Universiteit Leiden


Leiden University

Department of Humanities


# Personal Pronouns in Cuquila Mixtec


*Author:*

Zoi Gialitaki

*Supervisor:*

Martine Bruil


A thesis submitted for the degree of

*Master of Arts in Linguistics*

December 15, 2019

# Contents

**Acknowledgement** iv

**Glossary** v

**1 Introduction** vi

**2 Background Information** 1

2.1 The town and its speakers . . . 1

2.2 Language Use . . . 3

2.3 Language Attitudes . . . 4

**3 Previous Work on Mixtec** 7

3.1 Classification of Mixtec Languages . . . 7

3.2 Studies in Cuquila Mixtec . . . 11

3.3 Mixtec Grammars . . . 11

3.4 Mixtec Pronoun Studies . . . 12

**4 Methodology** 13

4.1 Speakers . . . 13

4.2 Data Gathering Strategies . . . 14

**5 Grammar Sketch of Cuquila Mixtec** 17

5.1 Phonology . . . 17

5.1.1 Vowel Inventory . . . 17

5.1.2 Consonants . . . 19

5.1.3 Loan Consonants . . . 20

5.1.4 Consonant Clusters . . . 21

5.1.5 Syllable Structure and the Couplet . . . 21

5.1.6 Glottal Stop . . . 22

5.1.7 Tone System . . . 23

5.1.8 Orthography . . . 25

5.2 Morphology . . . 26
5.2.1 Noun Morphology . . . 26
5.2.2 Nominal Modifiers . . . 31
5.2.3 Adverbs . . . 33
5.2.4 Verb Morphology . . . 34
5.3 Syntax . . . 37
5.3.1 Typological remarks . . . 37
5.3.2 Constituent Order . . . 37
5.3.3 Non-verbal predicates . . . 39
5.3.4 Coordination . . . 40
5.3.5 Subordination . . . 41

**6 Person Marking in Cuquila Mixtec** **43**
6.1 Overview . . . 43
6.2 Bound Clitics . . . 46
6.2.1 Clitics as Subject . . . 47
6.2.2 Clitics as Direct Object . . . 48
6.2.3 Clitics as Possessors . . . 48
6.2.4 Clitics as Objects of Prepositions . . . 49
6.3 Independent Forms . . . 49
6.3.1 Independent Forms in Comparative Constructions . . . 51
6.4 Plural Person Marking . . . 52
6.5 Pragmatically marked use of personal markers . . . 54
6.6 Semantic distinctions on the third person pronouns . . . 55
6.7 Lack of Person Markers . . . 59
6.8 Third Person Pronouns and Gender . . . 60

**7 Grammaticalization of Person Markers in Cuquila Mixtec** **63**
7.1 Person Marking in relation to Mixtec languages . . . 63
7.2 Theoretical Background . . . 65
7.3 Grammaticalization of Cuquila Mixtec Person Markers . . . 68
7.3.1 First and Second Person Pronouns . . . 68
7.3.2 Third Person Pronouns . . . 72

**8 Personal Pronouns In Use** **74**
8.1 Pronoun Usage and Politeness Distinctions . . . 74
8.1.1 Theoretical Background . . . 74
8.1.2 Interaction between Pronouns in Cuquila Mixtec . . . 76

8.1.3 Sociological Variables in Pronoun Use of Cuquila Mixtec . . . . . 78

**9 Conclusion** **82**

# Acknowledgements

First of all, I would like to thank Dagmar for inspiring me to go on this journey. His emotional support and cooking made the writing of this thesis possible. I would also like to thank Martine, whose helpful comments and empathy made this thesis readable.

I am deeply thankful to the community of Santa María Cuquila for accepting me, opening their world to me and teaching me their language. Living in the community for three months was an extraordinary adventure, but the attentiveness and generosity of the people made for an unforgettable experience. I especially want to thank Carolina and her whole family for always having a chair available for me on their table. Carolina's friendship was the best remedy against culture shock. Last but not least, I am most grateful to Lucero for offering her time and knowledge, for patiently answering all my questions, for always smiling and laughing.

# Glossary

| | |
|---|---|
| 1 | first person |
| 2 | second person |
| 3 | third person |
| ADD | additive |
| ADULT | adult |
| AFF | affirmative |
| ANML | animal |
| CAUS | causative |
| CHILD | child |
| COMP | complementizer |
| COMPL | completive |
| COP | copula |
| DEI | deity |
| DIST | distal |
| EXCL | exclusive |
| EXIST | existential |
| F | feminine |
| FAM | familiar |
| FLOW | flower |
| FOC | focus |
| GEN | generic |
| HUM | human |
| INCL | inclusive |
| INTSF | intensifier |
| ITER | iterative |
| LIM | limiter |
| LIQ | liquid |
| M | masculine |
| MX | mixed |
| NEG | negative |
| NMLZ | nominalizer |
| PL | plural |
| POT | potential |
| PRED | predicative |
| PROG | progressive |
| PROX | proximal |
| QUES | question marker |
| REFL | reflexive |
| RES | respect |
| SG | singular |
| SPEC | specifier |
| SPHER | spherical |
| TREE | tree |
| WATER | water |
| WOOD | wood |

# Chapter 1

# Introduction

The linguistic category of personal pronouns has been the subject of extensive research and debate for many years (Siewierska 2004 offers an overview of the discussion). Every aspect of the category, from the mere definition to the morphosyntactic properties to its semantic functions and cultural implications, has undergone many analyses in the past and continues to do so. Even though the personal pronouns, and person markers in general, are a ubiquitous category in all the languages of the world, scholars continuously try to understand them and classify them. However, definite answers seem elusive and even the fundamental question of "what constitutes a personal pronoun" are more complex than we originally thought. Perhaps it is precisely due to the universality of the person category that the our understanding of it is so problematic. The languages of the world exhibit great variation in their expression of person; the number of persons, the gender or politeness distinctions, the very lack of person marking, all act as variables in the matrix of possibilities where two languages can show variation in the person category. The study of previously undocumented languages further complicates these definitions, while at the same time enriches our understanding of how language works.

Specifically, looking into how the societal differences and hierarchies within a culture, or lack thereof, are expressed in the world's languages we can observe a considerable amount of variation. On a fundamental level, a division can be made between the languages that express this hierarchy using personal pronoun markers, like many European languages such as Spanish, Dutch and Greek, and the ones that use other terms of address to encode these differences, where kinship terms, titles, patronyms and other terms are used, like the English 'Mr/Mrs'. The distinctions branch out further within these groups. Focusing on the first group, we find languages that exhibit binary politeness distinctions towards the addressee, using what Brown & Gilman (1960) called the T/V pronouns, or threefold distinctions in languages like Portuguese or German which indicate more nuanced social parameters regarding the use of each pronoun. Furthermore, politeness can also be expressed on the part of the speaker, typically involving humbling forms which represent an inequality on the social roles of the

speech act participants. The complex pronoun systems found in languages of South Asia provide such examples and are often used to showcase the influence that the culture has in the linguistic expression. All of these cases, together with many other strategies, such as the Japanese pronoun avoidance, have been thoroughly documented, analysed and form part of large typological studies on personal pronouns in general (Cysouw 2003; Heine & Song 2011; Helmbrecht 2004; Siewierska 2004) and, more specifically, politeness (Agha 1994; Brown & Levinson 1987; Brown & Gilman 1960; Helmbrecht 2003).

The Mixtec languages have offered great insights in the study of personal pronouns. The person marking system exhibits characteristics that are typologically unusual and are not common in the geographical area where these languages belong. The noun categorisation mechanisms and their reflection on the pronoun system is one such feature. Additionally, the politeness distinctions that appear in all three persons, including a humbling form on the first person, is another trait that is not typically found outside of South Asia.

However, many aspects of the pronoun system have been understudied. What grammaticalisation path did the personal pronoun forms take? How did the respect forms emerge, and what are the social parameters that affect the use of the familiar and polite forms today? Through the analysis of a previously undocumented variety, Cuquila Mixtec, this study attempts to provide some answers to such questions. In order to understand the politeness distinctions in Cuquila Mixtec, many other aspects of the culture and the language are worth mentioning. Therefore, chapter 2 offers some insights on the speakers' way of life, as well as their attitudes towards their language. Chapter 3 discusses the linguistic classification of Cuquila Mixtec and its relation to other languages of the same family, so that a more complete picture of the background of this language can be formed. At the same time, it provides an overview of previous studies focused on personal pronouns that have been carried out in other Mixtec varieties, in an attempt to establish the relevant context for the present research. Chapter 4 presents the techniques used to gather the data that this research is based on, as well as some information on the language consultants. Following this, a brief grammar sketch is given in Chapter 5, which provides all the relevant information needed in order to better understand how the personal pronoun system is embedded in the language as a whole. The main chapters, where the pronoun system is analysed, come next. Chapter 6 deals with the person marking system as a whole, explaining its morphosyntactic, as well as semantic properties. Chapter 7 delves deeper into the grammaticalisation of these forms. Finally, having seen how the pronoun forms emerged in the past, chapter 8 deals with the social variables that are involved in the use of the familiar and the polite forms nowadays.

# Chapter 2

# Background Information

Cuquila Mixtec is a language spoken in south-central Mexico, in the town of Santa María Cuquila. The name 'Mixtec' is an Nahuatl exonym deriving from *miš* 'cloud' + *-teka* 'inhabitant of place of' (Campbell 1997: 402). However, many speakers prefer to refer to the language using the endonym *tu'un savi*, literally meaning 'the word of the rain'. Mixtec is spoken by approximately 490.000 people in Mexico according to the 2010 national census (Instituto Nacional de Estadística 2011). However, large communities have been established in USA, especially in California, exhibiting a continuous flow of migration since the 1970's (Kresge 2007). Within Mexico, the speakers of Mixtec are concentrated in the region known as 'La Mixteca', which spans across three states in south-central Mexico: Oaxaca, Puebla and Guerrero. The region is characterised by diverse geographical contrasts and climates. There are three main areas. The highlands (Mixteca Alta), located in Oaxaca and Guerrero, have high mountain ranges reaching up to 3.000 meters in altitude with relatively cold, dry winters and rainy summers. The lowlands (Mixteca Baja) of Oaxaca and Puebla are located at approximately 1500 meters altitude, while the coastal region (Mixteca de la costa) offers a warm, tropical climate along the coast of Oaxaca. The regions are visualised in figure 2.1.

## 2.1 The town and its speakers

The town of Santa María Cuquila is located in western Oaxaca. The town belongs to the municipality of Tlaxiaco, a major town of approximately 17.000 people located 25 km. away from the nucleus Santa María Cuquila. It belongs to the region of Mixteca Alta, sitting at approximately 2100 meters of altitude. The Mixtec name of the community is *ñuu kuiñi*, meaning 'the land of the tiger'. There are approximately 10.000 inhabitants in the community, spread over the town centre and seven hamlets (Ruiz Medrano 2015: 125). The nucleus of the town has 596 inhabitants, according to the 2010 census (Instituto Nacional de Estadística 2011).

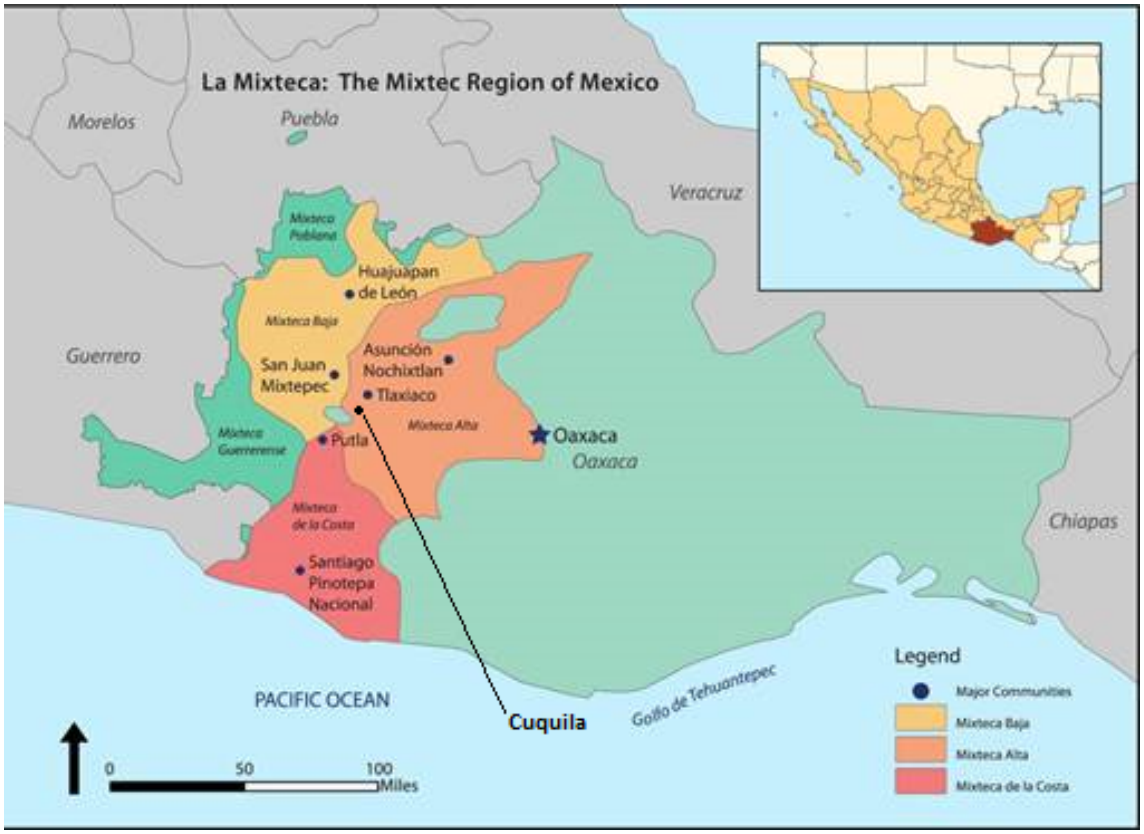


**Figure 2.1:** The region of La Mixteca (Rieger 2019, edited to show the location of Cuquila)

The community's main activities consist of agriculture and livestock farming. The cultivation of corn and beans is an important part of the locals' life around which many traditions are centered, as everyone has a piece of land on which they grow corn. In addition, some inhabitants have sheep, used for their wool and meat, as well as bulls which are used in the fields. Traditionally, Santa María Cuquila was known for the high quality ceramic utensils that the community produced. Many skilled *alfareros* (clay pot makers) learned the trade from their parents and would make ceramic bowls, mugs and vases, which would then be exchanged for food and other goods at the market in the town of Pinotepa Nacional, close to the coast. In the last years, the craft has experienced a steep decline; nowadays only a few people are left who know how to work the clay of the area. Similarly, the traditional art of textile making is slowly disappearing. As in every community of la Mixteca, Santa María Cuquila has its own, characteristic *huipil*, a woollen tunic-like dress that many people still wear on special occasions. The younger generation does not learn the art of the backstrap loom anymore, as it is perceived not to be a financially feasible means of substinence.

The way of life is communal; decisions that affect everyone are made in meetings where the head of every household is present. These meetings occur in average once a month in order to discuss important matters and decide on future actions. The town centre and the seven hamlets each have their own body of authorities. Approximately 20 people are elected every

year to take different positions in the authorities, which are in charge of implementing the decisions made during the general assemblies and include bodies such as an education council and a small police force. All the adult residents of the community are required to take part in the authorities. Even though these positions are not remunerated and require a considerable amount of financial and time investment, most of the people regard it as their ethical duty to serve the community work for the collective benefit. Furthermore, very few people privately own land, as it is owned by the community as a whole. As mentioned previously, most of the people have a plot of land, however they are not the legal owners of it. Communal ownership of the land is a common practice among the Mixtecans, as it is found in 86,5% of the communities in the Mixteca Alta region (Balderas n.d.).

## 2.2 Language Use

It is difficult to assess the vitality of Cuquila Mixtec. The language does not appear in the typical language status lists, such as Ethnologue (2012), the Endangered Languages Project (2019) or the UNESCO Atlas of the World´s Languages in Danger (Moseley 2010). The Endangered Languages Project lists most other varieties of Mixtec as "Vulnerable" or "Threatened", but no mention is made specifically on Cuquila Mixtec. Thus, the information that I will provide below derives from the conversations I had with the locals during my fieldwork, as well as my personal observations during my time in the community.

Most of the speakers are bilingual in Mixtec and Spanish, with varying degrees of proficiency of both languages. The majority of the people over 70 years old are mostly monolingual in Mixtec. The following generations, speakers of approximately 40-70 years old are bilingual with Spanish and equally fluent in both languages. In many cases, the way they learned Spanish was through migration to the big cities. Due to high levels of unemployment in the community, many of these people had to migrate to Mexico City or to Oaxaca to find work. Having stayed there for a number of years, they learned Spanish in order to establish themselves in the local society. Upon their return to the community, they retained a high fluency degree in both languages. The children and grandchildren of these generations often were not taught Mixtec. They attended an all-Spanish school and spoke Spanish to their parents. Nonetheless, many of them learned Mixtec through their interactions with their monolingual grandparents. Their children, today's adolescents, barely speak any Mixtec. Some of them have passive knowledge of the language, but their daily interactions rarely include Mixtec.

Language use at the hamlets of Santa María Cuquila is somewhat different from the centre. There, children are more acquainted with Mixtec and learn it from a young age, as it is frequently used at home and in the daily interactions. The reason for it might be

geographical; while many of the hamlets are located in relatively remote areas in the mountains, the centre of Cuquila sits on the main highway leading to the coast and the access to Tlaxiaco is very easy from there. Tlaxiaco, which once had its own variant of Mixtec but is now extinct, might have a strong influence on the population surrounding it, as it is the administration and commercial centre of the area. Perhaps the ease of access to an urban centre discourages the speakers from retaining their language and, as a consequence, changes their attitude towards it. It is a well known fact that urbanisation is a major factor in language shift; as pgcitetgrenoble35 write: "[...] the more isolated a community, the more likely it is to maintain use of the local language. Urbanization has the opposite effect: by bringing people into contact, it facilitates language shift [...]".

Distinct domains of Mixtec and Spanish language use can be observed. In events that require the attendance of many people, such as festivities or general assemblies, the announcements and speeches are done in both languages; if the speaker uses Spanish, then an interpreter translates the speech in Mixtec, and vice versa. Lastly, religious ceremonies in the catholic church always occur in Spanish, but traditional rituals, such as ones which involve asking the earth for permission to sow crops or pleading for rain, mainly happen in Mixtec (even though this might be because, usually, these rituals are performed by the elders.) If an elder is present, even if he/she is not directly addressed, the speakers predominantly use Mixtec to communicate. When children are either directly addressed or they are part of a conversation, Spanish is almost always used. In a setting of purely bilingual speakers, the language of choice is always Spanish, regardless of the environment.

In terms of education, the centre of Santa María Cuquila has a bilingual primary school. For children aged 5 to 11 a Mixtec language course is included in the curriculum, alongside the rest Spanish-taught subjects. In order to facilitate the learning process, the government has developed educational material in Mixtec. Every year, the students are provided with a storybook and an exercise book. Several universities in Mexico offer specialised programs where teachers receive training in bilingual education.

## 2.3 Language Attitudes

The chances of survival of a language or a speech variety largely depend on the speaker's attitudes towards it, as explained in Grenoble (2011). The negative view of a language can lead to its abandonment and a shift towards another dominant language. Usually, such negative attitudes are the result of years of suppression, marginalisation and stigmatisation, among other factors. This is certainly the case for the speakers of the Mixtec languages, as well as many other indigenous languages of Mexico. The effects of the Spanish colonisation are deeply rooted within the Mixtec consciousness, but more recent cases of racism and

stigmatisation have also had major influence on the language shift towards Spanish. Patronising and neo-colonial attempts by the government at 'modernising' the way of life of the indigenous people in the Mixteca region throughout the 1950s until the 1970s included, among other initiatives, the promotion of all-Spanish education (Nagengast & Kearney 1990). Centuries of such suppression, racism and vast economic inequality has resulted in the adoption of the belief that the Mixtec identity and language is inferior and useless by the Mixtecans themselves. Indeed, it has been reported that every year 200 speakers of Mixtec variants abandon their language (Caballero Morales 2013: 2).

The effects of these attitudes are tangible in Santa María Cuquila. For many years, parents have chosen not to transmit their children Cuquila Mixtec, as they view it as an inferior language. This attitude is the result of many factors: at school, they were punished for speaking Mixtec and during their professional life, it was impossible to find a job outside of the community if they didn't speak Spanish. For them, Mixtec does not offer any pragmatic value; it will not help them find a job and escape poverty. Instead, they encourage the children to learn English, as it will provide more opportunities for the future and they will not have to experience the same difficulties as their parents.

However, throughout my conversations with the community, there was another salient reason for not teaching Mixtec to the children. Apart from the inferior status that the language has, many speakers alluded to the fact that Mixtec was a very difficult language to learn. According to the speakers, there were mainly three factors which made the language difficult to learn and to teach in a formal setting: the tone system, the orthography and the grammar. They often referred to the tone system as a being almost impossible to assimilate, providing examples of tone word pairs to show that they were both strenuous to explain and to comprehend. At the same time, many speakers believe that the current orthography does not correctly reflect the way Mixtec is spoken. I was often provided with examples of words from the school textbooks which, according to the speakers, were confusing to read."If you try to write down what you speak out loud, it won't make sense when you read it back", one speaker told me. Additionally, the prevailing idea about Mixtec is that it had very complex grammar rules, with plenty of exceptions and puzzling principles, which they believe are beyond their capacities to teach to their children so that they can speak 'proper' Mixtec, as they reported. These reasons result in a 'why bother' attitude which, coupled with the low social status of the language, creates an environment where Mixtec is not being passed down to the next generations.

But where does this conviction that Mixtec is so difficult stem from? Mandarin has a complex tone system, and the English orthography does not reflect the spoken language any more, but the speakers of these languages do not use these factors as reasons not to pass on their language to their children. I believe that this attitude in Santa María Cuquila is the

result of the way the bilingual educational system is structured, which results in the perception that learning to speak Mixtec is a perplexing task. Teachers of Mixtec are required to do several rotations in different communities of the Mixteca Alta before they can work in their own community. This means that the teachers that arrive at Santa María Cuquila often speak a different variant of Mixtec than the one spoken in the community. Thus, there are discrepancies between the language that the children are taught at school and the language that they hear at home. The difficulties intensify through the use of the educational material. As mentioned previously, children are taught Mixtec with the help of a storybook that serves the whole Mixteca Alta region. Even though the variants spoken in the area are mutually intelligible, they exhibit great variation in phonology, lexicon and grammar. In order to create a generalised textbook for all these variants, many compromises needed to be made. Even though information about the development of these textbooks is scarce, it seems that a combination of several Mixtec variants is used in the stories. Looking into one of the books with some speakers, it seemed that they could understand the texts, but they would often times provide different pronunciation for some words, or entirely distinct words for some concepts. The use of different variants in the texts might be the reason behind their conviction that the current orthography does not reflect the spoken language. Furthermore, the tones are not transcribed, thus the speakers never learn to how they are represented in text, which makes these books even more difficult to read. Lastly, the students are not taught the grammar rules of Mixtec, instead learning the language through reading stories. Even though educational grammars written in Spanish have been published for several variants (cf. Hills 1990; Hollenbach 2013), it seems that they do not make their way into the educational system. All these factors impede the process of rendering the implicit knowledge explicit and reinforce the idea that learning their language is not worth the effort.

However, these attitudes seem to be changing. In the last years, efforts have been made in the community to salvage their traditions and their language. The organisation Yuku Savi ('Mountain of the Rain'), which was established by members of the community, promotes the local traditions through activities such as the planting of traditional seeds and trees and the re-establishment of the weekly marketplace where the exchange of goods is strongly encouraged. They also plan on offering Mixtec classes and creating educational material in Cuquila Mixtec. Additionally, Some young parents, between 20 and 30 years old, have decided to facilitate their children in learning Mixtec. They view their language as being part of their identity, and so they realise that they cannot afford to lose it. They now speak Mixtec to their children and try to have a closer contact with the elders, so that they can grow up in a Mixtec-speaking environment.

# Chapter 3

# Previous Work on Mixtec

In order to place this research in the context of the study of person markers, it is important to be aware of the body of knowledge that have been gathered so far on this subject within the Mixtec family. Additionally, we need to understand this variety's position within the Mixtec language family and the complex interactions of all the different varieties within, so that a more general linguistic setting can be revealed.

As mentioned in the before, Cuquila Mixtec is a previously undocumented variety of Mixtec. As such, purely linguistic works have not been published in the past. There is, however, some material from other disciplines that include information on Cuquila Mixtec, which I will present below. Furthermore, previous studies in other Mixtec varieties prove helpful in examining the structures of the variant in question, as many of them show a great degree of overlap due to the common historical background. In this chapter I will mainly present works that have been published in the past regarding the personal pronoun system of Mixtec, as this is the focus of this thesis. The chapter is organised as follows:

I will first discuss the position of the Mixtec languages within the Otomanguagean family, as well as the internal classification in section 3.1. Then, the works published in Cuquila Mixtec will be summarised in section 3.2. Section 3.3 will provide an overview of the grammars published in some Mixtec varieties organised according to the audience they serve. In section 3.4 I will discuss the works that focus specifically on the study on the personal pronouns.

## 3.1 Classification of Mixtec Languages

Linguistically, Cuquila Mixtec belongs to the Otomanguean language family, which is very diverse and relatively understudied, in relation to other Mesoamerican language families (Campbell 2017). Due to this diversity, but also because of the great internal variation, it has been difficult to say with precision how many languages belong to this family. As Campbell

(2017) explains, some groupings consist of only one language (for example, Ixcatec), while others include several variants, for many of which there is no consensus on whether they should be considered languages or dialects (for example, the variants of Zapotec and Mixtec). Diagram 3.1 presents the classification that Kaufman (1988) proposes for the Otomanguean family:

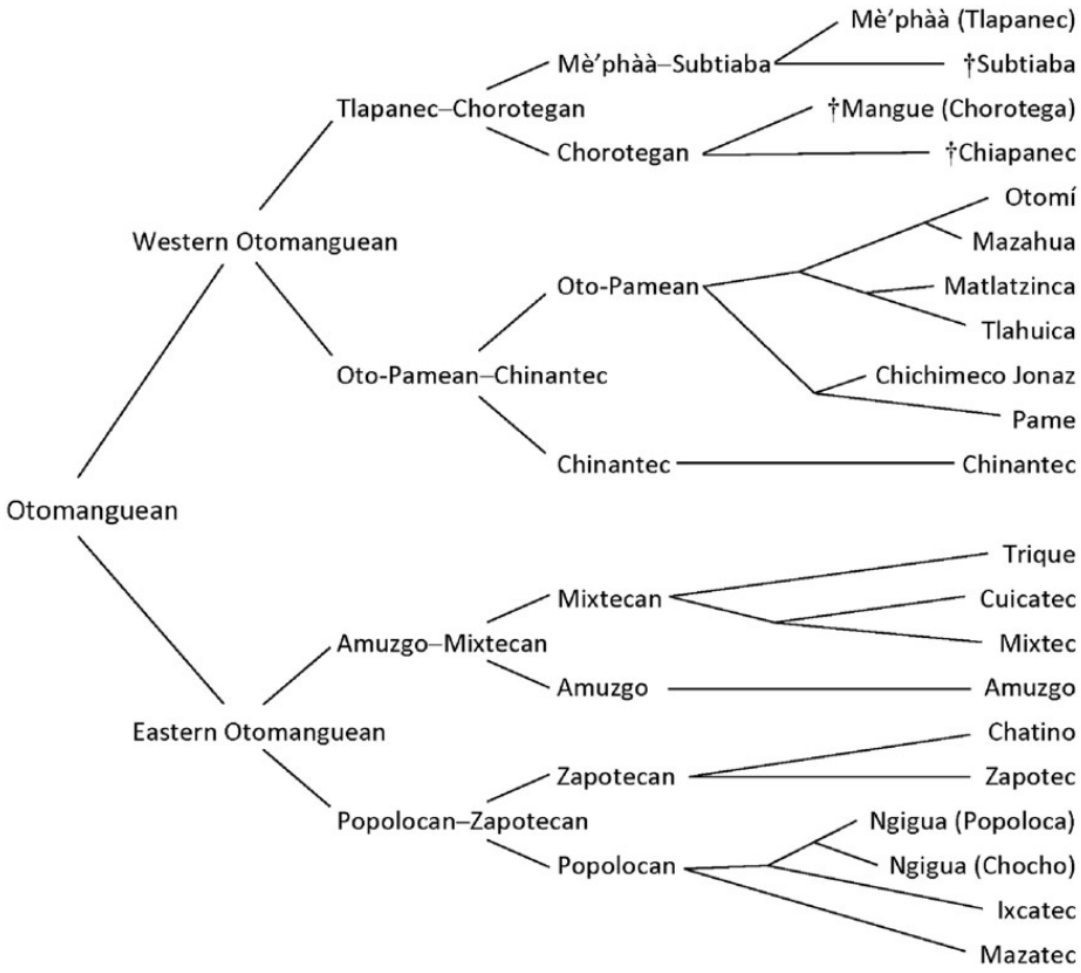


**Figure 3.1:** Otomanguean Classification per Kaufman (1988) (cited in Campbell (2017))

Within Mixtec, questions such as the amount of languages that are included in this branch of the Otomanguean family, and whether Mixtec should be considered one sole language with many dialects or a language family have been the subject of debates for many years. Some researchers (Caballero Morales 2013; Leon Pasquel 1988; Macaulay 1997; Perry 2017) refer to the 'Mixtec language' as a whole [1], while many others (among others, Lastra 1992; Marlett 1992) talk about different Mixtec languages. Even within the group of researchers who recognise that Mixtec is a language family, consensus is still not reached on

[1] even though Macaulay (1997) acknowledges the problem of internal variation, she prefers to use the term 'dialect' to refer to the variants

the amount of languages that it constitutes; the number of Mixtec languages often cited varies from 29 (Suárez 1983) to 51 (Eberhard, Simons & Fenning 2012). This difficulty lies in the fact that Mixtec comprises of complex dialect areas, where the variation is at times too gradual to pinpoint where one variant stops and the other one begins. Geographical distance is not a good indicator of language separation, either. Through a process which Macaulay (1997: 7) calls "the leapfrogging nature of Mixtec territorial expansion", there are cases of variants that are mutually intelligible but geographically distant and vice versa [2]. Recognising this complex system of dialect continua, many researchers have tried to provide an internal classification of the Mixtec variants using different methods and sample sizes. Holland (1959) (cited in Josserand (1983: 134)) compared 22 towns in a glottochronological study, which she then separated into different languages based on the benchmark that Swadesh (1956) established of 86% shared cognates. In an extensive study carried out by SIL linguists in the '60s and again in the '70s, 84 Mixtec towns were surveyed and language groupings were established based on the mutual intelligibility level, which was set to at least 70%. Lastly, Josserand's (1983) dialect study is, to my knowledge, the most complete one, drawing information and cognate sets from 120 Mixtec towns. Based on phonological, morphological, syntactical and lexical variations, she distinguishes five major dialect areas with many sub-groupings within. All of these studies resulted in dialect maps of the Mixtec language family which exhibit some overlap, but are still different enough to hinder any consensus. The map shown in figure 3.2 shows the dialect continua drawn by Josserand (1983), whose analysis I have chosen to follow and often refer back to in this thesis. Note that Cuquila Mixtec does not appear in this map, but it would most probably belong to the Western Alta area. Seeing that Cuquila Mixtec has not been previously documented, it is difficult to say with certainty which dialect area it belongs to. Further research is needed in order to confirm the hypothesis that it belongs to the Western Alta area. However, some information seems to point out this connection. Firstly, Santa María Cuquila geographically belongs to the Western Alta area as is delimited by Josserand (1983). Furthermore, data from the fieldwork suggest strong grammatical and lexical similarities between Cuquila Mixtec, Ocotepec Mixtec and Magdalena Peñasco Mixtec. Additionally, many Cuquila Mixtec speakers reported a higher level of intelligibility with these variants than the ones that belong to other dialect areas.

[2] Josserand (1983) also mentions many cases of linguistically isolated Mixtec towns surrounded by speakers of completely distinct languages such as Mazatec or Nahua

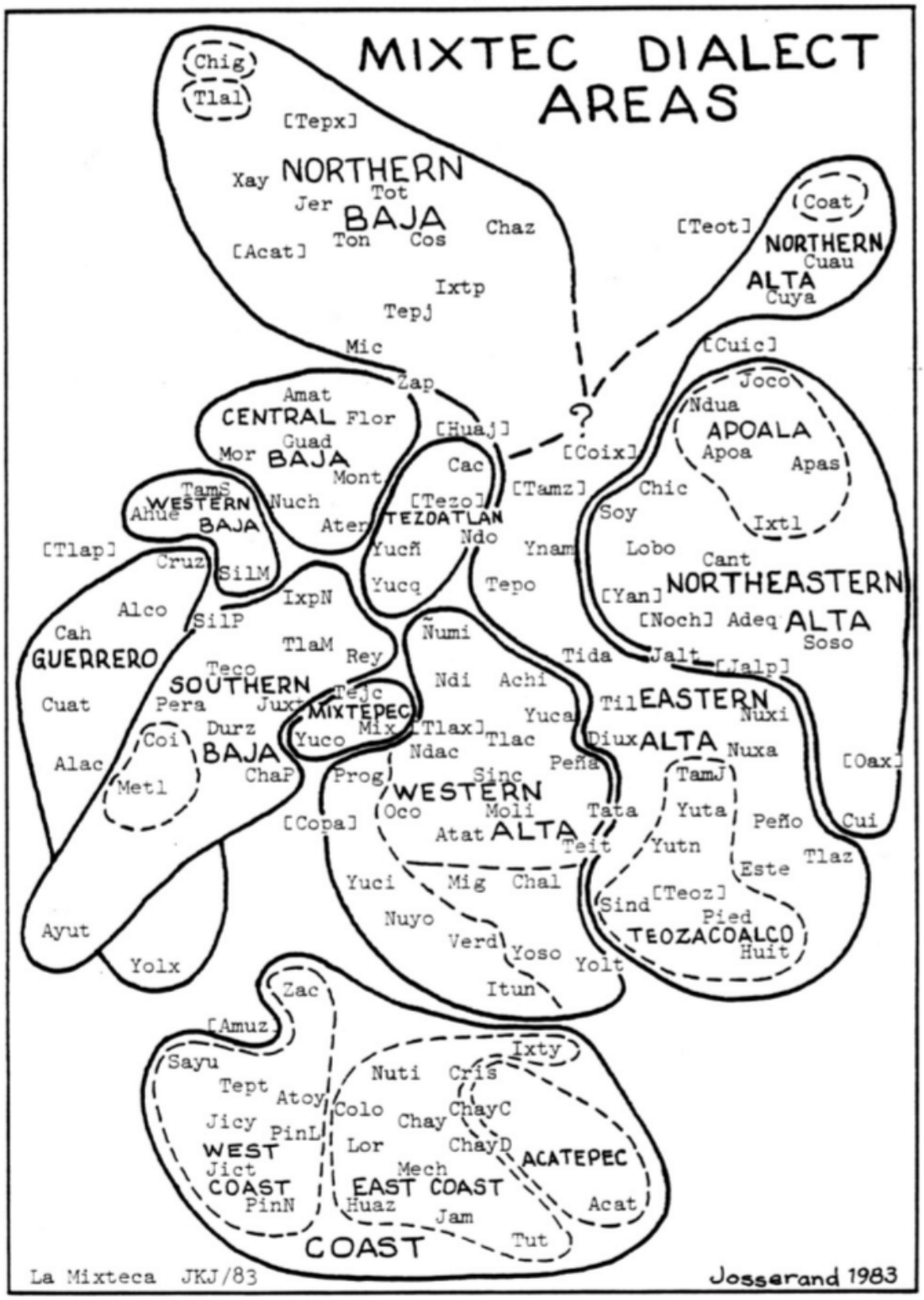


**Figure 3.2:** Mixtec dialect areas and continua (Josserand 1983: 470)

## 3.2 Studies in Cuquila Mixtec

The only work published in the variety of Cuquila Mixtec is a pedagogical study. Rojas Lopez's dissertation (2015) is a pedagogical proposal regarding the teaching of Cuquila Mixtec numerals to students in primary schools. In this study, the author provides a brief overview of the state of the language as well as the problematic of indigenous language teaching in Mexico. Furthermore, he explains how the numeral system works in Mixtec, and presents a prescriptive proposal which includes various practical exercises that teachers can use in order to teach simple mathematics in Mixtec.

## 3.3 Mixtec Grammars

Some of the earliest descriptive grammars carried out in Mixtec were published during the 1970's and 1980's by SIL, as part of a series on Mixtecan studies. The series comprises of short grammar sketches, mainly focused on syntax, which all follow the same structure. For many varieties of Mixtec, these grammar sketches are the only or most complete information that has been produced. Examples include Alexander's (1988) grammar sketch of Ocotepec Mixtec, and Farris' (1992) sketch of Yosondúa Mixtec. The benefit of these works is that they facilitate the comparison of several parts of grammar and syntax across varieties because they follow the same format. However, due to their size, the information included is usually incomplete and there is little argumentation regarding the analysis. The description of the pronoun system is usually no more than two pages long (cf. Alexander (1988: 263-265)) and is limited to providing the grammatical forms along with some information on the syntactic positions they can take, but further analysis is lacking. Apart from the section on pronouns, they usually include some information on the classification of the nouns into gender categories, again providing some lists with a few examples but without delving deeper into the particulars of these distinctions.

More recent works published in the 1980's until the 2010's by the SIL branch in Mexico include several grammars intended to be used by the language speakers themselves (Ferguson de Williams 2006; Gittlen 2016; Hollenbach 2013; Towne 2011; Zylstra 2012). These works are all written in Spanish, and Alexander's *Gramática Mixteca de Atatlahuca* (1980) is written in both Spanish and Mixtec. These grammars tend to focus more on the forms and use of the pronouns, but face the same argumentation issues that the previous grammar sketches had. However, these works clearly base their analysis on the Spanish language in terms of the grammatical categories used, which often do not correspond to the ones found in Mixtec. [3].

Apart from SIL, Macaulay (1997) has produced a detailed reference grammar of

[3] This analysis from an etic perspective is reflected, for example, on the fact that the explanation of the multiple third person pronouns that occur in Magdalena Peñasco Mixtec is based on the masculine/feminine gender distinction in Spanish (Hollenbach 2013: 51)

Chacaltongo Mixtec. This grammar is clearly intended for linguists and follows a generative approach. After providing a description of the grammatical forms of the pronouns, the author goes into detail on some of their uses and their syntactic constrains. Macaulay contributes to the study of the pronouns by providing extensive argumentation for her analyses. This is also possibly the reason why many typological and more general works on pronouns that include information on Mixtec have referenced her grammar (cf. Helmbrecht (2004), Corbett (2013)).

## 3.4 Mixtec Pronoun Studies

Elena Hollenbach has written various works which focus on the personal pronoun system of Mixtec. Her paper on the honorific pronouns (Hollenbach 2003) is mainly a comparative work of various different pronoun systems in Mixtec, which also provides some basic information on the historical development of these systems. Her later work on the reconstruction of the Mixtec pronouns (Hollenbach 2015a) analyses further the grammaticalisation of the pronoun forms, attempting to reconstruct them in Proto-Mixtec. To my knowledge, this is the only work published regarding the reconstruction of the Mixtec pronouns. In the same year, she published the *Mixtec pronoun database* (2015), a collection of the pronoun paradigms in all the Mixtec varieties documented until then. This is a salient work which facilitates the comparison of the pronoun forms across varieties. However, the terminology and glosses used are rather cryptic at times, and so the reader has to refer back to other works of her to better understand them.

As this chapter has shown, more research needs to be carried out in Mixtec varieties from an emic perspective. The work done on the personal pronoun system of the Mixtec languages is still rather limited and, specifically, no previous linguistic work has been done in Cuquila Mixtec. Before I delve deeper into the analysis of this variety, I will provide some information on the methodologies used during this study.

# Chapter 4

# Methodology

In order to ensure the transparency of the data upon which this research is based, this chapter contains information on the methodologies used during the fieldwork. Section 4.1 provides an overview of the speakers of Cuquila Mixtec that I collaborated with, giving some basic sociolinguistic information. Section 4.2 focuses on the strategies followed during the data-gathering process, as well as some information on the transcription and the translation process.

## 4.1 Speakers

In the course of this fieldwork I worked with a total of 10 speakers. Going through the grammars of other Mixtec varieties, I noticed that the majority of language consultants were male. Due to this over-representation of male speech, I made the decision to include slightly more female speakers for this study. Additionally, as a female researcher it was easier to approach and interview the women of the community. This resulted in three male speaker and seven female. All of them are bilingual in Mixtec and Spanish, many of them, however, learned Spanish during adolescence when they migrated to Mexico City or other big cities in search of work. Until then, they would only speak Mixtec. Most of the speakers are between the ages of 35 and 70, while two consultants are in their 20s. Even though I tried to have a broader age variation, it proved difficult to find younger speakers who were fluent in Mixtec. Lucero Ilario provided the vast majority of translations. Table 4.1 provides an overview of all the consultants.

**Table 4.1: Information on the speakers of this study**

| Name | Gender | Age | Place of Origin |
|---|---|---|---|
| Margarita Morales Ortiz | F | 49 | Santa María Cuquila |
| Carolina Ilario Morales | F | 27 | Santa María Cuquila |
| Lucero Ilario | F | 21 | San Juan Escutia, Cuquila |
| Juan Sebastian Rojas | M | 39 | Santa María Cuquila |
| Margarita Maria Sanchez Coronel | F | approx. 60 | Santa María Cuquila |
| Ermiña Carmen Lopez Ortiz | F | approx. 70 | Santa María Cuquila |
| Dominga Atila Vejiá | F | approx. 65 | Santa María Cuquila |
| Eleuteria Santiago Aguilar | F | approx. 45 | Santa María Cuquila |
| Fermin Cruz Ilario | M | approx. 35 | Santa María Cuquila |
| Abraam Santiago Lopez | M | approx. 40 | Santa María Cuquila |

## 4.2 Data Gathering Strategies

This thesis is based on data that I collected during a 3-month fieldwork trip to Santa María Cuquila, Oaxaca, conducted from June until September, 2019. The data gathered consist of a variety of genres, using several different elicitation strategies. Specifically, the data consist of: 1) conversations, both guided and spontaneous, 2) prompted monologues, 3) songs and 4) elicited sentences and other elicitation tasks.

Most of the conversations were guided, meaning that I would agree on a speaking prompt together with the consultants. I would usually provide them with two or three different subjects to choose from, and they would then pick the one they felt the most comfortable to talk about. Some examples of subjects were: "How is the rainy season going so far?", "What do you think of the re-established market?", or "What did you do yesterday?". I would also ask them to keep the conversations under 3 minutes, because I wanted to transcribe all the recordings while I was still in the community. The data also contains a 13-minute recording of a spontaneous conversation between two women who are working side-to-side. I was given permission to record this conversation, in which they discuss various subjects such as their plans for the next days, the Mixtec language, sickness etc.

The prompted monologues followed the same style as the guided conversations; I would provide the speaker with two or three subjects to choose from, and ask them to talk about it for two to three minutes. However, I also gave the speakers the freedom to choose their own subject if they felt more comfortable. The data collected from these monologues include recipes, local traditions and celebrations and personal stories. Some examples of speaking prompts are: "How is the election system of the local authorities organised?", "How do you take care of your animals?" and "What advice would you give to the adolescents of the community?"

The elicitation strategies used included the translation from Spanish into Cuquila Mixtec and vice versa, substitution tasks, felicity judgements and a picture task. The translation methods mainly involved phrases such as 'How would you say *Those trees have dried up* in Mixtec?' or 'What does the phrase *teéya ñu'ude xini* mean?' In some cases, such questions brought about issues, as some of the speakers who were very eager to teach me the language would either give me a simplified translation in Mixtec or they would explain an aspect of the language that was different from the one I wanted to focus on at that moment. To overcome these issues, later on during the fieldwork I realised that posing a question such as 'How would you answer *a chindeesani?*' or 'What would you say if I told you *luli ñukuã ũũ kuiyai*?' would focus their responses to the data that I wanted to elicit. This type of questions create a more realistic scenario, since the speakers are not asked to analyse a sentence, perhaps using metalanguage that they don't possess, but instead they answer as if simply continuing a conversation. I used similar techniques for the substitution tasks, most of which would come about spontaneously during the transcription and translation of monolingual material with a consultant. Whenever I found an interesting structure, I would change various aspects of it (for example word order, person marker or verbal aspect) and ask the consultant what the difference between the two forms are, or in which scenario they would use each one. The picture task was spontaneously developed by me during the fieldwork, when the need arose to investigate contrastive focus and comparative clauses. By then, I was aware that a simple translation of sentences would not work because of the reasons mentioned above, thus I developed the following task: I asked a consultant to look at different photos of herself, myself and someone else and create sentences based on similarities and differences she would find in the photos (for example, 'I am sitting and you are standing', or 'We both have long hair'). This task, coupled with substitution tasks based on the sentences that were produced, proved very fruitful, possibly because there was a great degree of freedom involved for the consultant to be creative and come up with her own sentences.

I tried to keep the elicitation of sentences to a minimum, only using this method when there were gaps in the data that was gathered during conversations and in order to better understand certain structures. Regarding the sentences that needed to be translated, I made a conscious effort to create phrases that would fit in the culture of the community, referencing local traditions and customs.

The transcription and translation took place with the aid of several speakers. Carolina Ilario Morales helped with many recordings, but most of them were transcribed and translated with the help of Lucero Ilario. Lucero was a fluent bilingual speaker who knew how to write in Mixtec and was aware of plenty of the metalanguage, which accelerated the process.

In our meetings, I would play some seconds of the recording and I would ask her to repeat what she heard, while I was writing it down. She would often check my text to make sure I

was writing it correctly. At times, I would ask her to make transcriptions herself, in order to better understand the word barriers. This process proved helpful in this aspect, but the lack of transcription of tones was a drawback.

After the transcription was finished, we would translate the text sentence-by-sentence. Lucero would provide plenty of details about the sentence, and not just a generic translation. I would frequently ask her questions about the meaning of specific words, which she would very often provide, along with other information such as minimal pairs and synonyms. When she was not able to provide a specific meaning, I would use some of the elicitation techniques explained above to ask for different contexts in which the word appears. Furthermore, I would often elicit alternative versions of the sentence would would focus on, in order to better understand certain details in morphosyntax or to fill in paradigm gaps.

The fieldwork resulted in over 22 hours of recordings, with approximately 71 minutes of monolingual data (monologues and conversations), 2 hours of elicitations and 17 hours of translations, analyses and transcriptions.

The examples used in this thesis are all extracted from the corpus obtained during the fieldwork. Since they usually form part of a longer discourse, I have tried to make sure that the sentences chosen can stand on their own and I provide further context whenever is needed. The few elicited sentences that have been used as examples are clearly marked as such.

Having established the research methods used in this study, in the following chapter I will provide some basic linguistic information on Cuquila Mixtec.

# Chapter 5

# Grammar Sketch of Cuquila Mixtec

In order to better understand how the personal markers and the personal pronoun system work in Cuquila Mixtec, a short grammar sketch is in order. Due to the scope of this thesis and the space constraints, not all grammatical structures of Cuquila Mixtec will be discussed. Instead, the basic language structures and the parts of the language that are relevant for the analysis of personal markers will be explained. At times, analyses of other varieties of the Mixteca Alta will be discussed, in order to provide some relevant context for Cuquila Mixtec.

Section 5.1 provides a phonological sketch with information regarding the consonant and vowel inventories, the prosodic features and the syllable structure, as well as a note on the orthography systems proposed and the one used in this thesis. Section 5.2 shows the relevant morphological structures of the noun phrase, the verb phrase and other parts of speech. Lastly, section 5.3 analyses some aspects of the syntax with respect to the word orders found in Cuquila Mixtec, the non-verbal predicates and the clause-linking patterns.

## 5.1 Phonology

### 5.1.1 Vowel Inventory

Cuquila Mixtec exhibits a symmetric system of 5 oral and 5 nasal vowels. Table 5.1 gives an overview of the inventory. The vowels contrast in height and position; high, mid, low and front, back, respectively. Long vowels are present in the language, however, they do not form part of the vowel inventory but instead are double vowels.

**Table 5.1: Vowel Inventory**

| | Oral | | Nasal | |
|---|---|---|---|---|
| | Front | Back | Front | Back |
| High | i | u | ĩ | ũ |
| Mid | e | o | ẽ | õ |
| Low | a | | ã | |

The oral vowels can occur in any position of the word. When they are followed by a nasal stop, the oral vowels acquire a slight nasalisation. Below are some examples of vowels in different positions of the word:

(1) [ān**ū**] 'heart' [nd**ū**tè] 'water'
[k**ā**tʃ**ī**] 'say' [jík**ā**] 'ask'
[**ō**k**ō**] 'twenty' [**ī**tʃ**ī**] 'road'

The nasal vowels do not appear with the same frequency as the oral vowels. There are only a few instances of nasal vowels in word-initial position, but they are more frequently found word-medially and word-finally. Due to the slight nasalisation that oral vowels acquire when followed by a nasal stop, the oral/nasal contrast may be neutralised before a nasal stop.

(2) [tũʔũ] 'paper' [ʃ̃ũʔũ̃] 'money'
[ñúkuà̰] 'there (close to speaker)' [ʃtéẽ̄] 'to teach'

However, the existence of minimal pairs contrasting oral and nasal vowels shows that the latter are, indeed, distinct phonemes. Below are some examples of minimal pairs:

(3) [kū̃ū̃] 'to fall' [kūū] 'to be'
[ī̃ī̃] 'one' [íí] 'nine'
[kʷá'ã̄] 'to go' [kʷāà] 'to strech out'

Vowels can also be doubled, which results in the utterance of a lengthened vowel. As mentioned previously, vowel length is a distinctive feature in Cuquila Mixtec, as shown in the following minimal pairs:

(4) [nā] ‘any’ [náā] ‘yes’
[ndē] ‘until’ [ndēē] ‘to sit’
[ʒū] ‘side, edge’ [ʒúʔú] ‘fear’

### 5.1.2 Consonants

The consonant inventory of Cuquila Mixtec contains 16 consonants, as shown in Table 5.2. Interestingly, there is no voiced bilabial plosive [p] inherited from Proto-Mixtec, so this phoneme does not appear in Mixtec words. It is important to note that it is not always easy to determine the underlying form of a phoneme. In what follows, I will only discuss the phonemes that are realised with several surface forms and may therefore present issues in the analysis of the inventory, due to space constraints. It can be assumed that the remaining phonemes show no environment restrictions and always maintain their underlying forms.

**Table 5.2: Consonant Inventory**

| | Bilabial | Dental | Alveolar | Post alveolar | Palatal | Velar |
|---|---|---|---|---|---|---|
| Plosive | b | | t ⁿd | | | k kʷ |
| Nasal | m | | n | | ɲ | |
| Trill | | r | | | | |
| Fricative | | ð | s | ʃ ʒ | x | |
| Affricate | | | | ʧ | | |
| Lateral | | | l | | | |

The bilabial voiced stop [b] presents great variation in its surface form. Word-initially, it often, but not always, receives pre-nasalisation and is pronounced as [mb] and intervocalically it is realised as [β]. However, there does not seem to be a predictable pattern in the pre-nasalisation of the phoneme, as it occurs in other Mixtec varieties. Below is an example of a word with the three possible realisations of the bilabial stop:

(5) [bāʔā] / [mbāʔā] ‘good’ [xāβāʔā] ‘something good’

The alveolar voiced stop [ⁿd] is always realised as a pre-nasalised stop, whether word-initially or intervocalically.

(6) [ⁿdāáβā] ‘so that’

(7) [tʃīⁿdēʔé] ‘to help’

The voiceless velar stops [k] and [kʷ] can occur in all positions of the word, but [kʷ] never

precedes the back rounded vowels [o] and [u]. Often times the former is pre-nasalised word-initially. When it does, it is commonly pronounced as voiced velar stop, as the example below demonstrates. No examples have been found in the corpus of a pre-nasalised [kʷ].

(8) [ŋkw̃ũ̃ʒā] / [ŋgw̃ũ̃ʒā] 'it is raining'

Furthermore, they are frequently pronounced as voiceless fricatives [ɣ] and [ɣʷ] respectively, both word initially and word-medially. These allophones occur as the result of rapid speech. For example:

(9) [kā] / [ɣā] 'more'

(10) [tūkū] / [tūɣū] 'again'

(11) [kʷītī] / [ɣʷītī] 'very'

The nasal [n] is velarised before the voiceless velar consonants and their voiced allomorphs, resulting in the sequences [ŋk], [ŋkʷ], [ŋg] and [ŋgʷ].

(12) [ŋkējāā] / [ŋgējāā] 'start'

(13) [ŋkūū] / [ŋgūū] 'was'

Two morphemes are attested where the dental fricative [ð] is used: [ðē], third person clitic and [ðē] 'and'. In the first case, the consonant is found word-medially, as the morpheme is a clitic and always attaches to the phonological unit it follows. In the case of the coordinating conjunction, the consonant appears word-initially, as the word is, more often than not, a phonological unit by itself. The nature of this phoneme is problematic and its origin has been the subject of several theories, as it is found in many other varieties of Mixtec (cf. Macaulay (1997: 20) for a relevant discussion). Seeing that the clitic [ðē] is phonetically reduced form of the noun [tēē] 'man', it is possible that the phonemes [t] and [ð] are diachronically related. Additionally, Macaulay (1997: p. 21) cites Kaufman's reconstruction *tææ 'man' in Proto-Otomanguean, mentioning that [ð] appears among the different developments of the first segment of this word.

### 5.1.3 Loan Consonants

There are several consonants that appear in loan words from Spanish. These include: [p], found in words such as [pérō] 'but', [prīmárīā] 'elementary school' and [pōlīsíā] 'police'; [f] and [fʷ], which appear in words like [gārráfō] 'water bottle' and [fʷérā] 'outside'; and the trill [rr] which is found in words like [kārrētérā] 'road'. It is also worth noting that many Spanish words and proper names that end with [o] are pronounced in Mixtec with a word-final [u], as can be

observed below:

(14) [karro] - [kárrū] 'car'

(15) [tiempo] - [tiēmpū] 'time, weather'

### 5.1.4 Consonant Clusters

Generally speaking, consonant clusters are not very common in Cuquila Mixtec. Most clusters involve the consonant [ʃ], as shown in the examples below [1]:

(16) [ʃⁿdīkī] 'bull'

(17) [ʃtʃūʔū] 'chicken'

Word-medial consonant clusters are rarely found and, again, they always include the phoneme [ʃ]:

(18) [īʃtā] 'tortilla'

### 5.1.5 Syllable Structure and the Couplet

All traditional analyses of the Mixtec varieties agree that words in Mixtec are formed by two syllables, which has been called a 'couplet'. Indeed, with the exception of clitic morphemes, all the words in Cuquila Mixtec follow the same two-syllable pattern. Longer words, with 3 or 4 syllables, are either the result of cliticisation/affixation, or compounding, and can be analysed as such either synchronically or diachronically.

Mixtec has a strong preference for open syllables, as there are no consonants that occur in coda position. One or more consonants (due to the consonant clusters) can optionally appear in onset position. The following couplet types can be formed:

[1] In theory, more clusters with an initial [s] can be found, but they always involve the causative prefix *s-* on a verb that begins with a consonant. The two words shown here fit Josserand's (1983:231) analysis that such words are the result of morphophonemic reductions of pre-couplet morphemes, where the initial [ʃ] derives from a morpheme that has lost its following vowel.

(19)

| | |
|---|---|
| VV | [ū.ū] ‘two’ |
| CVV | [kū.ū] ‘to be’ |
| CCVV | [ʃⁿdī.kī] ‘bull’ |
| CCVCV | [ská.sū] ‘to toast’ |
| VCV | [ī.kū] ‘yesterday’ |
| VCCV | [īʃtō] ‘uncle’ |
| CVCV | [ñánī] ‘brother’ |

Consonant clusters can be found in both word-initally and word-medially, at the beginning of each syllable of the couplet. Interestingly, this seems to be an innovation that Cuquila Mixtec has developed, as both Macaulay (1997) and Alexander (1988) mention that consonant clusters can only occur word-initially. As mentioned in previous sections, many words are found with a word-medial consonant cluster, that do not seem to be loans from other languages. In the following examples I present words with both word-initial and word-medial consonant clusters:

| | | |
|---|---|---|
| Word-initial: | [ʃkótʃī] ‘pig’ | [ʃlīlū] ‘person from Ocotepec’ |
| Word-medial: | [īʃtō] ‘uncle’ | [īʃtā] ‘tortilla’ |

Indeed, comparing the word [īʃtā] to other varieties, it seems that they do not involve this consonant cluster in the onset of the second syllable[2]:

| Cuquila | Ocotepec | Magdalena Peñasco | Chalcatongo |
|---|---|---|---|
| [īʃtā] | [stàà] | [ʃītā] | [stāà] |

No words have been found in the corpus which have consonant clusters in both syllables. Therefore, the couplet type CCVCCV seems to not be attested in Cuquila Mixtec.

### 5.1.6 Glottal Stop

The glottal stop seems to always appear word-medially: intervocalically, either between two identical or different vowels (as seen in (20) and (21) respectively), or followed by a consonant (as shown in 22):

(20) [βēʔē] ‘house’ [sāʔā] ‘to make’

(21) ʃūáʔū] ‘coyote’

[2]The data for these examples came from Alexander (1980: 162) for Ocotepec Mixtec, Hollenbach (2017) for Magdalena Peñasco and Macaulay (1997: 278) for Chalcatongo Mixtec. I transcribed the words phonetically using the conventions I use throughout this thesis in order to make the comparison easier.

(22) [xīkātáʔβi] ‘to ask for’ [sāʔmā] ‘clothes’

The glottal stop [ʔ] in Mixtec is a topic that has raised a lot of questions regarding its status as a phoneme. Three major analyses have been proposed over the years. Longracre (1957), in his study of Proto-Mixtec, among many other linguists, has analysed the glottal stop as a consonant. Bradley (1975) and many others have analysed the glottal stop as a feature of the vowel for the variety of Jicaltepec Mixtec. Macaulay & Salmons (1995) propose a different analysis of the glottal stop as a feature of the root, based mostly on data from the variety of Chalcatongo Mixtec. This analysis is based on the fact that glottalization is restricted to the initial syllable of the couplet in most varieties, and so the feature [+/- constricted glottis] is attached to the leftmost vowel of the root. This theory seems to address forthcomings that arise from the other two proposals, namely the introduction of a new syllable structure only to account for the glottal stop as a consonant, and the unusually large vowel inventory that the second theory creates of more than 20 vowels.

### 5.1.7 Tone System

Cuquila Mixtec, similarly to other Mixtec varieties makes use of three level tones, high, mid and low. It is a contrastive feature of the words, as can be seen from the minimal pairs below:

(23) [ndūkú] ‘search (completive) [ndúkú] ‘search (potential)’
[bīkō] ‘cloud’ [bíkó] ‘party’
[īyō] ‘to exist’ [īyó] ‘some times’

Long vowels can have multiple combinations of tone, either the same (for example, mid-mid) or different ones (for example, mid-low). Since there are no contour tones in Cuquila Mixtec, it seems that the tone-bearing unit is either the syllable or the mora. Pike (1944) analyses words with long vowels as disyllabic, whereas others (for example, McKendry (2013: 67)) analyse a long vowel as a heavy syllable with two moras. In any case, discussing the two analyses further is not pertinent to the discussion of tone in this thesis and lies outside of the scope of this grammar sketch, as none of the two analyses creates issues when dividing the syllables. Below I will provide some examples of words with long vowels and contrasting tone for reference.

(24) [téè] ‘attach’ [tēé] ‘man’
[ɲūʔū] ‘village’ [ɲūʔù] ‘fire’
[kʷáʔã̄] ‘to go (progressive)’ [kʷàʔã̄] ‘to give’

The languages of the Mixtec family often exhibit tone sandhi, whereby a word carries a tone that does not surface within the word, but instead it affects the prosody of the next word. For example, in Cuquila Mixtec the word *nûi* 'obtain' has a HL tone profile, but it carries an additional mid tone which affects the following word. When the clitic *nì* 'I', which has a low tone, attaches to this verb, its tone changes to mid:

(25) *nûi=nī*
acquire=1.FAM
'I acquire.'

Pike (1944) was among the first to describe how the tone of a word is affected by its environment in what he called 'tone perturbation'. Later scholars have used the term 'floating tone' (Hinton et al. 1991) to describe the same phenomenon. It seems that tone sandhi in Mixtec is notoriously difficult to describe and involves a set of complex rules which vary per language. Due to the subject of this study, and the fact that there is no other data available for Cuquila Mixtec, the subject of tones in this variety was not studied to a sufficient degree in order to draw any conclusions. Further research needs to be carried out in order to clarify how tone sandhi works in Cuquila Mixtec.

### 5.1.8 Orthography

The table below gives an overview of the orthographic conventions used in this thesis:

**Table 5.3: Orthographic conventions followed in the thesis**

| Phoneme | Transcription |
|---|---|
| /b/ | <b> |
| /t/ | <t> |
| /ⁿd/ | /nd/ |
| /ð/ | <d> |
| /k/ | <k> |
| /kʷ/ | <ku> |
| /m/ | <m> |
| /n/ | <n> |
| /ɲ/ | <ñ> |
| /r/ | <r> |
| /s/ | <s> |
| /ʃ/ | <x> |
| /ʒ/ | <y> |
| /x/ | <j> |
| /l/ | <l> |
| /tʃ/ | <ch> |
| /ʔ/ | <’> |

Ever since the establishment of *Ve’e Tu’un Savi* (Academy of the Mixtec Language) in 1997, the orthography of the Mixtec languages has been somewhat conventionalised and adopted by the communities. Throughout this thesis I will mostly use the orthography proposed by them for two main reasons. Firstly, I believe that it sufficiently matches the phonemes found in Cuquila Mixtec, without any need to make up new letters. Secondly, this orthography has been widely used both in the academic circles as well as in the communities. However, the transcription of the nasal vowels that I follow in this thesis differs from the established orthography. The Academy prefers to transcribe them as a cluster of an oral vowel plus a nasal stop (for example, <an>). However, I believe that this can cause issues in the interpretation of the position of the nasal stop: for example, in a word such as *kuanko* ‘go (plural)’ does the letter <n> represent a nasal vowel [ã] or does it stand for the pre-nasalisation of the following velar stop, as in [ŋk]? For this reason, I have decided to maintain the transcription of the nasal vowels with a tilde: <ã>, <ẽ>, <ĩ>, <õ> and <ũ>. Lastly, I will not transcribe the

allophones of the phonemes, but instead I will use the underlying forms [3].

The pre-nasalised stop [ⁿd] will be transcribed as [nd], for example in [ⁿdē] <nde> 'until'. The glottal feature of the root will be represented by an apostrophe following the vowel, as in [ɲūʔū] <ñu'u> 'village'.

Only the high tone (eg. á) and the low tone (eg. à) will be transcribed. Whenever there is no tone written, it can be assumed that the vowel has a mid tone. Lastly, because the tone sandhi rules in Cuquila Mixtec are not well established, I will always transcribe the underlying tones of each word. However, as mentioned previously, many of these words change their tone pattern due to the environment they appear in.

## 5.2 Morphology

In this section, some aspects of the nominal and verbal morphology will be analysed. Section 5.2.1 involves the morphology of the noun phrase and includes certain aspects of the noun, such as compounding, nominalisation and plural marking, among others. Section 5.2.4 presents some aspects of the verb phrase, such as TAM and number marking and the copulas. Section 5.2.2 includes information on adjectives and deictics. Lastly, section 5.2.3 introduces the properties of adverbs.

### 5.2.1 Noun Morphology

Nouns in Cuquila Mixtec are an open class, characterised by their ability to take modifiers and possessive markers. They can be marked for number through the use of the plural word and for possession via a personal marker. There is no grammatical gender in Cuquila Mixtec. Instead, certain nouns are grouped into specific categories, which are made apparent in the choice of an anaphoric third person pronoun.

#### 5.2.1.1 Noun Categorisation

Nouns are divided into several categories, depending solely on the third person pronoun that is used to co-reference them. Since there is no agreement in other parts of the grammar, such as in the nominal modifiers, the third person pronouns offer the only opportunity to understand how the nouns are categorised. There are seven categories that nouns can be part of: masculine, feminine, general human, deity, animal, liquid and tree. The rest of the nouns are grouped together, without any further distinctions being made. Below are some examples of nouns that generally fit these categories:

[3]The Academy's proposal includes the letter *ν* to transcribe the sound [β]. However, seeing that this is not a phoneme but an allophone of [b], I have decided not to follow the Academy's proposal.

(1) Masculine:

*teé* ‘man’
*kua’a* ‘brother’
*ixto* ‘uncle’

(2) Feminine:

*ña’a* ‘woman’
*yuúvá* ‘elderly woman’
*xixi* ‘aunt’

(3) General Human:

*suchi* ‘child’
*ñayìì* ‘person’
*sesi’i* ‘daughter’
*seyi’i* ‘son’

(4) Animal:

*xuáu* ‘coyote’
*chi’ili* ‘chicken’
*skochi* ‘pig’
*xkabayu* ‘horse’

(5) Deity:

*yandiuxi* ‘God’
*chu’nchi* ‘Jesus’
*santu* ‘saint’
*savi* ‘rain’

(6) Liquid:

*ndutè* ‘water’
*ndeuva* ‘beer’
*ndevixi* ‘soda’

(7) Tree:

| | |
|---|---|
| *ñutu* | 'tree' |
| *ti'ti* | 'juniper tree' |
| *tiundu* | 'arbutus' |

At first glance, this categorisation is rather straight-forward. All the animals are grouped together, all the trees are members of the same category and so on. However, certain nouns can be members of different categories, which is determined by the context and the pragmatic functions of the discourse. For example, a woman can belong to the category 'feminine' as well as the group of 'general humans'. This membership into different categories is reflected on the choice of the third person pronoun that the speaker makes. Similarly, *sutu* 'priest' can either belong to the 'masculine' category or be grouped together with other deities, depending on whether the speaker wants to emphasise their human or religious nature.

Outside of these categories, the rest of the nouns are grouped together into a generic category. That is, anything that is not an animate (masculine, feminine, general human, deity or animal), liquid or tree is not specifically categorised and cannot replaced by a pronoun.

#### 5.2.1.2 Compounding

Cuquila Mixtec has a very productive system of NP + NP constructions which have a narrower meaning than the parts. For example:

(26) *tú'ũ nda'bi*
word poor
'indigenous language'

(27) *ndo'o xtó'o*
adobe foreign
'wall'

Due to rapid speech, many times the first word is contracted to one syllable only. Through time, certain constructions that are frequently used become lexicalised as (often trisyllable) compounds, where the first syllable of the word is a contracted NP. The individual parts of these compounds are usually easily distinguishable, and speakers are fully aware of their origin. This is often the case with buildings, whereby frequently used words are often times trisyllable. In this case, the first syllable of the word is *be*, a reduced form of the word *be'e* 'house', as seen in (28) and (29). The fact that the word is reduced to its first syllable can be well observed in (30): the syllable *ki* originates from the word *kivi* 'day', which then attaches to the word *kumi* 'four' to create the word 'Wednesday'.

*be-* from *be'e* 'house'

(28) *be-ñu'u*
house-soil
'church'

(29) *be-tiñu*
house-task
'town hall'

*ki-* from *kibi* 'day'

(30) *ki-kumi*
day-four
'Thursday'

#### 5.2.1.3 Classifiers

A restricted form of a possible classifier system is found in most Mixtec languages, where some varieties show a higher degree of grammaticalisation than others. In the Western Alta varieties the classifiers have gone through a process of fossilisation, where remnants of a former noun classification system can be observed (Cassiano 1982: 87). This process of fossilisation has led scholars to debate the existence of a classifier system in Mixtec. Admittedly, recognising these classifiers is not as straightforward as, for example, the numeral classifiers in Mayan languages. The Mixtec varieties exhibit different degrees of fossilisation and grammaticalisation. Due to this reason, we find many differences in the number and shape of the classifiers per variety.[4].

In Cuquila Mixtec remnants of a possible classifier system can be found in certain words, but they already form part of the noun and they are not as productive [5]. Below are some examples of trisyllabic words that include a morpheme which could be analysed as a fossilised classifier.

A few nouns referring to animals begin with *ti-*, which is a contraction of the word *kiti* 'animal':

[4]For example, the variety of Coatzospan, spoken in the north-eastern part of the Mixteca Alta, has classifiers which are morphologically free in the noun phrase and are productive in terms of new referents (Leon Pasquel 1988: 137), which fit well into the prototypical definition of the classifiers.

[5]Furthermore, some of them are morphologically and semantically related to the third person markers, which will be discussed in the following chapter. However, the relation between the fossilised classifiers and the person markers is not clear. Passer (2016: 28), analysing Chalcatongo Mixtec, treats the latter as pronominal elements which do not form part of the nominal classification system, whereas Leon Pasquel (1988: 137) demonstrates that, in Coatzospán Mixtec, the noun classifiers perform pronominal functions, and are thus related.

(31) *tina* ‘dog’

(32) *tisuma* ‘scorpion’

The same syllable can be found in the names of certain vegetables and round objects:

(33) *tinana* ‘tomato’

(34) *tilúu* ‘ball’

Deities, mostly from the catholic religion, usually include the morpheme *ya*, which is related to the word *yaa* ‘deity’:

(35) *yandiuxi* ‘God’ (from Spanish *Dios*)

The names of some trees are trisyllabic words whose first syllable is *tu*, a contraction of *ñutu* ‘tree’:

(36) *tuyuja* ‘pine tree’

(37) *tuiña* ‘encina tree’

Liquids often times include the particle *nde-*, which derives from *ndutè* ‘water’:

(38) *nde-bixi*
water-sweet
‘soft drink’

(39) *nde-uba*
water-bitter
‘beer’

As we can see, the formation of these words resembles the construction of the aforementioned compounds. They also include a contracted form of a noun which occurs in word-initial position. However, an important difference in the patterns can be observed: in the compounds, the contracted noun is usually reduced to its initial syllable, while the particles described here result from a contraction of the noun to the second syllable. Some cases, though, are not as straight-forward; the last category shown here, that of liquids, involves the morpheme *nde*, which does not correspond to the second syllable of the word *ndutè* ‘water’. Furthermore, many of the examples shown here cannot be separated into their parts, as the rest of the word does not have a separate meaning. For example, the word *tina* ‘dog’ cannot be analysed as *ti-* ‘animal classifier’ + *na*, as the latter is not a word by itself.

Summarising the above, it is rather difficult to distinguish the compound nouns from the nouns that include a fossilised classifier, as they both involve the phonetic reduction of a noun

and its fusion to another noun. However, the different ways in which the nouns are reduced in each case hint at the theory that Cuquila Mixtec had a classifier system whose traces can be seen in several parts of the grammar. The fact that the vast majority of the grammars of other Mixtec varieties with the same characteristics make reference to classifiers reinforces the idea that such system also exists in Cuquila Mixtec. Nevertheless, it is important to note again that the classifiers found in this variety are by no means productive, but instead are fossilised.

#### 5.2.1.4 Nominalisation

The nominalizer *ja-* derives nouns from verbs (40) and adjectives (41). Macaulay (1997: 65) argues that this nominaliser prefix is related to the complementiser *ja* which introduces subordinate clauses, so "nominalised adjectives could conceivably be analysed as headless relative clauses". The following examples show nouns derived from a verb and an adjective respectively:

(40) *ja-chuná'a*
NMLZ-pay
'payment'

(41) *ja-ba'a*
NMLZ-good
'the thing/person that is good.'

The example below demonstrates that the derived word is, in fact, a noun, as it is preceded by a noun modifier, namely a numeral:

(42) *sa'a=nú ĩ̃ĩ̃ **ja-ba'a** ndaba kuu chinde=nú nu ñuu=nú*
make=2.RES one **NMLZ-good** so.that COP help=2.RES at village=2.RES
'Do something good so you can help your village.'

#### 5.2.1.5 Plural Marking

Plural marking on nouns occurs with the plural marker *ndáa*. Due to rapid speech contraction, it is frequently pronounced as *ndá*. The marker always immediately precedes the noun:

(43) *kasiki **ndáa** tée kasiki tilúu*
play **PL** man play round
'The men play basketball.'

### 5.2.2 Nominal Modifiers

The parts of speech that can modify a noun in Cuquila Mixtec are: adjectives, demonstratives, numerals, quantifiers, possessive pronouns and other nouns. The possessive pronouns will be

separately analysed in chapter 6. The discussion of the numerals, the quantifiers and the nouns as modifiers lies outside of the scope of this chapter. Instead, I will briefly discuss the adjectives and the demonstratives. It is important to note, however, that there is no agreement with the noun. Other than that, the modifiers in Cuquila Mixtec are not inflected for number or gender and, generally, do not carry any inflectional information.

#### 5.2.2.1 Adjectives

In their attributive use, adjectives follow the noun and do not take any plural marking:

(44) *ñúkuà̀ ká-ñu ndáa kiti* ***na'nu***
there PL-be.inside PL animal **big**
'The big animals are in there.'

Adjectives can also be used predicatively without the need of a verb. A copula can appear in such clauses (as seen in (47)), but its use seems to be optional (45 and 46):

(45) ***xẽẽ*** *túni ndáa ndixi=u*
**aggressive** very PL pimple=DIST
'Those pimples are dangerous.'

(46) *luu kuiñi tu xiyo=a*
pretty thin ? fabric=PROX
'This skirt is thin and pretty.'

(47) *kuu* ***xuchi*** *nùnì=u*
COP **smashed** corn=DIST
'That corn is smashed.'

Alexander (1980: 253) does not include adjectives in her grammar for Ocotepec Mixtec, but instead calls these words "stative verbs". Indeed, the adjectives is a traditionally problematic part of speech to define (Dixon 2010). It seems that in Mixtec the confusion stems precisely from the predicative use of the adjectives, which has led many SIL linguistics to treat them as verbs, and others (for example, Macaulay (1997), McKendry (2013)) to consider them as a separate word class. There are several reasons why adjectives form a class of their own in Mixtec: first of all, as (47) above shows, the adjectives can be used with a copula, a feature which cannot occur with verbs. Secondly, adjectives do not seem to take the same affixes as verbs. Let us look at example 45 again, repeated here:

(48) ***xẽẽ*** *túni ndáa ndixi=u*
**aggressive** very PL pimple=DIST
'Those pimples are dangerous.'

In this case, if *xẽẽ* was a verb in (45), we would have expected to see the plural prefix *ká-* which

is used with the verbs (cf. 60). However, this would produce an non-grammatical sentence.

(49) ***ká-xẽẽ** *túni ndáa ndixi=u*
**aggressive** very PL pimple=DIST
'Those pimples are dangerous.'

Therefore, adjectives can be best analysed as a separate word class in Mixtec as well.

#### 5.2.2.2 Demonstratives

Two morphemes are used as demonstrative clitics in Cuquila Mixtec, *a* 'this' and *u* 'that', which makes for a two-way distinction in terms of distance from the speaker. These clitics are the morphologically reduced forms of the demonstrative adverbs *yáa* 'here' and *ñúkuà̀* 'there' respectively. They attach to the last element of the noun phrase, usually a noun or an adjective.

(50) *de kibi=a ndatekuu yandiuxi*
and day=PROX resurrect god
'And on this day, God resurrects.'

(51) *nuù=sã ichi nu bakoo ndáa ña'a=**u** ñúkuà̀*
descend=1.RES road at come.PL PL woman=**DIST** there
'I would descend the road there, where these women are coming from."

Apart from distance from the speaker in the physical space, they are also used anaphorically, when a referent is re-activated or topicalised. In the example below, the speaker was providing the recipe for *pozole*, a food prepared from corn. After she explained how she prepares the corn and lets it cook overnight, she uttered this sentence:

(52) *tankuniji ja n-cho'ó nùnì=**u***
dawn already COMPL-cook corn=**DIST**
'At dawn, the corn is already cooked.'

### 5.2.3 Adverbs

Adverbs in Cuquila Mixtec are frequently found inside the VP as locative, temporal and manner adjuncts or as intensifiers. They can appear by themselves or can be preceded by a preposition. They usually follow the verb, but they can also go in preverbal position when they receive focus. Below are some examples of adverbs which function as adjuncts:

Location:

(53) ***ñúkuà̰*** *jinu=ti máa ichi ká'nu*
**there** run=3.ANML SPEC road big
'There they (horses) run, on the big road.'

Time:

(54) *iyo ba'a=yó* ***mitañu***
COP well=INCL **now**
'We are doing alright now.'

Manner:

(55) *jáki'i* ***ka'á****=yó lulu biti ndáa yu itu*
sow **in.vain**=INCL a.little INTSF PL edge cornfield
'We sow the corn in vain.'

The locative adverb *ñúkuà̰* 'there' is also often used with a temporal meaning, to denote succession of events, as shown in the example below:

(56) ***ñúkuà̰*** *ni keja'a jánini*
**there** LIM start evening
'Then, the evening starts.'

Intensifying adverbs such as *kuiti* and *neé* always follow the verb. A personal pronoun clitic can appear in the adverb, as these clitics attach to the last element of the verb phrase. This fact also demonstrates that such adverbs are part of the VP.

(57) *kunde'e* ***kuiti****=ña*
sit **very**=3.F.RES
'She is staying (here) for a long time.'

### 5.2.4 Verb Morphology

Verbs in Cuquila Mixtec are also a major open class, distinctly separate from the word class of nouns in a number of ways. Morphologically speaking, they take mostly prefixes and very few suffixes. Prefixes include derivational information, such as causative and repetitive markers, as well as inflectional information, such as TAM marking. Suffixes include some forms of plural marking, as well as a focus marker. Person marking is done through the use of pronominal clitics and independent pronouns, as will be shown in chapter 6. In what follows, I will give a brief overview of the TAM marking strategies and the plural marking, as well as information on negation.

#### 5.2.4.1 TAM Marking

Verbs is Cuquila Mixtec are not marked for tense, but only for aspect and mood, as in the vast majority of Mixtec varieties. The highest-level distinction is between the irrealis and realis mode (following the definition provided by Payne (1997: 244): "mode describes the speaker's attitude toward a situation, including the speaker's belief in its reality, or likelihood"). Irrealis is used for possible or potential events or events occurring in the future. It is also the form that the verb takes in imperatives and conditionals. It is often used in subordinate clauses to denote temporal relativity to the matrix clause, such as an event which occurs in a future relative to the time frame of the main verb (which might still refer to a past event). The irrealis only includes one aspect: potential. Below is an example showing its use to mark a future event:

(58) *kóo ĩ̋ĩ̋ kibi de* ***káka****=yó* ***kii****=yo nu ñuyibi=a*
exist.POT one day and **walk.POT**=INCL **come.POT**=INCL at earth=PROX
'There will come a day when we will walk this earth.'

Realis mode includes two aspects: completive and incompletive. The completive aspect is used for past events, or events that occur or have occurred non-habitually. It is often times marked on the verb through the use of the completive prefix *ni-*, through tone changes in relation to the potential aspect, or through a combination of both. The completive prefix *ni-* is often contracted to *n-*. (59) below demonstrates a verb with tone alternation between completive and incompletive, and (60) shows the use of the prefix *ni* with an incompletive verb stem:

(59) Incompletive: *jikáà=nì* 'I ask for'
Completive: *jikáá=nì* 'I asked for'

(60) ***n****-kii bi ndáa=de be'e*
**COMPL**-enter AFF PL=3.LIQ house
'The water entered the house.'

The incompletive aspect denotes events in progress, regardless of their location in time. An event or action can be interpreted as continuous both in the past or the present, a distinction which is made clear contextually, through the use of adverbs or other time expressions. Furthermore, it is frequently used to mark the habitual character of an event. The incompletive aspect can be marked through changes in the tone patterns of the potential verb stem, or through stem suppletion; several verbs have a different incompletive stem, which can also be used in combination with the prefix *ni-* to create the completive root. Some examples of stem suppletion are shown below:

| | eat | exist | walk |
|---|---|---|---|
| potential | *ka'á* | *kóo* | *káka* |
| incompletive | *ya'á* | *iyo* | *jíka* |

Lastly, two verbs have been found which show stem suppletion to denote further distinctions in the incompletive aspect, between habitual and progressive. Therefore, these verbs have four forms in total: potential, completive, habitual and progressive.

| | come | go |
|---|---|---|
| potential | *kii* | *ki'i* |
| completive | *n-kii* | *ja'a* |
| habitual | *kíi* | *já'à* |
| progressive | *vaji* | *kuá'ã* |

#### 5.2.4.2 Number Marking

Verbal agreement with a plural subject is marked in the form of affixes. There are two different plural affixes used, depending on the semantics of the verb. Verbs of motion take the suffix *-koò*, as shown in the example below:

(61) *tá sinko ké'e-**koò**=sã skuela de kuano-**koò**=sã*
when *cinco* exit-**PL**=1.RES *escuela* and go.home.PROG-**PL**=1.RES
'Around five o'clock we would leave school and return home.'

This suffix is found in the corpus with verbs such as *já'à* 'go', *vaji* 'come(PROG)', *kuá'ã* 'go(PROG), *ké'e* 'exit', *jinu* 'reach' etc.

The rest of the verbs take the prefix *ká-*[6]. Below are some examples of the use of the plural prefix. (63) demonstrates its use with a verb in potential aspect:

(62) *tángù **ká**-kejá=nì **ká**-té'e=nì ndo'o xtó'o*
later **PL**-start=1.FAM **PL**-attach=1.FAM adobe foreign
'Later, we start attaching the walls.'

(63) *tu ndáa ma=yó ma=**ká**-ka'a=yo*
if PL SPEC=INCL NEG.POT=**PL**-speak=INCL
'If we don't speak (Mixtec).'

[6]It seems that in other varieties of the Western Alta, (Macaulay (1997: 73) for Chalcatongo and Hollenbach (2013: 126) for Magdalena Peñasco, among others) the prefix is not used in potential aspect, but instead another prefix is used. However, Cuquila Mixtec uses the same prefix in all aspects.

#### 5.2.4.3 Negation

Verbs can be negated through the pre-verbal clitic *ndu=*. This form seems to be related to the negative copula *ndúu*. Verbs in potential aspect take the clitic *ma=* instead. Examples of both cases are illustrated below:

(64) *de ndáa kibi ñúkuà* ***ndu****=ká-yaji=nì kuñu*
and PL day there **NEG**=PL-eat.PROG=1.FAM meat
'And the following days we don't eat meat.'

(65) *nduna ndakani ká ndáa tú'ũ tu ndáa ma=yó* ***ma****=ká-ka'a=yo*
nobody narrate ADD PL word if PL SPEC=INCL **NEG.POT**=PL-speak=INCL
'Nobody will tell our stories any more if we don't speak (Mixtec).'

## 5.3 Syntax

### 5.3.1 Typological remarks

Typologically speaking, Cuquila Mixtec follows Greenberg's (1963) correlates for a VSO language: the adjective and the possessor both follow the noun, there are prepositions instead of postpositions, most affixes are prefixed (even though some suffixes also exist) and the question words and particles are sentence initial.

### 5.3.2 Constituent Order

In order to discuss word order in Cuquila Mixtec, I will follow Lambrecht's (1996: p. 17) definition of a 'neutral' or 'pragmatically unmarked' constituent order: "given a pair of allosentences, one member is pragmatically unmarked if it serves two discourse functions while the other member serves only one of them". That is, if a certain clause can answer only one specific wh- question, then it is pragmatically marked, containing an element which either receives focus or is a contrastive topic [7]. In this context, the most neutral word order in Cuquila Mixtec is VSO, as shown in the example below:

(66) *sa'a=sã texa=sã de chu'u=sã ajo*
make=1.RES salsa=1.RES and put=1.RES *ajo*
'I make my salsa and I put garlic.'

In the example above we can see how, in both clauses, the verbs are in initial position, along with the clitic pronoun *sã* 'I', which functions as a subject, followed by the object. It could be argued that, since the subject in this case is a pronominal clitic, it will always be in post-verbal

[7] following Lambrecht (1996) definitions of focus and Büring (2015) definition of a contrastive topic.

position. However, in the following example we see that a noun phrase functioning as subject also follows the verb:

(67) *ká-kaà* ***ndáa xtó'o*** *ja* *ma=kũũ=ya*
PL-say **PL foreign** COMP NEG.POT=fall=3.DEI
'The people on the radio said that it would not rain.'

Apart from the fact that VSO word order is also the most frequent one, another way of confirming that this word order is the most pragmatically unmarked is to analyse the SVO or OVS cases. Looking at the corpus, it can be observed that, whenever a constituent is fronted, whether it is the subject, the object or any other element of the clause, it is because it receives the focus of the proposition or it functions as the contrastive topic. To illustrate this, we can look at the example below. Two women are discussing the storm of the previous day. One of them asks the other where she was last night, and whether she got rained upon. The other woman replies:

(68) *ndúu chi* ***be'e*** *ịyo=nì* *de* *nde-ja'a=ya*
NEG because **house** EXIST=1.FAM and ITER-begin=3.DEI
'No, because I was at home when it started raining again.'

In this case, the locative adjunct *ve'e* 'house' is found in pre-verbal position. This sentence is a direct reply to the question 'Where were you last night?', which is one of the tests that Wal (2016: 265) provides to diagnose focus. This word order would be infelicitous as a reply to questions such as 'What did you do last night?' precisely due to the fronting of the locative adjunct, demonstrating that it cannot serve more than one discourse function.

That being said, there are many instances in the corpus where word orders other than VSO occur. However, because the fronting of a constituent serves many different functions, this is an expected effect of how discourse works; most of these alternative word orders occur in the recorded conversations, where the speaker wants to draw the hearer's attention to a specific element in the discourse, and so the fronted element will either be topicalised or focused.

As mentioned previously, any element of the sentence can occur in pre-verbal position. Left dislocation is the main focus and topic strategy, and in most cases it is difficult to distinguish among the two without the context they occur in. However, the distinction becomes clearer when personal markers are used, as will be explained in section 6.5. For now, I will provide some examples of different elements in pre-verbal position:

Subject:

(69) ***nduna*** *yi'i skuela* *xi'ina*
**nobody** go *escuela* before
'*Nobody* would go to school in the past.'

Object:

(70) *de* ***ndutè*** *jíso=ña* *saa*
and **water** carry=3.F.RES AFF
'She was carrying *water*, indeed.'

Subject Complement:

(71) ***uni*** *kuu ki'i sabado*
**three** COP day *sabado*
'It's the third (of the month) that is a Saturday.'

### 5.3.3 Non-verbal predicates

Kroeger (2005: 173) analyses the copulas cross-linguistically as words that, grammatically, function as a verb (meaning that they take TAM and agreement inflection) but they are semantically empty, barely contributing to the meaning of the sentence.

There are two main copulas in Cuquila Mixtec: *kuu* and *ìyo*. The former is used with both nominal and adjectival predicates, even though the adjectives can also appear without a copula. Based on the data, it seems that the use of the copula is optional, as there is no semantic difference between adjectival predicate constructions with and without a copula. The examples below demonstrate the use of the copula with a nominal and an adjectival complement.

(72) *de uni kibi=u kuu biernes*
and three day=that COP *viernes*
'And that third day is a Friday.'

(73) *kuu xuchi nùnì=u*
COP smashed corn=DIST
'That corn is smashed.'

The verb *ìyo* (*kóo* in potential aspect) functions both as a copula and as an existential verb. As a copula, it is often used with an adverbial complement or a prepositional phrase, or to talk about someone's age. The examples below demonstrative these uses

(74) *ba'a ni ìyo*
good LIM COP
'It is going alright.'

(75) *ká-ìyo=yó chiji netiñu ñuu kuiñi*
PL-COP=INCL under townhall village tiger
'We are under the townhall of Cuquila.'

(76) *iyo=de kuarenta i nuebe kuiyà*
COP=3.RES *cuarenta y nueve* years
'He is 49 years old.'

An example of the existential use of the copula is shown below:

(77) *kibi ñukua su ni iyo nu jinu nda xkabayu*
day there AFF LIM COP to run PL *caballo*
'The next day there is horse racing.'

Furthermore, the negative copula *ndúu* provides the meaning of 'not be'. When this form appears in a sentence, no other copula can be used with the negative marker, which confirms the fact that this form functions as a negative copula by itself.

(78) *Jua* ***ndúu*** *teé kuña'anu*
Juan **NEG.COP** man president
'Juan is not the president.' (elicited)

(79) *Xuetu* ***ndúu*** *ndeba'a anu=i*
Roberto **NEG.COP** bad heart=3.GEN.HUM
'Roberto is not a bad person.' (elicited)

As Dixon (2010: 180) mentions, because a copular verb lacks meaning but instead denotes a semantic relation between the subject and the subject complement, it is often times omitted when that relation can be easily inferred. This is also the case in Cuquila Mixtec, where verbless clauses are fairly common. As we saw in (45), an attributive clause does not necessarily require a linking verb, but instead can be expressed by an adjective in subject complement position. However, in all these cases a copula can still optionally appear.

Equative and possessive clauses can be expressed via juxtaposition of two noun phrases:

(80) *sã'ã nanì ixi xi'ni=sã de ndi'i kuiti ixi xini=ní*
1.RES long hair head=1.RES and 2.RES short hair head=2.RES
'My hair is long and your hair is short.'

(81) *Maria naxte'e*
Maria female.teacher
'Maria is a teacher.' (elicited)

### 5.3.4 Coordination

The main coordinating conjunction in Cuquila Mixtec is *de* 'and'. It is used to coordinate clauses, but also noun phrases, verb phrases and locative or temporal expressions. The examples below illustrate some of its uses:

(82) *ndáa tuku=sã ichi=a ni **de** jínu ndeti=sã kuano'o=sã*
climb again=1.RES road=PROX LIM **and** run INTSF=1.RES return.home=1.RES
'I would climb up this road and would run home very fast.'

(83) *ká-chu'u=sã **de** ká-ya'a=sã*
PL-pour=1.RES **and** PL-eat=1.RES
'We pour (the food) and we eat.'

(84) *ká-na-nduku ĩ̋ĩ̋ tesorero **de** ĩ̃ĩ̃ secretario*
PL-?-search one *tesorero* **and** one *secretario*
'They search for a treasurer and a secretary.'

However, many times the conjunction can be omitted, and coordination is expressed by simple juxtaposition:

(85) *bánti-kóo ndáa comité jíso=de síbi=sã*
come-PL PL committee bring=3.M.RES name=1.RES
'The committee came and brought my name.'

In narratives, sentences are often introduced by the marker *desu* 'and, and then':

(86) ***desu** ndii=sã skuela*
**and.then** arrive=1.RES *escuela*
'And then I arrived at school.'

Antithesis is expressed via the markers *mitu* and the Spanish loan *pero*, as illustrated below:

(87) *yaa iñi ĩ̋ĩ̋ ñutu **mitu** ja xíchí=tu*
here stand one tree **but** COMP dry=3.TREE
'There was a tree here, but it dried up.' (elicited)

(88) *sũkuã ká-ka'a ndáa xtó'o **pero** jana'a ua kuu yaa ndiuxi*
like.so PL-say PL foreign ***pero*** remember what.is COP deity god
'That is what they said on the radio, but remember who God is.'

Disjunction is expressed via the use of the question marker *a*. As the rest of the coordination markers, *a* can be used to link NPs, VPs, adjuncts or clauses. Below is an example of the coordinator used to join NP objects:

(89) *xinañu'u ndixi sacho'o=yó **a** niki **a** naña*
first corn.cob cook=INCL **or** pumpkin **or** chayote
'The first corn cob or pumpkin or chayote that we cook.'

### 5.3.5 Subordination

The main subordination marker is the complementiser *ja*. It is used to introduce subordinate clauses that function as object of the verb. It is often found with the verb *kúni* 'want', as shown

in the example below:

(90) *kúni=sã* ***ja*** *sũkuã n-ká-sa'a=o* *ndi'i=yó*
want=1.RES **COMP** like.so COMPL-PL-make=INCL all=INCL
'I want all of us to do it this way.'

Purpose clauses are introduced by the marker *ndaaba* 'so that'. The verb in the subordinate clause is always in potential aspect.

(91) *ká-katábi=o* *nu ñu'u* ***ndaaba*** *kuu kee ja-ka'à=yó*
PL-plead=INCL at earth **so.that** COP exit NMLZ-eat=INCL
'We plead the earth so that our food is harvested (well).'

Cause is expressed by the marker *chi* 'because', as shown below:

(92) *de su n-guta=yó* ***chi*** *baji=ní*
and AFF COMPL-thank-INCL **because** come=2.RES
'And we thank you because you came (here).'

Hypothetical clauses are introduced by *tu* 'if':

(93) *chuná'a tata=sã ĩ́ĩ́ multa de* ***tu*** *ma=ndii=sã* *skuela*
pay dad=1.RES one *multa* and **if** NEG.POT=return=1.RES *escuela*
'My father would pay a fine if I didn't return to school.'

Lastly, temporal clauses are expressed via the marker *tá* 'when'.

(94) ***tá*** *n-ká-janina=i* *de ká-sa'a sana=i*
**when** COMPL-PL-name=3.GEN.HUM and PL-make doubt=3.GEN
'When they put my name up (for election), the didn't take it seriously.'

In conclusion, there are a few aspects of the grammar of Cuquila Mixtec that we need to consider when analysing the pronoun system. Firstly, tones are a contrastive feature of the words. We can find plenty of examples of minimal pairs in the language in which the words are distinguished solely on the tones the carry. Looking into the morphology, Cuquila Mixtec groups certain nouns in a rather flexible system of noun categorisation. Seven categories are distinguished from the rest of the nouns, which are all grouped together in the same class. Some of these distinct categories overlap with the fossilised classifier system that we find in many words. In terms of verb morphology, it is important to remember that subject number is marked on the verb through affixes. Furthermore, adverbs syntactically belong inside the verb phrase. Lastly, the neutral word order of Cuquila Mixtec is VSO. Almost any element can appear in pre-verbal position in order to indicate the topic or the focus of the utterance.

# Chapter 6

# Person Marking in Cuquila Mixtec

## 6.1 Overview

Having seen the basic structures of the language, this chapter will provide some detailed information on the personal pronoun system. Table 6.1 shows the full paradigm of the person markers. The first and second columns provide the glosses. The third column gives the enclitic forms, and the fourth column gives the corresponding independent forms.

**Table 6.1: Person Marking Paradigm**

| Gloss | | Clitic | Independent Form |
|---|---|---|---|
| 1 PERSON | | | |
| | 1.FAM | *nì* | *ndu'u* |
| | 1.RES | *sã* | *sã'ã* |
| 2 PERSON | | | |
| | 2.FAM | *nu* | *ndo'o* |
| | 2.RES | *ní* | *ndí'i* |
| 3 PERSON | | | |
| | 3.GEN.HUM | *i* | (*ñayìbì*) |
| | 3.M.RES | *de* | (*teé*) |
| | 3.F.RES | *ña* | (*ña'a*) |
| | 3.ANML | *ti* | (*kiti*) |
| | 3.TREE | *tu* | (*ñutu*) |
| | 3.LIQ | *de* | (*ndutè*) |
| | 3.DEI | *ya* | (*yaa*) |
| INCL | | *yó / ó* | *yóo / ndáyo'o* |

As we can see from this table, the personal pronouns can appear in two forms: as clitics and as independent markers. The clitics attach to a VP usually marking the subject, or to a noun marking the possessor:

(95) *s-kásu=**sã*** *xá'á* *sa'a=**sã*** *texá'á=**sã***
CAUS-toast=**1.RES** chile make=**1.RES** salsa=**1.RES**
'I toast the chile and I make the salsa.'

(96) *a* *ñi'i=**nu*** *ñayìì* *s-kuchi=**nu*** *ñí'í* *a* *nduu*
QUES obtain=**2.FAM** person CAUS-bathe=**2.FAM** steam.bath QUES NEG
'Did you find people to bathe in the steam bath or not?'

The morphosyntactic characteristics of the bound clitics are further analysed in section 6.2. The independent forms also mark different arguments of the VP, but they do not attach to any word. Instead, they appear as free morphemes inside the VP.

(97) *kúni=nì* *chindee=nì* ***ndo'o***
want=1.FAM help=1.FAM **2.FAM**
'I want to help you.' (elicited)

In the example above, the independent personal pronoun *ndo'o* 'you' is used to mark the object of the verb. The syntactic properties of the independent forms are further explained in section 6.3.

The pronouns of the first and second person, as well as the inclusive, can appear as clitics or as independent forms. However, this is not the case for the third person pronouns. These markers, which mostly have an anaphoric function, only appear as clitics, but they are reduced forms of specific nouns. In the paradigm above, I have chosen to provide these nouns in order to show their diachronic relation. They appear in parenthesis in order to separate them from the independent forms, as they cannot be used as such. The clitic forms of the rest of the persons are also historically related to the independent forms. Even though their relationship is not as transparent as in the case of the third person pronouns, the clitics are the result of several changes in the morphological shape of the free forms.

The person markers in Cuquila Mixtec have a four-person distinction. Apart from the three main persons, the inclusive has its own form.

The first and second person pronouns mark the speech act participants. An interesting distinction is made in both persons between a familiar and a respect form. On the first person, the humbling form *sã* is obligatorily used to show respect towards the addressee, in the same contexts that the speaker will use the second person respect form *ní* to address the hearer. It is important to note here that this latter form is not homophonous to the first person familiar

form *nì*, as this receives a low tone, whereas the respect form *ní* has a high tone.

(98) ***sa̱'a̱*** *ñu'u=**sa̱*** *jatu de **ndi'i** niji=**ní*** *xiyo*
**1.RES** wear=**1.RES** pants and **2.RES** be.covered=**2.RES** cloth
'I am wearing pants and you are wearing a skirt.'

The example above, taken from the photo elicitation task, shows the parallel use of the humbling and the polite forms. The speaker has chosen to show respect towards the addressee, so she is using the humbling forms *sa̱'a̱* and *sa̱* to refer to herself and the respect forms *ndi'i* and *ní* to address the hearer.

The third person pronouns exhibit a seven-way distinction based on several properties of the referent. These distinctions seem to be based on the noun categorisation that Cuquila Mixtec makes as it was analysed in section 5.2.1. The familiar/respect distinction is maintained in this person, as the forms for a masculine and feminine referent, *de* and *ña* respectively, are specifically used to mark politeness. The differences among these pronouns are examined in section 6.6, and their classification as grammatical genders is evaluated in section 6.8. The primary function of the third person pronouns is the anaphoric reference of a previously introduced non-speech act participant, as the example below demonstrates.

(99) *nùnì xuchi taá=nì ja yaá=**ti***
corn broken throw=1.FAM COMP eat=**3.ANML**
'I put broken corn so that they (the chickens) can eat.'

The nouns that cannot be assigned to any of these seven categories do not correspond to any pronoun, so they do not receive any anaphoric reference. This is further explained in section 6.7.

The inclusive pronouns *yó* and *ndáyo* are used when the addressee is encompassed within the speaker's discourse. Here, the familiar/respect distinction that appears in the other three persons is missing, as the same form is used in both contexts. The clitic form *ó* is also attested as an alternative pronunciation of *yó*. This secondary form does not seem to be triggered by any phonological rules, nor is it restricted to specific speakers. In fact, the same speaker will often alternate between the two forms in the same discourse. Therefore, it is possible that *ó* is a contracted form produced during rapid speech. The examples below show instances where both forms are attested in the corpus:

(100) *kachi=ó* *also found: kachi=yó*
say=INCL
'We say'

(101) *anu=ó* *also found: anu=yó*
heart=INCL
'Our hearts'

Finally, the personal pronouns do not exhibit number distinctions. Instead, the nouns or verbs are marked for number following the strategies shown in sections 5.2.1.4 and 5.2.4.2 respectively.

(102) ***ká**-chituu=nu* *nu ñu'u-nu*
**PL**-support=2.FAM on village=2.FAM
'You support your village.'

Since the pronouns are used, the familiar/respect distinction also appears in the plural. On the first person, using a plural marker creates the exclusive form. Interestingly, this means that the two plural forms are marked for familiarity or politeness.

(103) ***ká**-taá=**nì*** *yù*
**PL**-put=**1.FAM** stone
'We put the stone.'

(104) ***ká**-chũ'ũ=**sã*** *de* ***ka**-yaá=**sã***
**PL**-pour=**1.RES** and **PL**.eat=**1.RES**
'We pour (the food) and we eat.'

This chapter is organised as follows: First, the morphosyntactic properties of the bound clitics are described in section 6.2. Next, the syntax and uses of the independent forms are described in section 6.3. Section 6.4 mentions the ways in which the plural of the pronouns is expressed and the differences therein. Having established the morphology of the person marking forms, I will move on to analyse other aspects. In section 6.5 I examine the ways in which the pronouns are pragmatically marked when they receive focus or when they are topicalised. The semantic distinctions among the forms of the third person pronouns are discussed in section 6.6. Since in many instances no personal pronoun is found within the VP, the cases when this occurs are analysed in section 6.7. Finally, the section, 6.8 will evaluate whether Cuquila Mixtec exhibits grammatical gender.

## 6.2 Bound Clitics

As mentioned previously, the clitics are related to the independent pronouns (and full nouns, in the case of the third person) in that they are the contracted, monosyllabic forms of the full couplet found in the independent forms. Since contraction is often made during rapid speech in couplets of the CV?V type, it could be argued that these clitics, are, in fact the independent forms which have been contracted due to rapid speech. However, a simple substitution test demonstrates that this is not the case.

(105) *iña ni ndáa=**sã̄** nde lomo ni*
there LIM. climb=**1.RES** until *lomo* LIM.
'I would climb all the way up the hill.'

(106) *ndáa ñutu ñúkuà̄ n-ká-kayu=**tu***
PL tree there COMPL-PL-burn=**3.TREE**
'The trees over there are burnt.' (elicited)

In the (105) the verb is used with the first person respect pronoun *sã̄*, which could be argued, theoretically, to be the rapid-speech realization of the independent pronoun *sã̄'ã̄*. However, (106) demonstrates that this is an incorrect analysis. If these clitics were to be the full forms, then we should expect to see 'ñutu', since the clitic *tu* is a reduced form of the full noun *ñutu*. Substituting the clitic *tu* with the word *ñutu* 'animal' results in an ungrammatical sentence:

(107) **ndáa ñutu ñúkuà̄ n-ká-kayu **ñutu***
PL tree there COMPL-PL-burn **tree**
'The trees over there are burnt.'

Rapid speech contraction does not explain the reduced form of this pronoun. As we saw in section 30, when nouns are phonetically reduced, they retain the first syllable and lose the second part of the couplet. This process is reflected in the word *kibi* 'day', which often contracts to *ki*. As a result, if the word *ñutu* were to be contracted, we would expect the appearance of the morpheme *ñu* instead of *tu* as we find in the example above.

(108) **máa kete tichanko liyii ku ndáa=**ki** nu toto*
SPEC dig opossum old go climb=**?** at rock
'The old opossum digs (holes) and the animal climbs up to the rock.'

### 6.2.1 Clitics as Subject

Clitics most commonly occur as subjects of verbs, in post-verbal position. They attach to the last element of the verb phrase, whether that is a verb or another part of speech. The examples below illustrate two cases where the clitic is attached to different elements.

(109) *be'e iyo=**nì** de nde-ja'a=**ya***
house COP=**1.FAM** and ITER-begin=**3.DEI**
'I was at home, and it began raining again.'

(110) *ndúù kúni ká ka'a ndí'i kuiti=**o** mita bi*
NEG can ADD speak smooth very=**INCL** NOW AFF
'We really cannot speak pure (Mixtec) anymore.'

In (109) the clitics *nì* and *ya* are attached directly to the verbs *iyo* and *ja'a*. In (110), however, the inclusive clitic *o* is attached to the adverb *kuiti* 'very', the last element that belongs to the

verb phrase. These two examples also demonstrate that these morphemes are, in fact clitics, and not inflectional suffixes of the verb. If that were the case, (110) would not be possible and the morpheme would need to immediately follow the verb.

Outside of the VP, they can also appear bound to quantifiers and numerals in subject (or object) position. In these cases, a clitic usually does not attach to a VP element:

(111) ***ndúù*=*yó*** *kasiki tilúu*
**both**=**INCL** play round
'The two of us play basketball.' (elicited)

(112) *ká-sa'a tiñu* ***ndaka*=*yó*** *ja ma=náa tú'ũ ká-ka'a=sã*
PL-make task **all**=**INCL** COMP NEG.POT=disappear word PL-speak=1.RES
'All of us work so that the language that we speak does not disappear.'

### 6.2.2 Clitics as Direct Object

Clitics can be attached at the end of the verb phrase after the subject clitic. The examples below show two different clitics in direct object position:

(113) *já'à=de jikáà=de=**sã** nu nana=sã*
go=3.M.RES ask=3.M.RES=**1.RES** at mom=1.RES
'He went to my mom to ask me (to marriage).'

(114) *jito=nì=**ti***
take.care=1.FAM=**3.ANML**
'I take care of it (the animal).'

As can be seen from the examples above, the position of the clitics seems to follow the pattern VERB=SUBJ=OBJ. [1]

### 6.2.3 Clitics as Possessors

As mentioned in the introduction, the same person markers are used to mark possession:

(115) *a ñi'i=nu ndutè=**nu***
QUES obtain=2.FAM water=**2.FAM**
'Did you buy your water (bottle)?'

Similarly to the behaviour of the clitic in the verb phrase, here also the clitic attaches to the last element of the noun phrase. Since the adjective always follows the noun in Mixtec, the pronominal clitic may attach to the adjective instead of the head of the NP, as illustrated in the example below:

[1] This seems to be an innovation that Cuquila Mixtec has developed. Several grammars of other varieties of the area Alexander; Macaulay (1988; 1997: m)ention that the clitics cannot be used to indicate the object.

(116) *yá'á ndeé=nì chi jito=nì chi'ili luli=a=**nì***
here sit=1.FAM because take.care=1.FAM chicken small=this=**1.FAM**
'I live here, taking care of my little chickens'.

### 6.2.4 Clitics as Objects of Prepositions

The same clitics can attach to a preposition, denoting its object. The following examples demonstrate this use:

(117) *ñuyibi=a náa ĩ=**yó***
world=PROX disappear with=**INCL**
'This world (ie. culture) will disappear with us.'

(118) *ndu-ku-ta'bi=yó nu=**ya***
ITER-receive-holy=INCL to=**3.DEI**
'We thank Him (God).'

## 6.3 Independent Forms

The free forms fill roughly the same argument positions as the clitics. That is, they can appear as subjects or objects of the verb phrase or as possessors inside the noun phrase. However, their use as objects of prepositions has not been attested. The following examples illustrate these uses:

**1) Subject of verb:**

(119) ***ndu'u** nee=nì de **ndo'o** iñi=nu*
**1.FAM** sit=1.FAM and **2.FAM** stand=2.FAM
'I am sitting and you are standing.'

**2) Object of verb:**

(120) *a kuu chindee=nu **ndu'u***
QUES COP help=2.FAM **1.FAM**
'Can you help me?'

**3) Possessor of noun:**

(121) *tina **ndu'u***
dog **1.FAM**
'My dog.'

In regards to syntactic environments, when the independent forms function as subjects, they can appear both pre-verbally as well as post-verbally. When the independent forms are present, clitics sometimes appear within the verb phrase. However, the inclusion of the clitics varies

depending on the two positions that the independent forms take. Theoretically, there are four combinations that could appear:

1. [Independent form] [VP]=enclitic

   a. *ndi'i iñi=nu*
      2.FAM sit=2.FAM
      'You are sitting.'

2. [Independent form] [VP]

   b. *ndi'i iñi*
      2.FAM sit
      'You are sitting.'

3. [VP]=enclitic [Independent Form]

   c. *iñi=nu ndi'i*
      sit=2.FAM 2.FAM
      'You are sitting.'

4. [VP] [Independent Form]

   d. *iñi ndi'i*
      sit 2.FAM
      'You are sitting.

The four schematics above show the different combinations of independent pronouns and enclitics; If the free form appears preverbally, the corresponding enclitic can either appear in the VP (combination 1) or the free form can appear by itself (combination 2). Similarly, if the free form appears in post-verbal position, the enclitic can either attach to the VP (combination 3) or not (combination 4).

Looking into my data, it seems that when the free form occurs before the verb, the clitic always appears within the VP (combination 1). Post-verbally the free form always appears by itself, without attaching the enclitic (combination 4). Below I provide an example of each position respectively:

**Comb. 1: [Independent form] [VP]=enclitic**

(122) ***yóo*** *jakuita=**o*** *sakatuni=**o***
**INCL** start=**INCL** mix.up=**INCL**
'We start to mix it up (Mixtec with Spanish).'

**Comb. 4: [VP] [Independent Form]**

(123) *sa'a ni ndo'o* ***yóo*** *ja kuu=yó ñayìbi*
so LIM suffer **INCL** COMP COP=INCL person
'[When our spirit animal suffers] We also suffer the same way, us the people.'

The other two combinations, 2 and 3, do not seem to occur. Constructed sentences where the independent form precedes the verb and no clitic appears, or where the independent form follows the clitic post-verbally, were deemed ungrammatical:

**Comb. 2: [Independent Form] [VP]**

(124) ****ndï'ï*** *iñi*
**2.FAM** sit
'You are sitting.'

**Comb. 3: [VP]=enclitic [Independent Form]**

(125) **iñi=**nu** **ndï'ï***
sit=**2.FAM 2.FAM**
'You are sitting.

Therefore, it seems that, when the independent form appears in pre-verbal position, the clitic needs to be present within the VP. Similarly, when the independent form follows the verb, the bound clitic is deleted.

However, the use of independent forms is very restricted in relation to the use of the enclitics. This is because when the independent forms are used, the clause is marked pragmatically, and the independent pronoun appears in focus position.

### 6.3.1 Independent Forms in Comparative Constructions

The independent person markers can also be used in comparison clauses. This seems to be the only case where an enclitic cannot be used to substitute the free form, instead the independent pronoun is always used. Below is an example of this use:

(126) *yáa=ká=nu ixta masu* ***ndu'u***
eat=ADD=2.FAM tortilla than **1.FAM**
'You ate more tortillas than me.'

In this type of constructions, if the standard of comparison is pronominal, then the free form must always be used. This is also the case for clauses with a predicate adjective:

(127) *kuachi ká=sã masu ká* ***ndi'i***
small ADD=1.RES than ADD **2.RES**
'I am younger than you.'

The data seems to suggest that comparative of likeness, when the subordinate phrase is a fragment clause, also requires the use of the independent form. Below I present the only example that has appeared in the corpus:

(128) *kata keja'a ndáa ma=yó kaxiko=yo ta* ***ndi'i***
immediately start PL SPEC=INCL sell=INCL like **2.RES**
'We are just starting to sell like you.'

Since what follows the conjunction *ta* is only the subject of a clause, it seems logical that only an independent pronoun can occupy that slot, as there is no NP or VP for the clitic to attach to, and there is no evidence that the enclitic can also attach to conjunctions.

## 6.4 Plural Person Marking

Since the bound clitics do not encode any information on number, plurality is marked through the plural affixes on the verb (see section 5.2.4.2), or the plural marker *ndáa* for possessive marking on nouns (as shown in section 5.2.1.5). Therefore, in order to form the first person plural, the appropriate clitic is attached to the VP and the plural affix is added on the verb, as shown below.

(129) ***ká****-sunde=sã*
**PL**-remove=1.RES
'We remove.'

The difference between this type of first person plural and the inclusive clitic *=yó* is that the former encodes an exclusive type of plural. The distinction is made clear in the examples below:

(130) *chi cosa buena kua sa'a=****yó***
because *cosa buena* is.what make=**INCL**
'Because what we do is a good thing.'

(131) *ñúkuà̀* ***ká****-sa'a=****sã*** *nduja*
there **PL**-make=**1.RES** pozole
'Then, we make the pozole.'

In the two examples above, the same verb, *sa'a* 'make', is used. Example (130) is extracted from a conversation between the two women who are sewing. The speaker says that they are not doing anything wrong, meaning both the speaker and the hearer, which licenses the use of the

inclusive marker. In (131), a speaker was providing several recipes, explaining certain traditional dishes are prepared. In all cases, she used the exclusive form of the person marker, as I (the hearer) did not form part of the process.

Seeing that the inclusive marker *=yó* already encodes plurality, we can assume that it does not require the appearance of the plural marking on the verb. Indeed, this is the case in most attestations of the marker in the corpus. However, there are also quite a few instances where the verb is marked for plural and the inclusive clitic is attached to the VP. An example of this use is shown below:

(132) *ká-iyo ba'a=yó kuu kue'e* ***ká****-iyo=****yo***
PL-COP good=INCL COP slow **PL-COP=INCL**
'We are doing alright.'

No pattern has been found on the use of the *ká-VP=yó* construction and its semantic difference with the *VP=yó* construction. The speakers translate the clitic in both cases in the same way, without pointing out any distinctions. Additionally, in all the elicitation tasks and translations, the speakers never provided this construction, but instead always used the verb without a plural affix. Further research is needed to find out if there is a difference between the two, or if both forms are equally accepted by the speakers.

Plural marking does not occur on the independent forms. These forms cannot take the plural marker *ndáa*, as it occurs with nouns. Instead, the clitic attaches to the plural marker, which can then occupy all the same slots as the independent forms. The plural word is often contracted to *ndá*.

(133) *su baji=nì chi yú'ú ndetì=nì* ***ndáa=nu*** *jikáà ña'a*
AFF. come=1.FAM because fear a.lot=1.FAM **PL=2.FAM** ask woman
'I came because I was very afraid of you (all) who ask women (to marriage).

(134) *kúni=i chindee=i* ***ndáa=yó***
want=3.GEN.HUM help=3.GEN.HUM **PL=INCL**
'He wants to help us.'

This is also the case for the third person pronouns, where the clitic can attach to the plural marker and occupies the same slots as the independent pronoun:

(135) *kana'a nduku=de na-nde kuá'ã no'o xunu* ***ndáa=de*** *inka kuiya*
again search=3.M.RES any-? go return instead **PL=3.M.RES** other year
'They are looking for someone to go instead of them next year.'

## 6.5 Pragmatically marked use of personal markers

Personal markers are very often used in pragmatically marked positions. Two possibilities exist to mark constituents: independent pronouns, or the use of the specifying marker *máa* (often contracted to *ma*) with a bound clitic.

Independent pronoun:

(136) ***ndu'u*** *nee=nì* *de* ***ndo'o*** *iñi=nu*
**1.FAM** sit=1.FAM and **2.FAM** stand=2.FAM
'I am sitting and you are standing.'

*Máa* marker + clitic:

(137) ***ndáa máa=i*** *kana-nduku ñayìì* *no'o nduna* *na* *nduku ká*
**PL-SPEC=3.GEN.HUM** ?-search people go nobody any search ADD
'Only they look for people to go, nobody else searches.'

Looking into the instances of both types of pronouns in the corpus, it seems that the independent pronouns are mostly used to mark the topic of a clause. For convenience, I will repeat here (122):

(138) ***yóo*** *jakuita=ó* *sakatuni=ó*
**INCL** start=INCL mix.up=INCL
'We start to mix it up (Mixtec with Spanish).'

Admittedly, it is often times difficult to distinguish between focus and topic. Mixtec uses the same basic strategy for both: the fronting of the argument in a pre-verbal position. However, we can look into the other parts of the clause to get some hints on whether the fronted element is focused or topicalised. As Lambrecht(1994) explains, topicalised elements which are found in extra-clausal positions usually require an intra-clausal pronominal morpheme. Logically, if the constituent is found outside of the clause, it follows that a referent of it should be found within the clause itself. In (138) we can observe such a case: an independent pronoun is found in preverbal position (*yóo*), and a clitic pronoun is attached to the verbs (*ó*). The context of the conversation does not provide any reason for the subject to be focused; the speaker was referring to the Triquis, a community nearby, who, when they get together, speak fluent Mixtec whereas the people in Cuquila already start mixing Mixtec with Spanish. Pragmatically, the sentence does denote contrast between 'us' (the people from Cuquila) and 'them' (the Triquis) but the subject above is not necessarily focused. In order to confirm this hypothesis, a wh-question test could be carried out, in order to find out whether this sentence could answer the question 'Who starts mixing Mixtec with Spanish'. If the same sentence as in (138) is produced,

then the hypothesis is invalid. However, if a different structure is used, it means that *yóo* is not in focus position but is, in fact, a topicalised subject. As mentioned above, an independent pronoun in preverbal position without a clitic attached to the verb is deemed ungrammatical by the speakers. It seems that, when independent pronouns appear in pre-verbal position, they always mark the topic of the conversation.

On the other hand, as Macaulay (1993: 19) points out, "focused constituents [...] get to pre-verbal position by movement". This is contrary to the case of the topics, which are doubled outside of the clause. Focused elements retain their status as arguments of the clause, even if they have moved to a different position within it. For this reason, pronouns with the same referent are not found within the same clause (Macaulay 1997: 105). Indeed, this seems to be the case when the exclusive marker *máa* with a bound clitic is found in preverbal position, as seen in (137). Note in this example that no clitic is attached to the verb *kananduku* 'to search', but instead *máa=i* 'only they' seems to function as its subject.

Consequently, the preliminary hypothesis is the following: If an independent pronoun appears in pre-verbal position, it always denotes the topic of the discourse, with the added pragmatic effect of contrast. If the exclusive marker *máa* appears in preverbal position followed by a bound pronominal clitic, it denotes a focused argument.

The issue that arises here is the fact that clitic pronouns are sometimes omitted in the third person or if the subject is clear enough from the context. Thus, it could be argued that, in fact, a clitic could in fact appear in the sentence above, which would invalidate this hypothesis. The task of testing the hypothesis proves even more difficult if we take into account the fact that the marker *máa* is widely used to denote exclusivity (approximately meaning 'only X', or 'X himself'), regardless of pragmatical markedness. In these cases, the construction of the exclusive marker and a bound clitic often appears with verbs which have a co-referential pronominal clitic attached to them (in most of these cases, however, the marker *máa* with the clitic appear in post-verbal position). In order to test this hypothesis, a grammaticality judgement test could be carried out to see how the same sentence with the clitic *i* attached to the verb would be judged by the speakers. Alternatively, a substitution test could be carried out, replacing the third person pronoun with a first or second person, as the first and second person clitics are rarely (if ever) omitted.

## 6.6 Semantic distinctions on the third person pronouns

In Cuquila Mixtec certain nouns can be anaphorically referred to using seven different pronouns: *i*, *de*, *ña*, *ti*, *tu*, *de* and *ya*. When a noun is to be replaced, the speaker needs to make a choice among these forms, based on the semantics of the noun. As we have already seen, these pronouns overlap with the classes of nouns that Cuquila Mixtec has. In what follows, I

will provide more information on the context in which each of the third person pronouns are used.

**1) *i***

The pronoun *i* is used when referring to adults with whom the speaker has a certain degree of familiarity. Its use is often attested among young people of the same age and friends. It is also commonly used to refer to children and babies. The example below illustrates the use of this pronoun when the speaker was referring to her 5-year-old granddaughter:

(139) *Aurora tá'á tuni anu=**i*** *kũ'=**i*** *bestido jíté lúù*
Aurora like very heart=**3.GEN.HUM** wear=**3.GEN.HUM** *vestido* wide pretty
*na'=**i*** *kachi=**i***
seem=**3.GEN.HUM** say=**3.GEN.HUM**
'Aurora likes to wear the wide dress (because) she says she looks pretty.'

The devil and the angels are also referred to with this pronoun. Even though all deities are referenced using the deity pronoun *ya*, these two seem to be regarded humans. One speaker told to me that the angels are regarded as babies, which would explain the use of *i* to refer to them. In the example below, the speaker was retelling the time when an angel visited her:

(140) *n-kíi* *kundee=**i*** *ndu'u*
COMPL-come see-**3.GEN.HUM** 2.FAM
'It (the angel) came to see me.'

Interestingly, references to the earth also make use of the generic human pronoun *i*. The following example comes from a speaker who was explaining the earth rituals. Before sowing the corn or building a house, an elder from the community asks permission from the earth in order to start working. The ritual involved pleading the earth to bless the work and giving it food and drink:

(141) *xinañu'u kúni kuá'ã=yó na ndutè ko'o=**i***
first need give=INCL any water drink=**3.GEN.HUM**
'First, we need to give it (the earth) something to drink.'

**2) *de* / *ña*:**

The markers *de* and *ña* are used when the speaker refers to a male or a female person respectively. Hollenbach (2015) glosses these pronouns in all Mixtec varieties as "adult", as opposed to *i* which she glosses as "child". In Cuquila Mixtec, however, the choice of these pronouns over the generic pronoun *i* is not based on age alone. It is used when referring to persons of a higher social status than the speaker, or for elders. As one speaker explained, if

the polite form *ní* 'you' will be used to address someone, then the pronouns *de / ña* will be employed to refer back to that same person. The difference between the use of *de / ña* and *i* can be observed in the examples that follow:

(142) *mitañu ndu=bi=**de** chi kuá'ã=**de** satiñu=**de**.*
now NEG=COP=**3.M.RES.** because go=**3.M.RES.** work=**3.M.RES.**
'He is not here now because he went to work.'

(143) *ìyo uú se'é=nì ĩ̃ĩ̃ naáni=**i** Roberto ĩ̃ĩ̃ naáni=**i***
COP two child=1.FAM one be.called=**3.GEN.HUM** Roberto one be.called=**3.GEN.HUM**
*Carolina.*
Carolina
'I have two children, one is called Roberto, the other is called Carolina.'

In (142), the speaker is talking about her husband, who is 40 years old. She refers to him using the male pronoun *de* to show that she has respect for him. When asked if she could have used the generic pronoun *i* she immediately rejected it, as it would have been impolite. In the second example, the same speaker talks about her children, using the generic pronoun *i*. It should be noted here that the children are adults (32 and 27 years old), however she has enough familiarity with them that she can use this pronoun. When asked if she could have used *de* and *ña*, she agreed, but added that it would sound too formal. This use of the pronouns contrasts Hollenbach's (2015) analysis that *de and ña* are used with adult referents; the use follows the same rules of politeness as the rest of the respect pronouns.

**3) *ti*:**

The pronoun *ti* is used to refer to animals, and it is reduced form of *kiti* 'animal'. The following example is extracted by an interview, where the speaker was explaining how she takes care of her animals:

(144) *taá=nì nùnì xuchi ja yasíni=**ti** ja kusaáma=**ti***
throw=1.FAM corn broken COMP eat.breakfast=**3.ANML** COMP eat.lunch=**3.ANML**
*ja kuxíni=**ti***
COMPL have.dinner=**3.ANML**
'I put broken corn for them to eat for breakfast, lunch and dinner.'

**4) *tu*:**

The marker *tu*, which originates from the word *ñutu* 'tree', is used to refer anaphorically to trees. Example (145) shows its use in context:

(145) *yaa iñì ĩ̃ĩ̃ ñutu mitu ja xíchí=**tu***
here stand.COMPL one tree but COMP dry=**3.TREE**
'There used to be a tree here, but it dried out.'

**5) *de*:**

The marker *de* is used to make reference to any sort of liquid. It is a contracted form of the word *ndutè* 'water'. It is homophonous with the third person male clitic *de* 'he'.

(146) *kúni=nì ko'o=nì ndutè ji'ini mitu ja ndiko=**de***
want=1.FAM drink=1.FAM water warm but COMP cold=**3.LIQ**
'I wanted to drink the coffee, but it got cold.' (elicited)

**6) *ya*:**

The marker *ya* is used to refer to deities such as God, Jesus, the Virgin Mary etc[2]. The full noun that the marker is related to it *yandiuxi* 'God' (from Spanish *Dios*). Example (147) is extracted from a conversation about the rainy season. The speaker makes reference to the year before, when there was not enough rain and the crop did not turn out well:

(147) *tá kuiya akua desu x-ndó'o na'ni=**ya** lulu=**ya***
when year old and CAUS.-suffer a.little=**3.DEI** small=**3.DEI**
'Like last year that He made (us) suffer a little bit.'

The same marker is also used to refer to the rain, as it is of vital importance to the community. The vast majority of the community has corn fields which are necessary for their substinence. If the rainy season does not yield enough water, they run the risk of not having enough food during the dry season. In the same conversation as (147), the interlocutors discuss this year's rain season:

(148) *tu kũũ=**ya** sũkua kũũ=**ya** mitañu su ni'i=o ja-kuteku inka tiempo*
if fall=**3.DEI** like fall=**3.DEI** now AFF obtain=INCL NMLZ-live other *tiempo*
'If it continues raining as it is raining now, we will obtain food for another season.'

One speaker also mentioned that the clouds are referred to using the deity pronoun *ya*, giving the phrase in (149). She explained that both the clouds and the rain are above us, just like the deities, and so the same marker is used (also in contrast to the earth which, as mentioned above, takes the generic human pronoun *i*). However, other speakers did not recognise this phrase as correct, using instead no marker at all to refer to the clouds.

(149) *ndeé=ya*
sit=3.DEI.
'It is cloudy.' (Lit. 'it is sitting')

[2]I was told that it is also used to refer to priests, but I was not able to obtain any examples of this use.

## 6.7 Lack of Person Markers

In many cases person marking on the verb is entirely omitted. In what follows I will analyse the four different circumstances under which a clitic marker is not used.

Firstly, we have just seen that in Cuquila Mixtec the assignment of a third person marker to nouns is based on specific noun categories. However, many nouns are not referenced back with any overt anaphoric marking on the verb. In fact, anything that does not fit the seven separate categories that correspond to the third person markers is not marked on the verb; anything that is not an animate (masculine, feminine, general human and animal), a liquid, a tree or a deity. The following two examples demonstrate such cases. In (150), the speaker was making a *tlacoyal* (hair tie) and was complaining that it was taking too long to finish. So, the speaker said:

(150) *su n-kúní jinù numí*
AFF COMPL-want finish quick
'It doesn't want to finish quickly.'

The sentence in (151) came up during a conversation about the newly re-established market, which had not taken place in the last 10 years:

(151) *tá ká xeo kí kibi ndukaba kí*
when ADD fifteen recently day restart recently
'(The market) resumed just 15 days ago.'

In the examples above we can observe that the subjects (the hair tie and the market, respectively) do not fit any of the seven categories that have a separate person marker. Therefore, there is no reference to the subject within the VP and it can only be inferred by the context in which the sentence appears. If the subject needs to be referenced again in the discourse, the full NP appears, always followed by the demonstrative clitic *=u* 'that'.

(152) *tankuniji ja n-cho'o* ***nùnì=u***
dawn already COMPL-cook **corn=DIST**
'At dawn, the corn is already cooked.'

Secondly, when a nominal subject or an independent pronoun follows the verb, no clitic appears in the VP, as we have already seen in section 6.3:

(153) *kéja'a* ***ndáa teé*** *kasiki tiluú*
start **PL man** play round
'The men start playing basketball.'

In the example above, the nominal subject *ndáa teé* appears immediately after the verb, and

so the pronominal clitic is omitted.

Next, a verb with a third-person human subject can sometimes lack a clitic. Analysing the cases where this occurs in the corpus, it seems that the lack of person marking happens when the speaker wants to intentionally leave the subject vague, or to speak generically.

(154) ***ká-ki'i=ta*** *ká-ki'i=ta* *su* *ba'a* *ne'e*
**PL-buy=PRED.FOC** PL-buy=PRED.FOC AFF good very
'They do buy indeed, that is very good.'

(155) *kúni* *ká-kata'bi* *xinañùu* *nu* *ñu'u* ***ká-kachi***
want PL-plead first at earth **PL-say**
'It is important to first ask the earth (for permission), they say.'

These two examples show that the speaker intentionally does not make the male/female and familiar/respect distinction. If the speaker were to use a pronoun, they would need to make a choice among the three options: *i*, *de* and *ña*. The use of *i* immediately reflects a level of familiarity with the referents that the speaker might not have. Similarly, choosing *de* or *ña* to refer to a group of people (as is the case in the examples above) shows that the speaker is respectful towards everyone. Therefore, by omitting the use of the pronoun, the speaker is intentionally being vague about the referent.

Lastly, when the subject of the clause is easily deduced by context, or has been mentioned in the immediate co-text, the pronominal clitic can be deleted. The dialogue that follows forms part of a conversation between two speakers who were making jokes about sitting on the side of the street:

(156) *nde* *yu'u* *carretera* *ja* *ku* *nde'e=**yó***
until edge *carretera* COMP go sit=**INCL**
'We'll end up sitting at the edge of the road.'

(157) *nde* *yu'u* *conasupo* *kuá'ã* *nde'e*
until edge *conasupo* go sit
'We will end up sitting at the super market.'

The second iteration of the joke (157) follows the same structure as the first, denoting a location even further away from where the speakers were sitting. By this time, it was already clear that they were talking about themselves, so the pronominal clitic could be deleted.

## 6.8 Third Person Pronouns and Gender

Many researchers have analysed the third person pronouns as exhibiting grammatical gender. So, Macaulay (1997: 82) talks about a 'fairly limited gender system', Hollenbach (2013: 51) says that 'Mixtec includes various genders' (own translation), and Alexander (1988: 257) analyses the

pronouns as having 'gender classes.' Drawing on this, WALS (Corbett 2013) adds Chalcatongo Mixtec on the map as having more than 7 genders. However, this analysis has two problematic areas.

Firstly, for languages that have grammatical gender, each noun are obligatorily assigned one of the genders. Dixon (1986: 106), while discussing the difference between noun class systems (of which gender is a sub-group) and classifiers, mentions: "noun classes involve a grouping of *all* the nouns of a language into a smallish number of classes [...]" (emphasis in the original). In contrast, Cuquila Mixtec only assigns certain nouns into one of the categories characterised as 'gender'. The semantics of these categories, as were explained previously, are quite specific and do not apply to all the nouns of the language. Admittedly, it is not easy to find out whether membership in these categories is obligatory in Cuquila Mixtec. There is no agreement between the noun and its modifiers (the adjectives and demonstratives always have the same morphology), there are no verb cases or other inflectional elements that would indicate noun class membership. As such, the only place we can look for such an agreement is on the pronoun use. As we saw in section 6.7, in many cases there is complete lack of a clitic or an independent pronoun. In the vast majority of instances, this occurs when the known referent does not fit in any of the above categories and cannot be cross-referenced as such. The nouns that fit into the categories described in 5.2.1 are almost always assigned a pronoun. The few cases where a noun is not co-referenced, while it would be possible to do so, are instances where the referent has a highly active status (per Chafe (1987)), having appeared in the immediate co-text. Secondly, this approach creates a typological rarity that is difficult to explain. If we assume that Mixtec has grammatical gender, then it is the only language in Mesoamerica to have this category (Corbett 2013). Furthermore, it is the only language family to have upwards of seven gender distinctions in the whole of the Americas (Hollenbach (2015b) shows up to 14 distinctions in some Mixtec varieties) [3].

The use of these pronouns simply reflects the system of noun categorisation that Cuquila Mixtec exhibits. It would be interesting to understand the origin of the semantic basis of these noun categories. Many languages in the world assign nouns into categories based on distinctions such as animate vs. inanimate, human vs. non-human, masculine vs. feminine and so on. However, these binary distinctions do not seem to fully explain the system in Cuquila Mixtec. Animacy certainly plays a crucial role in the noun categorisation, as four out of the seven categories apply to animates. Nevertheless, it is not the most salient characteristic, as the remaining categories include *certain* non-animates, deities, liquids and trees, but not the rest of inanimate things. The nouns that belong to these categories are clearly of cultural importance to the speakers. Water is a sacred element, found in many

[3]The closest languages in terms of number of gender in the continent are Pirahã and Paumarí in Brazil, which only have four genders (Corbett 2013).

stories and rituals. Several types of liquids are offered to the earth as gifts during the ritual that is performed before the ground is manipulated, through sowing or the building of a house. The sacred nature of the water can be further observed by the fact that the rain belongs to the same semantic category as the religious deities and is co-referenced using the deity pronoun. At the same time, seeing that the nature in and around the town forms a significant part of the residents' life, it is easy to understand why trees would be categorised separately. Trees provide warmth and energy for cooking, while at the same time they protect from floods and supply vital elements for their life. Thus, it seems that cultural relevance plays a major role in the semantic distinctions of the noun categories in Cuquila Mixtec.

In summary, The pronoun system is characterised by the existence of four persons (first, second, third and inclusive). Number is not marked on the pronouns themselves, but instead on the verbs or the head nouns. Politeness is marked through the use of respect forms in first, second and third person. The special humbling form found on the first person reflects respect towards the addressee. The pronouns appear as clitics, attaching at the end of the VP or NP, or as independent forms. These free forms are mostly used to mark the topic of the utterance in sentences with pragmatically marked word orders. Finally, the seven third person pronouns reflect the way nouns are categorised in the language, showing the prominence of cultural relevance.

# Chapter 7

# Grammaticalization of Person Markers in Cuquila Mixtec

This chapter focuses on the grammaticalization of the personal pronouns in Cuquila Mixtec. Specifically, the historical development of the pronominal forms will be analysed. In order to provide the relevant context for this discussion, pronoun systems found in other Mixtec varieties will be described in section 7.1. After providing some background on the theory of grammaticalization in section 7.2, I will provide some information on the development of the Cuquila Mixtec pronouns based on information available for other Mixtec varieties in section 7.3.

## 7.1 Person Marking in relation to Mixtec languages

The pronoun system shows significant diversity among the Mixtec varieties. The main difference between the dialect areas is the respect/number distinction. Highland Mixtec is characterised by the familiar/respect distinctions in first and second persons, with some varieties exhibiting this contrast also in the third person. The system used by Cuquila Mixtec shows many similarities with the systems of the same dialect area. All the varieties spoken in the Western Highland region make use of two separate pronouns for the familiar and the respect forms both on the first and on the second person. On the third person, the distinction between human and non-human referents persists in all varieties. At least two forms for human (masculine and feminine) and several forms of non-human are attested everywhere. In the non-human category, all varieties distinguish at least 3: animal, wood and liquid. However, some exhibit more distinctions for inanimate referents than others. In many varieties we find separate pronouns for the deceased, as well as for flowers. The following table shows the third person pronoun paradigm from San Juan Teita Mixtec, a variety spoken in the same area as Cuquila Mixtec, the Western Alta region.

**Table 7.1: Third person pronoun paradigm of San Juan Teita Mixtec (Hollenbach 2015b: 13)**

| Gloss | Clitic | Full Noun | Translation |
|---|---|---|---|
| 3.M.ADULT | *te* | *tee* | man |
| 3.F.ADULT | *ña* | *ña'a* | woman |
| 3.CHILD | *i* | *duchi* | child |
| 3.MX | *i* | *ñayi̱vi* | person |
| 3.ANML | *ti* | *kiti* | animal |
| **3.SPHER** | ***ti*** | ***kiti*** | **animal** |
| 3.DEI | *ya* | *i̱ya* | deity |
| 3.WOOD | *nu* | *xujnu* | tree |
| 3.WATER | *te* | *ndute* | water |
| **3.FLOW** | ***ta*** | ***ita*** | **flower** [1] |

In the table above, we can find a few pronouns that do not appear in the paradigm of Cuquila Mixtec. The pronoun *ta* is found to be used for flowers, which is a contracted form of *ita* 'flower'. Furthermore, spherical objects are referenced using the pronoun *ti*, which originates from the word *kiti* 'animal', just like the animal pronoun does. This is, in fact, found in several other varieties of the area, such as Magdalena Peñasco, Yosoñama and Ocotepec Mixtec, where it is also used to refer to fruit [2] . De Leon Pasquel (1988: 135) mentions several semantic extensions in the elements used as pronouns in Mixtec:

"The nouns also undergo semantic extensions to include general kinds of referents:

[...]

Animal —> Round Fruit —> Round shape."

In contrast to Highland Mixtec, the Lowland and Coastal varieties show a number distinction in the first and second pronouns, rather than the politeness distinctions that are characteristic of the Highland areas. Therefore, languages such as Silacayoapán Mixtec (North & Shields 1978) and Ayutla Mixtec (Hills 1990) in the Lowlands and Santa María Zacatepec (Towne 2011) in the coastal region have different forms for singular and plural on the first and second person, instead of familiar and respect. The following table shows the first and second person paradigm of Alacatlatzala Mixtec, spoken in the Lowland region of Guerrero (Hollenbach 2015b: 37):

[2] Alexander (1988) does not mention this particular use of the animal pronoun, however data previously collected by a consultant from Santo Tomas Ocotepec clearly show that =ti is used anaphorically to refer to fruits and vegetables, as well as spherical objects like balls. Furthermore, people from Cuquila mentioned that they found it "funny" when vegetable vendors from Ocotepec would use this pronoun referring to produce like tomatoes and oranges in the Saturday market of Tlaxiaco, which provides evidence that it is actively used in Ocotepec, but not in Cuquila.

**Table 7.2: Third person pronoun paradigm of Alacatlatzala Mixtec**

| Gloss | Clitic | Independent Form |
|---|---|---|
| 1.SG | *ì* | *yì'ì* |
| 1.PL.EXCL | *ndì* | *ndì'ì* |
| 2.SG | *ún* | *yó'ó* |
| 2.PL | *ndó* | *ndó'ó* |
| INCL | *yó* | NONE |

From the table above we can observe that, instead of the separate familiar and respect forms, the varieties of the Lowland and Coastal areas exhibit distinct forms for singular and plural. Furthermore, the inclusive person persists these languages, meaning that the four-person distinction is common across all varieties of Mixtec. Therefore, the first person plural pronoun *ndì* in the variety described above acquires an exclusive connotation. In the third person, roughly the same distinctions mentioned previously seem to be retained everywhere.

An interesting system seems to have developed in San Juan Coatzospán Mixtec in the Eastern Highland region, where different pronouns are used for the third person male reference depending on the gender of the speaker; men use a different pronoun to refer to other men than women do. Hollenbach (2015: 19) proposes that, due to the town's geographical isolation from other Mixtec-speaking communities, this system seems to have developed in a different way. Unfortunately, there is not much information regarding the origins of these separate pronouns, which hinders the task of finding cognates of these forms in other varieties.

The next section will focus the historical development and relationship between the two paradigms found in the Mixtec languages: respect/familiar and singular/plural.

## 7.2 Theoretical Background

Grammaticalization is defined as the process of the development of grammatical forms from lexical sources, as well as the shift of items from a less grammatical to a more grammatical status (Lehmann 1995: 13). Even though the study of grammaticalization of various forms has been established for a long time now, with many theoretical frameworks developed, the domain of personal pronouns in terms of grammaticalization has been somewhat neglected. However, the past few years have seen some major studies on this domain (cf. Baht 2004, Heine & Song 2011, Helmbrecht 2004). Heine & Song (2011) provide a good overview of the main conceptual sources which serve to develop personal pronouns across a variety of languages from a typological perspective, as well as the diachronic sequence of changes that

take place during this grammaticalization process.

There are four parameters in the grammaticalization process that also play a role in the development of the pronouns. The diagram below summarises these processes in their diachronic sequence:

**Extension** ⟶ **Desemanticization** ⟶ **Decategorialization** ⟶ **Erosion**

**Figure 7.1:** Sequence of changes during the grammaticalization process (based on Heine & Song (2011))

The grammaticalization process begins when words extend their scope to encompass new concepts. During this change, which is pragmatic in nature, a group of speakers innovate the use of the original linguistic expression into new contexts. This new use of the word is then gradually adopted by the speech community. An example of this stage is the extension of some body part terms into spatial expressions.

This initial stage triggers another change in semantics. Through the process that has been called 'desemanticization', the original meaning of the words is reduced or entirely lost, giving way to the new meaning acquired through extension. This is the case of the English definite article *the*, which derives from the demonstrative *that*. At some point during the grammaticalization, the original component of deixis was lost, retaining only the meaning of definiteness.

After a word has lost some of its semantic content, certain grammatical properties might not be useful anymore. At this stage decategorialization occurs, whereby morpho-syntactical changes take place. The word might lose its ability to be inflected or to be a free-standing morpheme. For example, when the English demonstrative *that* was extended into a relative marker, it lost its ability to be inflected for plural (*those*), as this grammatical property was not relevant in the new context.

Finally, words that acquire new meanings and lose morpho-syntactic properties often undergo changes on the phonetic level as well. Through the loss of segments or phonetic autonomy, these words can be reduced into affixes or clitics. Therefore, in the life cycle of grammaticalization of personal pronouns all domains can potentially be affected, from the pragmatic down to the phonetic level.

An example which covers all the stages of the grammaticalization process is the case of the Spanish polite pronoun *usted*, which is historically related to the honorific title *Vuestra Merced* 'Your Grace' (Heine & Song 2011: 606). Even though it was originally used to address the king, through extension and desemanticization its use was expanded to include members of the elite and the bourgeois society. Having lost both grammatical and phonetic content,

this expression was reduced to *usted*, which is nowadays used as the polite form in every social level. Furthermore, the grammaticalization process has continued in Latin America, where the plural form *ustedes* has lost its politeness meaning and is used as the general second person plural form ('you all').

This grammaticalization path is very often found in the development of the personal pronouns. Reconstructing the origins of the pronoun forms of different languages across the world, several patterns emerge regarding the linguistic expressions that are commonly used as sources. Before delving deeper into this subject, a distinction needs to be made between the third person pronouns and first/second person pronouns. As Helmbrecht (2004: 313) explains, the latter two, and especially the first person pronouns, are not reconstructable any more in the vast majority of languages, as these forms are old and fairly stable from a diachronic point of view, hindering the use of the historical-comparative method or their internal reconstruction. Indeed, both Helmbrecht (2004) and Heine & Song (2011) in their typological studies discuss the origin of the polite and humbling forms much more in depth than the familiar forms for the first and second person.

Third person pronouns are, typologically speaking, often derived from demonstrative pronouns and nouns. Furthermore, Helmbrecht (2004) shows that verbs of saying can also function as a source for such pronouns, as is the case for some Chadic languages. Romance languages developed the third person pronouns on the basis of demonstrative pronouns in Latin, a process which, according to Diessel (1999) is commonly found across the world. Lehmann (1995: 40) explains this grammaticalization process in some languages, whereby the demonstrative pronoun loses its strict semantic component of deixis and its meaning is reduced to that of definiteness, thus functioning as a definite article. From there, it is further reduced to a class marker, and finally to a free personal pronoun. Through the process of phonetic erosion explained above, it is easy to see how the free forms then become clitics and affixes.

The next largest source for personal pronouns on the third person are nouns. Abstract nouns such as 'man', 'woman' and 'person' often serve as the basis for these pronouns. The path that Helmbrecht (2004) suggests is the following: such abstract nouns that, in the beginning, function as modifiers to other nouns may undergo changes in their morphosyntactic properties and serve as classifiers and, later on, as determiners. At this stage, the determiner, which is semantically reduced to marking definiteness and gender values, may be used without the head noun anaphorically. This last stage gives rise to the anaphoric pronouns. Heine & Song (2011) also mention that certain nouns which signify social status may also be used as a source for politeness distinctions in third person pronouns. Terms such as 'royalty', as well as 'family' or 'creature' can serve as the basis for distinctions of honorification levels in several languages.

Second person polite pronouns very often derive from other pronouns, demonstrative

pronouns or certain nouns. Third person plural and second person plural pronouns are the most common sources for the second person polite form in many languages. In a process that Heine & Song (2011) call 'plurification', the use of the plural pronouns, whether second or third person, is extended to single referents in order to express honorification. Various explanations have been proposed for the use of plural forms in polite speech. Brown & Levinson (1987) argue that the use of the plural forms are less threatening to the addressee than the singular forms. Similarly, Helmbrecht (2004) mentions that plural forms allow for a more indirect reference to the addressee.

Apart from plural person pronouns, the third person singular, as well as demonstrative pronouns are often times used, either directly or as sources for the second person polite forms. Furthermore, nouns may serve as the basis for the formation of both second person polite and first person humbling forms. Status nouns and kinship terms, such as 'your honour' or 'brother' are used to elevate the status of the addressee, and nouns denoting notions such as 'servant' or 'slave' may give rise to humbling forms which lower the social status of the hearer.

## 7.3 Grammaticalization of Cuquila Mixtec Person Markers

### 7.3.1 First and Second Person Pronouns

The origin of the first and second person pronouns in the Mixtec languages is not well known. As mentioned previously, difficulties in reconstructing the origin of these pronouns is not rare across languages. To my knowledge, Hollenbach (2015a) has been the only study so far to attempt to reconstruct the personal pronouns in Proto-Mixtec. According to the typological studies and grammaticalization theories mentioned above, third person pronouns and plural pronoun forms would be a good place to start in search of the sources for the proto-forms. As we saw 6.6 the third person pronouns clearly derive from nouns and do not show any connection to the first and second person forms. In terms of plural forms, as we have seen (section 6.4) there is no number distinction in the personal pronouns, as plural is marked either on the verb or the head noun, or via the plural word *ndáa*.

However, Lowland Mixtec varieties, such as Santa María Zacatepec (Towne 2011) and Ayutla (Hills 1990), do make number distinctions. Looking at the first and second person paradigm from the variety of Ayutla (Hollenbach 2015b), we can see certain similarities with the Cuquila Mixtec paradigm:

**Table 7.3: Personal Pronouns Paradigm in Ayutla Mixtec**

| Gloss | | Clitic | Independent Form |
|---|---|---|---|
| 1 PERSON | | | |
| | SINGULAR | *ì* | *yu'ù* |
| | PLURAL | *ndù'* | *ndu'ù* |
| 2 PERSON | | | |
| | SINGULAR | *ùn'* | *yo'ó* |
| | PLURAL | *ndò* | *ndo'o* |

From the paradigm above, we can observe that the plural forms bear a resemblance with the familiar forms of Highland Mixtec. The table below provides a side-to-side comparison of the familiar forms of Cuquila Mixtec with the plural forms of Ayutla Mixtec:

**Table 7.4: Comparison of Pronoun Forms in Ayutla and Cuquila Mixtec**

| Cuquila Mixtec | | | Ayutla Mixtec | | |
|---|---|---|---|---|---|
| Gloss | Clitic | Independent | Gloss | Clitic | Independent |
| 1.FAM | *nì* | *ndu'u* | 1.PL | *ì* | *yu'ù* |
| 2.FAM | *nu* | *ndo'o* | 2.PL | *ùn* | *ndo'ó* |

Indeed, Hollenbach's (2015a) proposal is that the familiar forms of Highland Mixtec originate from the plural forms of Lowland Mixtec. In what follows, I will provide a short overview of the grammaticalization process of the first and second familiar forms in Cuquila Mixtec.

She proposes that the earliest enclitic pronouns that can be reconstructed are the following[3]:

**Table 7.5: Earliest reconstructed pronouns in Proto-Mixtec**

| Form | Gloss | Meaning |
|---|---|---|
| **yu* | 'I' | first person |
| **yo* | 'you' | second person |

Then, the plural forms for first and second person were introduced, which involved the morpheme *ndɨ meaning 'all'. Therefore, the pronoun system was expanded:

[3]In the explanation that follows, I have omitted some steps in the grammaticalization process for the sake of clarity. The omissions involve the rise of forms that are used in other Mixtec languages but are not relevant for Cuquila Mixtec, such as the second person singular form. I have also omitted the stage of differentiation between the second person and the inclusive forms, as it was not relevant for this discussion. I have also decided not to discuss the grammaticalization of the independent forms, as the stages they went through are the same as the enclitics, with the only difference being the phonetic substance.

**Table 7.6: Introduction of plural forms**

| Form | Gloss | Meaning |
|---|---|---|
| **yu* | 'I' | first person singular |
| **ndɨ-yu'* | 'all-I' | first person plural |
| **yo* | 'you' | second person singular |
| **ndɨ-yo'* | 'all-you | second person singular |

At a later stage, the plural forms of the first and the second person were extended to address singular referents indicating respect. This process resulted in a mixed system of both singular/plural and familiar/respect distinctions. The form *ndɨ-yu'* was used both to refer to multiple participants ('we') and as a humbling form for the speaker to refer to himself ('I'). Similarly, *ndɨ-yo'* referred to multiple addresses ('all of you') or to a single hearer with respect ('you'). Lastly, the singular forms of the first and second person were dropped, and the use of the plural forms to refer to a single person were generalised and lost their number and respect connotations. This left the following system:

**Table 7.7: Loss of singular forms and semantic extension of plural forms**

| Form | Gloss | Meaning |
|---|---|---|
| **ndɨ-yu'* | 'all-I' | first person |
| **ndɨ-yo'* | 'all-you | second person |

In Western Highland Mixtec specifically, these pronouns underwent further changes. The two forms lost phonetic substance and were simplified to *ndi and *ndo respectively:

**Table 7.8: Phonetic changes in Western Alta Mixtec**

| Form | Gloss | Meaning |
|---|---|---|
| **ndi* | 'I' | first person |
| **ndo* | 'you | second person |

Further phonetic reductions gave rise to the Cuquila Mixtec forms of *nì* and *nu* [4]. There is no information on the rise of the tone patterns that we see for the first and second person pronouns.

It seems that personal pronouns have shown very little variation since the 16th century. From what we can gather from works published between 1567 and 1593 on varieties of the Western Alta, both the enclitics and the independent forms show a striking resemblance to

[4] Hollenbach (2015a) uses this last step to explain the emergence of the enciltic pronouns of Ocotepec Mixtec, which are identical to the ones of Cuquila Mixtec. Seeing their semantic and morphological resemblance, I do not see a reason why this theory could not be applied to Cuquila Mixtec also, which is why I am adopting it.

the ones found nowadays. Below are the forms for first and second person pronouns found in a catechism published by Hernández (1567) for the variety of Tlaxiaco-Achiutla (as cited in Hollenbach (2015: 3)). For ease of comparison, I have included the paradigm of Cuquila Mixtec.

**Table 7.9: Comparison of Personal Pronouns Paradigm between Tlaxiaco-Achiutla in 1567 and Cuquila Mixtec**

| | Tlaxiaco-Achiutla Mixtec (1567) | | | Cuquila Mixtec | |
|---|---|---|---|---|---|
| Gloss | | Clitic | Independent Form | Clitic | Independent Form |
| 1 PERSON | | | | | |
| | 1.FAM | *ndi* | *ndu'u* | *nì* | *ndu'u* |
| | 1.RES | *sa* | *ñasaña* | *sã* | *sã'ã* |
| 2 PERSON | | | | | |
| | 2.FAM | *ndo* | *ndo'o* | *nu* | *ndo'o* |
| | 2.RES | *ni* | *ndiji* (*ndijï*)[5] | *ní* | *ndi'i* |

In terms of the use of the familiar and the polite forms, Reyes (1593: 14-16) informs that the polite forms are used when speaking to the Spanish elite, the priests and religious figures. It seems that their use was much more restricted than nowadays, seen that they were only reserved to address people with the highest social status. Therefore, Hollenbach (2015a) hypothesizes that, due to the highly stratified society of the post-colonial era, new pronouns were needed in order to address the nobility. Sometime before the 16th century, the forms *san and *ni were introduced to fill this void. Since these forms cannot be reconstructed any further and no antecedent is found within the pronoun system, it seems plausible that these forms originate from nouns. Hollenbach (2015: 20) proposes the word *dzana* (*dana*), found in the compound *dzaya dzana* (*da'ya dana*) 'slave born in the household' as the source for the first person respect pronoun *sã*, and the classifier *ndi* 'late, deceased' for the second person respect pronoun *ní*.

However, all of the above is highly speculative. Firstly, there are no descriptions of the pronoun system in pre-colonial times, and the context in which the data from the 1500's is unknown. As a result, it is difficult to assess whether the polite forms already existed before or, as Hollenbach proposes, these forms arose with the arrival of the Spaniards, since. Furthermore, the source of the second person respect form *ní* is rather dubious. The classifier *ndi* originates from the word *ndiyi* 'corpse', which seems highly unlikely to be the source of a honorific form.

Setting Hollenbach's proposal in the theoretical grammaticalisation context described above, it seems that the parameters provided by Heine & Song (2011) fit the historical development of the familiar pronouns. Firstly, when the plural forms arrived in the Western

Alta region, they underwent an extension of their pragmatic meaning to include single referent. Then they gradually lost their semantic connotation of number, while still maintaining the person reference. At this point, the old plural forms were used to denote single referents. Next, these forms were subject to phonetic changes and became shortened. Following this, another semantic change took place, whereby the use of the polite forms was extended into more familiar contexts. This final stage of the process gave rise to the familiar forms that are found in Cuquila Mixtec today.

### 7.3.2 Third Person Pronouns

The third person pronouns offer a much clearer image of grammaticalisation. As mentioned in 6.6, all third person enclitics derive from nouns. As explained above, nominal concepts are a common source for the third person pronouns. This is also the case in many Mesoamerican languages, and especially within the Otomanguean family (Operstein 2003). The sources of the three pronouns used for human referents, *de* 'he (respect)', *ña* 'she (respect)' and *i* 'he/she (general)' are the abstract nouns *teé* 'man', *ña'a* 'woman' and 'ñayìbì' respectively. Abstract nouns such as the ones seen here appear are the most common sources for third person pronouns, according to Heine & Song (2011: 597). What is interesting in this language, however, is that these same abstract nouns also give rise to the respect forms *de* and *ña*. This goes beyond the cases discussed in Heine & Song (2011), where these sources are used to derive pronouns of "neutral social status".

Helmbrecht (2004: 384) demonstrates how such abstract nouns are often grammaticalised into personal pronouns. According to his analysis, which is explained in section 7.2 in more detail, these modifying nouns first become classifiers and are eventually used as anaphoric pronouns. Through semantic bleaching, they are reduced to only express gender values. This process seems to accurately explain what I believe has happened in Cuquila Mixtec. Figure 7.2, provides a summary of grammaticalization process of the third person pronouns, which I will expand on below:

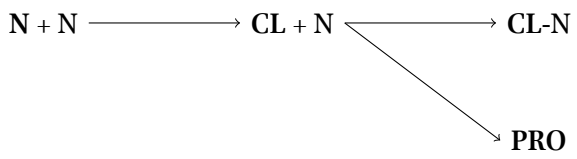


**Figure 7.2:** Grammaticalisation process of third person pronouns in Cuquila Mixtec

As shown in previous chapters, NP + NP compounds are very often found in the language, and such constructions are still very productive. Therefore, it seems reasonable that these abstract nouns would at first be used to modify another noun and restrict its meaning. At a

later stage, these nouns were phonetically reduced and became classifiers, used alongside the nouns they were classifying. Next, they fused with the head nouns into the fossilised forms that are found nowadays (cf. section 30). At the same time, these phonetically reduced classifiers grammaticalised into personal pronouns, maintaining their definiteness and gender-like properties. Leon Pasquel (1988: 139), while explaining the connection between classifiers and personal pronouns in varieties of Lowland Mixtec, arrives at this conclusion: '[...] those morphemes resulting from the nouns referring to 'man', 'woman', 'soil', 'god', 'animal', 'tree', and 'water' grammaticalize into a set of classifiers and into personal pronouns [...]'. Similarly, Helmbrecht (2004) argues that third person pronouns that derive from nouns often go through a stage when they are used as nominal classifiers.

Both Helmbrecht (2004: 384) and Heine & Mechtild (1984: 225-227) propose that this type of abstract nouns end up being semantically reduced to denote gender distinctions. Looking into the languages of the Western Alta, such as San Pedro Molinos (Merrifield & Stoudt 1967) or Yosoñama (Gittlen 2016), we can observe that the third person human pronouns indeed make gender distinctions between male and female. Interestingly, in Cuquila Mixtec, we see that the same pronouns that serve to make the distinction of gender, additionally mark politeness. This strategy is used in other varieties of Mixtec, such as the Diuxi variety from the Highland region (Kuiper & Pickett 1974). Other varieties of both the Highland and the Lowland regions mark politeness on the third person through the use of a separate pronoun. It would be interesting to study why at least some of the Mixtec languages have made this politeness distinction using abstract nouns as a source, and whether this is a historical innovation or rather a trait that other languages in the family lost over time.

Summarising, this chapter has shown how the personal pronouns of Cuquila Mixtec have grammaticalised into the forms that are found nowadays. Setting the relevant theoretical context, it was shown that both the first/second and the third person pronouns have gone through the grammaticalisation stages that are most frequently found cross-linguistically: the lexical sources were semantically extended and generalised, then lost some of their morphosyntactic and phonetic properties, and finally acquired today's forms. Furthermore, the sources of these pronouns were examined. It was found that the familiar forms for the first and second person derive from plural forms, in accordance with the 'plurification' proposal found in Heine & Song (2011). Lastly, the third person pronouns historically derive from abstract nouns, which developed in parallel to the old classifier system that Cuquila Mixtec had, which is now fossilised.

# Chapter 8

# Personal Pronouns In Use

This chapter focuses on the grammaticalization and use of the personal pronouns in Cuquila Mixtec. Specifically, the historical development of the pronominal forms will be analysed and the contexts in which the polite forms are used in contrast to the respect forms will be investigated. In order to provide the relevant context for this discussion, pronoun systems found in other Mixtec varieties will be described in section 7.1. After providing some background on the theory of grammaticalization in section 7.2, I will provide some information on the development of the Cuquila Mixtec pronouns based on information available for other Mixtec varieties in section 7.3. Section 8.1 analyses the variables which affect the pronoun use in Cuquila Mixtec within the theoretical context of the familiar/respect distinction.

## 8.1 Pronoun Usage and Politeness Distinctions

### 8.1.1 Theoretical Background

Several theories have been proposed in order to account for the social dimensions of the distinction between polite and familiar pronouns and their use. Starting in the 1960's, the classic work of Brown & Gilman (1960) provided the two macro-sociological dimensions of 'power' and 'solidarity' in the analysis of the pronoun systems of several European languages. According to this study, 'power' refers to the social distance between the speech act participants, and is defined as the "ability to control the behavior of the other" (Brown & Gilman 1960: 255). This is an inherently asymmetrical relation, so the use of personal pronouns and honorifics which are based on such a relation is also assymetrical. A person with a relatively higher social status will use the familiar forms with someone who is socially inferior, while the latter will in turn use the polite forms of address towards the socially superior. On the other hand, solidarity is a horizontal parameter, one which refers to the psychological distance between interlocutors. Because the speech act participants are

relatively close hierarchically, it is expected that the use of the pronouns will be symmetrical, whereby both speakers will use either the polite or the familiar forms of address (Brown & Gilman 1960: :256; Tannen & Kakava 1992: :3). The choice between the two forms will depend on the social distance between the speakers, but it will always result in a symmetrical use of the same forms. For example, two siblings will refer to each other using the familiar forms, as they are both close in terms of social hierarchy, and the social distance between them is small. In turn, work colleagues might address each other using the polite forms; even though they are hierarchically close to each other, the social distance is large enough in order not to grant for the use of the familiar forms.

This theory has received criticism by many scholars (Friedrich 1972; Morford 1997; Paulston 1976; Silverstein 2003), mostly for the fact that it oversimplifies the role that sociolinguistic phenomena play in this context. It does not account for many other parameters which are also relevant for the choice of terms of address and personal pronouns, such as the topic of the discourse or the context in which they appear (Friedrich 1972). Moreover, Morford (1997) points out that social distance is not only reflected in the terms of address, but that the very choice of one pronoun over the other can help define social relationships. Therefore, there is a two-way interaction of the language choices that the speakers make, which reflect these social parameters while at the same time influence and establish them. A speaker who chooses to address his interlocutor using the polite form immediately creates a social distance which the other person is expected to respect.

The influential work of Brown & Levinson (1987) incorporated the ideas of power and solidarity presented in Brown & Gilman (1960), but widened the scope of linguistic politeness. According to this theory, the public self-image of an individual, their 'face', dictates their language use in terms of politeness, expressed not only through the choice of pronouns, but also other grammatical constructions such as the use of passive voice or imperatives. The two aspects of the theory can be summarised as follows: 'positive face' is the wish of an individual that their self-image and their wants be appreciated and desirable by others. On the other hand, 'negative face' is their wish that their actions are not impeded by others, his freedom to action (Brown & Levinson 1987: 61-63).

When uttering a face-threatening act (an utterance which potentially threatens the wants of the addressee) such as a request, the individual needs to make a choice between a more indirect strategy or a more unambiguous, bold request. In the former case, they may choose for the positive politeness strategy, whereby the speaker appeals to the wants of the addressee, or they may go for negative politeness, with the intention of reducing the threatening nature of the proposition as much as possible. In terms of the choice of pronouns, the use of the polite forms are considered to belong to the negative politeness strategy, as they are more indirect and less threatening. Other approaches to negative politeness include the use of polite terms of

address and the all-together avoidance of direct reference. Brown & Levinson (1987) offer three major sociological variables that play a role in the choice of the above-mentioned strategies: the social distance between speaker and hearer, the relative power of each one and the weight that a request or an imposition has in a particular culture (Brown & Levinson 1987: 74). As they mention, these variables are not the only factors that are relevant to this choice, but they are the top-level variables which include others, such as status, authority, friendship etc. While these three dimensions are pan-cultural, the specific parameters that are included within are culture-specific.

Friedrich (1972), in his analysis of pronominal usage in Russian, lists ten different parameters that influence the choice between polite and familiar forms, which by themselves directly indicate social distance. Four of them have to do with the biological traits of the participants: their age, the generation they belong to, their sex and their kinship status. Next to these, he adds the relative political or jural authority of the individuals, as well as their group membership. The topic of the discourse and the context of the speech event are also relevant for the pronoun choice, as explained above. Finally, he mentions the dialect that the speakers use and the emotional solidarity between the speech-act participants as decisive factors in this choice.

In the following section, I will detail which sociological parameters influence the use of the polite pronouns in Cuquila Mixtec within the context of the local society, as well as the interaction between them.

### 8.1.2 Interaction between Pronouns in Cuquila Mixtec

Before delving deeper into the parameters, it is worth looking into the relationship between the pronoun forms themselves, in order to understand how they interact. The humbling form of the first person *sã* is directly related to the respect form of the second person *ní*. That is, if a speaker wants to use the respect form to address the hearer, they must also use the humbling form to refer to themselves. Otherwise, if the speaker mixes the respect and familiar forms, not only is this considered infelicitous by the speakers, but also ungrammatical. In the example below, the sentence (158-b) was deemed ungrammatical by the speakers, since the first person humbling form is combined with the second person familiar form.

(158) a. *kuachi ká ndáa=sã masu ká ndi'i*
small ADD PL=1.RES than ADD 2.RES
'We are younger than you.'

b. **kuachi ká ndáa=**sã** masu ká **ndo'o***
small ADD PL=**1.RES** than ADD **2.FAM**
'We are younger than you.'

Furthermore, even if the speaker does not address the hearer at all and only talks about himself, the use of the humbling form *sã* denotes respect towards the addressee. For example, when an elder woman was asked to provide me with some recipes, she did not address me at all in her talk, but instead was talking about the actions that she would perform during the preparation of the food. She used verbs such as *skuachi=sã* 'I cut', *chu'u=sã* 'I add', *sacho'o=sã* 'I heat up' etc. Throughout her speech, she consistently used the humbling form *sã* to refer to herself, thus choosing to show respect towards me. In contrast, when a speaker was asked to explain how they build a house, he consistently used the more direct (exclusive) plural of the familiar form, as he considered us equals. He used forms such as *kua-ko=nì* 'we go', *ká-taá=nì* 'we throw', *ká-sa'a=nì* 'we make'.

In contrast, the third person polite forms *de* and *ña* are not directly related to either the first or the second person respect forms. The speaker may choose to speak about someone with respect in a conversation where he uses the familiar forms with his interlocutor. The example below comes from an elicitation task where the speaker was given some photos and was asked to compare the people in them. One of the photos was of herself with an older lady, and she produced the following utterance:

(159) *ña'nu ká=**ña*** *masu ká* ***ndu'u***
big ADD=**3.F.RES** than ADD **1.FAM**
'She is older than I.'

Because the lady that she referred to is older than her, she chose to use the third person polite pronoun *ña* to show her respect for her. At the same time, she chose to use the familiar form to talk about herself as the power distance between her and the addressee (myself) was minimum.

Even though the form used for a referent is independent from the first and second person forms, the relationship between the addressee and the person being referred to does influence the speaker's choice of pronoun. Depending on the social distance of the hearer and the referent, the speaker may choose to change the pronoun form that they would normally use in order to accommodate for the hearer, taking on their perspective. This is not uncommon cross-linguistically. An adult speaker may change their choice of pronouns when talking to a child about another adult and use the polite form, as it is expected that the child would also use the polite form and different terms of address if they were to speak directly to the referent. Let's look at an example from the corpus:

(160) E.: *de nana efren ndee*
and mother *Efren* sit
'Where is Efren's mother?'

D.: *ndúu vi na ndee ki chi kuá'ã=**ña** chi*
NEG ? any sit now because go=3.F because
'Nobody is here because she left.'

E.: *michi kuá'ã=**i**?*
where go=3.GEN.HUM
'Where did she go?'

D.: *kuá'ã=**i** pasia kachi=**i***
go=3.GEN.HUM *pasear* say=3.GEN.HUM
'She went for a walk, she said.'

This dialogue is a good example of such a case, where the speaker takes on the perspective of the hearer. Two ladies, D. and E., are talking about an older lady. The elders in the community of Santa María Cuquila have a high level of social authority and consequently are always referred to with the polite form. However, E. is a close relative of the referent and, even though she is much younger, she can refer to her using the familiar form. In contrast D. who is closer in age to the person referred to, begins the conversation using the polite form, since they don't know each other very well and thus the emotional distance between them is greater. We can see that D. starts off by referring to Efren's mother using *ña*, to which E. replies using the familiar form *i*. Seeing that the familiar form is licensed by her interlocutor, D. then proceeds to also adopt this form to refer to this person.

### 8.1.3 Sociological Variables in Pronoun Use of Cuquila Mixtec

Moving on to the sociological parameters, we can observe that, in Cuquila Mixtec a variety of factors influence the choice between the familiar and the respect pronoun. Most of them apply to all three persons, but as we saw above, an extra parameter is added for the third person pronoun choice: the social relationship between the hearer and the referent. Age seems to be one of the most decisive factors in the choice of pronoun, as it was observed and also as it was reported by the speakers. This can be either absolute age (what Friedrich (1972) calls 'relative generation') or relative age. Absolute age refers to the elders; people over 70 years old are highly respected in the community. In most cases, the factor of absolute age takes priority over all others. An elderly person will be addressed and referred to using the polite forms, even by their children and grandchildren. Similarly, a child or adolescent will always be referred to using the familiar form. The example below comes from a speech where a history teacher gives life advice to his adolescent students. Because they are not adults, but also due to his relative social authority as a teacher, he addresses them using the familiar form *nu*:

(161) *chinde=nu nu ñuu=nu nu ndeé=nu nu n-kaku=nu nu*
help=2.FAM at village=2.FAM at sit=2.FAM at COMPL-be.born=2.FAM at
*ja'a=nu*
be.raised=2.FAM
'(so that you can) help your village, where you live, where you were born, where you were raised.'

Relative age refers to the age difference between the speaker and the addressee. If the hearer is younger or of the same age as the speaker, then the familiar forms *nì* 'I' and *nu* 'you' will most probably be used. For example, a 27-year-old speaker would consistently use the familiar form *i* to refer to a woman passing by who seemed to be the same age as her. This factor, however, is often times overwritten by other parameters, such as marital status. If a woman is married and has children, no matter how old she is, she will always be referred to using the polite form. A 40-year-old speaker was telling to me that an elder woman from the community always addresses her using the respect form *ní*. When she asked her to use the familiar form, the elder woman explained to her that she cannot do that, because she (the younger lady) already had kids, so she should address her politely. Married men are always referred to with the respect pronouns, also by their wives. The following utterance comes from a 49-year-old lady who talks to me about her family. She refers to her husband using the polite form *de*, even though she refers to everyone else, herself, her children and her grandchild, using the familiar forms:

(162) *mitañu ndúù bi=de chi kuá'ã=de satiñu=de*
now NEG ?=3.M.RES because go=3.M.RES work=3.M.RES
'He's not here now because he went to work.'

Even though he is exactly the same age as her, she uses the respect form because he is her husband and she is possibly protecting his positive face.

For men, the relative age factor can also be overwritten by the sociopolitical authority parameter. The community of Santa María Cuquila, like other communities in the Mixteca region, has a local body of authorities that take many decisions in the village. The people who take a seat in the authorities are elected by the community and are highly respected, as this role has many responsibilities and requires plenty of personal sacrifices. As soon as a person becomes an adult, they can potentially be asked to take a seat. Therefore, if a 20-year-old male has a position in the local council, he will be addressed with the polite form by everyone in the community, regardless of the speaker's age.

Finally, group membership also affects the choice of pronouns. The notion of 'group' is used here to denote any type of social community whose members are related by biological factors, a certain activity, role. For this reason, close kinship licenses the use of familiar forms among the members. Even though several speakers reported that, in the past, the children would always address their parents with the respect pronouns, this is not the case any more; unless they

are elders (in which case, the factor of absolute age takes priority over the group membership parameter) both children and parents, as well as siblings, address each other using the familiar forms. Furthermore, people who help each other in the corn fields develop a close enough relationship to license the use of the familiar pronouns. Being part of the same group elevates the level of solidarity, and at the same time diminishes the social distance. Being an outsider to the community, and thus not sharing the same group membership with the speakers, could explain why all of the elder consultants, as well as some speakers between 40 and 50 years old, addressed me with the respect pronoun. A hypothesis is that, since I was an outsider, these speakers did not base our interaction on the level of solidarity, but instead on the level of power (as used by Brown & Gilman (1960)). As mentioned before, in many cases this was not directly shown through the utterance of the second person polite form, but it was to be assumed via the use of the humbling form to refer to themselves. However, it is interesting to note that the same speakers would use the familiar form *tu* when speaking to me in Spanish.

Whether the interlocutor will reciprocate a familiar or respect form depends on the same factors and their complex interactions. Looking into the symmetry of familiar forms, interlocutors who belong in the same group will use the same pronouns symmetrically, as they are close in terms of social hierarchy. Group membership, similar social authority or age proximity license the symmetrical use of the familiar pronouns. The following dialogue was produced during a conversation between two men who have been tending the corn fields together and are close in terms of relative age:

(163) A.: *suni ndiki lulu viti xũ'ũ nasa jinda'a=**nì***
so spend small a.little money what carry=**1.FAM**
'This way I spend the little money that I'm carrying [...]'

F.: *kuã́'a=**nì** de ndaki ta'a=yo inka ichi*
go=**1.FAM** and see REFL=INCL another path
'I will leave, we will see each other another time.'

Among speakers over the age of 60, this symmetrical use of the respect forms was very common. In one occasion, two women were discussing about the community's market. Ermiña, who is approximately 70 years old, said the following to Margarita, who is 60 years old:

(164) *kata keja'a ndáa ma=yó kaxiko=yó ta **ndi'i***
just start PL SPEC=INCL sell=INCL like **2.RES**
'We are just starting to sell like you.'

Here, the factor of relative age does not play a role in the choice of pronoun, as the speaker is actually older than the addressee. Furthermore, the social distance between them is minimum, as they have close daily contact. However, it seems that, in these cases, these two factors are

not significant enough to grant the use of the familiar form. It could be argued that the context of the speech event, being a formal recording, and thus a semi-scripted conversation, had an effect of the choice of the pronominal forms. However, this is not the case either. In their daily interactions they would also reciprocate the respect forms.

This might be explained in two ways: on the one hand, it might portray a sociolinguistic change in the way the pronouns are used. Just like the case that was explained above, where up to one generation ago the speakers would address their parents with the respect forms, perhaps this generation of older speakers exhibits a use of the polite pronouns that was not adopted by the following generations. On the other hand, it might be due to the parameter of absolute age. Even though these speakers belong to the same generation, they are considered elders by the community and, because of that, are highly respected. In turn, perhaps they consider other elders as respected people, and as such they address them politely.

Asymmetrical use of the pronominal forms is granted when one of the speech act participants is in a greater position of power than the other. The only factor observed to consistently license this asymmetry was absolute age. Elder speakers would use familiar forms with younger hearers, often times belonging in the same group. Therefore, D., of approximately 60 years of age and E., a 45-year-old speaker, exhibited asymmetry in their conversation: D. would use the familiar forms to address E., whereas E. would consistently address D. with the respect forms. It should also be noted that the two speakers are cousins, spending a lot of time together on a daily basis. The following is a short dialogue which occurred during their conversation, and clearly portrays the asymmetrical use of the pronouns from the two speakers:

(165) D.: *a n-kunde'e=nu yu'u itu na?*
QUES COMPL-finish=2.FAM edge corn.field
'Did you finish with your corn filed?'

E.: *ndúù chi n-kunde'e ta=sã ndi'i*
NEG because COMPL-finish indeed=1.RES end
'I did finish it, it is done.'

In conclusion, the following sociological parameters have been found to affect the choice between the familiar and the respect pronoun forms: absolute age, relative age, marital status, relative social authority, and group membership. Even though some factors (like absolute age) appear to take priority over others, a more careful observation of the social contexts often suggests otherwise. Due to the complex interactions between the social variables, the choice of the familiar versus the respect form and vice versa is frequently overwritten when another parameter plays a more significant role in a certain context (for example, group membership).

# Chapter 9

# Conclusion

This study has focused on several aspects of the pronoun system in Cuquila Mixtec, serving as the first research carried out in this Mixtec variety. In order to better understand this system, first information on several other parts of the language and culture was provided. In chapter 2, some characteristics of the way of life in Santa María Cuquila were described, followed by an overview of the speaker's attitudes towards Cuquila Mixtec. The chapter showed that the inhabitants of Santa María Cuquila favour a communal way of life in which decisions are taken collectively. It also showed that, although the status of the language has been in decline for years, recently some revitalisation attempts are being made by young adults. Chapter 3 established the background of this study. Cuquila Mixtec was set within the context of the Mixtec language family, in relation to other Mixtec varieties. It detailed the problematic internal classification of Mixtec. Additionally, the chronological overview of the previous works published in other Mixtec varieties showed that the study of the personal pronoun systems is still rather limited. In order to understand how this research was conducted, chapter 4 focused on the data gathering techniques and provided some details on the language consultants. The grammar sketch in chapter 5 served as an introduction to Cuquila Mixtec, through the description of the language structures relevant to the study of the personal pronouns. The seven specific noun categories were defined and their relation to the fossilised classifier system was established. Furthermore, it was shown that subject number is marked on the verb through affixes, and that the neutral word order is VSO. Chapter 6 detailed the morphosyntactic and semantic properties of the pronoun system. In addition, it outlined some of its most interesting aspects, such as the politeness distinctions in all three persons through the use of humbling or respect forms, and the multiple third person pronouns which reflect the noun categorisation system of the language. The following chapter attempted to answer how these pronouns emerged. The grammaticalisation path of the different pronouns was set in the relevant theoretical context and was analysed. It was shown that the historical development of the pronoun forms follows the common path

established in the literature. The pronouns have undergone changes on the semantic, morphological and phonetic level since the reconstructed proto-forms. Additionally, the sources of the pronouns concur with the theory: plural pronouns gave rise to today's familiar forms of the first and second persons, and abstract nouns were used as a source for the third person pronouns. Seeing how the use of the forms has changed over the years, chapter 8 described the sociological parameters affecting their use today. The purpose of this final chapter was to create a snapshot of the current situation before it undergoes possible further changes. It was found that five variables influence the pronoun use: absolute age, relative age, marital status, relative social authority, and group membership. The complex interactions among these parameters result in a context-specific use of the familiar and respect forms.

The Mixtec languages are characterised by many typologically unusual features. Some of them were established and analysed in this study, such as how politeness is expressed, what the origins of the forms are and in which circumstances they are used. Further research is needed in order to better understand how the language works. A complete grammar description will clarify the tone system and the rules that govern tone sandhi. Such description would also shed light into many other characteristics of the morphosyntax, such as the specific features of TAM marking.

The current study serves as an expansion on the Mixtec literature, adding another previously non-described variety on the linguistic map. Offering new analyses and perspectives, this research broadens our understanding of how the languages of this family work, and it provides new insights in the pronoun studies as a whole.

# Bibliography


Agha, Asif. 1994. Honorification. *Annual Review of Anthropology* 23. 277–302.

Aikhenvald, Alexandra Y. 2000. *Classifiers. A Typology of Noun Categorization Devices.* Oxford: Oxford University Press.

Alexander, Mary Ruth. 1988. A Syntactic Sketch of Ocotepec Mixtec. In C. Henry Bradley & Barbara E. Hollenbach (eds.), *Studies in the Syntax of Mixtecan Languages*. Arlington: Summer Institute of Linguistics.

Alexander, Ruth María. 1980. *Gramática Mixteca de Atatlahuca*. CDMX: Instituto Linguistico de Verano.

Baht, D. N. Shankara. 2004. *Pronouns*. Oxford: Oxford University Press.

Balderas, Juan Jaime Hernández. N.d. Análisis social y de procesos organizativos en la Región Mixteca. `http://siteresources.worldbank.org/MEXICOEXTN/Resources/analisissocial.pdf`.

Bradley, C. Henry. 1975. *A linguistic sketch of Jicaltepec Mixtec*. Norman: Summer Institute of Linguistics.

Brown, Penelope & Stephen C. Levinson. 1987. *Politeness: Some universals in language usage*. Cambridge: Cambridge university press.

Brown, Roger & Albert Gilman. 1960. The Pronouns of Power and Solidarity. In Thomas A Sebeok (ed.), *Style in Language*, 253–276. MIT Press.

Büring, Daniel. 2015. (Contrastive) topic. In *The Oxford Handbook of Information Structure*. New York: Oxford University Press.

Caballero Morales, Santiago Omar. 2013. On the development of speech resources for the Mixtec language. *The Scientific World Journal* 2013.

Campbell, Eric W. 2017. Otomanguean historical linguistics: past, present and prospects for the future. *Language & Linguistics Compass* 11.

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

Cassiano, Vicente. 1982. *Análisis sintáctico del mixteco de Coatzoquitengo, Gro.* Mexico D. F.: SEP.

2019. Catalogue of Endangered Languages. `http://www.endangeredlanguages.com` (7 November, 2019).

Chafe, W. 1987. Cognitive contraints on information flow. In Russell S. Tomlin (ed.), *Coherence and Grounding in Discourse: Outcome of a Symposium, Eugene, Oregon, June 1984*, 21–51. Amsterdam/Philadephia: John Benjamin's.

Corbett, Greville G. 2013. Number of Genders. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. `https://wals.info/chapter/30`.

Cysouw, Michael Alexander. 2003. *The paradigmatic structure of person marking*. Oxford: Oxford University Press.

Diessel, Holger. 1999. *Demonstratives: Form, function and grammaticalization*. Amsterdam: John Benjamins Publishing.

Dixon, R. M. W. 1986. Typological variation of classifier systems. In Colette Craig (ed.), *Noun classes and categorization*, 105–112. Amsterdam/Philadephia: John Benjamin's.

Dixon, R. M. W. 2010. *Basic Linguistic Theory. Volume 2*. Oxford: Oxford University Press.

Eberhard, David M., Gary F. Simons & Charles D. Fenning. 2012. Ethnologue: Languages of the World. `https://www.ethnologue.com/credits` (6 November, 2019).

Farris, Edwin R. 1992. A syntactic sketch of Yosondúa Mixtec. In C. Henry Bradley & Barbara E. Hollenbach (eds.), *Studies in the syntax of Mixtecan languages*. Arlington: Summer Institute of Linguistics.

Ferguson de Williams, Judith. 2006. *Gramática popular del mixteco del municipio de Tezoatlán, San Andrés Yutatio, Oaxaca*. CDMX: Instituto Lingüístico de Verano.

Friedrich, Paul. 1972. Social Context and Semantic Feature: The Russian Pronominal Usage. In *Directions in Sociolinguistics: The Ethnography of Communication*, 270–300. New York: Holt, Rinehart & Winston.

Gittlen, Laura. 2016. *Gramática popular, mixteco del norte de Tlaxiaco*. CDMX: Instituto Lingüístico de Verano.

Greenberg, Joseph H. 1963. Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. In *Universals of Language*, 73'113. Cambridge, Massachusetts: MIT Press.

Grenoble, Lenore A. 2011. Language ecology and endangerment. In Peter K. Austin & Julia Sallabank (eds.), *The Cambridge Handbook of Endangered Languages*, 27–44. Cambridge: Cambridge University Press.

Heine, Bernd & Reh Mechtild. 1984. *Grammaticalization and Reanalysis in African Languages*. Hamburg: Buske.

Heine, Bernd & Kyung-An Song. 2011. On the grammaticalization of personal pronouns. *Journal of Linguistics* 47(3). 587–630.

Helmbrecht, Johannes. 2003. Politeness distinctions in second person pronouns. In Friedrich Lenz (ed.), *Deictic Conceptualisation of Space, Time and Person*, 185–202. Amsterdam: John Benjamin's.

Helmbrecht, Johannes. 2004. *Personal pronouns-form, function, and grammaticalization*. Erfurt: Univ. Erfurt PhD thesis.

Hernández, Fray Benito. 1567. *1567. Doctrina christiana en lengua mixteca*. Mexico: Casa de Pedro Ocharte.

Hills, Robert A. 1990. A Syntactic Sketch of Ayutla Mixtec. In C. Henry Bradley & Barbara E. Hollenbach (eds.), *Studies in the Syntax of Mixtecan Languages*. Arlington: Summer Institute of Linguistics.

Hinton, Leanne et al. 1991. Preliminary analysis of Chalcatongo Mixtec Tone. In James Redden (ed.), *Papers from the American Indian Languages Conferences: Occasional Papers on Linguistics 16*, 147–55. Carbondale: Southern Illinois University.

Holland, William R. 1959. Dialect variations of the mixtec and cuicatec areas of oaxaca, mexico. *Anthropological Linguistics* 1(8). 25–31.

Hollenbach, Barbara E. 2003. Los pronombres honorificos en Mixteco. *Cuadernos del Sur* 9(19). 1–14.

Hollenbach, Barbara E. 2013. *Gramática del mixteco de Magdalena Peñasco* (*Sa´an ñuu savi*). CDMX: Instituto Linguistico de Verano.

Hollenbach, Barbara E. 2015a. *A preliminary reconstruction of Mixtec pronouns*. SIL.

Hollenbach, Barbara E. 2015b. *Mixtec pronoun database*. SIL.

Hollenbach, Barbara E. 2017. *Diccionario mixteco de Magdalena Peñasco* (*Sa´an ñuu savi*). CDMX: Instituto Linguistico de Verano.

Instituto Nacional de Estadística, Geografía e Informática. 2011. Censo de Población y Vivienda, 2010 (Informe nacional y estatales), México. `http://www.censo2010.org.mx`.

Josserand, Judy Kathrin. 1983. *Mixtec Dialect History*. Tulane University PhD thesis.

Kaufman, Terrence. 1988. Otomanguean tense/aspect/mood, voice, and nominalization markers.

Kresge, Lisa. 2007. Indigenous oaxacan communities in California: An overview. *California Institute for Rural Studies* 1107.

Kroeger, Paul R. 2005. *Analysing grammar*. New York: Cambridge University Press.

Kuiper, Albertha & Velma Pickett. 1974. Personal pronouns in diuxi mixtec. *Summer Institute of Linguistics: Mexico Working Papers* 1. 53–58.

Lambrecht, Knud. 1996. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Lastra, Yolanda. 1992. The present-day indigenous languages of Mexico: an overview. *International Journal of the Sociology of Language* 96(1). 35–43.

Lehmann, Christian. 1995. *Thoughts on grammaticalisation*. Munich: Lincom Europa.

Leon Pasquel, Maria Lourdes de. 1988. *Noun and Numeral Classifiers in Mixtec and Tzotzil: A Referential View*. Sussex: University of Sussex.

Longracre, Robert E. 1957. Proto-mixtecan. *International Journal of American Linguistics* 23(4).

Macaulay, Monica. 1993. Argument Status and Constituent Structure in Chalcatongo Mixtec. In *Proceedings of the Nineteenth Annual Meeting of the Berkeley Linguistics Society: Special Session on Syntactic Issues in Native American Languages*, 73–85. Berkeley: Berkeley Linguistics Society.

Macaulay, Monica. 1997. *A Grammar of Chalcatongo Mixtec*. Berkeley: University of California Press.

Macaulay, Monica & Joseph C. Salmons. 1995. The phonology of glottalization in Mixtec. *International Journal of American Linguistics* 61(1). 38–61.

Marlett, Stephen A. 1992. Nasalization in Mixtec Languages. *International Journal of American Linguistics* 58(4). 425–435.

McKendry, Inga. 2013. *Tonal Association, Prominence and Prosodic Structure in South-eastern Nochixtlán Mixtec*. Edinburgh: University of Edinburgh PhD thesis.

Merrifield, William R & Betty J Stoudt. 1967. Molinos Mixtec clause structure. *Linguistics* 5(32). 58–78.

Morford, Janet. 1997. Social indexicality in French pronominal address. *Journal of Linguistic Anthropology* 7(1). 3–37.

Moseley, Christopher. 2010. Atlas of the World's Languages in Danger. `http://www.unesco.org/culture/languages-atlas/en/atlasmap.html` (7 November, 2019).

Nagengast, Carole & Michael Kearney. 1990. Mixtec Ethnicithy: Social Identity, Political Consciousness, and Political Activism. *Latin American Research Review* 25(2). 61–91.

North, Juanita & Juliana W. Shields. 1978. *Gramática popular del mixteco: mixteco de Silacayoapan*. Mexico D. F.: Instituto Lingüístico de Verano.

Operstein, Natalie. 2003. Personal Pronouns in Zapotec and Zapotecan. *International Journal of American Linguistics* 69(2). 154–185.

Passer, Matthias Benjamin. 2016. (What) Do verb classifiers classify? *Lingua* 174. 16–44. http://www.sciencedirect.com/science/article/pii/S0024384115002557.

Paulston, Christina Bratt. 1976. Pronouns of address in Swedish: social class semantics and a changing system. *Language in Society* 5(3). 359–386.

Payne, Thomas E. 1997. *Describing morphosyntax*. Cambridge: Cambridge University Press.

Perry, Elizabeth. 2017. *The Declining Use of the Mixtec Language Among Oaxacan Migrants and Stay-at-Homes: The Persistence of Memory, Discrimination, and Social Hierarchies of PowerThe Declining Use of the Mixtec Language Among Oaxacan Migrants and Stay-at-Homes: The Persistence of Memory, Discrimination, and Social Hierarchies of Power*. UC San Diego: Center for Comparative Immigration Studies.

Pike, Kenneth L. 1944. Analysis of a Mixteco Text. *International Journal of American Linguistics* 10(4). 113–138.

Reyes, Antonio de los. 1593. *Arte en lengua mixteca*. Reprint, Mexico: Pedro Bali.

Rieger, Ivy Alana. 2019. La tradición camaleónica en las prácttticas festivas de una comunidad mixteca tradicional. *Reflexiones* 98(1). 111–129.

Rojas López, Gerardo Crisanto. 2015. *Propuesta didáctica, para la enseñanza de la numeración tu´un savi, para escuelas de multigrado, indígenas, variante Ñuu Kuiñi, de Cuquila, Tlaxiaco, Oaxaca, región Mixteca*. Ciudad de México: Universidad Pedagógica Nacional PhD thesis.

Ruiz Medrano, Ethelia. 2015. Un pueblo milenario con infraestructura del siglo XXI: Santa María Cuquila, ex distrito de Tlaxiaco. In *El pasado lejano de la Mixteca*, 125–138. Huajuapan de León: Universidad Tecnológica de la Mixteca.

Siewierska, Anna. 2004. *Person*. Cambridge: Cambridge University Press.

Silverstein, Michael. 2003. Indexical order and the dialectics of sociolinguistic life. *Language & Communication* 23(3). 193–229. http://www.sciencedirect.com/science/article/pii/S0271530903000132.

Suárez, Jorge A. 1983. *The Mesoamerican Indian Languages*. Cambridge: Cambridge University Press.

Swadesh, Morris. 1956. *Conceptos geográfico-cronológicos de cultura y lengua* (Estudios Antropológicos publicados en homenaje al doctor Manuel Gamio). Mexico D. F.

Tannen, Deborah & Christina Kakava. 1992. Power and solidarity, in modern greek conversation:) disagreeing to agree. *Journal of Modern Greek Studies* 10(1). 11–34.

Towne, Douglas. 2011. *Gramática popular del tacuate* (*mixteco*) *de Santa María Zacatepec, Oaxaca*. CDMX: Instituto Linguistico de Verano.

Wal, Jenneke van der. 2016. Diagnosing focus. *Studies in Language* 40(2). 259–231.

Zylstra, Carol F. 2012. *Gramática del Tu un Savi* (*la lengua mixteca*) *de Alacatlatzala, Guerrero*. CDMX: Instituto Linguistico de Verano.