ] ‘amaranth sp. (*A. gracilis*)’, [ˈⁿdeɽɔwe] ~ [ˈⁿdɔɽɔwe] ‘common sowthistle (*S. oleraceus*)’, etc. (de La Fontinelle, 1976, p. 138).

Nasality is noted to spread progressively when following a nasal consonant, or regressively when preceding a nasal consonant or prenasalized stop. Nasality may also spread progressively across syllable boundaries in vowel sequences and across morpheme boundaries, so that /ˈⁿgʷã-e/ ‘its head (head-3SG.POSS)’ is realized as [ˈⁿgʷãẽ], as well as across word boundaries, so that /pũ ɯ/ ‘trunk of mangrove (trunk-mangrove)’ is realized as [pũ ũ] (de La Fontinelle, 1976, p. 139). Nasality is also noted to spread progressively across some consonants, e.g. [ˈᵐbʷãɽãwe] ~ [ˈᵐbʷãɽãwẽ] ‘animal’, but the limits of this process are not discussed by de La Fontinelle (1976).

2.3.1.2 Consonants

de La Fontinelle (1976) distinguished 25 consonants in Ajië, as presented in table 2.[8] The inventory is organized here according to four places of articulation and six primary manners of articulation, where labialization is distinctive as a secondary articulation for bilabial and velar consonants. The labialized bilabials are produced with simultaneous velarization, i.e. raising of the tongue against the velum or soft palate, and are more accurately described as labiovelarized consonants. The consonants described as alveolars may be slightly postalveolar (de La Fontinelle, 1976, p. 40).

Table 2: Consonants of Ajië.

		Labial			Coronal		Palatal	Velar	
		Bilabial	Labiodental	Labialized	Alveolar	Retroflex			Labialized
Stop	voiceless	p		pʷ	t		c	k	kʷ
	voiced	ᵐb		ᵐbʷ	ⁿd		ⁿɟ	ᵑg	ᵑgʷ
Nasal	voiced	m		mʷ	n		ɲ	ŋ	
Vibrant	voiced				r	ɽ			
Fricative	voiced		v		ɹ		j	ɣ	w
Lateral	voiced				l				

There are three series of occlusive consonants in Ajië, voiceless stops, voiced stops, and voiced nasal consonants. The voiced stops are always preceded by a brief, voiced nasal portion, and may alternatively be described as prenasalized stops (de La Fontinelle, 1976, p. 28). As such, both

8 The transcription I use here differs slightly from de La Fontinelle (1976) in favor of standard IPA practices.

voiced stops and nasal consonants trigger the spread of nasality, as described above. Occlusive consonants are distinctive at four places of articulation, in addition to secondary articulation for bilabial and velar consonants, though there is notably no labialized velar nasal.

The voiced fricatives are pronounced with varying degree of friction. /j/ is often realized as a voiced palatal approximant [j] (de La Fontinelle, 1976, p. 54). /v/ is realized as a voiced labiodental fricative [v], though de La Fontinelle (1976) also reported it as a voiced bilabial fricative [β] among older speakers at the time of her study (p. 40). /ɣ/ and /w/ are described as voiced velar fricatives, of which the latter forms the labialized counterpart (de La Fontinelle, 1976, pp. 61–62). de La Fontinelle (1976) also recorded a labiovelarized fricative [vʷ] as a variant of /w/ in some words in the older population (p. 39).

There are four liquid consonants in Ajië, including three rhotics /ɹ, r, ɽ/[9], and one lateral /l/. The lateral is phonemically distinct from the rhotics, compare /kalaʔ/ ‘boogeyman’ vs /kaɽaʔ/ ‘beautiful’ (de La Fontinelle, 1976, p. 78). On the other hand, /l/ varies with /n/ in many words, e.g. [luɔ] ~ [nuɔ] ‘fog’, [lɯ] ~ [nɯ] ‘island’, etc. (de La Fontinelle, 1976, pp. 42, 49). Some near-minimal pairs still exist, e.g. /ⁿɟaluuɽi/ ‘to flood (with water)’ vs /ⁿɟanuɽi/ ‘to attach’ (de La Fontinelle, 1976, p. 78).

de La Fontinelle (1976) described complex variation regarding the realization of the three rhotics across both age groups and regional groups. At the time of her study, many older speakers maintained a twofold distinction in initial position between an alveolar approximant /ɹ/ and trill /r/, as well as a threefold distinction in intervocalic position between an alveolar approximant /ɹ/, an alveolar trill /r/, and retroflex flap /ɽ/ (de La Fontinelle, 1976, pp. 68, 84). Other older speakers had reduced this opposition by one distinction, so that the approximant /ɹ/ occurred in allophonic distribution with the trill /r/ in all contexts, while the intervocalic flap /ɽ/ varied freely between either an alveolar tap [ɾ] or a retroflex flap [ɽ], where the latter realization was preferred in nasal contexts, where it was typically nasalized as well (de La Fontinelle, 1976, pp. 82, 84). Some younger speakers had further reduced this system to a single distinction in both positions, with different speakers preferring a trill, tap, flap, or approximant, depending on phonological context and regional background (de La Fontinelle, 1976, pp. 69, 98). However, this three-way contrast is reflected in the orthography used by Aramiou & Euritein (1995), and for that reason, I have chosen to present this system in table 2.

9 de La Fontinelle (1976) distinguished three rhotics in Ajië: an approximant /ɹ/, which she transcribed as ⟨rh⟩; a trill /r/, which she transcribed as ⟨rr⟩; and a vibrant characterized by a single contact, which she transcribed as ⟨r⟩ (pp. 40–41). The latter had two variants, the first likely a tap [ɾ], described as “a weak vibrant” (p. 41), and the second a retroflex flap [ɽ], which de La Fontinelle (1976) marked as ⟨ṙ⟩ for disambiguation (p. 81).

2.3.1.3 Phonotactics

The Ajië syllable follows a (C)V structure. The language allows open syllables only, and the onset is not obligatory filled. The nucleus may be filled with any vowel, which may be either short or long, creating a total of 32 distinct vowel nuclei, though long nasal vowels are quite rare.

Mono-morphemic words are generally one or two syllables, and rarely three. Words with more than three syllables are nearly nonexistent, and are always multimorphemic (de La Fontinelle, 1976, p. 132). The following examples are lifted from de La Fontinelle (1976, p. 347):

CV	/kʌˀ/	‘pot’
CVCV	/ˈka.ɹã/	‘duck’
CVCVCV	/ˈpã.ⁿga.ɽa/	‘European, white person’
CVCVCVCV	/ka.ˌpɔˀ.ka.ˈrɛˀ/	‘sacred kingfisher (*T. sanctus*)’

Sequences of two vowels are very common in the language. These typically involve close vowels /i, u/, mid-open /ɛ, ɔ/, and open /a/ in various combinations. Unstressed /i/ and /u/ may also be reduced to [j] and [w] after a vowel, producing closed diphthongs. Sequences of three heterorganic vowels are rare, and do not seem to be highly productive. Compare the following examples (de La Fontinelle, 1976, pp. 358–359):

/paɽawiɛ/	‘sea’	/piɔ/	‘cutting of plant’
/pɔi/	‘to bind’	/pɛi/	‘sick’
/pɛu/	‘yam sp.’	/pɛɔ/	‘calcite’
/põɹea/	‘plant sp. (*Deplanchea*)’	/poaˀ/	‘to bear fruit’

The distribution of consonants is largely the same in initial and intervocalic position. Only the flap /ɽ/ is not found in initial position (de La Fontinelle, 1976, p. 71). The velar fricative /ɣ/ is rare in initial position, where it is also lost in the coastal variety (de La Fontinelle, 1976, p. 61). Voiceless stops are frequent in initial position, but rare in intervocalic position, and mostly occur in known borrowings (de La Fontinelle, 1976, p. 82). Some stops are not found in this position at all, including /pʷ, c, kʷ, ⁿgʷ/ (de La Fontinelle, 1976, p. 98). On the other hand, the trill is rare in initial position, but frequent in intervocalic position. The voiced stop /ⁿgʷ/ and the nasal consonants /ɲ, ŋ/ are only known from a handful of lexical items (de La Fontinelle, 1976, pp. 54, 61).

Labialized consonants are only found before front and central vowels (de La Fontinelle, 1976, p. 32). These are nevertheless differentiated from sequences of bilabial and velar consonants followed by close or mid-close back vowels, for which de La Fontinelle (1976) cites the following (near-)minimal pairs, /ᵐbui/ ‘hard aspen (*A. laevis*)’ vs /ᵐbʷiˀ/ ‘cowrie (*Cypraea*)’ (p. 33), /kʷã/

‘boat’ vs /koã/ ‘eye of awl’ (p. 59). In some words, the two are observed in free variation however, e.g. ‘to arrive’ is recorded as [pʷaˀ] or [poaˀ] (de La Fontinelle, 1976, p. 31).

2.3.1.4 Stress

Stress is phonologically distinctive in Ajië, and usually falls on the first syllable of monomorphemic forms, as in /ˈkɔwi/ ‘hand, arm’, though examples of monomorphemic forms with non-initial stress are also found, as in /kaˈɹɛ/ ‘sun’, etc. As a rule, most forms with non-initial stress are multi-morphemic, and contain unstressed prefixes, e.g. /neˈkɔ/ ‘sky’ and /neˈɹʌɣa/ ‘river, creek’, are both formed with the unstressed prefix /ne/ (de La Fontinelle, 1976., pp. 125–133). On the word level, there is a ban on two consecutive stresses in multimorphemic forms, where the addition of a stressed affix or component stem triggers a stress shift, which causes the stress to move to the added syllable. In compounding and prefixing, the first stressed component stem or prefix will preserve the stress, while the second loses its stress (de La Fontinelle, 1976, p. 94). In the same vein, stress placement can also be conditioned by suffixing, for which there are several suffixes that are stress shifting, as illustrated in (1) and (2), as cited in de La Fontinelle (1976, p. 130):

(1) /na ˈtɔ-a/ → [na ˈtɔa]
3SG.SBJ exist-demonstrative_suffix
‘it is there (French: *il est là (visible)*)’

(2) /na ˈtɔ -ˈa/ → [na tɔˈa]
3SG.SBJ exist-inversive_suffix
‘it is elsewhere, it is misplaced (French: *il n’est pas à sa place, c’est déplacé*)’

If a word carries two stressed syllables which are separated by one or more unstressed syllables, both stresses remain intact, where the second syllable takes the primary stress, e.g. /ˌaˀpɛcɔˈɽɛ/ ‘to lisp’ (de La Fontinelle, 1976, p. 134).

2.3.1.5 Glottalization

There is a suprasegmental glottalization in Ajië, which de La Fontinelle (1976) calls the “glottal accent” (*l’accent glottalisé*). This glottalization is marked by de La Fontinelle (1976) with a glottal stop after the vowel, which is realized primarily through a difference in phonation. This is characterized by a sharp interruption to the vibration of the vocal cords on the stressed vowel, with the glottal stop itself being barely audible (de La Fontinelle, 1976, p. 140). This distinction exists independently from nasalization, where contrasting modal vowels are characterized by a tonal peak on the stressed syllable, compare the following minimal pairs (de La Fontinelle, 1976, p. 345):

/ⁿɲa/ ‘to step forward’

/ⁿɲã/ ‘hand of bananas’

/ⁿɲaˀ/ ‘to run out, drain off (of water)’

/ⁿɲãˀ/ ‘limestone formation’

The glottalization is generally located on the final syllable in multisyllabic words, as in /kalaˀ/ ‘boogeyman’, but it may also occur on the first syllable, as in /kaˀɽɔ/ ‘seashell’. In monomorphemic forms, it appears to correlate with the placement on stress (de La Fontinelle, 1976, p. 140). In multimorphemic forms, affixation and compounding affect glottalization, where a stressed glottalized syllable may not occur before a stressed modal syllable. First, there is a phonological rule that applies when two stressed glottalized syllables occur in succession, which causes the first syllable to lose the glottalization, while the following syllable preserves the glottalization, resulting in primary modal stress on the first syllable, as illustrated in (3):

(3) /aˀ/ ‘lid’ + /kʌˀ/ ‘pot’ → [ˈakʌˀ] ‘lid of pot’ (not *[aˀkʌˀ])

Second, there is a phonological rule that applies when a stressed glottalized syllable precedes a stressed modal syllable, which causes the glottalization to move to the following syllable, again resulting in primary modal stress on the first syllable, as illustrated in (4):

(4) /aˀ/ ‘lid’ + /je/ ‘oven’ → [ˈajeˀ] ‘lid of oven’ (not *[aˀˈje])

However, if any combinations of combinations above are separated by an unstressed syllable, both stresses remain intact (de La Fontinelle, 1976, pp. 141–142). What I hope to show here is that the glottalization is not an inherent feature of the vowel but is best described a suprasegmental unit whose target is the nucleus, as illustrated by its movement in relation to stress.

2.3.2 Tîrî

The phonological description I present here is based primarily on Osumi (1995), who mainly describes the La Foa variety of the language.

2.3.2.1 Vowels

There are eight oral and six nasal vowels in Tîrî, as shown in table 3, using the transcription proposed by Osumi (1995). As in Ajië, all vowels may occur as short and long, and differ only in duration, which is analyzed an aspect of the syllable structure. Minimal pairs include /mi/ ‘hiccup’ vs /mii/ ‘watermelon’, /ĩ/ ‘body’ vs /ĩĩ/ ‘to fly’, etc. (Osumi, 1995, p. 15).

Table 3: Vowels of Tîrî.

	Oral			Nasal		
	Front unrounded	Central unrounded	Back rounded	Front unrounded	Central unrounded	Back rounded
Close	i	ɯ	u	ĩ	ɯ̃	ũ
Mid-close	e		o	ɛ̃		ɔ̃
Mid-open	ɛ		ɔ			
Open		a			ã	

The system described by Osumi (1995) differs slightly from that of Grace (1976), who distinguished between three pairs of central oral and nasal vowels, close /ɨ/ and /ɨ̃/, mid-close /ə/ and /ə̃/, and mid-open /ʌ/ and /ʌ̃/. The close and mid-open vowel pairs were very rare in the dictionary, and according to Grace (1976), this three-way contrast was characteristic of the Mea variety, while in the Grand Couli variety, only mid-close /ə/ and /ə̃/ were preserved, which correspond to the two vowels transcribed as /ɯ/ and /ɯ̃/ by Osumi (1995). Because Grace (1976) and Osumi (1995) focused on different areas at different times, this difference in transcription could also reflect regional variation or language change.

Nevertheless, the vowels described by Osumi (1995) show a great deal of allophonic variation. The close vowels /ɯ/ and /ɯ̃/ vary from central unrounded [ɨ] and [ɨ̃] to back unrounded [ɯ] or [ɯ̃]. The nasal mid vowels vary freely in their degree of closeness. Front unrounded /ɛ̃/ is realized as a mid-close [ẽ] or a mid-open [ɛ̃], while back unrounded /ɔ̃/ varies freely between mid-close [õ] and mid-open [ɔ̃], and may even have a central, unrounded pronunciation [ə̃] (Osumi, 1995, pp. 13–14).

The phonemic distinctions between the close and mid-close oral vowels is frequently neutralized in unstressed position, though some minimal pairs still exist, e.g /mʷaⁿgi/ ‘again’ vs /mʷaⁿge/ ‘play (v)’. Nasal vowels /ɛ̃, ɔ̃, ã/ are also interchangeable in many words, e.g. [mʷĩɛ̃] ~ [mʷĩã] ‘woman’, [hapã] ~ [hapɔ̃] ‘learn (v)’, etc. (Osumi, 1995, pp. 13–14). This is also reflected in regional variation. /ɔ/ and /a/ are interchangeable in some words regionally, typically when neighboring back rounded vowels, or when following labialized consonants (Osumi, 1995, pp. 4–5).

Nasality is noted to spread regressively before nasal consonants and prenasalized stops, where all vowels tend to become nasal in rapid speech (Osumi, 1995, p. 22).

2.3.2.2 Consonants

Osumi (1995) recognizes 30 consonant distinctions in Tîrî, as shown in table 4.[10] The inventory is organized here according to five places of articulation and six primary manners of articulation, where labialization is distinctive for labial and velar consonants. The labialized velars are characterized by simultaneous rounding of the lips, while the labialized bilabials and labiodentals are characterized by simultaneous raising of the back of the tongue towards the soft palate (Osumi, 1995, p. 17), and are more accurately described as labiovelarized.

Table 4: Consonants of Tîrî.

		Labial			Apical		Retroflex	Palatal	Velar	
		Bilabial	Labiodental	Labialized	Dental	Alveolar				Labialized
Stop	voiceless	p		p^w	t̪		ʈ	c	k	k^w
	voiced	^{m}ɓ		^{m}ɓ^w	^{n}d̪		^{n}ɖ		^{n}g	^{n}g^w
Nasal	voiced	m		m^w	n̪		ɳ	ɲ		
Vibrant	voiced					r	ɽ			
Fricative	voiceless		f	f^w			ʂ		h	h^w
	voiced		v		ð			j	ɣ	w

There are three series of occlusive consonants in Tîrî, voiceless stops, voiced stops, and voiced nasal consonants. Like in Ajië, the voiced stops are always preceded by a brief voiced, nasal portion, and as such, both voiced stops and nasal consonants trigger the spread of nasality. Some of the occlusives are rare, including labiovelar /k^w, ^{n}g^w/ and palatal /c, ɲ/, which may be borrowed from neighboring languages. /t̪, ^{n}d̪, n̪/ are realized as apicodental consonants, while /ʈ, ^{n}ɖ, ɳ/ are realized as retroflex consonants. The retroflex stops may be realized as palatoalveolars before close front vowels /i, ĩ/. The phonemic status of the palatal nasal consonant /ɲ/ is questionable. This consonant it is neutralized with /j/ in intervocalic position between nasal vowels, and with /n̪/ before close and mid-close front vowels /i, ĩ, e/ (Osumi, 1995, pp. 15–21). Grace (1976) noted an additional nasal consonant /ŋ/, which was not recorded by Osumi (1995) in the La Foa variety.

There are two vibrants in the language, which differ in both place of manner. /r/ varies freely between an alveolar trill [r] and alveolar approximant [ɹ], while /ɽ/ is always realized as a retroflex flap (Osumi, 1995, p. 19). There is some mix-up (or possibly regional variation) in the identity of the vibrants in individual words between Grace (1976) and Osumi (1995).

10 Note that /w/ is grouped with the labials in Osumi (1995). Because she describes this as a labiovelar consonant (pp. 18–19), I have placed it among the labialized velars in table 4 instead.

The fricatives show a notable degree of allophonic variation. Of the voiceless fricatives, /ʂ/ has a palatoalveolar allophone before close front vowels /i, ĩ/. /h/ is realized as palatal [ç] before close front vowels /i, ĩ/, as bilabial [ɸ] before close central and back vowels /ɯ, ɯ̃, u, ũ/, and varies freely between velar [x] and glottal [h] before non-close vowels. /hʷ/ varies freely between labiovelarized bilabial [ɸʷ] and labiovelar [xʷ] in all contexts (Osumi, 1995, pp. 19–20). /fʷ/ is replaced by /f/ in many words in the Grand Couli variety (Osumi, 1995, p. 4).

The voiced fricatives are pronounced with varying degrees of friction. /v/ and /ð/ are realized as a voiced labiodental and interdental fricative respectively. /j/ is typically realized as a palatal approximant [j] but may be pronounced with more friction before front vowels. /w/ is always realized as a labiovelar approximant [w] (Osumi, 1995, p. 18–19).

2.3.2.3 Phonotactics

The Tîrî syllable structure can be formalized as (C)V. Long vowels are pronounced roughly twice as long as short vowels and are analyzed as geminates. Each vowel may occur as both short and long, creating a total of 28 possible vowel nuclei. For the purpose of stress assignment, long vowels always behave as single syllables in Tîrî (Osumi, 1995, p. 25).

Words in Tîrî are typically mono- or disyllabic. Words with three or more syllables are overwhelmingly multimorphemic. Consider the following examples (Osumi, 1995, p. 21):

CV	/ha/	‘to speak’
CVCV	/me.ɣi/	‘be hot’
CVCVCV	/moo.fa.ⁿɖɔ/	‘present, souvenir’
CVCVCVCV	/ʂɯ.ve.ha.ɽu/	‘to like’

Vowel sequences are frequent in Tîrî but are not realized as diphthongs. These typically consist of two vowels, but sequences of three vowels are also found, though less frequently. The following examples are lifted from Osumi (1995, pp. 23–24):

/ae/	‘four’	/ai/	‘to uncover, unload’
/ãe/	‘where?’	/ʂãĩ/	‘to jump’
/puɽɯu/	‘dust’	/tuo/	‘to put’
/ãɯ̃u/	‘fog’	/mʷĩɛ̃/	‘woman’

The distribution of consonants is largely the same in initial and intervocalic position, but some consonants are rare in initial position, including voiced fricatives /ð, ɣ/, which are typically found in intervocalic position, but are attested initially in many grammatical words, but are for that reason

quite frequent in speech (Osumi, 1995, p. 22). Other rare consonants include voiceless stops /c, kʷ/, voiced stop /ⁿgʷ/, and nasal consonant /ɲ/, which mostly occur in identified borrowings from neighboring languages. Of the vowels, only /ɛ/ is noted to be rare (Osumi, 1995, pp. 13, 15).

The labialized consonants show varying degrees of productivity, but no labialized consonant is distinct before close back vowels /u, ũ/. /kʷ, ⁿgʷ/ are most restricted, and are only found before front vowels /i, ĩ, e, ɛ, ɛ̃/ and open central /a, ã/. /ᵐbʷ, mʷ, fᵛ, hʷ/ are found before all vowels except close and mid-close non-front vowels /ɯ, ɯ̃, u, ũ, o/.[11] /pʷ, w/ are least restricted, and are found before all vowels except close back vowels /u, ũ/ (Osumi, 1995, pp. 22–23). /w/ forms minimal pairs with /u/, e.g. /wa/ ‘to fall, go down’ vs /ua/ ‘to bear fruit’ (Osumi, 1995, p. 18).

2.3.2.4 Stress

Stress is always placed on the first syllable in multisyllabic words in Tîrî and is characterized by higher pitch than unstressed syllables (Osumi, 1995, p. 25).

2.3.3 Xârâcùù

The phonological description here is based on Moyse-Faurie & Néchérö-Jorédié (1989).

2.3.3.1 Vowels

Xârâcùù contrast ten oral and seven nasal vowel qualities, as presented in table 5 following the transcription used by Moyse-Faurie & Néchérö-Jorédié (1989). All vowels occur as short and long but differ only in duration. For this reason, only the short vowels are presented in table 5. Minimal pairs include /xa/ ‘white-breasted woodswallow (*A. leucorhynchus*)’ vs /xaa/ ‘paddle’, /ʃi/ ‘to build’ vs /ʃii/ ‘to fish with a line’, etc. (Moyse-Faurie & Néchérö-Jorédié, 1989, p. 14).

Table 5: Vowels of Xârâcùù.

	Oral			Nasal		
	Front unrounded	Central unrounded	Back rounded	Front unrounded	Central unrounded	Back rounded
Close	i	ɨ	u	ĩ	ɨ̃	ũ
Mid-close	e	ɤ	o			
Mid-open	ɛ	ʌ	ɔ	ɛ̃	ʌ̃	ɔ̃
Open		a			ã	

11 Osumi (1995) does not discuss the status of these consonants before mid-close /o/, but labialization is notably not marked before /o/ by either Grace (1976) or Osumi (1995).

In addition to the system in table 5, a mid-close nasal vowel /ɤ̃/ is attested in the variety spoken in Thio along the east coast. This vowel was also described by Grace (1976), who transcribed it as /ə̃/, in contrast with /ɨ̃/ and /ʌ̃/. This distinction is not observed in other varieties of the language, where /ɤ̃/ appears to correspond to any of the remaining interior nasal vowels /ɨ̃/ or /ʌ̃/ (Moyse-Faurie & Néchérö-Jorédié, 1989, p. 13).

Close and mid-close vowels are often interchangeable in vowel sequences, e.g. [xai] ~ [xae] ‘how’, [kaɨ] ~ [kaɤ] ‘interrogative particle’, [nɨi] ~ [nɤi] ‘island’, etc. (Moyse-Faurie & Néchérö-Jorédié, 1989, p. 17). There are also instances of variation in realization of vowels which appear to exemplify regressive assimilation across syllable boundaries, e.g. [ʃuamɛ] ~ [ʃuɛmɛ] ‘star’, [çɛmʷãça] ~ [çamʷãça] ‘grandmother’, etc. (Moyse-Faurie & Néchérö-Jorédié, 1989, p. 17).

Nasality spreads regressively before nasal consonants and prenasalized stops in monomorphemic words, but does not cross morpheme or word boundaries, so that /paa-nã/ ‘my leg (leg-my)’ is distinct from /pãã-nã/ ‘my tooth (tooth-my)’, etc. (Moyse-Faurie & Néchérö-Jorédié, 1989, p. 15).

2.3.3.2 Consonants

Xârâcùù distinguishes between 27 consonants, as shown in table 6, following (Moyse-Faurie & Néchérö-Jorédié, 1989). The inventory is organized here in four places of articulation and seven manners of articulation, where labialization is distinctive as a secondary articulation for labial and velar consonants. Labialized velars are characterized by simultaneous rounding of the lips, while labialized bilabials are articulated with simultaneous raising of the back of the tongue towards the soft palate and are more accurately be described as labiovelarized. The coronal consonants may be slightly postalveolar, note the use of /ʃ/ (Moyse-Faurie & Néchérö-Jorédié, 1989, p. 13).

Table 6: Consonants of Xârâcùù.

		Labial			Coronal	Palatal	Velar	
		Bilabial	Labiodental	Labialized				Labialized
Stop	voiceless	p		pʷ	t	c	k	kʷ
	voiced	ᵐb		ᵐbʷ	ⁿd	ⁿɟ	ᵑg	ᵑgʷ
Nasal	voiced	m		mʷ	n	ɲ	ŋ	
Vibrant	voiced				r			
Fricative	voiceless		f		ʃ	ç	x	xʷ
	voiced		v			j		w
Lateral	voiced				l			

There are three series of occlusive consonants in Xârâcùù, voiceless stops, voiced stops, and voiced nasal consonants. Like in Ajië and Tîrî, the voiced stops are always preceded by a voiced nasal portion and may be alternatively described as prenasalized stops. As previously noted, both voiced stops and nasal consonants trigger the spread of nasality in monomorphemic forms.

Moyse-Faurie & Néchérö-Jorédié (1989) discuss no notable allophonic variation in the language, apart from the vibrant /r/, which is commonly realized as an alveolar tap [ɾ] (p. 13).

There is however some regional variation, particularly of the voiceless fricatives, centered around the varieties spoken around Canala, and those spoken around Thio, where the latter neighbors Xârâgurè. This variation is only reflected on a word-level and does not involve the loss of any phonemic distinctions. The palatal fricative /ç/ in Canala corresponds to the alveolar fricative /ʃ/ in many words in Thio, e.g. Canala /çikʷɛ/ but Thio /ʃikʷɛ/ ‘withered’, etc. Likewise, both /ʃ/ and /ç/ in Canala correspond to /x/ in many words in Thio, e.g. Canala /ʃaa/ but Thio /xaa/ ‘one’, Canala /çĩⁿɲa/ but Thio /xĩⁿɲa/ ‘shell bracelet’, etc. (Moyse-Faurie & Néchérö-Jorédié, 1989, p. 16).

2.3.3.3 Phonotactics

Xârâcùù has a maximal (C)V syllable structure, with a non-obligatory onset. The nucleus may be filled by any vowel, which may occur as short or long, producing a total of 34 possible vowel nuclei (Moyse-Faurie & Néchérö-Jorédié, 1989, pp. 13–14).

Most words in Xârâcùù are mono- or disyllabic. Words with three or more syllables are less frequent and are typically multimorphemic. Consider the following examples (Moyse-Faurie & Néchérö-Jorédié, 1989, p. 30):

CV	/nɛ/	‘fire’
CVCV	/pɛ.ci/	‘book’
CVCVCV	/nɛ.ɲĩ.ŋɛ̃/	‘nest (of rat)’
CVCVCVCV	/xa.ʃɛ.pu.ru/	‘to shorten a word’

In addition to long vowels, Xârâcùù allows for many vowel sequences, typically of two vowels. These are not realized as diphthongs but form separate syllables. The following examples are extracted from Moyse-Faurie & Néchérö-Jorédié (1989, p. 30):

/pia/	‘to fight’	/panɛa/	‘paternal aunt’
/nɤa/	‘shell’	/nɛ̃ã/	‘thus’

The distribution of consonants in initial and intervocalic position is identical in Xârâcùù. That said, some consonants are quite rare, such as /v/, /j/, and /l/. The lateral is only known from recent

borrowings. The vibrant is frequent in intervocalic position, but rare in initial position. Labialization is only distinctive before front unrounded vowels and open central vowels /a, ã/ (Moyse-Faurie & Néchérö-Jorédié, 1989, p. 12–13).

2.3.3.4 Stress

Xârâcùù has a fixed stress pattern, with prominence on the first syllable in words with three syllables or less, and prominence on the first two syllables in words with four syllables or more (Moyse-Faurie & Néchérö-Jorédié, 1989, p. 15).

2.3.4 Summary of phonological systems

Table 7 compares the vowel inventories of the three languages. The main difference between the languages lies in the number of non-open central vowels (i.e. non-front unrounded vowels)[12]. Otherwise in terms of phonological oppositions, the system of oral vowels is the same in Ajië and Xârâcùù, while the system of nasal vowels is the same in Ajië and Tîrî.

Table 7: Comparative vowel inventory.

Oral vowels									**Nasal vowels**								
Front			**Central**			**Back**			**Front**			**Central**			**Back**		
AJE	**TIR**	**XAC**	**AJE**	**TIR**	**XAC**	**AJE**	**TIR**	**XAC**	**AJE**	**TIR**	**XAC**	**AJE**	**TIR**	**XAC**	**AJE**	**TIR**	**XAC**
i	i	i	ɯ	ɯ	ɨ	u	u	u	ĩ	ĩ	ĩ	ɯ̃	ɯ̃	ɨ̃	ũ	ũ	ũ
e	e	e	ə		ɤ	o	o	o									
									ẽ	ɛ̃	ɛ̃			ʌ̃	õ	ɔ̃	ɔ̃
ɛ	ɛ	ɛ	ʌ		ʌ	ɔ	ɔ	ɔ									
			a	a	a							ã	ã	ã			

The extensive vowel systems presented here are not only rare crosslinguistically, but are unusually large among Kanak languages. The oral-nasal contrast is found in all Mainland languages, but the addition of central vowels is only known in the Northern language Paicî (Haudricourt, 1971). The addition of a phonemic vowel length which is independent from vowel quality in all three languages is also quite unusual. In addition to this, Ajië also has a characteristic glottalization realized on the nucleus, which has no equivalent in the two neighboring languages.

Table 8 compares the consonant inventories of the three languages. Each language is characterized by quite large consonant inventories, where Tîrî stands out by differentiating between two series of

12 For sake of efficiency, I will use the term “central” for the non-front unrounded vowels throughout this thesis, but I will remain true the transcription proposed in the original sources.

coronal consonants, a dental and retroflex series, in addition to the labial, palatal, and velar consonant series found in all three languages.

Table 8: Comparative consonant inventory.

Labial			**Lab. labial**			**Coronal**			**Palatal**			**Velar**			**Lab. velar**		
AJE	**TIR**	**XAC**	**AJE**	**TIR**	**XAC**	**AJE**	**TIR**	**XAC**	**AJE**	**TIR**	**XAC**	**AJE**	**TIR**	**XAC**	**AJE**	**TIR**	**XAC**
p	p	p	pʷ	pʷ	pʷ	t	t̪ ʈ	t	c	c	c	k	k	k	kʷ	kʷ	kʷ
ᵐb	ᵐb	ᵐb	ᵐbʷ	ᵐbʷ	ᵐbʷ	ⁿd	ⁿd̪ ⁿɖ	ⁿd	ᶮɟ		ᶮɟ	ᵑg	ᵑg	ᵑg	ᵑgʷ	ᵑgʷ	ᵑgʷ
m	m	m	mʷ	mʷ	mʷ	n	n̪ ɳ	n	ɲ	ɲ	ɲ	ŋ		ŋ			
	f	f		fʷ			ʂ	ʃ			ç		h	x		hʷ	xʷ
v	v	v				ɹ	ð		j	j	j	ɣ	ɣ		w	w	w
						r ɽ	r ɽ	r									
						l		l									

The languages share some notable features, including the presence of prenasalization in voiced stops, and the presence of labialization as a secondary articulation for both labial and velar consonants. There are also some overlapping traits between the languages. Notably, Ajië and Tîrî both distinguish between a trill and a flap, while only a single vibrant is found in Xârâcùù. On the other hand, Tîrî and Xârâcùù have a series of voiceless fricatives, which have no equivalent in Ajië.

2.4 Previous comparative research

In the following section, I will summarize results of previous historical linguistic work Oceanic languages, as it relates to the reconstruction of Proto-Oceanic and the Melanesian languages of New Caledonia. The purpose of this section is to provide a background to which I will relate the results of the phonological reconstruction as presented in chapter 4.

2.4.1 Proto-Oceanic phonology

The reconstructed phonology of Proto-Oceanic (POc) is the result of decades of comparative work on sound correspondences in both Oceanic and non-Oceanic languages. The resulting inventory is fairly stable, but the realization of some phonemic distinction in the proto-language is less certain. Five vowels are reconstructed to POc, *i, *e, *a, *o, *u, which showed no difference in length or nasality (Ross, 1998, p. 15). Table 9 present the consonant distinctions commonly reconstructed to POc, following Ross (1998). The term “labiovelars” is often used for *pʷ, *bʷ, *mʷ, *w in the

literature, but the exact realization of these is not known. The proto-language is reconstructed with a set of prenasalized consonants, as reflected in many Oceanic languages (Ross, 1998, pp. 15–16).

Table 9: Reconstructed POc phonemes.

		Labiovelar	Bilabial	Coronal		Palatal	Velar	Uvular
				Dental	Alveolar			
Stop	oral	*p^{w}	*p	*t		*c	*k	*q
	prenasal	*b^{w}	*b		*d	*j	*g	
Nasal		*m^{w}	*m		*n	*ñ	*ŋ	
Trill	oral				*r			*R
	prenasal				*dr			
Fricative					*s			
Glide		*w				*y		
Lateral					*l			

The POc syllable followed a maximal CVC structure. Codas were only allowed in word-final position, and medially only (C)V syllables occurred. Reconstructed lexical forms are typically two to three syllables, while many grammatical forms are monosyllabic, and words with four or more syllables are typically multimorphemic. Stress was likely on the penultimate syllable in POc, as reflected in many daughter languages (Ross, 1998, pp. 17–18).

2.4.2 Phonological innovations in Kanak languages

Comparative work on sound correspondences in Kanak languages in comparison with POc reconstructions has resulted in the following list of shared phonological innovations attributed to the last common ancestor of all Kanak languages, Proto-New Caledonian (PNC), as illustrated below (Haudricourt, 1971; Ozanne-Rivierre, 1992; Lynch & Ozanne-Rivierre, 2001):

- Strengthening of POc *r and *dr to PNC *ʈ and *^{n}ɖ.
- Merger of POc *c and *s as PNC *s [c].
- Merger of POc *n, *ñ, and *l as PNC *n.
- Loss of POc *R and *y.
- POc *ai (*ai, *aRi, *ayi) to PNC *e (or *ee).
- Emergence of non-etymological prenasalization.

- Emergence of aspirated stops and nasal consonants.

The most characteristic innovations of the group include the emergence of an aspiration contrast in stops and nasal consonants. Haudricourt (1971) first noted sound correspondences between aspiration contrast in stops and nasal consonants with tonal distinctions in the languages of the center and far south, for which he proposed the reconstruction of a series of "hard consonants" in the common proto-language, whose origin he credited to reduplication (p. 384). Haudricourt (1971) hypothesized that reduplication produced long consonants through the loss of pretonic syllables, which later gave rise to aspirated consonants, as shown below:

$$*C_iV'C_iV \rightarrow *C_iC_iV \rightarrow *C^hV$$

This hypothesis has since found support in comparison with POc reconstructions containing such $*C_iVC_iV$-forms, which has favored the reconstruction of a series of geminate or aspirated voiceless stops and nasal consonants in PNC (Rivierre, 1993). This contrast is preserved in most Northern languages, where some varieties have evolved voiceless fricatives from earlier aspirated stops, while tonal distinctions have replaced the aspiration contrast in two languages of the center, as well as in the two languages of the far south. As for the latter, tone must have evolved independently, as the same innovation is not known in languages of the mid south. At least in Xârâcùù, aspirated stops became voiceless fricatives, while the aspiration contrast appears to have been lost with nasal consonants (Rivierre, 1993).

The prenasalization of voiced stops is consistently found in languages of mainland New Caledonia, and in so far that it can be traced back to POc, it was undoubtedly present in PNC as well. However, Kanak languages also frequently show evidence of non-etymological prenasalization, where prenasalized stops variably appear in words in descendant languages where voiceless stops are predicted on the basis of POc reconstructions. The loss of pretonic vowels has further been credited to this innovation, where it is hypothesized that the addition of a proclitic in the proto-language such as the POc nominal article *na could leave a fossilized trace in the form of prenasalization on the initial stop (Ozanne-Rivierre, 1992, pp. 196–197).[13]

It has since been noted is that non-etymological aspiration and prenasalization is incredible varied in related vocabulary between closely related languages, where it may also produce doublets in individual languages, where a difference in manner correlates with a difference in meaning or function. The identification of many such etymological doublets have caused researchers to

13 The attrition of this article has parallels in other languages of the Southern Oceanic linkage (Lynch, 1999, p. 429).

speculate that these contain fossilized remains of morphological processes in PNC, which must have remained productive after the languages diverged (Ozanne-Rivierre, 1992, p. 200).

Thus, previous research has concluded that the highest-order split in the New Caledonian group occurred between the languages of the mainland and those of the Loyalty Islands (Lynch & Ozanne-Rivierre, 2001; Lynch, 2003).[14] Within the Mainland group, Mid-Southern languages are known to share a number of phonological innovations with both the Far-Southern languages, as well as with the neighboring Northern language Paicî, which include the loss of final consonants and development of nasal vowels (Ozanne-Rivierre & Rivierre, 2004, pp. 147–150).

In the Northern languages, there is a widespread loss of final open syllables in unbound forms, which produced closed mono- or disyllabic forms from earlier di- or trisyllabic forms inherited from POc. This change is also reflected in the languages of the south, but here, the syllable has been further reduced by deleting the coda, as illustrated below. Where the deleted coda represented an earlier nasal or prenasalized consonant, the nasality is preserved on the nucleus, and has given rise to phonemically distinct nasal vowels (Ozanne-Rivierre & Rivierre, 2004, p. 147).

$$*\text{ˈCVCV} \rightarrow *\text{CVC}_{\text{oral}} \text{ or } *\text{CVN} \rightarrow \text{CV or C}\tilde{\text{V}}$$

This series of changes are also known to have occurred in the Northern language Paicî, as well as in one variety of the Northern language Yuanga. In each of these languages, this movement towards open monosyllables is correlated with an increased complexity of the vowel system (Ozanne-Rivierre & Rivierre, 2004, pp. 149). This correlation may have emerged in the Paicî due to contact with neighboring Mid-Southern languages but must have evolved independently in Yuanga in the far north. The problem with previous observations about sound correspondences in Mid-Southern languages is that while some innovations are known from individual languages, there is no accounting for which innovations are either regular in individual languages, or general to the group. Because Mid-Southern languages are known to share phonological innovations with the Far-Southern languages, is it easy to suspect that this is indicative of a closer genetic affiliation between the two. However, because there has been no systematic study into the phonological history of the Mid-Southern languages, it has not been possible to evaluate both the internal and external classification of these languages. What is missing is an intermediary stage which may fill this gap. I will return to these findings in connection to the phonological reconstruction in chapter 4.

14 Lynch & Ozanne-Rivierre (2001) credit only a merger of PNC *ŋ and *n to the Mainland group (p. 35).

3 Method

3.1 Material

In order to address my research question and reliably reconstruct the phonological system of the last common ancestor of the Mid-Southern languages, a large set of lexical and morphological items had to be available in each of the languages. The three languages Ajië, Tîrî, and Xârâcùù are the most well documented of the group and were chosen based on the accessibility of at least one bilingual wordlist or dictionary in each of the languages. This choice was also motivated from a comparative point of view, as the three languages have previously been classified into three separate subgroups within the Mid-Southern group, Houailou, La Foa, and Canala (Haudricourt, 1971), and should therefore provide a good starting point for historical linguistic research on the group.

The dataset used for the analysis in this thesis consisted of secondary lexical and morphological data extracted from multiple independent bilingual wordlists and dictionaries of the three languages Ajië, Tîrî, and Xârâcùù. The raw data was organized in a spreadsheet in the form of a wordlist, and consisted of 1070 words for Ajië, 781 for Tîrî, and 777 for Xârâcùù. The data on Ajië was primarily gathered from the bilingual English-Ajië wordlist by Sylvain Aramiou and Jean Euritein (1995), included in *The comparative Austronesian dictionary*. Additional data was also gathered from Leenhardt (1935) and de La Fontinelle (1976). The data extracted from Aramiou & Euritein (1995) was double-checked against de La Fontinelle (1976) and Lichtenberk (1978) regarding the phonological transcription and glossing. The data on Tîrî was gathered primarily from Grace's bilingual Tîrî-English dictionary, published in 1976, and was double-checked against Osumi (1995) regarding the phonological transcription and regional usage. The data on Xârâcùù was gathered from the bilingual Xârâcùù-French dictionary by Moyse-Faurie and Néréchü-Jorédié, published in 1989. Grace's bilingual Xârâcùù-English dictionary, published in 1975, was also consulted in a handful of cases.

From the start, criteria were set up for which words to include in the dataset. As such, both lexical and grammatical words were included in the dataset. This included all relevant lexical word classes (nouns, verbs, etc.) and grammatical word classes, including pronouns, prepositions, articles, and various affixes. Interjections were not included in the dataset, as these are extra-linguistic words,

which often deviate from the phonological structure of the language and may therefore not be informative about the phonological history of the languages. To the extent that it was possible, known borrowings of Polynesian or European origin were excluded during the data collection.

3.2 The comparative method

In order to reconstruct the phonological history of the three languages, I have relied on the comparative method of linguistic reconstruction. This method builds on the assumption that a selected set of languages are descended from a single hypothesized ancestor, a so-called proto-language. The method can be described as a bottom-up approach, in which the aim is to reconstruct as much as possible about the proto-language through comparison of sound correspondences in cognate vocabulary in the daughter languages (Campbell, 2013, p. 107).

The foundation of the comparative method is the "regularity principle", first formulated by the Neogrammarian scholars of the late 19th century. The Neogrammarian hypothesis about the regularity of sound change states that sound changes recur generally and apply indiscriminately whenever the relevant sound or sounds occur in phonological environments which condition the change. The fundamental paradigm shift which produced the comparative method is derived from the formulation of regular sound changes as "sound laws". That is, if sound change is predictable within the domain in which it is conditioned, by definition, it must allow for generating and testing of hypotheses about sound correspondences across related languages. Based on the regularity principle, exceptions to a sound law must therefore either be the result of an unpredictable or irregular sound change, or the result of a different sound law, which targets a different domain than the former (Campbell, 2013, p. 15).

The comparative method can be summarized in three steps, of which the first involves assembling cognate sets between the selected languages. A stated requirement for establishing regular and systematic sound reflexes in a set of daughter languages involves identifying any "noise" in the data, which misguide the linguistic reconstruction. The goal when assembling cognate sets is to only compare "true cognates" to the extent that this is possible, i.e. words in the descendant languages which are related by virtue of being inherited from a common proto-language, and not words which exhibit similarities due to either borrowing or chance (Campbell, 2013, p. 111).

The second step involves establishing sound correspondences based on the previously identified cognate sets. Correspondence sets are derived by comparing each of the segments in the

related words in a cognate set. A cognate sound in a daughter language that is believed to be derived from a proto-sound is called a “reflex” (Campbell, 2013, p. 110).

The third step involves reconstructing aspects of the hypothesized ancestor based on the established cognate sets and derived correspondence patterns. There is a difference between a phonological reconstruction, i.e. a reconstruction of proto-sounds, and a lexical reconstruction, i.e. a reconstruction of proto-forms. To an extent, all linguistic reconstruction is arguably phonological, from which a lexical reconstruction can be secondarily derived. In this thesis, I will only conduct a phonological reconstruction.

While there are no fixed rules for how to conduct the phonological reconstruction, there are some general principles that may aid the process, of which the most central is the concept of directionality in sound change. By observing known sound laws in other languages, it is possible to identify certain tendencies across independent languages where sounds are typically found to change in one direction (A → B), but rarely in another (B → A). These tendencies in the directionality of sound change are typically phonetically motivated and widely reflected both synchronically and diachronically. For example, it is well supported that allophonic processes more commonly exemplify sound change from less sonorous to more sonorous under defined conditions (so-called lenition or consonant weakening), than from more sonorous to less sonorous (so-called fortition or consonant strengthening) (Bybee & Easterday, 2019).

The directionality principle is by no means foolproof, as becomes clear where a known sound law either does not illustrate any preferred directionality crosslinguistically or may be phonetically motivated in either direction. In cases where multiple alternatives are available, the law of parsimony is relevant, which states that no more than is necessary should be assumed to derive the most rational conclusion, i.e. the reflex which requires the least number of independent sound laws to explain the sound correspondences is judged closest to the proto-sound. Otherwise, the reflex which is found in most daughter language may also be judged to be closest to the proto-sound based on the principle of majority rule. This falls back on the law of parsimony, i.e. it is more likely that one language has undergone a change, than that several, daughter languages have independently undergone the same change, provided that the majority languages are not more closely related to each other than to the rest (Campbell, 2013, pp. 114–117).

Once proto-sounds have been reconstructed, it is important to check the plausibility of the reconstructed segments against the phonemic inventory of the proto-language as well as against typological tendencies (Campbell, 2013, pp. 124–127).

3.3 Preparation of the data

After organizing the raw data in a spreadsheet, the material was prepared for cognate assessment and sound correspondence detection using EDICTOR (Etymological Dictionary Editor), a web-based tool designed for creating, editing, and analyzing etymological datasets (List, 2017). The EDICTOR tool takes as an input a spreadsheet with tab-separated values, where each row introduces a new entry in the etymological wordlist, and the respective columns specify additional information about the entry, as illustrated in Image 1 below.

ID	DOCULECT	CONCEPT	VALUE	FORM	TOKENS	COGID	SOURCE
1	AJE	and	mạ (var: ma); cf. mã (Aramiou …	mã	m ã	1	de La Fontinelle (1976), Arami…
3	TIR	and	mê	mɛ̃	m ɛ̃	1	Grace (1976), Osumi (1995)
2	XAC	and	mê	mɛ̃	m ɛ̃	1	Moyse-Faurie & Néchérö-Joré…
6	TIR	arm	bê-	ᵐbɛ̃	ᵐb ɛ̃	2	Grace (1976), Osumi (1995)
5	XAC	arm	mê	mɛ̃	m ɛ̃	2	Moyse-Faurie & Néchérö-Joré…
4	AJE	wing	baa-	ᵐbaa	ᵐb a a	2	de La Fontinelle (1976)

Image 1: EDICTOR menu.

The fields COGID and TOKENS were crucial for annotating cognate sets and generating sound correspondences, which I will explain shortly below. When preparing the data for the sound correspondence analysis, each entry was assigned a unique numeric ID, as well as a three-letter abbreviation which specified the language of the entry in the database (AJE = Ajië, XAC = Xârâcùù, TIR = Tîrî). The meaning or gloss of the word was either translated or directly lifted from the original source. Longer translations or glosses were in some cases reformulated to simplify the presentation of the data. These were also standardized to an extent so that morphological variants were eliminated, e.g. “ash” and “ashes”, etc.

In the raw data, the form of entry was presented as indicated in the original source, which allowed for complex entries with multiple variant forms, as included in the field VALUE. For the sound correspondence analysis, only a single entry could be chosen. This form was specified in the field FORM. When dealing with words that had multiple free or regionally conditioned variants occurred, the choice of form was largely arbitrary, as the choice could often not be motivated.

Because the languages are presented in different orthographies in every source, all forms were transcribed into IPA to enable the analysis. Each language was transcribed according to the transcription presented in the phonological background in 2.3. When transcribing the data in Ajië,

words containing any of the three rhotics written as ⟨r̃⟩ (=/ɹ/), ⟨r⟩ (=/r/), and ⟨ř⟩ (=/ɽ/) in Aramiou & Euritein (1995) were double-checked against de La Fontinelle (1976) to confirm their phonological identity. Glottalization was not consistently marked in Aramiou & Euritein (1995), and to the extent that it was possible, words were double-checked against de La Fontinelle (1976) or Lichtenberk (1978) to confirm the presence of glottalization. This meant that it was not possible to confirm the presence of glottalization in all cognates, and it is therefore possible the distinction is underrepresented in the dataset.

After transcribing and standardizing the data, the cognate sets were manually assembled in EDICTOR. Each cognate set was assigned a unique numeric ID which served the function of annotating which words were related in the database. Cognacy judgments were based primarily on sound similarity, and secondarily on meaning resemblance. This process was circular, where the initial identification of cognates resulted in an increased understanding of sound correspondences between the languages, which in turn enabled the identification of new cognates, and so on. Through this process, 266 potential cognates were identified across the three languages.

The next step in the process involved deriving sound correspondences, which were generated in EDICTOR through a method known as phonetic alignment, in which the transcribed cognates are space-segmented and aligned in a matrix, where gaps in the matrix are marked using a hyphen, as illustrated in Image 2. During this segmentation process, long vowels were consistently treated as disyllabic sequences of identical vowels, following the phonological analyses adopted in 2.3.

Image 2: EDICTOR alignment menu.

Cognate sets were automatically aligned in EDICTOR, but were later manually checked, after which sound correspondence sets could be automatically generated. This allowed non-cognate parts of multimorphemic words to be ignored in the alignment, which was necessary where only parts of a word could be attributed to a cognate set. Each generated sound correspondence set was assigned a unique identifier by the program, e.g. a_1, a_2, a_3, etc. A total of 127 consonants correspondence sets and 220 vowel correspondence sets could be derived based on the cognate sets.

4 Results and discussion

The following chapter is organized into three primary sections. Sections 4.1 and 4.2 are designed to address the main and secondary aim respectively. Thus, in 4.1, I discuss the results of the sound correspondence analysis and present my phonological reconstruction, while in 4.2, I discuss how the results of the phonological reconstruction relate to other Mainland New Caledonian languages in comparison with higher-order reconstructions. The chapter is concluded in 4.3 with a general discussion of the method and results.

4.1 Phonological reconstruction

The goal of the following section is to address my primary aim as guided by the research question:

RQ: What phonological distinctions can be reconstructed to the last common ancestor of the Mid-Southern subgroup?

To answer this research question, 266 potential cognate sets were collected between Ajië, Xârâcùù, and Tîrî, as explained in the preceding chapter. These generated 127 correspondence sets involving consonants and 220 correspondence sets involving vowels, which comprised a total of 381 occurrences of correspondences involving consonants and 481 occurrences of correspondences involving vowels. Many of these correspondence sets were quite marginal. Notably, 72 consonant sets and 148 vowel sets only occurred once in the dataset. This comprised around 18.8% of all occurrences of consonant correspondences and 30.7% of all occurrences of vowel correspondences. 19 consonants sets and 13 vowel sets were found which occurred at least five times in the dataset, which in total comprised 208 occurrences of correspondences involving consonants (54.5% of total sum) and 186 occurrences of correspondences involving vowels (38.6% of total sum). Thus, there were proportionally more marginal correspondences involving vowels than consonants.

In the following section, only the correspondence sets and patterns which was most crucial to the phonological reconstruction are presented and discussed. For each step of the reconstruction, I will exemplify and discuss the relevant sound correspondences, and motivate my reconstruction. Correspondences involving vowels and consonants are presented and discussed separately in 4.1.1 and 4.1.2 respectively. The glottalization is discussed in 4.1.3 in relation to the reconstruction.

4.1.1 Reconstruction of vowels

In the following section, vowels are presented and discussed in order of degree of backness, from front to back. Oral and nasal vowels overlap a great deal and are therefore presented together. Distance assimilation and vowel length are discussed separately in 4.1.1.4 and 4.1.1.5.

4.1.1.1 Front vowels

Table 10 exemplifies sound correspondences involving front oral vowels in the three languages.[15] Note that there is a notable degree of overlap between the correspondences, where the vowels tend to be more closed in Tîrî than in Ajië and Xârâcùù.

Table 10: Correspondences between front oral vowels

Set	Pattern			Matches	AJE	XAC	TIR	Concept
i17	i	i	i	25	**ni**ɽɛ	**ni**rɛ	ŋ**i**ɽe	snot
e12	e	ĩ	i	2	ⁿgɛr**e**	ĩr**ĩ**	har**i**	1PL.INCL (FOC/SBJ)
i3	e	i	i	5	ɽ**e**	r**i**	ɽ**i**	3PL (OBJ/POSS)
e3	e	e	i	5	p**e**	p**e**	v**i**	stingray
e2	e	e	e	3	ⁿd**e**	ⁿɟ**e**	ⁿɖ**e**	trap (n)
e6	e	ɛ	e	3	m**e**	m**ɛ**	m**e**	dead
ɛ5	ɛ	ɛ	e	16	p**ɛ**	p**ɛ**	v**e**	take
ɛ7	ɛ	ɛ	ɛ	4	ⁿg**ɛɛ**	ⁿg**ɛɛ**	ⁿg**ɛɛ**	grandmother

Starting from the top with set i17, note that a close vowel /i/ is found in all three languages in a comparatively large set of cognates, which must illustrate a regular development from *i in the proto-language. Below in set e3, /i/ is also found in Tîrî, whereas /e/ is found in the other two languages. Here Tîrî appears to have merged of two distinctions in the proto-language, which are distinctly retained in the other languages. This suggests the vowel reflected by set e3 was likely *e in the proto-language as well, which must have been regularly raised to /i/ in Tîrî.

Based on this reconstruction, the two intermediary sets e12 and i3 are ambiguous, where the correspondences in Ajië suggest *e, while the correspondences in Xârâcùù suggest *i.[16] Based on the reconstruction proposed above, the correspondences in Tîrî could suggest either *i or *e, which

15 Note in the following table that the set marks the identifier assigned by EDICTOR and not the reconstructed form. For each correspondence set, the correspondence pattern is exemplified followed by how many times it occurred in the dataset. Each correspondence is exemplified with one matching cognate set in the three languages.

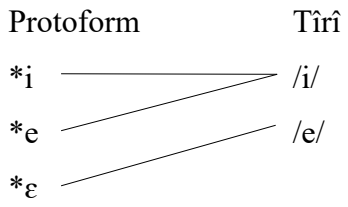
16 The nasality in this word is secondary, probably resulted from the loss of the initial consonant (see 4.1.2.2).

offers no further clarity about the identify of the proto-sound. The correspondence sets e12 and i3 are only find in word-final position in multisyllabic words, and may therefore be subject to a conditioned sound change. It is therefore possible that these sets reflect a parallel raising of *e in Xârâcùù. It is on the other hand not surprising to find diverging correspondences like this in unstressed position. For now, these correspondences remain ambiguous.

Moving on to the correspondences below in set ɛ5, mid-close /e/ is found in Tîrî, while mid-open /ɛ/ is found in the other two languages. This pattern is likewise widely attested in the dataset, and undoubtedly illustrates a regular development from the proto-language. Following the majority principle as explained in 3.2, a mid-open vowel *ɛ can be reconstructed in reference to this set, which must have been regularly raised to /e/ in Tîrî. As a result of this change, *ɛ took the place of former *e, which had shifted to /i/.

Based on this reconstruction, another ambiguous pattern can be identified in the neighboring set e6, in which Ajië suggests reconstructing *e, while Xârâcùù and Tîrî suggest reconstructing *ɛ. This correspondence is comparatively quite small, and therefore likely irregular. Because /e/ and /ɛ/ are neutralized in certain environments in Ajië (see 2.3.1.1), I suspect this could have driven a mix-up of the two in some words in the language. If so, this set likely reflects *ɛ as well, on account of the regular reflexes in Xârâcùù and Tîrî, where the vowel has been sporadically raised to /e/ in Ajië.

There is therefore solid evidence that the non-close vowels were regularly raised in Tîrî, rather than independently lowered in the other two languages. At the same time, the mid vowels *e and *ɛ have evidently escaped raising in a handful of words in Tîrî, as exemplified by sets e2 and ɛ7 respectively. There is no clear conditioning for these marginal correspondences in Tîrî, and as such they must be deemed sporadic. This is also the only attestation of /ɛ/ in Tîrî in the dataset. It is therefore likely that this irregular reflex is representative of borrowed material in Tîrî, introduced in loans from either Ajië or Xârâcùù, which have neither undergone a raising of the same vowel. The rarity of this sound in Tîrî further supports the hypothesis that the distinction was regularly lost in the language, after which it may have been reintroduced through borrowings from the other languages, which regularly preserve the distinction. Thus, the following chain shift can be regularly reconstructed to Tîrî, as illustrated below, in reference to sets i17, e3, and ɛ5 respectively.



With the reconstruction of the front oral vowels in mind, we may precede to correspondences involving front nasal vowels, as exemplified in table 11 below.

Table 11: Correspondences between front nasal vowels

Set	Pattern	Matches	AJE	XAC	TIR	Concept
ĩ4	i ĩ ĩ	3	jiwãʈ**i**	ⁿʝuxʷãr**ĩ**	t̪ufaʈ**ĩ**	spit (v)
ĩ5	ĩ ĩ i	2	mʷããⁿʝ**ĩ**	ɲ**ĩ**	n̪**i**	bow (n)
ĩ6	ĩ ĩ ĩ	2	pʷ**ĩ**ʈ**ĩ**	pʷ**ĩ**r**ĩ**	pʷ**ĩ**ʈ**ĩ**	roll up (a cord) (A, X), roll up (into a coil) (T)
ĩ1	ẽ ĩ ĩ	3	c**ẽẽ**	c**ĩĩ**	**ĩĩ**	fly (v)
ẽ14	ẽ ĩ i	1	ⁿʝ**ẽ**	ɲ**ĩ**	n̪**i**	vein, artery (A), vine sp. (*S. purpurata*) (X, T)
ɛ̃5	ẽ ɛ̃ ɛ̃	6	k**ẽ**	k**ɛ̃**	**ɛ̃**	burn (vi)
ɛ̃3	ã ɛ̃ ɛ̃	4	m**ã**	m**ɛ̃**	m**ɛ̃**	and

Starting with sets ĩ4, ĩ5, and ĩ6, note that there is a close vowel in all three languages, with differing nasality. These status of the nasality on the vowels in the proto-language in these sets is highly varies from cognate to cognate. Where no nasal consonant is present from which nasality could have hypothetically spread, it must at this point be concluded that a nasal vowel was also present in the proto-language. This suggests the presence of the close nasal vowel *ĩ in the proto-language, which formed the nasal counterpart to former *i.

Moving on to the correspondences in set ɛ̃5, all three languages show a mid vowel, but exemplify a minor difference in degree of closeness. The majority principle suggests reconstructing a mid-open vowel *ɛ̃ here, which must have been raised in Ajië. There is also a degree of overlap between the former two as exemplified in sets ĩ1 and ẽ14, where a mid-close vowel is found in Ajië, but a close vowel in the other languages. On account of the raising of *e in Tîrî, this could suggest reconstructing a mid-close vowel *ẽ instead. However, these correspondences are notably marginal, which makes it more likely that these exemplify irregular developments from either *ĩ or *ɛ̃.

It is also possible to identify a degree of overlap between mid /ẽ/, /ɛ̃/ and open /ã/, as exemplified by sets ã11 and ɛ̃3. In most of the words in question, it is not clear if the vowel was *ɛ̃ or *ã in the proto-language. Reconstructing a variation between the two is also not out of question, as the respective vowel qualities are quite similar, to which nasality further reduces the distinction. On a related note, de La Fontinelle (1976) also described a nasal vowel [ɛ̃] in the speech of older Ajië speakers, which younger speakers had replaced with [ã] at the time of her study.[17] It is not clear if

17 This vowel is found in one word included in the dataset, /kãi/ ‘eat (starches)’, for which she recorded the archaic form /kɛ̃i/. de La Fontinelle (1976) analyzes this word as containing the transitive suffix /-i/. Thus, the fronting of the vowel in this word could be the result of contact assimilation as well.

this vowel was related to the proto-vowel I reconstruct here, or if this vowel evolved as a variant of /ã/ at a later date. For this reason, most of these cases are best regarded as ambiguous for now.

4.1.1.2 Central vowels

Table 12 exemplifies sound correspondences involving central oral vowels in the three languages. As with the front vowels, there is a notable degree of overlap between the languages here as well.

Table 12: Correspondences between central oral vowels

Set	Pattern			Matches	AJE	XAC	TIR	Concept
ɯ6	ɯ	ɨ	ɯ	8	k**ɯɽɯ**	k**ɨtɨ**	**ɯɽɯ**	louse
ɯ1	ɯ	ɤ	ɯ	1	m**ɯ**	m**ɤ**	m**ɯ**	tree sp. ‘faux gaïac’ (*A. spirorbis*)
ə2	ə	ɤ	ɯ	7	ᵐb**ə**	ᵐb**ɤ**	ɔ̃p**ɯ**	buttocks
ʌ3	ʌ	ɤ	ɔ	1	tʌvʌ	tɛp**ɤ**	ʈev**ɔ**	speak, talk (A), speech (X, T)
ʌ7	ʌ	ʌ	ɔ	1	ᵐb**ʌ**ʔ	ᵐb**ʌ**	p**ɔ**	budge, move
ʌ6	ʌ	ʌ	o	1	p**ʌ**ʔ	p**ʌ**	v**o**	sacred pole
ɛ14	ɛ	ʌ	a	1	m**ɛ**ɽi	mʌtʌ	m**a**ɽi	dry
a23	ʌ	a	a	2	kʌɣʌ	ⁿg**a**k**a**	k**a**ɣ**a**	crow (*C. moledunoides*)
a5	a	a	a	58	j**a**ʔ	ç**a**	ʈ**a**	strike (v)

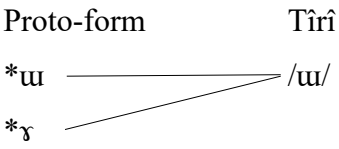
Starting from the top with the correspondences in set ɯ6, there is a slight difference in the degree of backness between the vowels, but this is likely only an artifact of the transcription. These correspondences suggest the sound was a close vowel in the proto-language as well, which may have been either central unrounded [ɨ] or back unrounded [ɯ]. Following the majority principle, I will represent this vowel as *ɯ, but the degree of backness of this proto-vowel is debatable.

In set ə2, a close vowel is also found in Tîrî, while mid-close vowels are found in the other languages. Here the phonemic distinction between a close and mid-close vowel in Ajië and Xârâcùù support a merger unique to Tîrî. For this reason, a mid-close vowel is best reconstructed to account of the correspondences in set ə2. Note again that there is a slight difference in degree of backness between the correspondences here, which may not be of consequence. The proto-vowel may therefore have been either a central unrounded [ə] or back unrounded [ɤ]. Following my previous choice of graph, I will represent this vowel as *ɤ.

With this reconstruction in mind, the intermediate set ɯ1 is ambiguous in its reconstruction, where Ajië suggests reconstructing *ɯ, while Xârâcùù suggests reconstructing *ɤ. Because both reconstructions predict /ɯ/ in Tîrî, neither reconstruction can simultaneously explain both reflexes

in the other two languages, and as such it must be deemed irregular for now.

The following merger can therefore be reconstructed for the close and mid-close central oral vowels in Tîrî, as illustrated below:[18]



In the correspondences in set ʌ7 and ʌ6, there is no difference between Ajië and Xârâcùù, which suggests the vowel in the proto-language may also have been *ʌ, following the majority principle. As with previous central vowels, the degree of backness of this proto-vowel is debatable. In Tîrî however, the sets in question show two unconditioned reflexes, /ɔ/ and /o/. These are both indicative of a rounding, but where the mid-close reflex /o/ must have been secondarily derived from the mid-open reflex /ɔ/, where the vowel was first rounded, then raised, as illustrated below:

1. Protoform	*ʌ
2. Rounding	/ɔ/
3. Raising	/o/

In consideration of this, *ʌ can be reconstructed to the intermediary set ʌ3 on the basis of Ajië and Tîrî, as while *ɤ must be reconstructed on the basis of Xârâcùù. This vowel may have been sporadically raised in Xârâcùù in this word. However, because there are no other instances of the same change in Xârâcùù, this reflex must be deemed irregular in the language in question.

/ʌ/ overlaps to a degree with /a/ as well, as exemplified by set ɛ14 and a23.[19] In these sets, it is not clear whether the vowel was *ʌ or *a in the proto-language, and in which direction the change has gone. For comparison, set a5 is the most widely attested correspondence pattern in the dataset, and thus offers solid evidence for an open central vowel *a in the proto-language as well. However, while *ʌ is only marginally reflected in the dataset, the fact that it maintains distinct back reflexes in Tîrî indicates that the inherited vowel must have been distinct from both *ɤ and *a in the proto-language. For this reason, I argue that the correspondences support the reconstruction of four central oral vowels, *ɯ, *ɤ, *ʌ, *a.

Thus, with the preceding reconstructions in mind, we may move on to correspondences involving the central nasal vowels, as exemplified in table 13 below.

18 Note that Grace (1976) transcribed this vowel as /ə/ in Tîrî, which suggests a different directionality.

19 In set ɛ14, the vowel has been fronted in Ajië under influence from the vowel in the following syllable. I will review this change in more detail in 4.1.1.4.

Table 13: Correspondences between central nasal vowels

Set	Pattern			Matches	AJE	XAC	TIR	Concept
ũɯ₃	ũɯ	ɨ̃	ũɯ	2	m**ũɯũɯ**	m**ɨ̃ɨ̃**	m**ũɯũɯ**	unripe (A), cold, damp (X, T)
ʌ₈	ʌ	ʌ̃	ɔ̃	1	v**ʌ**	f**ʌ̃**	f**ɔ̃**	beautyleaf (*C. inophyllum*)
ã₂₃	ã	ʌ̃	ɔ̃	2	t**ã**	t**ʌ̃**	ʈ**ɔ̃**	peace (A), calm (X, T)
ã₂₁	ã	ʌ̃	ã	3	v**ã**	f**ʌ̃**	f**ã**	build a wall
ã₇	ã	ã	ã	16	k^{w}**ã**	k^{w}**ã**	w**ã**	boat

Starting from the top with set ũɯ₃, there are again no notable differences between the languages, which suggests the sound was likely a close nasal vowel in the proto-language as well, for which one could reconstruct *ũɯ, as would form the nasal counterpart to former *ɯ. A problem here is that this reconstruction is only supported by a single cognate set, which notably contains a preceding nasal consonant in all three languages, where nasality is particularly unstable across the three languages synchronically. It can therefore not be excluded that this correspondence reflects *ɯ, where the vowel has been independently nasalized in all three languages.

In sets ʌ₈, ã₂₃, and ã₂₁, the mid-open nasal vowel /ʌ̃/, which is specific to Xârâcùù, is found to overlap with /ʌ/ in Ajië, /ã/ in both Ajië and Tîrî, as well as /ɔ̃/ in Tîrî. Here, there is good indication that the correspondences in Ajië derive from an identical vowel in the proto-language as well, which can be reconstructed as *ʌ̃, the nasal counterpart to former *ʌ.

First, de La Fontinelle (1976) recorded a phonemic opposition in Ajië between /ʌ̃/ and /ã/ in the older population, which younger speakers had merged as /ã/ at the time of her study (see 2.3.1.1). This sound change is corroborated by archaic forms recorded by de La Fontinelle (1976) for some of the cognates in Ajië included in sets ã₂₃ and ã₂₁. For example, Ajië /tã/ ‘peace’, has the archaic form /tʌ̃/, which corresponds to Xârâcùù /tʌ̃/ ‘calm’, etc. However, it should be noted that variant forms with /ʌ̃/ in Ajië could not be located for all cognates cited by de La Fontinelle (1976) which exemplify sets ã₂₃ and ã₂₁.

Further evidence that support reconstructing *ʌ̃ here can also be found in set ʌ₈, where the vowel in Ajië must have been sporadically denasalized, as there is no surrounding consonant from which nasality could have spread in the other languages. In this word, the denasalized reflex is /ʌ/, and not /a/, which confirms that the denasalized vowel in Ajië must derive from earlier *ʌ̃, and not *ã, and must therefore have preceded the above-mentioned merger of earlier /ʌ̃/ and /ã/.

Regarding the correspondences in Tîrî, two distinct reflexes can be identified in sets in question, /ɔ̃/ and /ã/ respectively. Here, /ɔ̃/ exemplifies the same rounding that affected its oral

counterpart *ʌ in Tîrî, which favors the preceding reconstruction, but also speaks for the regularity of this reflex. The presence of /ã/ notably cannot be the result of a conditioned change in Tîrî, as it forms a minimal pair with /ɔ̃/ in /fɔ̃/ ‘beautyleaf’ vs /fã/ ‘build a wall’ in Tîrî, which both correspond to /fʌ̃/ in Xârâcùù. Thus, if these correspondences reflect two distinctions in the proto-language, it is not possible to determine what this difference would be based on the data. For this reason, it is worth considering that this occurrence of /ã/ may be indicative of borrowed material in Tîrî. If so, Ajië appears to most likely source at first glance, but because the vowel /ʌ̃/ is not found in Tîrî, it may just as well have been replaced by /ã/ in borrowings.

Lastly in set ã7, the correspondences show no differences between the languages, and must largely exemplify a regular development from open nasal vowel *ã in the proto-language, which was phonemically distinct from former *ʌ̃, and formed the nasal counterpart to *a. This vowel has also been sporadically denasalized in the daughter languages, with no change in vowel quality.

4.1.1.3 Back vowels

Some notable correspondences between back vowels are exemplified in table 14. There is notably less overlap between the back vowels compared to their front and central counterparts.

Table 14: Correspondences between back oral vowels

Set	Pattern			Matches	AJE	XAC	TIR	Concept
u17	u	u	u	16	**u**ʔ	x**u**	h**u**	houp tree (*M. cauliflora*)
u4	o	u	u	2	ᵐb**o**	ᵐb**u**	p**uu**	smell
o2	o	o	o	3	ⁿɟ**o**ɽa	ⁿɟ**o**ra	ʈ**o**ɽo	giant clam (*Tridacna*)
o4	o	ɔ	o	3	ⁿd**o**ɽu	ⁿd**ɔ**ru	ⁿɖ**o**ɽu	coral tree (*E. variegata*)
ɔ8	ɔ	ɔ	o	13	ɔj**ɔ**	xɔj**ɔ**	haj**o**	marry
ɔ10	ɔ	ɔ	ɔ	1	ⁿd**ɔ**jaa	ⁿd**ɔ**jaa	ⁿɖ**ɔ**jaa	whale

Starting with the correspondences in sets u17 and o2, there are no notable differences between the three languages, which indicates the respective vowels were likely *u and *o in the proto-language as well. The intermediary set u4 shows overlapping correspondences between these two sets, and is therefore ambiguous in its reconstruction between *u and *o. It may not be possible to resolve this ambiguity here, but the correspondences nevertheless do not seem to exemplify a regular change. It can also be noted that a great deal of variation between /u/ and /o/ is found in vowel sequences of the kind exemplified in (5). In these cases, the identify of the vowel in the proto-language is difficult to determine, as there is no consistent correspondence pattern across the languages. After

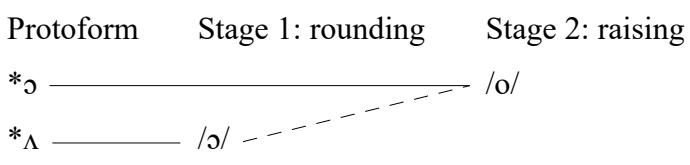
labials, the back vowels are also reduced to labialization on the preceding consonant in some words, where the earlier vowel sequence results in a long vowel instead.

	AJE	XAC	TIR	Concept
(5)	p^{w}**a**a	p**o**a	**u**a	grate (A), sharpen (X, T)
	p**o**aˀ	p**o**a	**u**a	bear fruit
	m^{w}ar**u**a	m^{w}ãt**ɔ**a	m^{w}ar**u**a	grandfather
	r**u**a	t**o**a	r**o**a	up (directional)

Moving on to the correspondences in set ɔ$_{8}$, there is a mid-open vowel /ɔ/ in Ajië and Xârâcùù, but a mid-close vowel /o/ in Tîrî. The law of parsimony therefore suggests this vowel was raised in Tîrî, and not independently lowered in the two other languages, therefore favoring the reconstruction of *ɔ for the correspondence in question. By virtue of this reconstruction, the mid-open vowel *ɔ has evidently escaped raising in one cognate in Tîrî, as exemplified in set ɔ$_{10}$. Note that the same pattern was observed with the front mid oral vowels in Tîrî, as noted in 4.1.1.1, which I suggested is the result of later borrowings. This reflex of *ɔ is only found in a single word in Tîrî. Therefore, it is worth considering that this also reflects a later borrowing in Tîrî, from either of the other two languages, as neither have undergone a regular raising of the same vowel.

With this in mind, consider the intermediary set o$_{4}$, where the same vowel appears to have been raised in Ajië as well. Note that a sporadic raising can also be reconstructed with the vowel *ɛ in Ajië (see 4.1.1.1), which may favor the reconstruction of mid-open vowel in this set as well, though this does not follow from the majority principle. Because this reflex shows no obvious conditioning it must be deemed irregular in Ajië. Thus, this correspondence remains ambiguous in its reconstruction for now.

At this point it can be noted that the raising of *ɔ in Tîrî also variably affected the rounded reflexes of *ʌ in the same language, causing a merger of the two in some words. The rounding of *ʌ to /ɔ/ in Tîrî must therefore have overlapped in time with the raising of /ɔ/, as some instances of *ʌ were raised to /o/ in Tîrî (note dotted line), while others were retained as /ɔ/, as illustrated below:



With this reconstruction in mind, we may precede with correspondences involving the back nasal vowels, as exemplified in table 15 below.

Table 15: Correspondences between back nasal vowels

Set	Pattern	Matches	AJE	XAC	TIR	Concept
ũ₆	ũ ũ ũ	3	**ũ**	x**ũ**	h**ũ**	work with pickax (A), tend a garden (X), plough (v) (T)
ũ₄	ũ u ũ	2	p**ũũ**	p**uu**	**ũũ**	cause, origin
ũ₂	ũ ũ ã	1	p**ũ**ɽũ	p**ũ**	w**ã**	hair
ũ₃	ũ ũ ɔ̃	1	k**ũ**ɽũ	k**ũ**	**ɔ̃**	seedling, shoot (A), cutting (X, T), head (T)
ɔ̃₁	õ ɔ̃ ɔ̃	3	k**õ**ɽõ	k**ɔ̃**^{n}dɔ	**ɔ̃**^{n}ɖo	blue, green

Starting with the correspondences in set ũ₆, the vowel is identical in all three languages, which suggests an identical reconstruction in the proto-language, *ũ, which serves as the nasal counterpart to former *u. This same vowel has also been sporadically denasalized in one cognate in Xârâcùù as exemplified in set ũ₄, with no change in vowel quality. In Tîrî, the vowel also appears to have been sporadically lowered in some words, as exemplified in set ũ₂ and ũ₃.

Regarding the correspondences in set ɔ̃₁, there is a minor difference in degree of closeness between the languages, which parallels the reflexes of *ɛ̃ as noted with front nasal vowels in 4.1.1.1. Based on the majority principle, I propose the proto-vowel was also a mid-open vowel *ɔ̃, which must have been regularly raised in Ajië. These correspondences therefore support the reconstruction of two back nasal vowels, *ũ and *ɔ̃, which exemplify regular reflexes in sets ũ₆ and ɔ̃₁ respectively. There is no evidence for a nasal counterpart to mid-close *o.

4.1.1.4 Distance assimilation

With our newfound knowledge about regular vowel changes in the three languages, it is now possible to identify instances where front vowels in the proto-language have changed due to assimilatory processes in one or several of the descendant languages, which have obscured these changes. In this regard, both progressive and regressive distance assimilation can be reconstructed in the descendant languages, where a vowel has changed in one or several of the descendant languages to become more like a vowel in either the preceding or following syllable.

Examples of progressive assimilation are shown in (6) and (7) below.

	Set	Pattern	AJE	XAC	TIR	Concept
(6)	i₁₀	i ɛ e	^{m}bɔɽ**i**	^{m}b^{w}ɛr**ɛ**	p^{w}eɽ**e**	some, a few
	i₂₆	i ʌ i	mɛɽ**i**	mʌtʌ	maɽ**i**	dry
(7)	a₉	a a o	^{n}ɟoɽ**a**	^{n}ɟor**a**	ʈoɽ**o**	giant clam (*Tridacna*)
	i₇	i a o	(cuɽi)	cur**a**	juɽ**o**	put on (clothing) (A), put among (X), wear (T)

In (6), *i can be reconstructed based on the final vowels in both words in Ajië. In ‘some, a few’, the vowel in the first syllable must have been a mid-open *ɛ, where the back vowel in Ajië may derive from the labialized onset. The final *i must therefore have been lowered to *ɛ in Xârâcùù and Tîrî, after which both vowels were later regularly raised in Tîrî. In ‘dry’, word-final *i is preserved in Ajië and Tîrî, but has been assimilated in Xârâcùù in reference to the preceding vowel /ʌ/.

In (7), the word-final vowel reflects *a in Xârâcùù, which must have shifted to /o/ in Tîrî under influence from the back vowel /u/ in the preceding syllable. In Ajië, *a is preserved in ‘giant clam’, but has been replaced by *i in ‘put on (clothing)’. This word-final /i/ in Ajië is likely a fossil of the transitive suffix /-i/. Similar morphological pairs exist in Tîrî, e.g. /ᵐbaɽa/ ‘be afraid (intransitive verb)’ vs /ᵐbaɽi/ ‘fear (transitive verb)’ (Osumi, 1995, p. 75).

Regressive assimilation appears to have been more frequent, and has targeted both front and central vowels, as exemplified in (8) and (9) below respectively.

	Set	Pattern			AJE	XAC	TIR	Concept
(8)	ɔ3	ɔ	e	o	n**ɔ**ɽ**ɔ**	n**e**t**ɔ**	ŋ**o**ɽ**o**	thunder
	i25	i	ʌ	e	k**i**u^{ʔ}	k**ʌ**ɨ	**e**vɯ	Malay apple (*S. malaccense*)
	ʌ1	ʌ	ɛ	e	t**ʌ**v**ʌ**	t**ɛ**p**ɤ**	ʈ**e**v**ɔ**	speak, talk (A), speech (X, T)
(9)	e14	e	ɨ	ɯ	n**e**ji	n**ɨ**i	ŋ**ɯ**i	island
	i28	ɯ	i	i	ⁿg**ɯ**ji	ⁿg**i**ci	ⁿg**i**ði	crushed (suffix) (A), crumpled (bound verb) (X, T)

Starting with (8), *e can be reconstructed in the first syllable in ‘thunder’ based on Xârâcùù. In Ajië and Tîrî, this vowel must have been assimilated under influence from the following back vowel. In Tîrî, both vowels have later been additionally raised, as described in 4.1.1.3. In ‘Malay apple’, we can reconstruct *ɛ based on the vowel /e/ in Tîrî (see 4.1.1.1), which could then have been centralized to /ʌ/ in reference to /ɨ/ in Xârâcùù. In Ajië, *ɛ must then have been raised to /i/ in reference to /u/ in the following syllable. In ‘speak, talk’, *ɛ can be reconstructed based on Xârâcùù and Tîrî. This vowel must have been centralized to /ʌ/ in Ajië in reference to the vowel in the following syllable. In (8), a close central vowel in the proto-language, either *ɯ or *ɤ, has been fronted under influence from *i in the following syllable, producing /e/ in Ajië, but /i/ in Xârâcùù and Tîrî. Note however in both (8) and (9) that the assimilatory processes are not regularly observed in either language.

Otherwise, regressive assimilation has most frequently targeted open vowels, as illustrated in (9) and (10) below, where an open vowel *a or *ã has been assimilated under influence from a front or back vowel in the following syllable.

	Set	Pattern			AJE	XAC	TIR	Concept
(9)	ɛ14	ɛ	ʌ	a	mɛɽi	mʌtʌ	maɽi	dry
	a22	ɛ	a	a	pɛii	paii	vajɯ	sick, dead
	a13	a	ɛ̃	e	maaɣe	mɛ̃ⁿgi	meɣi	warm
	ẽ4	ẽ	ã	a	wẽjẽ	xʷãĩ	iifaⁿɖe	path
	ĩ2	ĩ	ã	a	kʷĩɽĩ	kʷãrĩ	waɳe	tree sp. ‘acajou’ (*S. vitiensis*)
	i9	i	ɛ̃	a	miⁿdɛ	mɛ̃ⁿdɛ	maⁿɖe	orange (A), citrus (X, T)
	ɛ11	ɛ	ĩ	a	ⁿgɛre	ĩrĩ	hari	1PL.INCL (FOC/SBJ)
	e9	ɛ	ɛ̃	e	ⁿgɛve	ŋɛ̃ɛ̃	kevi	1PL.EXCL (FOC/SBJ)
(10)	a21	a	ɔ	a	kaɽaˀ	xɔru	haɽu	beautiful, good
	a21	a	ɔ	a	paɽi	pɔru	waɽu	peel (v)
	ɔ7	ɔ	ɔ	a	ɔjɔ	xɔjɔ	hajo	marry
	ɔ12	ɔ	ũ	a	ⁿgɔru	ũrũ	haru	1DL.INCL (FOC/SBJ)
	ɔ9	ɔ	ɔ̃	o	ⁿgɔvu	ŋɔ̃ɔ̃	komu	1DL.EXCL (FOC/SBJ)

The examples in (9) and (10) show either partial or complete assimilation. As can be seen, this assimilation is highly productive in Ajië and Xârâcùù, but largely absent in Tîrî. In the first-person exclusive pronouns, the change has evidently happened in all three languages, as the vowel has changed in a different direction depending on the following number suffix. In these words, the proto-vowel must therefore have been either *a or *ã.

4.1.1.5 Vowel length

All three languages notably distinguish vowel length, but long vowels are highly variable in the dataset, and appear to be the result of multiple different processes in the descendant languages. These can be organized according to four patterns, in which: (a) long vowels are found in all three languages; (b) long vowels correspond to vowel sequences; (c) long vowels variably correspond to short vowels; and (d) long vowels correspond to VCV-sequences.

Where long vowels are found in all three languages, a long vowel must also be reconstructed to the proto-language. Long vowels notably overlap to a degree with vowel sequences. Correspondences of this kind are challenging for the reconstruction as they do not illustrate any clear directionality either between the languages or within specific languages, as to which vowel sequences result in which long vowels, or vice versa. In some cases, there is reason to suspect there was a vowel sequence in the proto-language if there are no neighboring syllables that could have triggered a vowel breaking in either of the three languages. Examples of this are shown in (12) below.

	AJE	XAC	TIR	Concept
(12)	n**ãã**	n**ɔ̃ɔ̃**	ŋ**ãũ**	mosquito
	k**aa**ɽu	^{m}b**aa**ru	**au**ɽu	two
	taw**aa**ɽu	tap^{w}**aa**ru	ʈaw**au**ɽu	oar

In ‘mosquito’, there must have been a vowel sequence in the proto-language, as there are no neighboring syllables that could have triggered a vowel breaking in this word. Furthermore, the corresponding long vowels differ in Ajië and Xârâcùù, but are both explainable in reference to the vowel sequence in Tîrî. In ‘two’ and ‘oar’, the reconstruction is unclear. Note that ‘two’ contains a numerical prefix in each of the languages, /ka-/, /^{m}ba-/, /a-/ respectively. However, this does not clarify whether the numeral had the form *-uCu or *-aCu in the proto-language, or whether the vowel in the first syllable has been assimilated in reference to the preceding or following vowel.

There is also a great deal of variation in vowel length. In monosyllabic words, there is no clear tendency between the languages, and therefore it is often not clear whether one language has shortened an earlier long vowel or vice versa. In multisyllabic words, long vowels in word-final position in Ajië frequently correspond to short vowels in Tîrî, which is also sporadically seen in Xârâcùù. Notably, the reverse is never seen, but there are however many examples where short vowels are found in word-final position in all three languages. Thus, long and short vowels were likely distinct in this position in multisyllabic words in the proto-language as well. Thus, out of 17 identified occurrences of word-final long vowels in Ajië, 3 correspond to short vowels in Xârâcùù, while 11 correspond to short vowels in Tîrî. This therefore illustrates a clear trend in which long vowels are shortened at the word-end in Tîrî, as well as sporadically in Xârâcùù, while there are no traces of this change in Ajië. Because stress is always initial in Tîrî, this could have driven a reduction at the word-end, which applies to Xârâcùù as well. A few examples of the following are shown in (13) below:

	AJE	XAC	TIR	Concept
(13)	mãɹɯɽ**ii**	mãrã^{n}d**ii**	mãɽã^{n}ɖ**i**	sea-snake
	^{m}baɽ**əə**	^{m}bar**ɤ**	paɽ**ɯ**	also
	kɔɽ**ee**	k^{w}ar**ii**	kɔɽ**ii**	grasshopper

There are also examples where long vowels have resulted from the loss of an intervocalic consonant in one or several of the descendant languages, as exemplified in (14) below. I will discuss the reconstruction of these consonant correspondences in more detail in 4.1.2.

	AJE	XAC	TIR	Concept
(14)	ⁿdee	ⁿdee	ʈ**i**vi	flood
	ⁿgɔ**v**u	ŋɔ̃ɔ̃	ko**m**u	1DL.EXCL (FOC/SBJ)
	ĩĩ	xi**ɲ**i	hũ**j**ũ	cricket

4.1.1.6 Summary of reconstructed vowels

As outlined in the preceding sections, the correspondences support the reconstruction of the following ten oral and six nasal vowels to the proto-language, as shown in table 16. Long vowels are also supported by the reconstruction, at least oral vowels *i, *ɯ, *u, *ɤ, *ɛ, *ɔ, *a, and possibly nasal vowels *ĩ, *ũ, *ɛ̃, *ɔ̃, *ã, had short and long counterparts in the proto-language.

Table 16: Reconstructed vowel inventory of Proto-Mid-South

	Oral			Nasal		
	Front unrounded	Central unrounded	Back rounded	Front unrounded	Central unrounded	Back rounded
Close	*i	*ɯ	*u	*ĩ		*ũ
Mid-close	*e	*ɤ	*o			
Mid-open	*ɛ	*ʌ	*ɔ	*ɛ̃	*ʌ̃	*ɔ̃
Open		*a			*ã	

It is now possible to evaluate the plausibility of the reconstructed inventory. Note that not all oral vowels have a nasal counterpart, much like in the descendant languages. However, there is one clear gap in the phonemic inventory, where a close central nasal vowel *ũ would be present. While this vowel is not well supported by the correspondences, the distinction is present in all three languages. It is therefore possible that such a vowel was also present in the proto-language but that this distinction was overlooked in the dataset. On the other hand, all three languages could have independently evolved this sound through the spread of nasality from neighboring nasal and prenasalized consonants, which is highly productive in the respective languages. In these environments, nasality is also highly variable between the three languages, and as such it is often not clear if it was present in the proto-language.

4.1.2 Reconstruction of consonants

In the following section, correspondences involving consonants are presented and discussed in order of manner of articulation, respectively voiced continuants and vibrants (4.1.2.1), voiceless fricatives (4.1.2.2), voiceless stops (4.1.2.3), voiced stops (4.1.2.4), and nasal consonants (4.1.2.5).

4.1.2.1 Voiced continuants and vibrants

Correspondences involving voiced continuants and vibrants are quite numerous in the three languages, and are therefore organized below by place of articulation, respectively labial /v/, coronal /r/ and /ɽ/, palatal /j/, and labiovelar /w/.

To start, table 17 exemplifies correspondences involving /v/ in intervocalic position.

Table 17: Correspondences involving voiced labial fricatives

Set	Pattern			Matches	AJE	XAC	TIR	Concept
-₇	-	-	v	3	kiuˀ	kʌɨ	e**v**ɯ	Malay apple (*S. malaccense*)
v₂	v	-	v	1	ⁿgɛ**v**e	ŋɛ̃ɛ̃	ke**v**i	1PL.EXCL (FOC/SBJ)
v₁	v	-	m	1	ⁿgɔ**v**u	ŋɔ̃ɔ̃	ko**m**u	1DL.EXCL (FOC/SBJ)
v₄	v	p	v	2	tʌ**v**ʌ	tɛ**p**ɤ	ʈe**v**ɔ	speak, talk (A), speech (X, T)

Two correspondence patterns can be identified in Xârâcùù in the table 17 above, where one reflects the loss of an earlier intervocalic consonant, note the null correspondences in sets -₇ to v₁, while another has been retained as a voiceless stop /p/, which corresponds to a voiced fricative /v/ in the other languages, as exemplified in set v₄. Here, the latter suggests reconstructing a voiceless stop *p, which must have undergone a lenition process in Ajië and Tîrî, producing /v/ in both languages. This reconstruction is not applicable to the remaining sets, however.

In set -₇, the same consonant has been deleted in Ajië, but is retained in Tîrî as /v/. This correspondence must therefore reflect a similar consonant in the proto-language, which I will note as *v. The null correspondences in Xârâcùù also extent to the two pronominal forms which contain sets v₂ and v₁, where the loss of the earlier consonant has produced long vowels from earlier vowel sequences in Xârâcùù. Here, the consonant has instead been retained as /v/ in both forms in Ajië, while in Tîrî, it surfaces as /m/ in the dual form, but /v/ in the plural form. As evident by the nasal vowels in Xârâcùù, as well as the nasal consonant in Tîrî, there must have been a degree of nasalization in these words in the proto-language as well. The question here is whether the nasality was on the vowels or the intervocalic consonant in the proto-language, thus the correspondences remain ambiguous between either *v or *m. Reconstructing *m here would simultaneously account for the nasality, as well as sporadic retention seen in Ajië.

Table 18 exemplifies correspondences involving vibrants /r/ and /ɽ/ in intervocalic position. The two sets r₄ and r₂ also include notable correspondences in initial position in unstressed or bound grammatical words, to be discussed below.

Table 18: Correspondences involving voiced coronal vibrants

Set	Pattern			Matches	AJE	XAC	TIR	Concept
ɽ5	ɽ	r	ɽ	43	mɔ**ɽ**u	mu**r**u	mɔ**ɽ**ɔ	live (v)
ɽ4	ɽ	r	ɳ	1	kʷĩ**ɽ**ĩ	kʷã**r**ĩ	wa**ɳ**e	tree sp. ‘acajou’ (*S. vitiensis*)
ɽ6	ɽ	t	ɽ	5	nɔ**ɽ**ɔ	ne**t**ɔ	ɳo**ɽ**o	thunder
r4	r	t	r	4	pʷa**r**i	ᵐbʷa**t**i	pʷa**r**i	stomach
r2	r	r	r	5	ⁿgɔ**r**u	ũ**r**ũ	ha**r**u	1DL.INCL (FOC/SBJ)

Again, two correspondence patterns can be identified in table 18 in Xârâcùù, which show overlapping correspondences in Ajië and Tîrî. First, we find a vibrant /r/, exemplified in sets ɽ5, ɽ4, and r2, and second, we find a voiceless stop /t/, exemplified in sets ɽ6 and r4. As before, the presence of a voiceless stop favors the reconstruction of a lenition process in Ajië and Tîrî. This process must have targeted an earlier voiceless stop in intervocalic position, producing flaps as well as trills in both languages. For now, this proto-sound can be represented as *t, based on the reflex in Xârâcùù. However, I will return to the reconstruction of this consonant in more detail in 4.1.2.3.

Moving on to the correspondences in set ɽ5, we find a vibrant in all three languages, which surfaces as a flap in Ajië and Tîrî. This set involves correspondences between both oral and nasal vowels. Between nasal vowels, this vibrant overlaps with an intervocalic nasal in one word in Tîrî as well, as seen in set ɽ4. This overlap is notable as it appears to have been more widespread in the past. In fact, we find that many cognates in the dataset in Ajië which contain correspondences in set ɽ5 between nasal vowels are transcribed with ⟨n⟩ in Leenhardt (1935) instead. Thus, for /vã**ɽ**ã/ ‘count (v)’, /ᵐbeã**ɽ**ĩ/ ‘old’, /pʷẽviŋõ**ɽ**õ/ ‘scorpion’, Leenhardt (1935) wrote *va**n**a*, *bea**n**i*, and *poevinyo**n**o*. What is notable about this use of ⟨n⟩ is that it cannot solely be an artifact of the transcription, as Leenhardt (1935) also recorded intervocalic vibrants in some words, where he marked the nasalization on the vibrant rather than the surrounding vowels as done by later authors, so for /mẽ**ɽ**ẽ/ ‘hungry’ and /jiwã**ɽ**i/ ‘spit (v)’, Leenhardt (1935) wrote *me**r̃e*** and *siwa**r̃i***.[20]

This could suggest that the correspondences in set ɽ5 are indicative of two reconstructions, between nasal and oral vowels respectively. Between oral vowels, a vibrant is found in all three languages, and as such, a similar consonant must also be reconstructed to the proto-language. For now, this consonant can be reconstructed as *r, which may have been either a retroflex flap [ɽ], based on its reflexes in Ajië and Tîrî, or an alveolar tap [ɾ], based on its allophonic realization in Ajië and Xârâcùù. Between nasal vowels however, the consonants must have been an earlier coronal nasal consonant, sporadically retained as /ɳ/ in Tîrî, and historically as /n/ in Ajië. I will

20 The flap is allophonically nasalized between nasal vowels in Ajië according to de La Fontinelle (1976, pp. 82, 84).

return to the reconstruction of this and other nasal consonants in more detail in 4.1.2.5.

In set r₂, a vibrant is again found in all three languages, but this consonant surfaces as a trill in Ajië and Tîrî instead. This set involves correspondences solely in grammatical words, either in intervocalic position, or initial position in unstressed or bound forms. At first sight, this set appears to indicate a merger in Xârâcùù of two distinctions inherited from the proto-language, a flap and a trill. However, because of the distribution of this pattern, it appears unlikely that there was an additional phonemic distinction in the proto-language that only occurred in grammatical words. It is important to note here that a trill is also found in set r₄ in Ajië and Tîrî, which reflects earlier *t, as explained above. For this reason, it is more reasonable that this set also reflects a voiceless stop *t in the proto-language, where the stop has been sporadically lenited in Xârâcùù, producing /r/.

Table 19 exemplifies correspondences involving /j/ in intervocalic position.

Table 19: Correspondences involving voiced palatal approximants

Set	Pattern			Matches	AJE	XAC	TIR	Concept
j₄	j	j	j	3	ɔ**j**ɔ	xɔ**j**ɔ	ha**j**o	marry
-₃₅	j	-	-	3	ne**j**i	nɨi	ŋɯɨ	island
-₃	-	-	j	2	pɛii	paii	va**j**ɯ	sick, dead
-₂₀	-	ɲ	j	1	ĩĩ	xi**ɲ**i	hũ**j**ũ	cricket

In set j₄, the correspondences are identical across all three languages, thus a palatal approximant *j must likewise be reconstructed to the proto-language. The same consonant appears to have been sporadically deleted in all three languages, as illustrated in sets -₃₅ and -₃. In set -₂₀, the consonant has been deleted in Ajië as well, but is retained as /j/ in Tîrî, However, in this set, a palatal nasal consonant can be found in Xârâcùù instead, where nasal vowels are present in the other languages. Again, the question is whether the nasality was on the vowels or the intervocalic consonant in the proto-language, where the latter would suggest reconstructing *ɲ instead. For now, this set is ambiguous, but I will return to the reconstruction of nasal consonants in more detail in 4.1.2.5.

Lastly, we can address correspondences involving /w/ in intervocalic position, as shown in table 20.

Table 20: Correspondences involving voiced labiovelar approximants

Set	Pattern			Matches	AJE	XAC	TIR	Concept
w₈	w	w	w	4	ⁿdãã**w**ẽ	nã**w**ã	ⁿɖɛ̃ɛ̃**w**ɛ̃	sand
w₅	w	pʷ	w	1	ta**w**aaɽu	ta**p**ʷaaru	ʈa**w**auɽu	oar
w₁	w	kʷ	w	1	ⁿgʷẽẽ**w**ɛ	ᵐbʷaa**k**ʷɛ	ⁿgi**w**e	mountain

Starting with the cognates in Xârâcùù, three correspondences can be identified, /w/, /pʷ/, /kʷ/. As before, the presence of intervocalic stops in Xârâcùù must be indicative of a lenition in the other languages. The law of parsimony suggests reconstructing two earlier stop distinctions in reference to sets w5 and w1, *pʷ and *kʷ, which merged as /w/ in both Ajië and Tîrî. The fact that these both correspond to /w/ in the same position in Ajië and Tîrî excludes the possibility of the reverse, and a different reconstruction must be proposed for set w8. In set w8, the correspondences do not differ between the languages, and as such a labiovelar approximant *w can be reconstructed to the proto-language instead. Note that this set involves correspondences between both oral and nasal vowels.

In summary, the correspondences presented above support the reconstruction of one vibrant and three voiced continuants in the proto-language, respectively *r, *v, *j, *w. These consonants overlap with both voiceless stops and nasal consonants in intervocalic position, due to a lenition of voiceless stops in Ajië and Tîrî. The latter overlap with nasal consonants is notably sporadic and may have resulted from a gradual merger of nasal and non-nasal consonants in the descendant languages. I will return to this hypothesis when reconstructing nasal consonants in 4.1.2.5.

4.1.2.2 Voiceless fricatives

Table 21 exemplifies correspondences involving voiceless fricatives in Xârâcùù and Tîrî in initial position. As can be seen in the following sets, there are few direct matches between the languages. Because of this, it is helpful to look for shared phonological features between the correspondences instead and reconstruct by incorporating these into the proto-sounds.

Table 21: Correspondences involving voiceless fricatives

Set	Pattern	Matches	AJE	XAC	TIR	Concept
ɹ3	ɹ ʃ ʂ	8	**ɹ**aˀ	**ʃ**a	**ʂ**a	strike (v) (A), chop (X), cut (a board) (T)
j3	j ç ʈ	8	**j**aˀ	**ç**a	**ʈ**a	strike (v)
-27	- x h	12	aˀ	**x**a	**h**a	speak
-2	- - h	1	ẽ	ɛ̃rɛ̃	**h**ɛ̃ɽɛ̃	content
-26	- x -	1	ãju	**x**ãĩ	ai	red ant (A), ant (X, T)
-28	- x hʷ	2	ɔɽɔ	**x**uru	**hʷ**ɔɽɔ	flee
-29	- xʷ h	1	ɔˀ	**xʷ**ɛ	**h**o	meat (A), eat (meat) (X, T)
w11	w xʷ hʷ	1	**w**aˀ	**xʷ**ii	**hʷ**ii	drink from a sugarcane (A), chew sugarcane (X,T)
w10	w xʷ fʷ	2	**w**ii	**xʷ**i	**fʷ**i	exist, make, do
w9	w xʷ f	3	ji**w**ãɽi	ⁿɟu**xʷ**ãrĩ	ʈu**f**aɽĩ	spit (v)
f3	v f f	7	**v**ãɽã	**f**ãrã	**f**ãɽã	count (v)

In set ɹ3, we find an apicoalveolar (or apico-postalveolar) consonant in Ajië and Xârâcùù, but a retroflex in Tîrî, while we find a voiceless fricative in Xârâcùù and Tîrî, but a voiced approximant in Ajië. Taken together, the majority principle favors the reconstruction of a voiceless apicoalveolar (or apico-postalveolar) fricative in the proto-language in reference to this set, which I will note as *ʃ, following the transcription in Xârâcùù. This consonant must have shifted to a retroflex articulation in Tîrî, while it became a voiced approximant /ɹ/ in Ajië instead, which many speakers have later merged with /r/ (de La Fontinelle, 1976). Notably, a voiceless reconstruction was hypothesized by de La Fontinelle (1976) in reference to a bible translation from 1903, in which /ɹ/ was written as ⟨sh⟩ (p. 23).[21] My reconstruction confirms this hypothesis. If we take the bible translation into consideration, this voicing may have happened already, in the late 19th century.

In set j3, we find a voiced respectively voiceless palatal fricative in Ajië and Xârâcùù, but a voiceless dental stop in Tîrî. The majority principle favors the reconstruction of a palatal consonant, which must therefore have shifted to a dental articulation in Tîrî. Note that this proto-sound must also have been different from former *j, as this shift from palatal to dental did not affect the reflex of intervocalic *j in Tîrî (see voiced palatal continuants in 4.1.2.1). Thus, this correspondence favors the reconstruction of a voiceless consonant, on account of the reflexes in Xârâcùù and Tîrî. This consonant must have unconditionally shifted to a voiced approximant in Ajië, thus the ancestral form must have been *ç, which directly matches the reflex in Xârâcùù.[22] To explain the reflex in Tîrî, two independent changes must be reconstructed, where the voiceless fricative was on the one hand strengthened to a voiceless stop, and on the other hand shifted from palatal to dental. This reflex must therefore be derived via an intermediary form, where two options are available, as illustrated in (a) and (b) below:

(a)	1. Protoform	*ç	(b)	1. Protoform	*ç
	2. Manner shift	*c		2. Place shift	*θ
	3. Place shift	/t̪/		3. Manner shift	/t̪/

The correspondences in set -27 show the regular loss of a segment in Ajië, which has been retained as /x/ in Xârâcùù, and /h/ in Tîrî. Note that the realization of /h/ in Tîrî is conditioned by the following vowel, where it varies freely between [x] and [h] before non-close vowels, but is realized as [ç] before close front vowels, and [ɸ] before close central and back vowels (Osumi, 1995, pp.

21 St. Matthew's Gospel: *Visheshi i Jesus Keriso na sou na Matthieu*, published by Societe des Missions Evangeliques in 1903, was translated by Pastor Mathaia, from Ouvéa of the Loyalty Islands.

22 A voiceless reconstruction was also hypothesized by de La Fontinelle (1976) on the basis that /j/ was represented as ⟨s⟩ in many words in the 1903 Gospel (p. 55).

19–20). The sound must therefore have lost its earlier place of articulation in Tîrî, while it was completely elided in Ajië instead. It stands to reason that the reflex in Xârâcùù remains closest to the proto-sound, which supports reconstructing *x for the correspondences in question. In sets -2 and -26, the same consonant appears to have been sporadically deleted in both Xârâcùù and Tîrî.

In sets -28 to w11 we find overlapping correspondences with labialized consonants /w/, /xʷ/, /hʷ/ in the three languages. Note that the nonlabialized correspondences in sets -28 and -29 are only found before back vowels and can therefore be explained in reference to the phonotactic rules of the respective languages (see 2.3).[23] These must have been labialized in the proto-language as well but later merged with the reflexes of *x in all three languages. This sound must therefore have formed the labialized counterpart to *x, which gives the reconstruction *xʷ. In Tîrî, this consonant was debuccalized to /hʷ/, while in Ajië, it shifted to /w/ before front and central vowels.

However, the labialized consonants in Ajië and Xârâcùù also show overlapping correspondences with /f/ and /fʷ/ in Tîrî, as exemplified in sets w10 and w9 above. The two labiodental reflexes in Tîrî cannot be explained in reference to the phonological environment, thus it stands to reason that these reflect a different distinction in the proto-language. On account of this, the nonlabialized reflex in w9 must be secondary in Tîrî, on account of the presence of labialization in the other languages in the same set. While this delabialization is not predicted by the phonological context in Tîrî, it is previously known from regional variation, where Osumi (1995) noted that /fʷ/ is replaced by /f/ in many words in the variety in Grand Couli (p. 4). This delabialization can therefore not have been restricted to the variety spoken in Grand Couli but clearly occurred generally in Tîrî. The correspondences in sets w10 and w9 must therefore reflect a labialized labiodental fricative in the proto-language, which I will note as *fʷ. This consonant has independently merged with the reflex of *xʷ in both Ajië and Xârâcùù, as /w/ and /xʷ/ respectively. In Ajië, /w/ may have been derived via earlier *vʷ, which de La Fontinelle (1976) recorded as a variant of /w/ in the older population at the time of her study. However, I have not been able to establish a connection between the two.

In set f3, a plain labiodental fricative is found in all three languages. Thus, a voiceless fricative can be reconstructed here, following the majority principle, where the sound must have been regularly voiced to /v/ in Ajië. I will note this proto-sound as *f, which formed the plain counterpart to former *fʷ, and the voiceless counterpart to former *v.[24] Because the reflex in Ajië was also recorded as

23 In Ajië and Xârâcùù, labialized consonants only occur before front and central vowels, while in Tîrî, /h/ and /hʷ/ are phonemically distinct before mid-open /ɔ/ but not mid-close /o/, cf. /hʷɔɽɔ/ ‘flee’ respectively /ho/ ‘eat (meat)’.

24 ⟨f⟩ and ⟨v⟩ were both used in the 1903 Gospel but may not have reflected a phonemic distinction (de La Fontinelle, 1976, p. 40). For example, de La Fontinelle (1976) uses ⟨f⟩ to write /v/ in loanwords.

bilabial [β] in the speech of older speakers (de La Fontinelle, 1976, p. 40), this consonant may also have been a voiceless bilabial fricative [ɸ] in the proto-language.

To summarize, the correspondences support the reconstruction six voiceless fricatives in the proto-language, respectively *f, *f^{w}, *ʃ, *ç, *x, *x^{w}. These may only have occurred in initial position. Though the dataset includes a small set of reflexes of voiceless fricatives in intervocalic position in the descendant languages, these are overwhelmingly the result of productive compounding or derivation in the respective languages. Thus, there is no solid evidence for voiceless fricatives in intervocalic position in the proto-language. In Ajië, the correspondences illustrate a recurring voicing that may be reformulated as a sound law which predicts that all voiceless fricatives in the proto-language were regularly voiced in Ajië. By virtue of this sound law, the null reflex of *x in Ajië may have been derived via an intermediary voiced stage *ɣ, which was later elided in the language. However, it should be noted that a velar fricative /ɣ/ is found in Ajië, but primarily in intervocalic position. This reflects a different proto-sound, which I will return to in 4.1.2.3. In initial position, /ɣ/ is found in lexical words only in the inland variety, where it is still rare, while it is deleted in the same environment in the coastal variety (de La Fontinelle, 1976, p. 61). However, this sound is not attested in initial position in lexical words in the dataset.

4.1.2.3 Voiceless stops

Correspondences involving voiceless stops are very numerous and are therefore presented below in order of place of articulation, respectively bilabial, coronal, palatal, and velar.

To start, correspondences involving voiceless bilabial stops in initial and intervocalic position are exemplified in table 22 below.

Table 22: Correspondences involving voiceless labial stops

Set	Pattern			Matches	AJE	XAC	TIR	Concept
p$_{6}$	p	p	v	9	**p**ɛ	**p**ɛ	**v**e	take
p$_{5}$	p	p	p	8	**p**owee	**p**owee	**p**owee	round
p$_{4}$	p	p	m	2	**p**ãã	**p**a	**m**ãã	thigh (A, T), leg (X)
p^{w}$_{5}$	p^{w}	p^{w}	p^{w}	3	**p**^{w}ẽ	**p**^{w}ɛ̃	**p**^{w}ɛ̃	turtle
p^{w}$_{6}$	p^{w}	p^{w}	w	2	**p**^{w}e	**p**^{w}e	**w**ɯ	belly
p$_{3}$	p	p	-	2	**p**oaˀ	**p**oa	ua	bear fruit
v$_{4}$	v	p	v	2	tʌ**v**ʌ	tɛ**p**ɤ	ʈe**v**ɔ	speak, talk (A), speech (X, T)
w$_{5}$	w	p^{w}	w	1	ta**w**aaɽu	ta**p**^{w}aaru	ʈa**w**auɽu	oar

As noted with *v and *w in 4.1.2.1, plain and labialized bilabial stops have undergone a lenition process in intervocalic position in Ajië and Tîrî, producing voiced continuants /v/ and /w/ respectively, while they are regularly retained as voiceless stops /p/ and /pʷ/ in Xârâcùù in the same position, as exemplified in sets v_4 and w_5 above. Based on the correspondences in table 22, it is now evident that this lenition did not targeted bilabial stops in initial position in Ajië, where both Ajië and Xârâcùù invariably show a plain and labialized stop, respectively /p/ and /pʷ/, as exemplified in sets p_6 to p_3. However, the correspondences differ greatly in initial position in Tîrî, where up to six distinct correspondences patterns can be identified.

In sets p_6 and p^w_6, voiced continuants /v/ and /w/ are found in initial position in Tîrî as well, mirroring the two reflexes found in intervocalic position. This suggests a lenition process targeted the reflexes of bilabial stops in initial position in Tîrî as well, which favors the reconstruction of *p and *pʷ to the proto-language in reference to sets p_6 and p^w_6. In set p_4, the stop has instead been elided in Tîrî before a close back vowel /u/ or /ũ/. This reflex must be derived from *p via intermediary *w, where *w has been deleted before the close back vowel, as predicted by the phonological rules of Tîrî (see 2.3.2.3). In set p_4, we find a nasal consonant /m/ instead. This reflex is only found before nasal vowels, it is therefore reasonable to conclude that this is another conditioned reflex. This must exemplify a nasalization of earlier *p, possibly via intermediary *v.

This leaves the correspondences in sets p_5 and p^w_5, where bilabial stops are found in all three languages. Notably, these correspondences in Tîrî are about as frequent as the lenited reflexes in sets p_6 and p^w_6 respectively. This raises the question of why bilabial stops were only lenited in initial position in some words in Tîrî. In this regard, phonological context is not helpful, as the two share the same phonological environment. One could consider resolving this issue by reconstructing a second pair of bilabial stops in the proto-language. However, this would only complicate matters further, as secondary changes would have to be proposed in Ajië and Xârâcùù as well. Instead, there is reason to suspect that the non-lenited reflexes in initial position can be credited to borrowings in Tîrî, as the other two languages have not undergone a lenition in this position. Thus, if any of these words are borrowed, it is reasonable to hypothesize that exceptions to other regular changes should also be present in these words, where the language differ. Precisely this is observed in Tîrî /powee/ ‘round’, where we find /e/ in Tîrî, where /i/ would otherwise be expected from earlier *e (see front oral vowels in 4.1.1.1). Note that the word also contains a word-final long vowel in Tîrî, which is typically shortened in the language (see 4.1.1.4). On account of these exceptions, both Ajië and Xârâcùù are likely donor languages.

With this reconstruction in mind, we may precede with correspondences involving voiceless alveolar and retroflex stops in initial and intervocalic position, as exemplified in table 23 below.

Table 23: Correspondences involving voiceless coronal stops

Set	Pattern			Matches	AJE	XAC	TIR	Concept
t₂	t	t	ʈ	5	**t**a	**t**a	**ʈ**a	carve (A, T), write (X, T)
ɽ₆	ɽ	t	ɽ	5	kɯ**ɽ**ɯ	ki**t**i	ɯ**ɽ**ɯ	louse
r₄	r	t	r	4	p^{w}a**r**i	^{m}b^{w}a**t**i	p^{w}a**r**i	stomach
r₂	r	r	r	5	^{n}gɔ**r**u	ũ**r**ũ	ha**r**u	1DL.INCL (FOC/SBJ)

As noted with the vibrants earlier, I hypothesized that a voiceless stop *t was regularly lenited in intervocalic position in Ajië and Tîrî, producing either a flap /ɽ/ or a trill /r/, as exemplified in sets ɽ₆ and r₄ respectively. By now it can be concluded that this proto-consonant is regularly retained as a voiceless stop in initial position in all three languages, an alveolar stop /t/ in Ajië and Xârâcùù, a retroflex stop /ʈ/ in Tîrî, as exemplified in set t₂. Here, I will continue the reconstruction I proposed for the intervocalic reflexes in 4.1.2.1 and reconstruct a voiceless apicoalveolar or apico-postalveolar stop *t to the proto-language in both positions, which must have shifted to a retroflex stop in Tîrî, following the majority principle.

As mentioned above, this consonant produced either a flap or trill in intervocalic position in both Ajië and Tîrî. In lexical words, /t/ in Xârâcùù generally corresponds to a flap in both Ajië and Tîrî in this position, as exemplified in set ɽ₆. There is however an overlap in the distribution between the flap and trill in the two languages, where the trill is also found twice in intervocalic position in lexical words, included in set r₄. The remaining occurrences of the trill are restricted to grammatical words, either in intervocalic position, or in initial position in unstressed or bound forms, in sets r₄ and r₂. As noted with the vibrants in 4.1.2.1, there is reason to suspect that *t has been sporadically lenited in Xârâcùù in set r₂, which would otherwise predict /t/. This is because this correspondence pattern is solely found in grammatical words, and as such, it appears unlikely that a separate phonemic distinction (i.e. a trill) was present solely in these words. Thus, the general pattern in intervocalic position in Ajië and Tîrî shows that *t was typically lenited to a flap in lexical words, but a trill in grammatical words. However, it is less easy to explain this correlation. It is possible that the sounds were lenited at different stages in the languages' history, and therefore produced different reflexes.

Moving on, correspondences involving voiceless palatal and dental stops in initial and intervocalic position respectively are exemplified in table 24 below.

Table 24: Correspondences involving voiceless palatal stops

Set	Pattern	Matches	AJE	XAC	TIR	Concept
c_2	c c j	3	**c**uɛ	**c**uɛ	**j**uo	sit
c_1	c c -	1	**c**ẽẽ	**c**ĩĩ	ĩĩ	fly (v)
c_3	c ɲ t̪	1	**c**ãɽã	**ɲ**ẽrẽ	**t̪**ãɽã	Gould's petrel (*P. leucoptera*)
j_2	j c ð	2	pa**j**a	pa**c**a	pa**ð**a	loud, noisy
$ɲ_1$	ɲ c n̪	1	p^{w}ẽvi**ɲ**õɽõ	^{m}b^{w}ax**ĩc**ãrã	^{m}bɔhi**n̪**ãɽã	scorpion

In the correspondences in initial position in sets c_2 and c_1, voiceless palatal stops are found in both Ajië and Xârâcùù. In Tîrî, a palatal approximant /j/ is found in set c_2, while the consonant has been elided in set c_1 instead. Based on the majority principle, the proto-sound must have been a palatal consonant as well, for which I propose reconstructing a voiceless palatal stop *c to the proto-language on account of the correspondences in Ajië and Xârâcùù. This stop must have been lenited to /j/ in Tîrî. This lenition can also account for the null correspondence in set c_1 in Tîrî, which is found before a close front vowel /ĩ/, and as such must be a conditioned elision. This correspondence can be explained via intermediary *j from earlier *c.[25]

In set c_3, we find a voiceless palatal stop in Ajië as well. Based on the previous reconstruction, this must also reflect *c, but contrary to expectations, a palatal nasal consonant /ɲ/ is found in Xârâcùù, while a voiceless dental stop /t̪/ is found in Tîrî instead. This correspondences in this set are restricted before a nasal vowel in all three languages. Thus, there is a clear conditioning here in Xârâcùù, where by reconstructing *c for this set based on the reflex in Ajië, the earlier voiceless stop *c must have shifted to a nasal consonant /ɲ/ in Xârâcùù under influence from the following nasal vowel. However, because this change is not reflected in set c_1 in the same language, where the consonant also occurs before a nasal vowel, the regularity of this reflex is unclear.

Regarding the voiceless dental stop /t̪/ in Tîrî in set c_3, this reflex cannot be explained by the predicted lenition which targeted the reflexes of *c in sets c_2 and c_1 and produced /j/. Instead, this reflex must be derived from earlier *c by proposing a single change in place of articulation from palatal to dental, which also targeted the reflex of *ç in Tîrî as noted before in 4.1.2.2. The dental stop must therefore exemplify a reflex of *c which escaped the lenition that targeted the reflexes in sets c_2 and c_1 in Tîrî, which meant it remained free to undergo a shift in place of articulation from palatal to dental, alongside the reflex of the voiceless fricative *ç. The voiceless fricative must have merged with the former as *c at this point, following the chronology proposed in 4.1.2.2. However,

25 The loss of /j/ is not predicted by the phonotactic rules of the language (at least not synchronically speaking), cf. Tîrî /ji-ria/ 'divide land' vs /ii/ 'long time' (Osumi, 1995, p. 19).

unlike the reflex in Xârâcùù in the same set, this reflex is not self-evidently motivated in reference to the phonological environment. Note however that the presence of competing lenited and non-lenited reflexes in initial position mirrors the pattern observed with voiceless bilabial stops in Tîrî, which may be the result of borrowing from the other languages.

In intervocalic position, the correspondences in set j_2 are again indicative of a lenition process in Ajië and Tîrî, where *c has produced /j/ in Ajië, while *c has produced /ð/ in Tîrî. However, this intervocalic reflex in Tîrî cannot be explained in reference to the lenition process which targeted *c in initial position in sets c_2 and c_1 and produced /j/. Otherwise, /ð/ would be expected in initial position as well, where it is notably not found. Therefore, this reflex must have been secondarily lenited from earlier *t̪, which derives from *c through a single change in place of articulation. Thus, the shift from palatal to dental must therefore have occurred before the lenition process that produced /ð/ had applied.

In set $ɲ_1$, we find nasal consonants in intervocalic position in Ajië and Tîrî instead, a palatal and dental respectively. This correspondence is found before a nasal vowel in all three languages, and as such is clearly conditioned by the phonological environment. Here, the stop must have changed into a nasal consonant, possibly via a voiced continuant in both Ajië and Tîrî, where the reflex has also undergone the palatal-dental shift in Tîrî. It is unclear if this change would have preceded or followed the conditioned nasalization. Because these changes are only exemplified by a single cognate set, it is not possible to evaluate their regularity.

The correspondences in Tîrî therefore indicate that there must have been two separate lenition processes in the language, which occurred at different stages in the language's history and produced different reflexes in initial respectively intervocalic position. The oldest change must have produced /j/ from earlier *c, which is only attested in initial position. Reflexes of *c which had either been sporadically retained or reintroduced into the language after this lenition–presumably via later borrowings–were free to undergo a later lenition in intervocalic position, after first undergoing a shift in place of articulation, thus producing /t̪/ and /ð/ in initial and intervocalic position. A good candidate for a loan is Tîrî /paða/ 'loud, noisy', in which the presence of a non-lenited bilabial stop is further indicative of a borrowing. If this and other words containing the dental reflexes are borrowed, the shift from palatal to dental must have occurred recently, as the corresponding consonants in Ajië and Xârâcùù are both palatals.

Lastly, correspondences involving voiceless velar stops in initial and intervocalic position are exemplified in table 25 below.

Table 25: Correspondences involving voiceless velar stops

Set	Pattern			Matches	AJE	XAC	TIR	Concept
k₁	k	k	-	11	**k**ɛ̃	**k**ɛ̃	ɛ̃	burn (vi)
kʷ₂	kʷ	kʷ	w	5	**k**ʷã	**k**ʷã	**w**ã	boat
k₃	k	k	k	4	**k**ãˀ	**k**ʌ̃	**k**ã	breadfruit (A), papaya (X, T)
k₆	k	kʷ	k	1	**k**ɔɽee	**k**ʷarii	**k**ɔɽii	grasshopper
ɣ₁	ɣ	k	ɣ	4	kʌ**ɣ**ʌ	ⁿga**k**a	ka**ɣ**a	crow (*C. moledunoides*)
w₁	w	kʷ	w	1	ⁿgʷɛ̃ɛ̃**w**ɛ	ᵐbʷaa**k**ʷɛ	ⁿgi**w**e	mountain

In initial position in table 25, a plain and labialized stop /k/ and /kʷ/, are consistently found in Ajië and Xârâcùù. Again, the correspondences in initial position in Tîrî differ from the other languages, where three distinct patterns can be identified. The labiovelar approximant /w/ in set kʷ₂ can clearly be credited to the same lenition process which targeted *p and *c in initial position in Tîrî discussed before. Following the majority principle, *kʷ can therefore be reconstructed based on set kʷ₂. In set k₁, the consonant has been elided in Tîrî, based on which *k must be reconstructed to the proto-language in reference to the correspondences in the other languages. However, considering the proposed lenition discussed above, it can be hypothesized that the null correspondence in this set was derived via an intermediary stage *ɣ in Tîrî, which must have later been elided in this position.

As before, there are exceptions to this lenition in Tîrî, where voiceless stops are found in all three languages, as in sets k₃ and k₆. These two sets must therefore also reflect earlier *k and *kʷ, where the reflexes of the velar stops have somehow escaped lenition in Tîrî.[26] These reflexes are clearly sporadic, as is supported by their relative frequency compared to the lenited reflexes, which furthermore must be original to Tîrî, as they reflect a sound change that cannot be reconstructed to the other languages. Therefore, there is reason to suspect that the non-lenited reflexes in initial position have entered the language through later borrowings, as proposed with the non-lenited bilabial and palatal reflexes. A good candidate in favor of this hypothesis is Tîrî /kã/ ‘papaya’, where the diverging vowel reflex is further indicative of borrowing, here the expected reflex is /ɔ̃/ in Tîrî (see central nasal vowels in 4.1.1.2). On a semantic basis, this word must have been borrowed from Xârâcùù /kʌ̃/ ‘papaya’. Xârâcùù also may have been the donor language for Tîrî /kɔɽii/ ‘grasshopper’, where the vowel must have changed due to the preceding labialization in Tîrî.

Voiceless stops are also found in intervocalic position in Xârâcùù, which suggest reconstructing *k and *kʷ to this position as well, in reference to the reflexes in initial position. Ajië and Tîrî must

26 In set k₆, the reflexes in Ajië and Tîrî exemplify a conditioned delabialization before a back vowel (see 2.3).

therefore reflect a lenition of voiceless stops in this position, which produced /ɣ/ respectively /w/, as exemplified in sets ɣ1 and w1. The status of these consonants in the proto-language is debatable however, as both can be reconstructed to intervocalic position in monomorphemic forms only in one cognate set each. The remaining three occurrences of *k in set ɣ1 are restricted to initial position in bound forms, including one preposition and two bound verbs, which may not have necessarily been bound in the proto-language. The one occurrence of *k^{w} in set w1 is also a compound, where the first component part must be related to Ajië /^{n}g^{w}ã-/ ‘head’ and Xârâcùù /^{m}b^{w}a/ ‘head, top’. Thus, this word was likely not monomorphemic in the proto-language, if it is not a later compound.

Regarding the relationship between the initial and intervocalic reflexes in Tîrî, note that *k^{w} reflects /w/ in both initial and intervocalic position, while *k has been elided in initial position, but lenited to /ɣ/ in intervocalic position instead. By the hypothesis that *k was elided via *ɣ, we may entertain the idea that /ɣ/ was only elided in initial position. However, note that the voiced velar fricative occurs in initial position in a preposition, where it should have been elided if such a rule existed. Thus, it is more likely that this consonant was lenited at a later stage, from instances of *k that had been reintroduced after the regular loss of Proto-Mid-South *k in initial position in Tîrî.

Thus, to summarize, the correspondences suggest the reconstruction of six voiceless stops to the proto-language, respectively *p, *p^{w}, *t, *c, *k, *k^{w}. These stops have remained unchanged in initial position in Ajië and Xârâcùù, but show evidence of a lenition process in Tîrî, which targeted all voiceless stops in initial position except *t. Later borrowings in Tîrî appear to have obscured this sound law by introducing stop correspondences in new words, of which some have later undergone additional changes of their own. This has led to the presence of two competing correspondences patterns in initial position in Tîrî, as exemplified below:

Proto-form	*p	*p^{w}	*t	*c	*k	*k^{w}
	↓	↓	↓	↓	↓	↓
	/v/	/w/	/t̪/	/j/	Ø	/w/
borrowings? →	/p/	/p^{w}/		/t̪/	/k/	/k(^{w})/

Reflexes of all voiceless stops can be identified in both initial and intervocalic position in the descendant languages, though only *t can be frequently reconstructed to both positions, while the remaining voiceless stops are rare in intervocalic position.[27] Thus, only *t may have regularly occurred in this position in monomorphemic words in the proto-language, while the remaining

27 Voiceless stops can be reconstructed to intervocalic position in the following number of cognates: *t (14 cognates), *p (3 cognates), *p^{w} (3 cognates), *c (2 cognates), *k (1 cognate), *k^{w} (1 cognate).

could be the result of secondary morphological processes, either in the proto-language, or in the descendant languages, such as compounding, derivation, or reduplication, as evident with at least one cognate set. These intervocalic stops have later been regularly lenited in both Ajië and Tîrî. In Tîrî, this lenition parallels the lenition of stops in initial position, but because these processes have produces different results in the two positions, there is reason to suspect that the latter lenition of the non-coronal voiceless stops in intervocalic position occurred after the lenition which targeted the non-coronal stops in initial position. This intervocalic lenition must therefore only have been able to target reintroduced instances of voiceless stops, presumably via later borrowings, as illustrated below. With the reflexes of Proto-Mid-South *t in Tîrî, it is not possible to separate between inherited and borrowed reflexes, as this sound was never lenited in initial position in Tîrî.

Proto-form	*p	*pʷ	*t	*c	*k	*kʷ	
	↓	↓	↓	↓	↓	↓	
	/v/	/w/	/ɽ/	/j/	Ø	/w/	initial
			/ɽ, r/				intervocalic
borrowings? →	/p/	/pʷ/		/ʈ/	/k/	/k(ʷ)/	initial
	/v/	/w/		/ð/	/ɣ/	/w/	intervocalic

4.1.2.4 Voiced stops

Correspondences involving voiced stops are very numerous, and are therefore organized below in place of articulation, respectively bilabial, coronal, palatal, and velar.

To start, correspondences involving voiced bilabial stops are exemplified in table 26 below.

Table 26: Correspondences involving voiced labial stops

Set	Pattern			Matches	AJE	XAC	TIR	Concept
ᵐb₂	ᵐb	ᵐb	ᵐb	7	**ᵐb**aɽa	**ᵐb**ata	**ᵐb**aɽa	fear (v) (A, X), afraid (T)
ᵐb₃	ᵐb	ᵐb	p	5	**ᵐb**aɽəə	**ᵐb**arɤ	**p**aɽɯ	also
ᵐb₁	ᵐb	m	ᵐb	1	**ᵐb**aa	**m**ɛ̃	**ᵐb**ɛ̃	wing (A), arm (X, T)
ᵐbʷ₄	ᵐbʷ	ᵐbʷ	ᵐbʷ	1	**ᵐbʷ**iɣaa	**ᵐbʷ**exoo	fɔ**ᵐbʷ**ija	navel
ᵐb₅	ᵐb	ᵐbʷ	pʷ	1	**ᵐb**ɔɽi	**ᵐbʷ**ɛrɛ	**pʷ**eɽe	some, a few
ⁿgʷ₁	ⁿgʷ	ᵐbʷ	ⁿg	1	**ⁿgʷ**ɛ̃ɛ̃wɛ	**ᵐbʷ**aakʷɛ	**ⁿg**iwe	mountain
ᵐbʷ₁	ᵐbʷ	m	pʷ	1	**ᵐbʷ**ɛ̃	**m**ɔ̃	**pʷ**ɯ	night
ᵐbʷ₆	ᵐbʷ	ŋ	ⁿg	1	**ᵐbʷ**ɛ̃ɽɛ̃	**ŋ**ĩrĩ	pu**ⁿg**ɔ̃ɽɔ̃	black

As shown in table 26, voiced bilabial stops are found across all sets in Ajië. These form two overlapping correspondence patterns in Xârâcùù and Tîrî, where the corresponding consonants are either voiced stops or nasal consonants in Xârâcùù, and voiced stops or voiceless stops in Tîrî.

In Xârâcùù, the distribution of the two types is correlated with oral and nasal vowels respectively, where the corresponding nasal consonants in sets $^{m}b_{1}$, $^{m}b^{w}{}_{1}$, and $^{m}b^{w}{}_{6}$ are found before nasal vowels in Xârâcùù, while the corresponding voiced stop are found before oral vowels in the same language. On account of the correspondences in Ajië, the two correspondence patterns in Xârâcùù must be the result of a conditioned split, where voiced stops shifted to homorganic nasal consonants before nasal vowels.[28] Because voiced stops are always prenasalized in the languages, this change in manner only involves an extension of the nasal phase of the occlusive consonant, which suggests the voiced stops were prenasalized in the proto-language as well.

In Tîrî, the voiceless stops are only found before oral vowels, while the voiced stops are found before both oral and nasal vowels, though they are only marginally attested before the latter. Note that the voiceless stops cannot reflect lexical or morphological doublets in the proto-language, as they have not been affected by the lenition rule that produced voiced continuants from earlier voiceless bilabial stops, as demonstrated in 4.1.2.3. There must therefore have been a recurring devoicing or denasalization in Tîrî, which targeted earlier voiced bilabial stops before oral vowels. It is possible that this was the original conditioning behind this change in Tîrî, but that this pattern has since been overshadowed by later borrowings, which have introduced voiced stops before oral vowels, where they were previously eliminated through regular sound change.

Note that there also is a degree of overlap between bilabial and velar stops where labialization is involved, as exemplified in sets $^{n}g^{w}{}_{1}$ and $^{m}b^{w}{}_{6}$. Because /$^{m}b^{w}$/ can be found in both of these sets, it stands to reason that the proto-sound in these sets was *$^{m}b^{w}$, which must have sporadically resulted in velars in the three languages. Thus, in set $^{n}g^{w}{}_{1}$, this sound must have shifted to /$^{n}g^{w}$/ in Ajië, while in sets $^{n}g^{w}{}_{1}$ and $^{m}b^{w}{}_{6}$ in Xârâcùù and Tîrî, where a plain velar can be found instead, the resulted labialized velars must have been further delabialized. In Xârâcùù /ŋ̃r̃ɨ̃/ ‘black’, the resulted plain velar stop has additionally undergone a conditioned nasalization due to the following nasal vowel, resulting in /ŋ/ instead. Thus, the correspondences therefore only support the reconstruction of two bilabial stops to the proto-language, *^{m}b and *$^{m}b^{w}$, which indicates that /$^{n}g^{w}$/ must have evolved as a sporadic reflex of the latter.

28 The absence of labialization in Xârâcùù in set $^{m}b^{w}{}_{1}$ exemplifies a conditioned delabialization before a back vowel, as predicted by the phonotactic rules of the language (see 2.3.3.3).

With this in mind, we may precede with correspondences involving voiced alveolar and retroflex stops in initial position, as exemplified in table 27 below.

Table 27: Correspondences involving voiced coronal stops

Set	Pattern			Matches	AJE	XAC	TIR	Concept
ⁿd₂	ⁿd	ⁿd	ⁿɖ	16	**ⁿd**ɔjaa	**ⁿd**ɔjaa	**ⁿɖ**ɔjaa	whale
ⁿd₃	ⁿd	ⁿd	ʈ	6	**ⁿd**ɔ	**ⁿd**o	**ʈ**o	true
ⁿd₁	ⁿd	n	ⁿɖ	3	**ⁿd**ããwẽ	**n**ãwã	**ⁿɖ**ɛ̃ɛ̃wɛ̃	sand

Note here that voiced stops are found in all correspondence sets in Ajië. As with the voiced bilabial stops, these correspondences can instead be organized into two correspondence patterns in Xârâcùù and Tîrî in terms of manner. In Xârâcùù, the corresponding consonants are either voiced stops or nasal consonants. The corresponding nasal consonants in set ⁿd₁ are found before nasal vowels in Xârâcùù, while the corresponding voiced stops in set ⁿd₃ are found before oral vowels. This distribution corroborates the presence of a conditioned split in Xârâcùù, where voiced stops changed into homorganic nasal consonants before nasal vowels. However, it is possible to identify two correspondences in set ⁿd₂ which violate this pattern, as shown in (17). These words may retain a sound-symbolic form-meaning relation, which could have motivated an exception to the rule.

	AJE	XAC	TIR	Concept
(17)	ⁿdã	ⁿdɛ̃	ⁿɖʊ̃	swamp harrier (*C. approximans*)
	ⁿdũũ	ⁿdũũ	ⁿɖũũ	groan (A), snore, growl (X), whirling noise (T)

In Tîrî, the corresponding consonants in table 27 are either voiced stops or voiceless stops. There is an overlap between the two types, where the voiceless correspondences in set ⁿd₃ only occur before oral vowels in Tîrî, while the voiced correspondences in sets ⁿd₂ and ⁿd₁ occur before both oral and nasal vowels. As previously stated, it is possible that there was a phonological conditioning behind this change, but that this pattern has been overshadowed by later borrowings which have introduced voiced reflexes in phonological environments where they were previously eliminated by regular sound change. A good candidate for a borrowing is Tîrî /ⁿɖɔjaa/ ‘whale’. Note that the vowel in the first syllable in this word also retains the same quality as in Ajië and Xârâcùù, where /o/ is expected from earlier *ɔ in Tîrî (see 4.1.1.3). Likewise, the word also shows a long vowel in word-final position, which is rare in Tîrî (see 4.1.1.4). This word may therefore have been borrowed from either Ajië or Xârâcùù, as neither language is predicted to have changed either the consonant or vowel reflex in this environment. Correlations of this type cannot account for all correspondences,

however. Still, it remains most practical for now to reconstruct a single voiced prenasalized stop to the proto-language in reference to the correspondence in table 27. Based on the majority principle, I propose reconstructing a voiced apicoalveolar or apico-postalveolar stop *ⁿd here, which must have moved to a retroflex articulation in Tîrî, cf. the reflexes of *ʃ (see 4.1.2.2) and *t (see 4.1.2.3).

With this in mind, we may move on to correspondences involving voiced palatal and dental stops in initial position, as exemplified in table 28 below.

Table 28: Correspondences involving voiced palatal stops

Set	Pattern			Matches	AJE	XAC	TIR	Concept
ⁿɟ₂	ⁿɟ	ⁿɟ	ⁿɖ̪	4	**ⁿɟ**ɔ	**ⁿɟ**ɔ	**ⁿɖ̪**o	flow (v)
ⁿɟ₃	ⁿɟ	ⁿɟ	ʈ̪	3	**ⁿɟ**oɽa	**ⁿɟ**ora	**ʈ̪**oɽo	giant clam (*Tridacna*)
ⁿɟ₁	ⁿɟ	ⁿɟ	ɳ̪	4	**ⁿɟ**əˀ	**ⁿɟ**ɔ	**ɳ̪**oo	throwing spear
ⁿɟ₄	ⁿɟ	ɲ	ɳ̪	4	**ⁿɟ**uu	**ɲ**ĩ	**ɳ̪**õ	bone

As before, voiced stops are found in all correspondence sets in Ajië. Again, these sets can be organized into two correspondence patterns in Xârâcùù in terms of manner, where the corresponding consonants are either voiced stops or nasal consonants. The corresponding voiced stops in sets ⁿɟ₂ and ⁿɟ₃ are found before oral vowels in Xârâcùù, while the corresponding nasal consonants in set ⁿɟ₄ are found before nasal vowels. Thus, the correspondences indicate again the presence of a conditioned split in Xârâcùù, where voiced stops changed into homorganic nasal consonants when preceding a nasal vowel. However, one correspondence which violates this pattern can be identified in set ⁿɟ₁, as exemplified below in (18). This exception is unexplained.

	AJE	XAC	TIR	Concept
(18)	ⁿɟãˀ	ⁿɟɅ̃	ɳ̪õ	limestone formation (A), cave in cliff or rock formation (X, T)

Unlike before, the correspondences in Tîrî can be organized into three distinct patterns here, where the corresponding consonant is either a voiced stop /ⁿɖ̪/, a voiceless stop /ʈ̪/, or a nasal consonant /ɳ̪/. These correspondences are all indicative of the palatal-dental shift reconstructed for the reflexes of *ç (see 4.1.2.2) and *c (see 4.1.2.3), and as such, all three sounds must therefore reflect earlier palatal occlusives, respectively *ⁿɟ, *c, *ɲ. As established above, the correspondences in Ajië and Xârâcùù in all four sets favor the reconstruction of a voiced palatal stop *ⁿɟ. With the palatal-dental shift in mind, this consonant must therefore have been nasalized to *ɲ in some words in Tîrî, devoiced to *c in others, while it has been retained as a voiced stop in others more. However, the regularity and conditioning behind these changes is not directly evident. Notably, these reflexes

show overlapping distribution, where the voiced and voiceless stops are both attested solely before oral vowels, as exemplified in sets ⁿɟ₂ and ⁿɟ₃, while the nasal consonant is notably found before both oral and nasal vowels, as exemplified in sets ⁿɟ₁ and ⁿɟ₄. Because the nasal consonant is the only reflex of *ⁿɟ found before nasal vowels in Tîrî, it must be regular in that environment. On the other hand, /n̪/ overlaps with both /ⁿd̪/ and /t̪/ before oral vowels, making it unclear what the regular reflex is this environment. As before, there is reason to suspect that the voiced reflex /ⁿd̪/ may be a result of later borrowings before oral vowels. This is supported by Tîrî /t̪oɽo/ ‘giant clam’, which has a regional variant /ⁿd̪oɽa/ found in the Grand Couli variety, which resembles more closely the related forms in Ajië and Xârâcùù (see table 28). In this regional variant, the voiced stop has notably not been devoiced before an oral vowel, while the vowel in the second syllable has not been assimilated in reference to the preceding vowel (see 4.1.1.4). This still leaves the voiceless stops and nasal consonants as competing reflexes before oral vowels. If we assume that the nasal reflex exemplifies the oldest change in the language, the voiceless reflex must go back to reflexes of *ⁿɟ that were either not targeted by this nasalization, or later reintroduced into the language, after which they were free to undergo the devoicing which targeted all other voiced stops in Tîrî. There is nevertheless no indication that either of these reflexes occur in borrowed material in Tîrî.

Lastly, we can address correspondences involving voiced velar stops in initial position, as exemplified in table 29 below.

Table 29: Correspondences involving voiced velar stops

Set	Pattern			Matches	AJE	XAC	TIR	Concept
ⁿg₅	ⁿg	ⁿg	k	3	**ⁿg**ɛ	**ⁿg**ɛ	**k**e	2SG (FOC/SBJ)
ⁿg₆	ⁿg	ⁿg	ⁿg	5	**ⁿg**ɛɛ	**ⁿg**ɛɛ	**ⁿg**ɛɛ	grandmother
ⁿg₇	ⁿg	ŋ	k	2	**ⁿg**ɛve	**ŋ**ɛ̃ɛ̃	**k**evi	1PL.EXCL (FOC/SBJ)
ⁿg₈	ⁿg	ŋ	ⁿg	2	**ⁿg**ʌˀ	**ŋ**ũ	**ⁿg**ũ	move, fidget

As before, voiced stops are found in all correspondence sets in Ajië, whose correspondences can be organized into two patterns in Xârâcùù and Tîrî based on manner of articulation.

In Xârâcùù, the corresponding consonants are either voiced stops or nasal consonants, as observed with the preceding correspondence sets involving voiced stops. The corresponding voiced stops in sets ⁿg₅ and ⁿg₆ are likewise found before oral vowels in Xârâcùù, while the corresponding nasal consonants in sets ⁿg₇ and ⁿg₈ are found before nasal vowels. The distribution of these correspondences therefore confirms the presence of a conditioned split in Xârâcùù, which produced homorganic nasal consonants from earlier voiced stops when preceding a nasal vowel.

In Tîrî, the corresponding consonants are either voiced stops or voiceless stops, which, as before, overlap in their distribution. That is, /ⁿg/ is found before both oral and nasal vowels, as in sets ⁿg₆ and ⁿg₈, while /k/ is found solely before oral vowels in Tîrî, as in sets ⁿg₅ and ⁿg₇. Note as well that the voiceless reflexes cannot derive from a voiceless stop *k, as they have not been targeted by the sound change which regularly elided voiceless velar stops in Tîrî (see 4.1.2.3). This reflex must therefore be indicative of a conditioned devoicing in Tîrî, but its conditioning remains unclear. Here, there is reason to suspect that some of the cognates containing voiced stops in initial position before an oral vowel have been borrowed after the proposed devoicing rule in Tîrî. A good candidate for this is Tîrî /ⁿgɛɛ/ ‘grandmother’, where we also find an irregular vowel reflex in the form of /ɛ/, where /e/ from earlier *ɛ in Tîrî (see 4.1.1.1). Compare this word to Tîrî /ke/ ‘2SG (FOC/SBJ)’ in table 29, where the vowel is both regularly raised, and the voiced stops is predictably devoiced before an oral vowel. This former may therefore have been borrowed from either Ajië or Xârâcùù, as no changes are predicted in either language in this environment. Correlations of this type cannot account for all correspondences, however.

To summarize the results of the voiced stops, the correspondences support the reconstruction of five voiced stops in the proto-language, respectively *ᵐb, *ᵐbʷ, *ⁿd, *ⁿɟ, *ⁿg. The correspondences do not support the reconstruction of a voiced labialized velar stop in the proto-language, which is notably found in all three descendant languages. As far as it can be determined, this consonant emerged as a sporadic reflex of *ᵐbʷ, which is directly attested in one word in Ajië.

In Xârâcùù, the voiced stops have undergone a conditioned split, where voiced stops resulted in homorganic nasal consonants when preceding nasal vowels. Though this change is not frequently attested in the dataset, it can be reconstructed for voiced stops in every place of articulation, and as such is a good candidate for a regular sound change in the language.

In Tîrî, there is a recurring devoicing of voiced stops before oral vowels. This may have been the result of a conditioned change, though there is a notable overlap between voiceless and voiced stops before oral vowels. As I have argued, this overlap may be the result of later borrowings, which have reintroduced voiced stops before oral vowels, though this cannot be demonstrated for all words in question. On the other hand, there is also evidence of a nasalization, which solely affected the reflex of the voiced palatal stop *ⁿɟ. Unlike in Xârâcùù, this nasalization appears to have been an unconditioned change, which must have preceded the devoicing trend.

The sets presented in the preceding sections include primarily correspondences in initial position, though there are two notable occurrences of voiced stops in intervocalic position as well, included

in the same correspondence sets. On account of the two words in (19) below, it is possible to reconstruct *^{m}b^{w} and *^{n}d to intervocalic position in the proto-language, though there are no traces of the remaining voiced stops. Thus, if voiced stops existed in this position in the common proto-language, they must have been highly marginal, and left few traces in the descendant languages.

	AJE	XAC	TIR	Concept
(19)	ũ^{m}**b**^{w}aˀ	xũ^{m}**b**^{w}ɛ	hu^{m}**b**o	short
	mi^{n}**d**ɛ	mɛ̃^{n}**d**ɛ	ma^{n}**ɖ**e	orange (A), citrus (X, T)

4.1.2.5 Nasal consonants

Table 30 exemplifies correspondence sets involving nasal consonants, here primarily in initial position. Related correspondences in intervocalic position will be discussed separately below.

Table 30: Correspondences involving nasal consonants

Set	Pattern			Matches	AJE	XAC	TIR	Concept
m$_1$	m	m	m	21	**m**ã	**m**ɛ̃	**m**ɛ̃	and
m^{w}$_1$	m^{w}	m^{w}	m^{w}	4	**m**^{w}ã	**m**^{w}ã	**m**^{w}ã	house
n$_1$	n	n	ɳ	22	**n**ẽ	**n**ã	**ɳ**ɛ̃	fly (n)
l$_1$	l	n	ɳ	2	**l**ʌ	**n**ɨ	**ɳ**ɔ	landslide
ɲ$_2$	ɲ	n	ɳ	2	**ɲ**uu	**n**ɔ̃	**ɳ**ũ	saddled puffer (*C. valentini*) (A), fish (n) (X, T)
ɲ$_3$	ɲ	ɲ	ɲ	1	**ɲ**awã	**ɲ**ãwã	**ɲ**ãwã	Indian mackerel (*R. kanagurta*)

In sets m$_1$ and m^{w}$_1$ we find a recurring correspondence between bilabial nasal consonants /m/ and /m^{w}/ in the three languages. While bilabial nasal consonants have previously been shown to derive from *p in Tîrî and *^{m}b in Xârâcùù before nasal vowels, neither of these reconstructions are applicable here. These sets must therefore reflect nasal consonants in the proto-language as well, for which I propose reconstructing *m and *m^{w} respectively.

In set n$_1$, nasal consonants are likewise found in all three languages, here /n/ in Ajië and Xârâcùù, and /ɳ/ in Tîrî. This correspondence cannot reflect a voiced stop either, as this change is only regularly observed in Xârâcùù. This correspondence must therefore reflect a nasal consonant in the proto-language. Based on the majority principle, I propose reconstructing an apicoalveolar or apico-postalveolar nasal consonant *n for this set, which must have shifted to a retroflex in Tîrî, as observed with the reflexes of all other apicoalveolar obstruents in Tîrî.

In set l$_1$, the coronal nasal consonants overlap with a lateral approximant /l/ in Ajië. The lateral is known to reflect an earlier nasal consonant in some words in Ajië, where /n/ and /l/ are

sometimes interchangeable. This is likewise confirmed by consulting older transcriptions.[29] This is a recent or ongoing conditioned change in the language, where /n/ becomes /l/ in oral contexts in Ajië.

The correspondences in sets ɲ2 and ɲ3 involve palatal nasal consonants. In set ɲ2, we find a palatal nasal consonant in Ajië, but a coronal nasal consonant in Xârâcùù and Tîrî. This set appears to exemplify a sporadic palatalization of earlier *n, for which we find regular reflexes in Xârâcùù and Tîrî. However, this correspondence overlaps with /n/ before central and back vowels in Ajië, as exemplified by other correspondences in the neighboring set n1. Therefore, this palatalization cannot have been phonologically conditioned, and is thus unexplained.

In set ɲ3, a palatal nasal consonant is instead found in all three languages though this is only representative of a single cognate set in the dataset. This reflex must be irregular in Tîrî, as it has not undergone the palatal-dental shift which has regularly targeted all palatal obstruents in the language. Thus, the word in question is likely a recent borrowing from either of the other languages. Here, reconstructing a palatal nasal may seem most appropriate. However, the problem with this reconstruction–apart from the lack of reliable evidence due to likely borrowing in Tîrî–is that there is there is no additional evidence for *ɲ in initial position in the dataset. Thus, it is not possible to judge the regularity of this correspondence set. On the other hand, it could reflect earlier *c, which has sporadically shifted to /ɲ/ before nasal vowels in both Xârâcùù and Ajië (see 4.1.1.3).

With this reconstruction in mind, we can now return to the correspondences involving nasal consonants in intervocalic position as discussed in 4.1.2.1, here exemplified in (20–22) below.

	Set	Pattern			AJE	XAC	TIR	Concept
(20)	m1	m	m	m	ko**m**ãã	kɔ**m**ãã	ko**m**ã	yawn (v)
	v1	v	-	m	ⁿgɔ**v**u	ŋɔ̃ɔ̃	ko**m**u	1DL.EXCL (FOC/SBJ)
	v2	v	-	v	ⁿgɛ**v**e	ŋɛ̃ɛ̃	ke**v**i	1PL.EXCL (FOC/SBJ)
(21)	n1	n	n	ɳ	pɛ**n**ãã	pɛ**n**ãã	pe**ɳ**ãã	mast
	n2	n	n	ɽ	**n**ẽˀ	**n**ĩ	fa**ɽ**ĩ	name
	ɽ4	ɽ	r	ɳ	kʷĩ**ɽ**ĩ	kʷã**r**ĩ	wa**ɳ**e	tree sp. ‘acajou’ (*S. vitiensis*)
	ɽ5	ɽ	r	ɽ	vã**ɽ**ã	fã**r**ã	fã**ɽ**ã	count (v)
(22)	-20	-	ɲ	j	ĩĩ	xi**ɲ**i	hũ**j**ũ	cricket
	-35	j	-	-	ã**j**u	xãĩ	ai	red ant (A), ant (X, T)

29 A lateral approximant was also recorded by Leenhardt (1935), but many words transcribed with /l/ in de La Fontinelle (1976) and Aramiou & Euritein (1995) contain /n/ in Leenhardt (1935), e.g. /newe/ ‘inside’ (Leenhardt, 1935) but /lewe/ ‘inside, belly’ (de La Fontinelle, 1976; Aramiou & Euritein, 1995).

In reference to the reconstruction of nasal consonants in initial position, we can therefore reconstruct *m for the cognates in (20) and *n for the cognates in (21) above, which have been lenited to varying degrees in the descendant languages. Further support for this hypothesis can be found in Tîrî /faɽĩ/ ‘name’, where the nasal consonant has been lenited in a derived intervocalic position, which is not present in the other languages.[30] In (22), a palatal nasal consonant can be identified in one word, which may be a sporadic retention of an earlier *ɲ. The same reconstruction may be proposed for the correspondence in ‘red ant (A), ant (X, T)’, on account of the surrounding nasal vowels. The problem here is that the palatal nasal is not well supported in initial position in the proto-language, which makes it difficult to exclude that this is not a sporadic reflex of a different proto-sound as well, possibly *j or *n. Because *w can also be reconstructed to both oral and nasal contexts (see 4.1.2.1), it follows that this consonant may likewise reflect a later merger of earlier *m^{w} and *w. However, this hypothesized nasal reflex is not attested in the dataset in intervocalic position.

These correspondences therefore support the reconstruction of three nasal consonants, *m, *m^{w}, *n, in both initial and intervocalic position in the proto-language, which have variously merged with *v, *w, *r in intervocalic position in the three descendants. A palatal nasal consonant *ɲ may also have occurred marginally in both positions.

4.1.2.6 Manner doublets

Now that a fundamental understanding of the consonant distinctions in the proto-language has been established, as well as what changes these have undergone in the descendant languages, we can address some remaining correspondences which deviate from these patterns. Among these were 23 notable consonant correspondence sets in which the languages showed predictable correspondences in terms of place of articulation, but not in terms of manner. These sets were all marginal, of which 20 of 23 only occurred once. Thus, these correspondences most likely do not reflect additional distinctions in the proto-language but must be the result of other processes. Based on the patterns of change described in the preceding sections, these correspondences allow for the reconstruction of various doublets in the proto-language, which reflect alternations in manner of articulation.

The most widespread type involves an alternation between voiceless and voiced stops, which is attested in both initial and intervocalic position, as illustrated in (23) and (24) below. Note that because both voiced and voiceless stops may merge as voiceless stops in Tîrî, an initial voiceless

30 There is a nominalizing prefix /fa-/ in Tîrî (Osumi, 1995, p. 108), but it is not clear if this is correctly assigned here.

stop in Tîrî can potentially reflect both, and in such cases, it is therefore not possible to conclude what the earlier consonant was based on the reflex in Tîrî.

	Set	Pattern			Matches	AJE	XAC	TIR	Concept
(23)	p_1	^{m}b	p	p	1	**^{m}b**i	**p**ii	**p**ii	palm sp. (*C. circinalis*)
	$p^{w}{}_3$	p^{w}	$^{m}b^{w}$	p^{w}	1	**p^{w}**ari	**$^{m}b^{w}$**ati	**p^{w}**ari	stomach
	$^{n}d_5$	^{n}d	t	ʈ	1	**^{n}d**ɛ	xa**t**ɛ	**ʈ**a	ash
	k_8	k	^{n}g	k	1	**k**ʌɣʌ	**^{n}g**aka	**k**aɣa	crow (*C. moledunoides*)
(24)	$^{m}b^{w}{}_7$	w	$^{m}b^{w}$	$^{m}b^{w}$	1	^{n}da**w**a	^{n}dɔɔ**$^{m}b^{w}$**ɛ	^{n}ɖɔ**$^{m}b^{w}$**e	drum
	r_1	r	^{n}d	^{n}ɖ	1	mɛ**r**ɛ	mã**^{n}d**ɨ	ŋɛ̃ma**^{n}ɖ**i	bitter
	j_5	j	^{n}ɟ	ð	2	vi^{n}ga**j**u	ŋã**^{n}ɟ**u	^{n}ga**ð**u	sorcery (A), bewitch (X, T)
	$ɣ_2$	ɣ	^{n}g	ɣ	1	maa**ɣ**e	mɛ̃**^{n}g**i	me**ɣ**i	warm

In initial position, this pattern can be found in nouns as well as once in a pronoun. In intervocalic position, it is attested in both nouns and verbs. In intervocalic position, the reflexes are lenited in both Ajië and Tîrî. In Ajië, the lenited reflexes must reflect earlier voiceless stops, however in Tîrî, it is less clear. It should be noted that these have not been lenited according to the first lenition rule which regularly targeted *c and *k in initial position but have instead been lenited according to the second lenition rule, which can be associated with later borrowings (see 4.1.2.3). However, it is difficult to find good candidates for donor languages in the examples in (24). Seeing as voiced stops are frequently devoiced in Tîrî, it is not impossible that the lenited reflexes reflects voiced stops if this trend extended to intervocalic position as well.

The second type of doublet involves an alternation between prenasalized and nasal consonants, as exemplified in (25) below.

	Set	Pattern			Matches	AJE	XAC	TIR	Concept
(25)	$ʈ_2$	ʈ	^{n}d	^{n}ɖ	4	kõ**ʈ**õ	kõ**^{n}d**ɔ	õ**^{n}ɖ**o	blue, green

This type is only attested in intervocalic position, and is found in four cognate sets, where the consonants in Ajië reflect earlier *n, on account of the bilateral spread of nasality, while the consonants in Xârâcùù and Tîrî reflect earlier *^{n}d. This may not necessarily reflect a doublet, but could be the result of a secondary change, either in Ajië, or in both Xârâcùù and Tîrî.

The third type of doublet involves an alternation between voiceless stops and fricatives, as exemplified in (26) below. This pattern is attested in multiple parts of speech, but is only found in initial position, like voiceless fricatives in general.

	Set	Pattern	Matches	AJE	XAC	TIR	Concept
(26)	f₂	p f f	1	**p**ɛ	**f**a	**f**a	causative prefix
	k₉	k x h	1	**k**aɽa^{ʔ}	**x**ɔru	**h**aɽu	beautiful, good

The fourth type is only known from the first-person inclusive focal or subject pronouns, where the reflex in Ajië suggests a voiced velar stop *^{n}g, while the reflexes in Xârâcùù and Tîrî suggests a voiceless velar fricative *x. This doublet clearly reflects a morphological alternation in the common ancestor, perhaps relating to the focal and subject forms.

	Set	Pattern	Matches	AJE	XAC	TIR	Concept
(27)	^{n}g₂	^{n}g - h	2	**^{n}g**ɔru	ũrũ	**h**aru	1DL.INCL (FOC/SBJ)
				^{n}gɛre	ĩrĩ	**h**ari	1PL.INCL (FOC/SBJ)

These doublets are a positive addition to the reconstruction, as they confirm the relationship between the reconstructed consonants. Doublets with varying prenasalization and aspiration are known from other Kanak languages, which is hypothesized to reflect morphological variation in Proto-New Caledonian (Ozanne-Rivierre, 1992; Rivierre, 1993). Still, it is not clear how much is the result of later changes in the respective descendant languages.

4.1.2.7 Summary of reconstructed consonants

To summarize the results of the preceding sections, the correspondences support the reconstruction of 24 consonants in the proto-language, as organized in table 31 below. The status of the palatal nasal consonant is debatable, note the parenthesis. The plain labial fricatives were either labiodental or bilabial, as recorded in the speech of older speakers of Ajië (de La Fontinelle, 1976, p. 40).

Table 31: Reconstructed consonant inventory of Proto-Mid-South

		Labial		Coronal	Palatal	Velar	
		Plain	Labialized			Plain	Labialized
Stop	vl oral	*p	*p^{w}	*t	*c	*k	*k^{w}
	vd prenas	*^{m}b	*^{m}b^{w}	*^{n}d	*^{n}ɟ	*^{n}g	
Nasal	vd	*m	*m^{w}	*n	(*ɲ)		
Vibrant	vd			*ɾ			
Fricative	vl	*f	*f^{w}	*ʃ	*ç	*x	*x^{w}
	vd	*v			*j		*w

At this point, it is possible to check the plausibility of the reconstructed segments against the phonemic inventory of the proto-language. In this regard, there are some clear gaps in the phonemic inventory. Notably there is no voiced labialized velar stop, no velar nasal consonant, no labialized counterpart to *v, and no plain counterpart to *w.

The distribution of consonants in the proto-language differ greatly in initial and intervocalic position. All consonants except the vibrant *r and the voiced continuants *v, *j, and *w can be reconstructed to initial position in lexical words. The reconstruction indicates that the number of possible consonants in intervocalic position in monomorphemic forms was highly restricted. There is notably no support for voiceless fricatives in this position. Stops were likewise highly restricted. Only *t occurs repeatedly in intervocalic position in the proto-language, but the remaining voiced and voiceless stops are not well supported in this position. Three or four nasal consonants were likely phonemically distinct in intervocalic position in the common proto-language as well, which were all merged with the reflexes of vibrants and voiced continuants in the descendant languages.

4.1.3 Glottalization

While the glottalization in Ajië has no counterpart in the other two languages, it is possible to manually analyze the distribution of the glottalization. Based on the reconstruction outlined above, several cognate sets can be identified which form minimal or near-minimal pairs in the proto-language that differ only by the presence of glottalization in the cognate in Ajië, as exemplified in (28) and (29). Minimal pairs of this kind offer strong support for the presence of a phonemically distinct glottalization in Proto-Mid-South as well.

	AJE	XAC	TIR	Concept
(28)	poaˀ	poa	ua	bear fruit
	pʷaa	poa	ua	grate (A), sharpen (X, T)
(29)	mãˀ	mĩã	mã	urinate
	mii	mĩã	maa	red

4.2 Relating the reconstruction

In the preceding section, I presented and discussed the phonological structure of the last common ancestor of the three languages Ajië, Xârâcùù, and Tîrî, which we can call Proto-Mid-South. With this reconstruction in mind, the goal of the following section is to address the secondary aim of this

study as formulated in 1.2, i.e. to clarify the position of the Mid-Southern languages within the New Caledonian group, by comparing the results of the derived reconstruction with higher-order lexical reconstructions. However, because there is no extensive lexical reconstruction of Proto-New Caledonian (PNC), primarily lexical reconstructions of Proto-Oceanic (POc) or other higher-order ancestral forms must be consulted instead. In this regard, 122 of 266 cognate sets could be reliably connected with established POc or other higher-order lexical reconstructions.[31]

4.2.1 Canonical forms

As cited in the background in 2.4.2, previous research has revealed the presence of a widespread loss of final syllables in the languages of the southern mainland, which can be summarized in two consecutive steps (Ozanne-Rivierre & Rivierre, 2004, pp. 146–147):

Proto-Oceanic	1: Loss of final vowels	2: Loss of final consonants
*CVCV	> *CVC	> *CV

This observation is corroborated by my reconstruction, where it is possible to see that the loss of a final consonant has left a trace on the nucleus. Thus, note in (30) that where the resulted coda reflects an earlier nasal, lateral, or prenasalized consonant in POc, the loss of this consonant is reflected by the presence a nasal vowel in Proto-Mid-South.[32]

	AJE	XAC	TIR	Concept	POc
(30)	ne	nɛ̃	ŋɛ̃	taste	*ña**m**i- (Ross et al., 2016)
	ne^{n}dẽ	nɛ̃^{n}dɛ	ŋ̃^{n}ɖɛ̃	lake	*dra**n**o (Ross et al., 2003)
	p^{w}ẽ	p^{w}ɛ̃	p^{w}ɛ̃	turtle	*po**ñ**u (Ross et al., 2011)
	nẽ	nã	ŋɛ̃	fly (n)	*la**ŋ**o (Ross et al., 2011)
	kãˀ	kʌ̃	kã	breadfruit (A), papaya (X, T)	*ku**l**uR ‘breadfruit’ (Ross et al., 1998)
	kẽ	kɛ̃	ɛ̃	burn (vi)	*(k,g)a**b**u (Ross et al., 2003)
	k^{w}ã	k^{w}ã	wã	boat	*wa**g**a (Ross et al., 1998)

31 This number should be interpreted with caution. However, previous lexicographic studies on Melanesian language have indicated that Kanak languages are highly divergent lexically speaking (see Pawley, 2006; Skirgård, 2020).

32 Notations for reconstructions are as follows:

*(A) Reconstruction cannot determine presence of A

*(A,B) Reconstruction cannot determine A or B with certainty.

On the other hand, where the elided coda goes back to a non-nasal consonant in POc, the loss of this consonant is correlated with the presence of a glottalization in many cognates, as exemplified in (31) below.

	AJE	XAC	TIR	Concept	POc
(31)	anʌˀ	nɛ	ŋe	firewood (A, X), fire (X, T)	*na a**p**i (Ross et al., 1998)
	kɛˀ	kɛ	e	husk with teeth	*ka**t**i (Ross et al., 1998)
	nɛ̃ˀ	nĩ	faɽĩ	name	*na i(**s,c**)a- (Ross et al., 2016)
	kiuˀ	kʌɨ	evɯ	Malay apple (*S. malaccense*)	*kapi**k**a (Ross et al., 1998)
	poaˀ	poa	ua	bear fruit	*pua**q** (Ross et al., 2008)

It should be noted that this non-nasal coda is commonly reflected as a final voiceless stop in Northern languages (Ozanne-Rivierre, 1995). Therefore, it stands to reason that the glottalization also derives from earlier voiceless stops in coda position, as preserved in the north. However, in many cognates in my dataset, glottalization is not present where it would be expected following this hypothesis. And while the glottalization may be underrepresented in the dataset to an extent (see 3.3), it is also possible that the glottalization only evolved in certain environments in Proto-Mid-South, though I have not made a systematic analysis in this regard.

In the same vein, there are also cognate sets between the three languages in which the final vowel is sometimes retained in one of the descendant languages but lost in another, as exemplified in (32) and (33) below. What is interesting to note in these examples is that the presence of an intervocalic consonant is correlated with the trace of an earlier coda, in the form of a nasalization or glottalization on the preceding vowel.

	AJE	XAC	TIR	Concept	POc
(32)	pũɽũ	pũ	wã	hair	*pulu- (Ross et al., 2016)
	ẽ	ẽrẽ	hɛ̃ɽẽ	content	*kanoŋ ‘flesh’ (Ross et al., 1998)
(33)	kʌˀ	kɨrɛ	ɔ	pot	*kuron (Ross et al., 1998)
	kʷaˀ	kʷiɛ	wie	rain (n)	*qusan (Ross et al., 2003)

A variable syllable reduction of this kind was also noted by Ozanne-Rivierre (1995) in her study of the Northern languages, which surfaces either as a productive alternation, where the final vowel is lost in free forms, but preserved in bound forms, or through fossilized traces of a similar process, where the final vowel has been retained in one language but lost in another (pp. 48–49). It can therefore be concluded that the intervocalic consonants and the nasalization or glottalization must

derive from a common source, which goes back to voiceless stops respectively nasal consonants in PNC. The question is whether these processes were already fossilized in Proto-Mid-South, or if the reduction of the coda which produced nasality and glottalization is the result of a later drift. Notably, the coda is also lost in the Far-Southern languages, as well as in the Northern language Paicî (Ozanne-Rivierre & Rivierre, 2004, pp. 147–150), which may favor the latter scenario.

4.2.2 Origin of consonant distinctions

In the following section, phonological antecedents of the reconstructed consonant distinctions in Proto-Mid-South are discussed in order of manner of articulation. See 2.4.2 for a list of consonant changes attributed to PNC.

4.2.2.1 Voiceless stops

In initial position, it is possible to identify phonological antecedents for all voiceless stops except *c, as exemplified in (34). Note that Proto-Mid-South *t can only be related to a POc antecedent once in a doublet form, note the parenthesized cognates.

		AJE	XAC	TIR	Concept	POc
(34)	*p	**p**ɛ	**p**ɛ	**v**e	take	***p**a (Ross et al., 2016)
	*p^{w}	**p**^{w}ẽ	**p**^{w}ã	**w**ɔ	fruit	***pu**aq (Ross et al., 2008)
	*t	(^{n}dɛ)	xa**t**ɛ	(ʈa)	ash	***r**apu(R) (Ross et al., 1998)
	*k	**k**ãˀ	**k**ʌ̃	**k**ã	breadfruit (A), papaya (X, T)	***k**uluR (Ross et al., 1998)
	*k	**k**ũʈũ	**k**ũ	ɔ̃	shoot (A), cutting (X, T), head (T)	***q**ulu ‘head’ (Ross et al., 2003)
	*k^{w}	**k**^{w}aˀ	**k**^{w}iɛ	**w**ie	rain (n)	***q**usan (Ross et al., 2003)
	*k^{w}	**k**^{w}ã	**k**^{w}ã	**w**ã	boat	***w**aga (Ross et al., 1998)

As can be seen, all reconstructed stops go back to voiceless stops in POc, with two notable exceptions, where we find POc *r and *w. However, POc *r is reconstructed as a voiceless retroflex stop *ʈ in PNC (Ozanne-Rivierre, 1992), thus this sound continues the stops reflex in PNC. However, this consonant must have undergone a different change, a shift in place of articulation from retroflex *ʈ in PNC to alveolar *t in Proto-Mid-South. Regarding POc *w, which is reflected as *k^{w} in Proto-Mid-South, this fortition is unique to the Mid-Southern group (Ozanne-Rivierre & Rivierre, 2004, p. 152). While Ozanne-Rivierre & Rivierre (2004) already noted this innovation in members of the Mid-Southern group, my reconstruction indicates that this fortition occurred already in Proto-Mid-South, on account of the regular lenition of voiceless stops in Tîrî (see 4.1.2.3).

One important merger can be identified here, where Proto-Mid-South *k reflects a merger of POc *k and *q in initial position. This merger is likewise unique to the Mid-Southern group. In the Northern and Far-Southern languages, POc *k has changed in a different direction, but consistently retains a distinct reflex from POc *q (Ozanne-Rivierre, 1992, pp. 194–195). Note that POc *q has resulted in Proto-Mid-South *kʷ as well, where the labialization has transferred from the vowel.

In intervocalic position, my reconstruction indicates that the voiceless stops inherited from PNC underwent a lenition process in Proto-Mid-South, where all voiceless stops in Proto-Mid-South except *t were either lenited or deleted, as shown in (35). This lenition can account for all voiced continuants except Proto-Mid-South *w, for which no phonological antecedents can be identified.

		AJE	XAC	TIR	Concept	POc
(35)	*v	kiuˀ	kʌɨ	e**v**ɯ	Malay apple (*S. malaccense*)	*ka**p**ika (Ross et al., 1998)
	*t	kɯ**ɽ**ɯ	kɨ**t**ɨ	ɯ**ɽ**ɯ	louse	*ku**t**u (Ross et al., 2011)
	*r	ⁿdo**ɽ**u	ⁿdɔ**r**u	ⁿɖo**ɽ**u	coral tree (*E. variegata*)	*ra**r**ap (Ross et al, 1998)
	*j	ɔ**j**ɔ	xɔ**j**ɔ	ha**j**o	marry	*qa**s**awa (Blust & Trussel, ongoing)
	–	ii	xii	hii	grate, shave	*ko**s**i (Ross et al., 1998)
	–	kʷɛɛ	kʷɛɛ	wãã	root	*wa**k**aR (Ross et al., 2008)
	–	ⁿdaa	ⁿdaa	ⁿɖãã	day	*ra**q**ani (Ross et al., 2003)

This lenition therefore explains the notable exception of intervocalic *t, which regularly goes back to POc dental *t. This sound is reconstructed as a voiceless dental stop in PNC as well (Ozanne-Rivierre, 1992, p. 202), which is preserved in the two languages of the far south (Ozanne-Rivierre, 1995, p. 59). My reconstruction therefore indicates a manner shift in Proto-Mid-South, though notable, this sound was not lenited. This connection therefore supports the hypothesis that the remaining voiceless stops–which can only be marginally reconstructed to intervocalic position–were only restricted to morpheme-initial position in Proto-Mid-South (see 4.1.2.3).

Regarding the lenited and elided consonants, there are parallels in other languages of the mainland, including the lenition of POc *p, *r, and the elision of POc *k and *q in intervocalic position, which Ozanne-Rivierre (1995) reconstructs to Proto-North as well. It is therefore possible that this reflects a lenition process in an earlier ancestral form, Proto-Mainland New Caledonian. However, due to the commonality of such changes, it cannot be excluded that this is the result of a later drift.

First, we see that Proto-Mid-South *r goes back to POc *r in intervocalic position. Note that the same sound became *t in initial position instead, thus merging with the intervocalic reflex of POc *t. This merger may have extended to initial position as well, but this cannot be corroborated

here. Nevertheless, the fact that POc *t and *r retain distinct reflexes in intervocalic position in Proto-Mid-South means that the two must have remained distinct at an early stage of the proto-language. Because Ozanne-Rivierre (1995) also reconstructs an intervocalic lenition of POc *r in Proto-North, this lenition may have occurred already in Proto-Mainland, where POc *t was still dental. This may explain why the latter did not undergo a lenition in Proto-Mid-South.

My reconstruction also reflects a sporadic elision of POc *s in intervocalic position, where it is otherwise reflected as Proto-Mid-South *j in the same position. Note that this consonant is reconstructed as a voiceless palatal stop [c] in PNC. Thus, it is very likely related to *c in initial position, though this cannot be corroborated here in relation to POc. Nevertheless, it is difficult to say whether this sound was lost sporadically in Proto-Mid-South, or lost independently in the descendant languages, as noted with the reflexes of Proto-Mid-South *j (see 4.1.2.1).

In relation to POc, my reconstruction reflects a regular loss of POc *k and *q in intervocalic position. Because these consonants are merged in initial position, they may have been elided via a common intermediary form as well, e.g. *ɣ. This explains the absence of such a consonant in the reconstruction. It may likewise be noted that both POc *k and *q are elided in Proto-North in this position as well, though they retain distinct reflexes in initial position, respectively Proto-North *c and *k (Ozanne-Rivierre, 1995, pp. 57–58). Thus, it is possible that these were lenited or elided already in Proto-Mainland New Caledonian.

4.2.2.2 Voiced stops

In initial position, it is possible to identify phonological antecedents for all five voiced stops reconstructed in 4.1.2.4. As exemplified in (36), the prenasalization can often be traced back to POc, but in many cases, it must be secondarily attributed.

		AJE	XAC	TIR	Concept	POc
(36)	*ᵐb	ᵐ**b**ə	ᵐ**b**ɤ	ɔ̃**p**ɯ	buttocks	***b**(ʷ)oto- (Ross et al., 2016)
	*ᵐbʷ	ᵐ**b**ʷẽ	**m**ɔ̃	**p**ʷɯ	night	***b**oŋi (Ross et al., 2003)
	*ⁿd	ⁿ**d**u	ⁿ**d**u	**ʈ**u	clam (A), large black clam (X, T)	***t**u(q)e (Ross et al., 2011)
	*ⁿd	ⁿ**d**u	ⁿ**d**u	**ʈ**uu	dual prefix (A, T), dual article (X)	***r**ua ‘two’ (Ross et al., 1998)
	*ⁿd	ⁿ**d**ɔ	ⁿ**d**o	**ʈ**o	true	*(t,**d**)uqu (Ross et al., 2016)
	*ⁿd	neⁿ**d**ẽ	nẽⁿ**d**ɛ	ŋĩⁿ**ɖ**ɛ̃	lake	***dr**ano (Ross et al., 2003)
	*ⁿɟ	ⁿ**ɟ**əˀ	ⁿ**ɟ**ɔ	**ɲ**oo	throwing spear	***s**ao(t) (Ross et al., 1998)
	*ⁿɟ	ⁿ**ɟ**ẽ	ⁿ**ɟ**ɛ	ⁿ**ɖ**e	breeze (A), blow (of wind) (X, T)	***j**aŋi ‘wind’ (Ross et al, 2003)
	*ⁿg	ⁿ**g**i	ⁿ**g**iɛ	ⁿ**g**iwa	hatchet	***k**iRam (Ross et al., 1998)

This non-etymological prenasalization must also be responsible for the many voiced/voiceless doublets (see 4.1.2.6), which reflects an alternation found in other Kanak languages as well (Ozanne-Rivierre, 1992). It is nevertheless not clear how much of this prenasalization may be reconstructed to PNC, and how much evolved in Proto-Mid-South or the descendant languages.

Note in (36) that Proto-Mid-South *^{n}d confirms the predicted merger in Proto-Mid-South of dental and retroflex stops inherited from PNC, which goes back to POc *t/*d, respectively *r/*dr.

4.2.2.3 Voiceless fricatives

In comparison with POc reconstruction, it can be determined that the voiceless fricatives of Proto-Mid-South derive from a reduplication of initial voiceless stops, which produced voiceless aspirated stops in late PNC or Proto-Mainland via the loss of pretonic syllables (Ozanne-Rivierre, 1992, p. 203). This reduplication can be traced back to POc in one cognate set, as exemplified by 'mouth' in (37) below (note the *C_iVC_iV-form). Elsewhere, this reduplication must be inferred based on the reconstructed consonant in Proto-Mid-South.

		AJE	XAC	TIR	Concept	POc
(37)	*f	**v**ãɽã	**f**ɤ̃^{n}dʌ	**f**a^{n}ɖɔ	walk (v)	***p**ana/***p**ano (Ross et al., 2016)
	*f^{w}	ne**w**ã	nã**x^{w}**ã	ŋ̃**f**ɔ	mouth	***p^{w}ap^{w}**a(q) (Ross et al., 2016)
	*ʃ	**ɹ**u	**ʃ**ii	**ʂ**i	sew	***t**uRi (Ross et al., 1998)
	*ʃ	**ɹ**ɯɽɯɯ	**ʃ**ɨrɨɨ	**ʂ**ɯɽɯɯ	shake (v)	***r**uru (Ross et al., 2016)
	*ç	**j**aane	**ç**aa	**ʈ**aa	bad	***s**aqat (Ross et al, 1998)
	*x	ii	**x**ii	**h**ii	grate, shave	***k**osi (Ross et al., 1998)
	*x	ɔjɔ	**x**ɔjɔ	**h**ajo	marry	***q**asawa (Blust & Trussel, ongoing)
	*x^{w}	ɔˀ	**x^{w}**ɛ	**h**o	meat (A), eat meat (X, T)	***q**oda(q) (Ross et al., 2016)

Here, my reconstruction shows that earlier aspirated stops must have been spirantized already in Proto-Mid-South. A spirantization of earlier aspirates is also observed in many Northern languages but cannot be reconstructed to Proto-North (Ozanne-Rivierre, 1995). According to Rivierre (1993), this aspiration is reflected through a tonal distinction in the Far-Southern languages instead. This secondary reduplication can also account for the voiceless stop/fricative doublets, as noted in 4.1.2.6. However, it remains to be determined how much of this alternation can be credited to PNC.

Note that the reflexes in (37) continue the same mergers reflected by the corresponding voiceless and voiced stops, where Proto-Mid-South *ʃ reflects a merger of the reduplicated forms of POc *t and *r, while Proto-Mid-South *x reflects a merger of the reduplicated forms of POc *k and *q. Labialized *x^{w} evolved secondarily from POc *q before back vowels in some words.

4.2.2.4 Nasal consonants

Reconstructed nasal consonants derive from earlier nasal consonants in PNC as well, as exemplified in (38) and (39) in initial and intervocalic position respectively.

		AJE	XAC	TIR	Concept	POc
(38)	*m	**m**ã	**m**ɛ̃	**m**ɛ̃	and	***m**a (Ross et al., 2003)
	*m^{w}	**m^{w}**ã	**m^{w}**ã	**m^{w}**ã	house	*Ru**m**aq (Ross et al., 1998)
	*n	**n**eji	**n**ɨi	**ɳ**ɯi	island	***n**usa (Ross et al., 2003)
	*n	**n**ãã	**n**ɔ̃ɔ̃	**ɳ**ãũ	mosquito	***ñ**amuk (Ross et al., 2011)
	*n	**n**iɽɛ	**n**irɛ	**ɳ**iɽe	snot	***ŋ**orok (Ross et al., 2016)
	*n	**n**ɛ̃	**n**ã	**ɳ**ɛ̃	fly (n)	***l**aŋo (Ross et al., 2011)
(39)	*m	^{n}g**v**u	ŋɔ̃ɔ̃	ko**m**u	1DL.EXCL (SBJ)	*ka**m**i-rua (Lynch & Ozanne-Rivierre, 2001)
	–	nãã	nɔ̃ɔ̃	ɳãũ	mosquito	*ña**m**uk (Ross et al., 2011)
	*n	k^{w}ĩ**ɽ**ĩ	k^{w}ã**r**ĩ	wa**ɳ**e	tree sp. (*S. vitiensis*)	*wa**l**asi (Ross et al., 2008)

In initial position in (38), Proto-Mid-South *n continues a merger of POc *n, *ñ, and *l, as reconstructed to PNC (Ozanne-Rivierre, 1992, p. 202), but also reflects a later merger of PNC *n and *ŋ, which has been attributed to Proto-Mainland (Lynch & Ozanne-Rivierre, 2001, p. 35). Note that Proto-Mid-South *m and *m^{w} both go back to POc *m, where the labialization has evolved secondarily. In 'house', the nasality could have spread from the preceding back vowel, or the word may have undergone a metathesis, i.e. POc *Rumaq > PNC **uma(q)[33] > **mua(q) > **m^{w}a(q).

In intervocalic position in (39), the nasal consonants have typically merged with Proto-Mid-South *v and *r in the descendant languages, though there is reason to suspect these were also nasal consonants *m and *n in Proto-Mid-South, as discussed in 4.1.2.5. However, in relation to POc forms, it can be concluded that the reflex of POc *m was also sporadically elided. It is difficult to say if this sound was lost independently here, as *v is regular elided in both Ajië and Xârâcùù.

As with the voiceless stops, nasal consonants had split into aspirated and unaspirated forms in Proto-Mainland, which is hypothesized to reflect an earlier reduplication process in pre-PNC (see 2.4.2). This reduplication can be traced back to POc in one cognate set, as exemplified in (40) below (note the *C$_{i}$VC$_{i}$V-form).

		AJE	XAC	TIR	Concept	POc
(40)	*m	**m**ãã	**m**ãã	**m**ãã	chew (A, T), chew (leafs) (X)	***mam**aq (Ross et al., 2016)

33 Regular loss of *R in PNC (see 2.4.2).

This confirms that aspirated nasal consonants were unconditionally merged with unaspirated nasal consonants in Proto-Mid-South. In other cognate set, the presence of an earlier aspirate may only be inferred by analyzing cognate forms in other Kanak languages that regularly preserve the contrast. For comparison, the cognate of POc *mamaq retains a voiceless nasal consonant in many Northern languages, e.g. Bwatoo /m̥aa/ ‘chew’ (Rivierre et al., 2006).

4.2.3 Origin of vowel distinctions

Several factors make charting the origin of the vowel distinctions in Proto-Mid-South a difficult task. First, there is a tendency to lose all POc final vowels, as noted in 4.2.1. Second, there is frequent loss of intervocalic consonants, which has resulted in many vowel sequences, of which many have produced both short and long vowels in Proto-Mid-South, for which I have not been able to identify any clear regularity. In many cases where a likely etymology in POc can be identified, the vowels have changed irregularly in the descendant languages, and thus it has not been possible to determine what the vowel was in Proto-Mid-South. In other cases, it is possible to establish a clear sound change in comparison with POc, but because the correspondence is only marginally reflected, it is difficult to establish both the conditioning behind the change, as well as the regularity of the sound correspondence. This task would require a more systematic study, which I will not attempt here. Still, two conditioned splits can be identified in Proto-Mid-South, which are quite well supported in the dataset.

First, POc *u underwent an unrounding in certain phonological environments, as exemplified in (41). This split produced Proto-Mid-South *u and *ɯ, respectively *ũ and *ʌ̃ after the loss of nasal codas. Note that the rounding is preserved on the onset in ‘rain (n)’, which must reflect earlier *u.

		AJE	XAC	TIR	Concept	POc
(41)	*ɯ	k**ɯɯ**ɽɯ	k**ɨ**t**ɨ**	**ɯ**ɽɯ	louse	*k**u**tu (Ross et al., 2011)
	*ʌ̃	k**ã**ʔ	k**ʌ̃**	k**ã**	breadfruit (A), papaya (X, T)	*k**u**luR (Ross et al., 1998)
	*ʷi	kʷ**a**ʔ	kʷ**i**ɛ	w**i**e	rain (n)	*q**u**san (Ross et al., 2003)
	*ũ	k**ũ**ɽũ	k**ũ**	**ɔ̃**	shoot (A), cutting (X, T), head (T)	*q**u**lu ‘head’ (Ross et al., 2003)
	*u	p**o**aʔ	p**o**a	**u**a	bear fruit	*p**u**aq (Ross et al., 2008)
	*ũ	p**ũ**ɽũ	p**ũ**	w**ã**	hair	*p**u**lu- (Ross et al., 2016)

As illustrated in (41), POc *u has retained its rounding in Proto-Mid-South after reflexes of POc bilabials and uvulars, while it has become unrounded after reflexes of POc velars. This split shows

that POc velars and uvulars, which have otherwise merged in Proto-Mid-South as velars, must have remained distinct at an early stage of the proto-language, which left a trace on the following vowel. This split–and by extension the evolution of central vowels–must signal one of the earliest innovations that define the Mid-Southern subgroup as a whole.

Interestingly, Ozanne-Rivierre (1995) reconstructs a similar split of POc *u in Proto-North, in which POc *u is fronted to *i in contact with unrounded or non-uvular consonants. However, it is not clear if this split of POc *u in Proto-Mid-South followed the same conditioning after coronal and palatal consonants, thus it cannot be excluded that these are analogous changes.

Second, POc *a reflects a split in Proto-Mid-South, which produced *a and *ɛ, respectively *ã and *ɛ̃ after the loss of nasal codas, as exemplified in (42) below.

		AJE	XAC	TIR	Concept	POc
(42)	*ɛ	k**ɛ**ʔ	k**ɛ**	**e**	husk with teeth	*k**a**ti (Ross et al., 1998)
	*ɛ	p**ɛ**	p**ɛ**	v**e**	take	*p**a** (Ross et al., 2016)
	*ɛ̃	k**ɛ̃**	k**ɛ̃**	**ɛ̃**	burn (vi)	*(k,g)**a**bu (Ross et al., 2003)
	*a	k**a**	**a**	**a**	agentive nominalizer	*q**a**ta ‘person’ (Ross et al., 2016)
	*a (or *ã)	pʷ**ɛ̃**	pʷ**ã**	w**ɔ**	fruit	*pu**a**q (Ross et al., 2008)
	*ã	kʷ**ã**	kʷ**ã**	w**ã**	boat	*w**a**ga (Ross et al., 1998)

As illustrated in (42), POc *a has been raised and fronted in unrounded environments, producing Proto-Mid-South *ɛ and *ɛ̃ after plain *p, while it has been retained as Proto-Mid-South *a and *ã after labialized *pʷ and *kʷ. The split is also attested after Proto-Mid-South *k, where it can be correlated with POc *k and *q, which have both merged as velars in Proto-Mid-South. This conditioning must therefore have applied before the merger of POc *k and *q reflected in my reconstruction. This split has no known parallels in the Northern and Far-Southern groups.

4.2.4 Summary

According to the results of this study, the following divergent traits can now be reconstructed to the Proto-Mid-South language in comparison with POc reconstructions, which separate the group from the neighboring Northern and Far-Southern groups:

- Split of POc *u.
- Split of POc *a.

- Fortition of POc *w.
- Merger of POc *k and *q in initial position.
- Merger of POc *t/*d and *r/*dr in initial position.
- Spirantization of aspirated stops.
- Merger of unaspirated and aspirated nasals.

The Mid-Southern languages also share multiple traits with the Far-Southern group, including the loss of final vowels and final consonants, which produced nasal vowels from earlier nasal codas. It is possible that these traits evolved in an early ancestral form of the two groups, but this could also be the result of convergence between the ancestral forms, or the result of a later drift, as the same series of innovations have also independently evolved in the Northern language Paicî.

4.3 General discussion

4.3.1 Method and material

To begin, I should be stated that the phonetic alignment in EDICTOR was valuable as it significantly reduced the amount of work spent deriving sound correspondences manually. At the same time, some issues also became apparent with the tool which are worthy of consideration here. Because EDICTOR only aligns words linearly in a matrix, it was not helpful in identifying contextual changes between the languages, which primarily obscured sound changes involving long vowels and vowel sequences. Because EDICTOR forces the user to make a choice about the segmentation, there was no ideal way to segment long vowels and vowel sequences for the phonetic alignment, thus the contexts had to be manually located and checked in the dataset. Because both long vowel and vowel sequences were treated as disyllabic vowel sequences, these consistently generated two correspondence sets through the phonetic alignment. The problem with this is that neither of these correspondence sets can illustrate relevant differences between the languages, which relate to the vowel sequences as a whole and not their respective parts. This proved further problematic when vowel sequences corresponded to short vowels, where there was no possible alignment that was helpful in generating sound correspondences. These shortcomings showed the importance of relying on phonological context when analyzing sound correspondences, and if dealing with correspondences in isolation, there would be no way to derive reliable reconstructions.

Regarding the material, there was one important aspect which had an influence on the result, which related to artifacts of the transcription. In this regard, there was one notable inconsistency between the sources in the transcription of the so-called “central vowels”, which were variously transcribed as either central or back unrounded vowels depending on the author. Because this study relied on secondary data, it was not possible to challenge the transcriptions proposed in the original sources. This therefore had an impact on the reconstruction as well, as it was not possible to determine the degree of backness of the proto-vowels. As a compromise, all were marked as back unrounded.

In this thesis, I approached the topic from two angles when addressing my primary and secondary aim. For the primary aim, I relied on a bottom-up approach through the comparative method, which involved starting with synchronic data and reconstructing back in time, while for the secondary aim, I relied on what can be described as a top-down approach, which involved mapping the derived phonological reconstruction to established higher-order reconstructions and comparing the results with previous historical linguistic work on Kanak languages. In order for the two approaches to complement each other, it was critical to derive a phonological reconstruction before attempting to connect the descendant languages with higher-order reconstructions. As an example, consider POc *k, which has up to five reflexes in Tîrî: /k/, /ɣ/, /ⁿg/, /h/, and zero. Based on the intermediary phonological reconstruction we can conclude that POc *k was regularly elided in intervocalic position in Proto-Mid-South, while it produced *k, *x, and *ⁿg in initial position, the latter two via non-etymological aspiration and prenasalization. Proto-Mid-South *k was then regularly elided in Tîrî, though it was later reintroduced as /k/ and /ɣ/ in borrowings as well. Proto-Mid-South *x was regularly debuccalized to /h/, or even sporadically elided too, while Proto-Mid-South *ⁿg produced both /k/ and /ⁿg/ in Tîrî. Only by relying on a bottom-up reconstruction first was it possible to differentiate between these reflexes which reflect the same consonant in POc but derive from different intermediary forms in Proto-Mid-South.

Concerning the method, one of the main challenges facing the reconstruction regarded the cut-off point between regular and irregular correspondences in the descendant languages. The issue is that the smaller a correspondence set is, the more difficult it becomes to judge the regularity of the correspondence pattern, which increases the risk of irregularities misguiding the reconstruction. Because most of the correspondence sets did not occur more than five times in the dataset, many sound correspondences were not well supported in terms of quantity. To counteract this, it was important to identify recurring or correlated patterns of change, to strengthen the regularity of individual changes. This is exemplified by voiced stops changing into nasal consonants in Xârâcùù.

In this case, the specific changes are only marginally attested, but because this sound change can be reconstructed for voiced stops in all places of articulation, it was clearly a regular process.

Still in some cases where correspondences only occurred once in the dataset it was not possible to derive a reliable reconstruction, as there were no related correspondence sets to which a hypothesis could be tested. The presence of many marginal correspondences also implies that distinctions in the proto-language could be easily overlooked due to absence of data. This problem relates to the lack of negative evidence in the comparative method, which means that even if distinction is not reflected by the correspondences, it cannot be fully excluded that it was present in the proto-language. This also implies that if certain distinctions have been independently lost in each of the languages under consideration, there may be no way to reconstruct this distinction in the proto-language based on sound correspondences alone. In this regard, higher-order reconstructions were able to fill in some gaps in the phonological reconstruction.

4.3.2 Explaining the results

While there was a great challenge to identify regular correspondences between the three languages, the results of this study show that it is to a large extent still possible, through a qualitative analysis, to identify recurring sound changes and separate between borrowed and inherited material in the languages. At the same time, the preceding sections highlight many of the traits previously attributed to the "aberrant" languages of Melanesia (see Grace, 1992; Pawley, 2006). That is, in the three languages Ajië, Tîrî, and Xârâcùù, we find sound systems that radically depart from the systems reconstructed to earlier ancestral forms, here in comparison with POc, between which regular sound correspondences are notably difficult to work out. These two findings can be credited to a high rate of phonological innovation in the three languages, which has resulted in many regular and sporadic sound changes between the three languages, both in relation to their common ancestor, and between their common ancestor in relation to POc. The question is why the languages behaved this way.

The traditional explanation for linguistic diversification considers divergence over time. Based on the archaeological record, New Caledonia was settled around 1000 BCE (Sand, 1999, p. 155), thus the time span from the arrival of the last common ancestor to the evolution of the roughly 28 modern Kanak languages is around 3000 years. This time span is not notably different from that of many other Oceanic subgroups, whose members are undoubtedly less innovative, at least phonologically speaking. Therefore, there must have been other reasons behind this diversification.

Another factor that can be considered is language contact. This is evident in two ways in the phonological reconstruction, either by lexical borrowings, or by the diffusion of linguistic features. Regarding lexical borrowings, one of the main challenges for deriving regular sound correspondences came to the issue of separating between inherent and borrowed material in the descendant languages. The reason why potential borrowings could be identified in Tîrî was because these exemplify exceptions to regular sound changes in the language. That is, borrowings are more transparent in Tîrî where the language has undergone changes according to sound laws which did not occur in the neighboring languages. Considering where the languages are spoken in relation to each another, Tîrî could have directly borrowed vocabulary from both languages, which is also supported by the reconstruction, while it is less likely that Ajië would borrow vocabulary from Xârâcùù or vice versa, as neither are spoken in direct contact. On the other hand, languages size could also have had an impact, where Tîrî is notably less widely spoken than Ajië and Xârâcùù, and may therefore be more prone to borrowings. On the same note, some words which appear to be borrowed in Tîrî, are more likely borrowed in all three languages, either directly or indirectly from a common source. For example, Tîrî /ⁿɖɔjaa/ ‘whale’ shows signs of borrowing from either Ajië or Xârâcùù /ⁿdɔjaa/ ‘whale’. However, all three must be borrowed from Numèè (far south), cf. /ⁿɖo-jaa/ ‘sperm whale’, a compound formed from the noun /ⁿɖo/ ‘whale’ (Rivierre & Vandegou, 2020).

The phonological reconstruction also uncovered shared innovations between the languages that are the result of independent changes in either two or all three languages. This includes the lenition of intervocalic voiceless stops in Ajië and Tîrî, which have produced trills and flaps in both languages, and the lenition of intervocalic nasal consonants, which has occurred in all three languages. What is notable about some irregular sound correspondences between the languages is that these exemplify the introduction of marginal phonemic distinctions. For example, Proto-Mid-South *n has sporadically changed into both /ɲ/ and /l/ in Ajië, of which neither appear to be regularly inherited from the common proto-language. Another example is the voiced stop /ⁿgʷ/, which is distinctive in all three daughter languages, but is in neither language inherited via regular sound change from Proto-Mid-South. This sound is in fact so marginal that it is only observed in a single cognate in the dataset, where it appears as a sporadic reflex of *ᵐbʷ in Ajië. What is interesting to note about these innovations is they serve the function of filling gaps in the phonemic inventory. This role is also fulfilled by borrowing. The best example of this is the reintroduction of /ɛ/ in Tîrî, which had previously been deleted through the regular raising of Proto-Mid-South *ɛ to Tîrî /e/, which left a gap in the resulted phonemic system after Proto-Mid-South *ʌ was regularly backed to /ɔ/ in Tîrî. Thus, there is an interplay between convergence and internal pressure from the phonological

systems of the respective languages to fill gaps in the phonemic inventories.

This interplay may also account for rare phonemic distinctions in the descendant languages which did not occur in the dataset. This includes the velar nasal consonant /ŋ/ in Ajië, which did not occur in any cognate vocabulary, but the introduction of this consonants may also serve the function of filling a gap. On the other hand, some rare or even absent phonemes are evidently the result of later borrowings. This includes /l/ in Xârâcùù, which is only known from recent loans (Moyse-Faurie & Néchérö-Jorédié, 1989, pp. 12–13), or perhaps the absence of /v/ in Xârâcùù, which is likewise explained by its regular deletion in the language. Likewise, the rarity of palatal /c, ɲ/ and labialized velar /kʷ, ⁿgʷ/ in Tîrî has previously been credited to borrowings from neighboring languages (Grace, 1976; Osumi, 1995), which is confirmed by my reconstruction as well.

It can also be noted in relation to sporadic correspondences was that there were proportionally more marginal correspondences involving vowels than consonants. This suggests that vowels changed more irregularly than consonants, which is well supported both between the descendant languages, as well as in their common ancestor in relation to POc. This could relate to the nature of vowels themselves, which exist on a spectrum. This could have been further accentuated by the size of the vowel inventories, which reduces the distinction between the different vowel qualities.

5 Conclusion

The primary aim of this study was to reconstruct as much as possible about the phonological structure of the last common ancestor to the Mid-Southern languages of the New Caledonian mainland, by means of comparing lexical and morphological data in the descendant languages. The three languages Ajië, Tîrî, and Xârâcùù were chosen for this task, based in part on the accessibility of material, which was further motivated by a fundamental hypothesis about their internal classification. The goal of this phonological reconstruction was to clarify the position of the Mid-Southern languages within the New Caledonian subgroup, by comparing the results with higher-order reconstructions. As such, the study was driven by the following research question:

RQ: What phonological distinctions can be reconstructed to the last common ancestor of the Mid-Southern subgroup?

In response to this research question, 266 potential cognate sets were collected between Ajië, Xârâcùù, and Tîrî, from which 127 correspondence sets involving consonants and 220 correspondence sets involving vowels were generated. By analyzing these correspondences, the results supported the reconstruction of 16 vowels (ten oral and six nasal) and 24 consonants to the proto-language. As such, this study offers the first systematic reconstruction of the phonological history of the Mid-Southern group, which indicates that many of the characteristic traits of the Mid-Southern languages, such as the central vowels, evolved already in the common ancestor.

By further comparing the results of the phonological reconstruction with POc etymologies, the results present evidence that the three languages in question are defined by multiple phonological innovations that characterized their common ancestor, of which several cannot be reconstructed to the neighboring Northern and Far-Southern subgroups. This presents strong evidence in favor of a distinct Mid-Southern subgroup, which must form a sister group to the previously proposed Northern and Far-Southern subgroups.

At this point, we can return to the initial hypothesis about internal classification in the Mid-Southern subgroup as formulated in 1.2, which assumed equal distance between the varieties. In this regard, there is not sufficient ground to challenge the internal classification proposed by Haudricourt (1971), as there is no evidence from the phonological reconstruction that indicates that either of the three languages are more closely related to one another. However, lexical or morphological evidence may shine further light on this topic in future research.

While the phonological reconstruction has demonstrated that the three languages share a common ancestor, it remains to be determined how the remaining varieties relate to this reconstruction, and how these can be classified in relation to the three languages investigated here. This particularly concerns the more divergent ones, Orowe, Neku, and Sichëë, for which there is little to no documentation. This marks a gap in our new-found understanding of the Mid-Southern group. However, this also implies that there may be gaps in the phonological reconstruction, though this remains to be determined by future research.

There are other areas where this reconstruction can be expanded as well. There are still many unanswered questions about the details of the evolution from POc to Proto-Mid-South, particularly regarding to the evolution of the reconstructed vowel system. It should also be noted that this study focused solely on phonological history in the three languages, but there is still work to do focusing on lexical and morphological reconstruction, which is not only lacking on Mid-Southern languages, but on Kanak languages as a whole. The results of this study also have the potential to inform new lexicostatistical studies involving languages of the Mid-Southern group, where cognacy judgments can now be informed by sound correspondences.

References

Aramiou, S., & Euritein, J. (1995). Wordlist in Ajië. In D. T. Tryon (ed.), *Comparative Austronesian dictionary: An introduction to Austronesian studies*, part 2–4. Berlin/New York: Mouton de Gruyter.

Campbell, L. (2013). *Historical linguistics*. Edinburgh University Press.

Blust, R. & Trussel, S. (ongoing). *The Austronesian Comparative Dictionary, web edition*. Accessed at https://www.trussel2.com/ACD/

Bybee, J., & Easterday, S. (2019). Consonant strengthening: A crosslinguistic survey and articulatory proposal. *Linguistic Typology, 23*(2), 263–302.

de La Fontinelle, J. (1976). *La langue de Houailou, Nouvelle-Calédonie: description phonologique et description syntaxique (Vol. 17)*. Paris: Peeters Publishers.

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (eds.). (2019). *Ethnologue: Languages of the World*. Twenty-third edition. Dallas: SIL International. Accessed at https://www.ethnologue.com/

Grace, G. W. (1975). *Canala dictionary (New Caledonia)*. Linguistic Circle of Canberra.

Grace, G. W. (1976). *Grand Couli dictionary (New Caledonia)*. Linguistic Circle of Canberra.

Grace, G. W. (1992). How Do Languages Change? (More on "Aberrant" Languages). *Oceanic Linguistics, 31*(1), 115–130. University of Hawai'i Press

Grace, G. W. (1996). Regularity of Change in What?. In M. Durie & M. Ross (eds.), *The comparative method reviewed: Regularity and irregularity in language change*, 157–179. Oxford University Press.

Skirgård, H. (2020). *Multilevel dynamics of language diversity in Oceania* [Doctoral dissertation, The Australian National University]. Open research library. https://openresearch-repository.anu.edu.au/handle/1885/218982

Leenhardt, M. (1935). *Vocabulaire et grammaire de la langue houaïlou*. Paris: Institut d'ethnologie.

Lichtenberk, F. (1978). A sketch of houaïlou grammar. *University of Hawaii Working Papers in Linguistics 10*. 73–116. Department of Linguistics, University of Hawai'i.

List, J. M. (2017). A web-based interactive tool for creating, inspecting, editing, and publishing etymological datasets. In *Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, 9–12.

Lynch, J. (1999). Southern Oceanic linguistic history. In J.-C. Galipaud & I. Lilley (eds.), *Le Pacifique de 5000 2000 avant le present / The Pacific from 5000 to 2000 BP*, 443–449. Paris: Institut de recherche pour le developpement.

Lynch, J. (2003). The bilabials in Proto-Loyalties. In J. Lynch (ed.), *Issues in Austronesian Historical Phonology*. 153–173. Canberra: Pacific Linguistics.

Lynch, J., & Ozanne-Rivierre, F. (2001). Some shared developments in pronouns in languages of Southern Oceania. *Oceanic linguistics, 40*(1), 33–66. University of Hawai'i Press.

Haudricourt, A. G. (1965). Problems of Austronesian comparative philology. *Lingua, 14*, 315–329. Amsterdam: North-Holland Publishing Co.

Haudricourt, A. G. (1971). New Caledonia and the Loyalty Islands. In J. D. Bowen (ed.), *Linguistics in Oceania, 1*, 359–396. Walter de Gruyter GmbH & Co KG.

Moyse-Faurie, C. & Néchérü-Jorédié, M. A. (1989). *Dictionnaire Xârâcùù-Français (Nouvelle-Calédonie) (Vol. 286)*. Paris: Peeters Publishers.

Moyse-Faurie, C. & Ozanne-Rivierre, F. (1983). Subject case markers and word order in New Caledonia and Loyalty Islands Languages. In A. Halim, L. Carrington, & S.A. Wurm (eds.), *Papers from the Third International Conference on Austronesian Linguistics: Volume 4 Thematic variation*, 113–152. Canberra: Pacific Linguistics C77.

Osumi, M. (1995). Tinrin grammar. *Oceanic Linguistics Special Publications,* (25), i–304. University of Hawai'i Press.

Ozanne-Rivierre, F. (1989). Le développement des voyelles nasales dans les langues du nord de la Nouvelle-Calédonie. *Revue d'Ethnolinguistique,* (4), 83-100.

Ozanne-Rivierre, F. (1992). The Proto-Oceanic Consonantal System and the Languages of New Caledonia. *Oceanic Linguistics, 31*(2), 191–207. University of Hawai'i Press.

Ozanne-Rivierre, F. (1995). Structural changes in the languages of northern New Caledonia. *Oceanic Linguistics, 34*(1), 44–72. University of Hawai'i Press.

Ozanne-Rivierre, F., & Rivierre, J. C. (2004). Évolution des formes canoniques dans les langues de Nouvelle-Calédonie. *Faits de Langues, 23*(1), 141–153. Brill.

Pawley, A. (2006). Nayacakalou Medal Address: Explaining the Aberrant Austronesian languages of Southeast Melanesia: 150 Years of Debate. *The Journal of the Polynesian Society, 115*(3), 215–258. The Polynesian Society.

Pawley, A., & Ross, M. (1995). Chapter 3. The prehistory of Oceanic languages: a current view. In P. Bellwood, J. Fox, & D. Tryon (eds.), *The Austronesians: historical and comparative perspectives*, 43–74. Canberra: ANU E Press.

Rivierre, J. C. (1993). Tonogenesis in New Caledonia. *Oceanic Linguistics Special Publications*, (24), 155–173. University of Hawai'i Press.

Rivierre, J. C., & Vandegou, C. (2020). *Dictionnaire numèè-français*. HAL. https://halshs.archives-ouvertes.fr/halshs-02916886

Rivierre, J. C., Ehrhart, S., & Diéla, R. (2006). *Le Bwatoo et les dialectes de la région de Koné (Nouvelle-Calédonie): introduction (Vol. 435)*. Paris: Peeters Publishers.

Ross, M. (1998). Proto-Oceanic Phonology and Morphology. In M. Ross, A. Pawley, & M. Osmond (eds.), *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society. Volume 2 The physical environment*. Canberra: Pacific Linguistics. Accessed at https://digitalcollections.anu.edu.au/handle/1885/106908

Ross, M., Pawley, A., & Osmond, M. (1998). *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society. Volume 1 Material culture*. Canberra: Pacific Linguistics. Accessed at https://digitalcollections.anu.edu.au/handle/1885/106908

Ross, M., Pawley, A., & Osmond, M. (2003). *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society. Volume 2 The physical environment*. Canberra: Pacific Linguistics. Accessed at https://digitalcollections.anu.edu.au/handle/1885/106908

Ross, M., Pawley, A., & Osmond, M. (2008). *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society. Volume 3 Plants*. Canberra: Pacific Linguistics. Accessed at https://digitalcollections.anu.edu.au/handle/1885/106908

Ross, M., Pawley, A., & Osmond, M. (2011). *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society. Volume 4 Animals*. Canberra: Pacific Linguistics. Accessed at https://digitalcollections.anu.edu.au/handle/1885/106908

Ross, M., Pawley, A., & Osmond, M. (2016). *The lexicon of Proto Oceanic. The culture and environment of ancestral Oceanic society. Volume 5 People: body and mind*. Canberra: Asia-Pacific Linguistics. Accessed at https://digitalcollections.anu.edu.au/handle/1885/106908

Sallabank, J. (2015). Language ideologies, practices and policies in Kanaky/New Caledonia. In M. Jones (ed.), *Policy and Planning for Endangered Languages*, 31–47. Cambridge: Cambridge University Press.

Sand, C. (1999). Lapita and non-Lapita ware during New Caledonia's first millennium of Austronesian settlement. In J.-C. Galipaud & I. Lilley (eds.), *Le Pacifique de 5000 2000 avant le present / The Pacific from 5000 to 2000 BP*, 139–159. Paris: Institut de recherche pour le developpement.

Spriggs, M. (1995). Chapter 6. The Lapita Culture and Austronesian Prehistory in Oceania. In P. Bellwood, J. Fox, & D. Tryon (eds.) *The Austronesians: historical and comparative perspectives*, 119–142. Canberra: ANU E Press.