



January 2017

Tools For Assessing Relatedness In Understudied Language Varieties: A Survey Of Mixtec Varieties In Western Oaxaca, Mexico

Erin Padgett Padgett

Follow this and additional works at: <https://commons.und.edu/theses>

Recommended Citation

Padgett, Erin Padgett, "Tools For Assessing Relatedness In Understudied Language Varieties: A Survey Of Mixtec Varieties In Western Oaxaca, Mexico" (2017). *Theses and Dissertations*. 2306.
<https://commons.und.edu/theses/2306>

This Thesis is brought to you for free and open access by the Theses, Dissertations, and Senior Projects at UND Scholarly Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of UND Scholarly Commons. For more information, please contact zeinebyousif@library.und.edu.

TOOLS FOR ASSESSING RELATEDNESS IN UNDERSTUDIED LANGUAGE VARIETIES:
A SURVEY OF MIXTEC VARIETIES IN WESTERN OAXACA, MEXICO

by

Erin Padgett
B.A, University of North Carolina-Chapel Hill, May 2014

A Thesis
Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Master of Arts

Grand Forks, North Dakota
August
2017

© 2017 Erin Padgett

This thesis, submitted by Erin Padgett in partial fulfillment of the requirements for the Degree of Master of Arts from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

DR. JOHN M. CLIFTON, Chair

DR. JAMES S. ROBERTS

DR. STEPHEN A. MARLETT

This thesis meets the standards for appearance, conforms to the style and format requirements of the School of Graduate Studies of the University of North Dakota, and is hereby approved.

Grant McGimpsey
Dean of the School of Graduate Studies

Date

PERMISSION

Title Tools for assessing relatedness in understudied language varieties: A survey of
 Mixtec varieties in western Oaxaca, Mexico.

Department Linguistics

Degree Master of Arts

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work or, in his absence, by the chairperson of the department or the dean of the Graduate School. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Signature __Erin Padgett_____

Date __July 18, 2017_____

TABLE OF CONTENTS

LIST OF FIGURES	vii
LIST OF TABLES	ix
ACKNOWLEDGEMENTS	x
ABBREVIATIONS	xi
ABSTRACT	xii
CHAPTER	
1 INTRODUCTION	1
1.1 Geographic information	6
1.2 Demographic information	9
1.3 Government information	11
2 METHODOLOGY	14
2.1 General Wordlist	15
2.1.1 A brief history of wordlists	16
2.1.2 Using the General Wordlist	18
2.2 Tone Wordlist	22
2.3 Sociolinguistic Questionnaire	26
2.3.1 A brief history of questionnaires	27
2.3.2 Using the Sociolinguistic Questionnaire	30
2.4 Observation	34
2.5 Summary of methodology	34
3 FINDINGS	36
3.1 Analysis of similar forms	37
3.2 Grouping language varieties according to phonological correspondences using isoglosses	45
3.3 Tone	54
3.3.1 Analysis of tone patterns	55

3.3.2	Tone analysis of individual words	58
3.3.2.1	Rationale for a second tone analysis	58
3.3.2.2	Modified Levenshtein Distance: A second tone analysis	60
3.4	Reported intelligibility	64
3.5	Summary of findings	69
4	REFLECTION ON METHODOLOGIES AND DIRECTIONS FOR FURTHER STUDY	73
4.1	Use of isoglosses to determine relatedness	73
4.2	Use of tone to determine relatedness	75
4.3	Reservations about reported intelligibility findings	77
4.4	Rapid Appraisal surveys	78
4.5	Directions for further study	80
	REFERENCES	82
	APPENDICES	85

LIST OF FIGURES

Figure	Page
1 Map of the Mexican state of Oaxaca	2
2 Location of Santiago Juxtlahuaca, Oaxaca	3
3 Map of communities visited for language survey	7
4 Roads connecting communities surveyed	8
5 LL forms in Praat	26
6 MM forms in Praat	26
7 Tree of relatedness for Group 1	41
8 Relative location of COI and SJM	43
9 Isogloss 1—[k ^w] vs. [β]	47
10 Isogloss 2—[ʃ] vs. [s/ʒ/ks]	47
11 Isogloss 3—[ji] vs. [iʃ]	48
12 Isogloss 4—[tʃ̃] vs. [t]	48
13 Isogloss 5—[ʒ] vs. ∅	49
14 Isogloss 6—[ʒ] vs. [ʃ]	49
15 Isogloss 7—[ⁿ dʒ̃] vs. [ⁿ d]	50
16 Isogloss 8—[ⁿ d] vs. [l]	50
17 Isogloss 9—[n] vs. ∅	51
18 Isogloss 10—[i] vs. [u]	51
19 Isogloss 11—[a] vs. [e]	52

20 Composite map of isoglosses	52
21 Map of reported intelligibility	67
22 Map of relatedness groups	71

LIST OF TABLES

Table	Page
1 Population and geographic information for villages surveyed	9
2 General Wordlist	21
3 Tone Wordlist	24
4 Sociolinguistic Questionnaire	32
5 Summary of tools and village visits	35
6 Guidelines for determining similar forms	38
7 Percentage of similar forms	40
8 Isoglosses	46
9 Tone patterns	56
10 Additional tone correspondences	59
11 Examples of tone distance calculations	62
12 Percentage of dissimilar tones	63
13 Sociolinguistic Questionnaire questions about intelligibility	65

ACKNOWLEDGMENTS

Many, many thanks to Dr. John M. Clifton, my committee chair, for his tireless help in revising and improving my thesis with guidance, countless detailed comments, and answers to questions. Thank you to Dr. Stephen A. Marlett and Dr. James S. Roberts, whose expertise and feedback as committee members greatly enhanced my thesis. I am also grateful to Dr. Inga McKendry for her guidance in designing my language survey, as well as her help with my analyses of isoglosses and tone. Thank you to Bruce and Candice Beatham, Millie Nieves, and Gisela Beckmann for providing additional Mixtec language data and advice, and to my teammate, Isme, for accompanying me on many trips. Finally, I am grateful to Dr. Juanita Watters for her help in polishing my work, and to Jesse Atkinson for providing unwavering support throughout the thesis process.

ABBREVIATIONS

COI	Coicoyán de las Flores
GNA	Guadalupe Nundaca
RA	Rapid Appraisal
RTT	Recorded Text Test
SCA	Santa Catarina Noltepec
SIL	Summer Institute of Linguistics
SJM	San Juan Mixtepec
SMC	San Miguel Cuevas
SMD	San Martín Durazos
SPC	San Pedro Chayuco
SRC	Santa Rosa Caxtlahuaca
SRT	Sentence Repetition Test
YUC	Yucunicoco

ABSTRACT

This thesis presents findings of research conducted on the relatedness of seven Mixtec varieties spoken in indigenous language communities in western Oaxaca, Mexico. Mixtec varieties vary widely from one community to the next, and it is necessary to determine the relatedness of Mixtec varieties in order to best serve the language development needs of communities. Understanding the relatedness of these varieties is also an important step in measuring their intelligibility.

I used three research tools to gather data: a General Wordlist, a Tone Wordlist, and a Sociolinguistic Questionnaire. I present five analyses: percentage of phonologically similar forms, displaying phonological correspondences using isoglosses, two analyses of tone patterns, and reported intelligibility. Taken together, the first four analyses provide a clear picture of the linguistic relations of the Mixtec varieties studied. The analyses of tone and use of isoglosses are of particular note, as they present new strategies for analyzing unstudied tonal languages and language families. Findings on linguistic relatedness are then compared to the reported intelligibility of native speakers from the Questionnaire. With minor exceptions, the proposed relatedness matches up closely with intelligibility reported by survey participants.

I then clarify how preexisting linguistic designations for this region could be improved, based on my findings. The Ethnologue currently includes all seven of the language varieties surveyed under a single designation, but my findings show that it is necessary to list YUC in a separate designation from the other six communities. The *Instituto Nacional de Lenguas Indígenas* (INALI, National Institute of Indigenous Languages) needs to revise its current designations so that YUC is left under its current designation, the *mixteco del oeste alto* (High Western Mixtec),

while all of the six varieties surveyed should be under the *mixteco del oeste* (Western Mixtec) designation.

CHAPTER ONE

INTRODUCTION

One of the most persistent challenges of linguistic work has been defining where languages begin and end, both linguistically and geographically. Determining the level of linguistic similarity required to qualify one language as the “same” as another, as well as how widely those languages are spoken, requires an understanding of the overall linguistic situation of an area. This understanding is usually accomplished through language survey or language assessment.

While all language families present this reality, traditionally it has been especially challenging to distinguish among varieties of the Mixtecan family of languages. The Mixtecan language family, or genus, belongs to the Otomanguean language family of Mexico, which consists of ten language genera (Dryer & Haspelmath 2013). Mixtec varieties are known for their complex tone systems and especially tone perturbation (Hollenbach 2003:2). From one Mixtec community to the next, there may be large linguistic differences despite close geographic proximity (Egland 1978:25). The tonal complexities and linguistic diversity of Mixtec varieties make it difficult to delineate where each one begins and ends, resulting in a wide range of estimates of the number of Mixtec varieties spoken in Mexico, as well as conflicting ways of delineating languages. While some estimate there are 52 Mixtec *languages* (Simons & Fennig 2017), others propose that there are as many as 85 distinct Mixtec *varieties* (Instituto Nacional de Lenguas Indígenas 2015:132-147). Given this range of estimates and

ways of classifying language communities, I use the broader term “varieties,” rather than “languages” in this thesis.

In addition to the linguistic complexities and diversity of Mixtec varieties, there are also remaining geographic pockets in the areas where Mixtec varieties are spoken that remain understudied. Clearly, there is not a thorough understanding of the linguistic and geographic boundaries among Mixtec varieties.

Varieties of Mixtec are spoken in the Mexican states of Puebla, Guerrero, and Oaxaca, though the majority are found in Oaxaca; the *Instituto Nacional de Lenguas Indígenas* (National Institute of Indigenous Languages; henceforth “INALI”) of Mexico estimates that of the 85 Mixtec varieties that it recognizes, fifty are spoken in Oaxaca, while twenty-two are spoken in Guerrero and three are found in Puebla (2015:132-147). Additionally, there are diaspora communities of Mixtec speakers ranging from several hundred to several thousand people living throughout the United States.

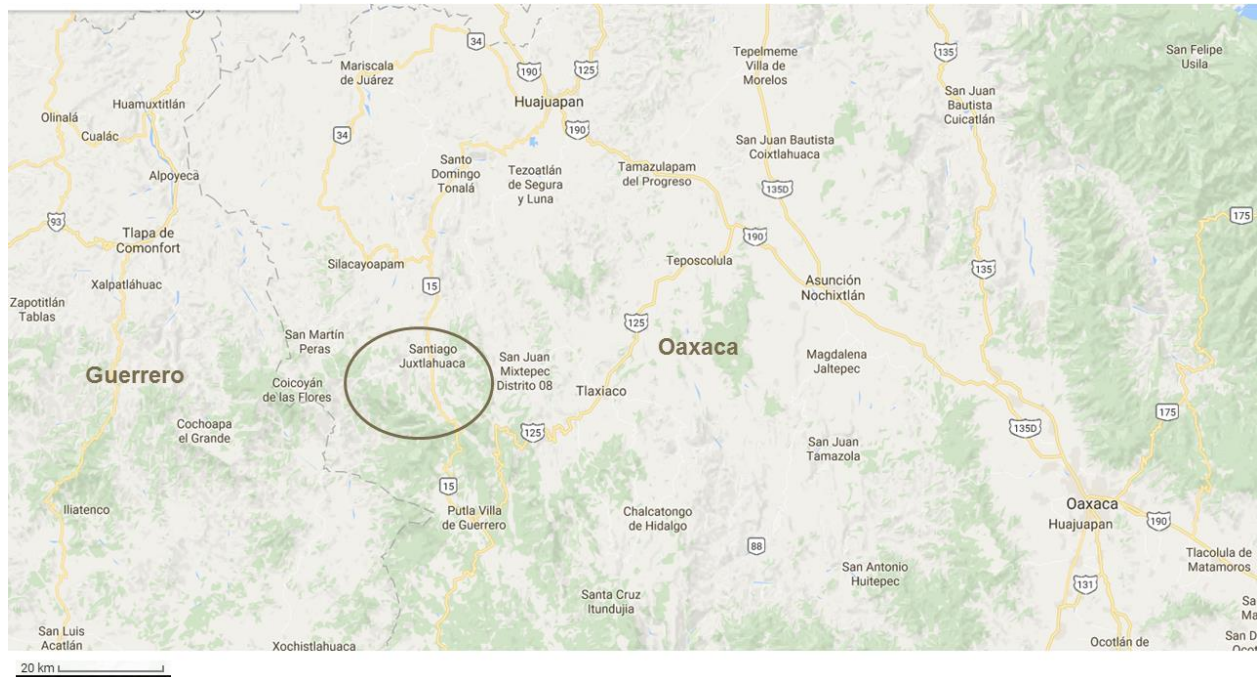
The focus of this thesis is establishing the linguistic relationships among the Mixtec varieties that are spoken in seven communities in the state Oaxaca, which is shown in Figure 1.

Figure 1: Map of the Mexican state of Oaxaca (Wikipedia Commons)



Mixtec varieties of Oaxaca are spoken in the western and southwestern portions of the state (Instituto Nacional de Lenguas Indígenas 2015:132-147). One of the understudied geographic pockets where Mixtec varieties are spoken is the region surrounding Santiago Juxtlahuaca (17.3353, -98.0124), shown in Figure 2.

Figure 2: Location of Santiago Juxtlahuaca, Oaxaca (Google Maps)



These communities are designated by INALI as either *mixteco del oeste* (Western Mixtec) or the *mixteco del oeste alto* (High Western Mixtec) (2015:140, 143). The Ethnologue designates them all as Juxtlahuaca Mixtec [vmc] (Simons & Fennig 2017). However, based on reports from SIL colleagues working in the region, there was reason to believe that these communities should be grouped separately from each other, even before I began work in the region. In January of 2016, I began working as a linguist in this area with the Summer Institute of Linguistics (SIL-Mexico), with the hopes of beginning language learning and conducting a language survey to better understand language development needs of this area.

Before beginning work, it seemed likely that the communities in the area spoke Mixtec varieties distinct from nearby varieties that had already been identified by members of SIL-Mexico.¹ The unclear linguistic relationships between these communities' varieties, both among the communities surveyed and neighboring language varieties, was a key motivation for my language survey work. My work was further motivated by apparent interest among community members in language development efforts, reported by SIL-Mexico colleagues based on their interactions with local people.² Therefore, SIL-Mexico needed to better understand the linguistic diversity of the area in planning for future language development projects.

I began work in the area by learning the Mixtec variety known as San Martín Durazos (SMD),³ which is spoken in various communities belonging to the San Sebastián Tecomaxtlahuaca and Santiago Juxtlahuaca municipalities.⁴ These communities include San Martín Durazos, Yucuyi, and Río de Hielo, the largest and best known of which is San Martín Durazos. (San Martín Durazos is located on the major road that runs through the area.) I chose these communities to learn Mixtec due to friendships with several people living in Juxtlahuaca who had personal connections to the community. I spent six months learning about the language and culture of these communities, with a special focus on tone and phonology. This background knowledge provided a basis for my subsequent language survey work.

¹ Information from personal communication with David Riggs, former Director of Field Programs for SIL-Mexico.

² Information from personal communication with Bruce and Candice Beatham, SIL-Mexico colleagues.

³ Some people living in the community also refer to this Mixtec variety as *mixteco de Yucuyi* (Yucuyi Mixtec). The SMD designation refers to both this variety and the group of communities where it is spoken.

⁴ Municipalities are the second smallest governmental designation in Mexico.

I was accompanied in this work by a young Mexican woman,⁵ who joined me in the language learning process and language survey work as a travel companion and fellow language learner. In the fall of 2016, we conducted a language survey in which we visited seven Mixtec communities in the region.

The survey's primary goals, determined by SIL-Mexico administrators and me, were to assess regional language development needs and understand the linguistic situation of this largely undocumented area. In order to accomplish these main two goals, there were several secondary goals of the survey: first, to determine the language vitality of SMD and related Mixtec varieties in the region; second, to estimate the relatedness of these Mixtec varieties; and third, to determine if SMD is distinct enough from Yucunicoco Mixtec, a known language development need on the eastern side of the region, to warrant its own project (David Riggs, p.c.). This thesis focuses on the second of these goals: the relatedness of seven Mixtec varieties surveyed, introduced below.

When first deciding where to collect data for this survey, I picked the geographic extremes of the region where I thought SMD or a closely related Mixtec *might* be spoken. These decisions were based on my conversations with local people over the preceding months of living in the area. In addition to SMD (which consists of the villages of San Martín Duraznos, Río de Hielo, and Yucuyi), the communities included Guadalupe Nundaca, Santa Rosa Caxtlahuaca, San Pedro Chayuco, San Miguel Cuevas, and Yucunicoco. Part way through the survey, I decided to add Santa Catarina Noltepec to the list of towns to visit, as it is centrally located but quite isolated. (It is the only town on its road, and it only has one road to and from the community.)

⁵ Her name had to be omitted for reasons of security.

In initial data collection trips, many survey participants mentioned Santa Catarina Noltepec as speaking a Mixtec similar to theirs, so it seemed like a relevant place to investigate.

We also attempted to gain permission from town authorities to conduct survey work in Zaragoza Alacranes, since residents of SMD reported that their language was also spoken there. Zaragoza Alacranes is located on the opposite side of the mountain from Yucuyi (part of the SMD designation). However, we were never given permission by the town leaders to conduct language survey in their community.

The final list of communities visited included:

- San Martín Durazos (including Yucuyi and Río de Hielo) (SMD)
- Guadalupe Nundaca (GNA)
- Santa Catarina Noltepec (SCA)
- Santa Rosa Caxtlahuaca (SRC)
- San Pedro Chayuco (SPC)
- San Miguel Cuevas (SMC)
- Yucunicoco (including seven small villages) (YUC)

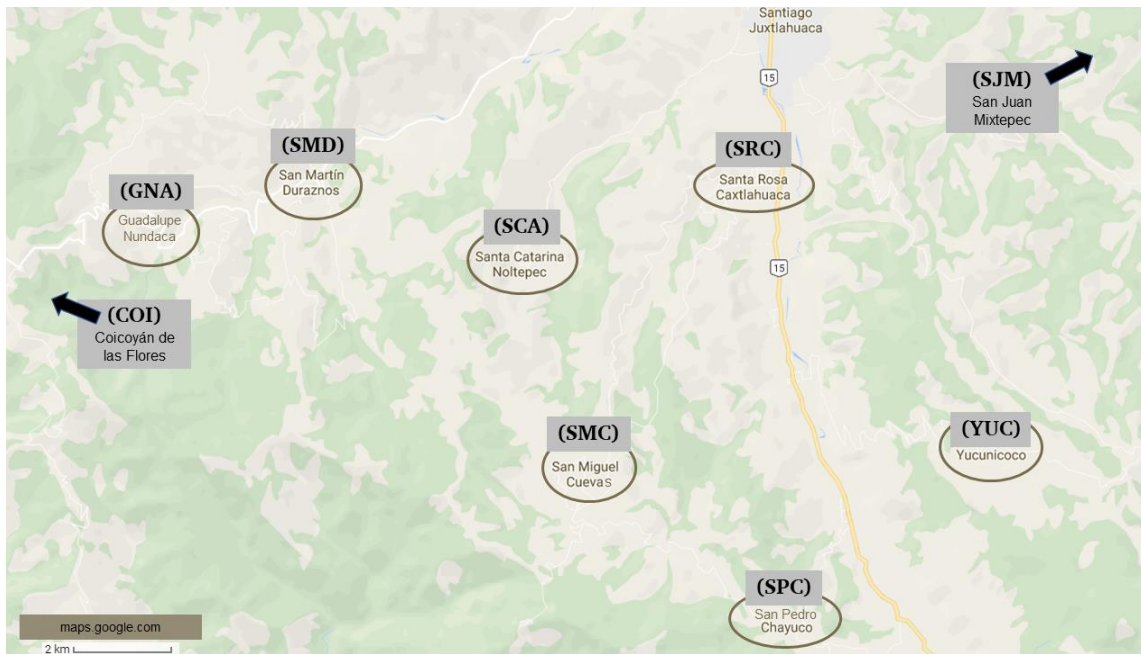
In the rest of this chapter, I present geographic, demographic, and government information that is important for understanding the local context of this research.

1.1 Geographic information

Beginning with geography, there is a range of sociolinguistic factors that informed my language survey work. The survey area is quite remote by most standards. The market town, Santiago Juxtlahuaca (also referred to as simply “Juxtlahuaca”), is about a five hour drive from Oaxaca City, the state capital. Figure 2 (above) shows the location of the survey area within the context of the state of Oaxaca.

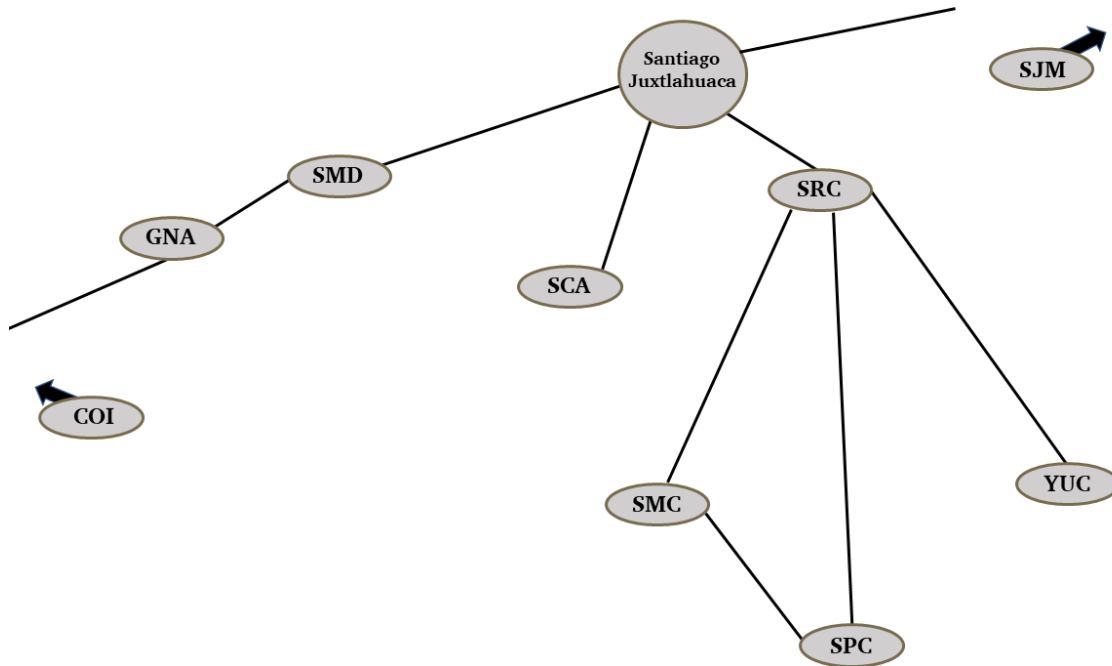
The road from Oaxaca City to the area is in poor condition (there are many potholes and landslides) and winds through mountains with many treacherous curves. From Juxtlahuaca, only Santa Rosa Caxtlahuaca (SRC) is accessible by entirely paved roads. The other villages are accessible primarily via gravel or dirt roads, with the occasional stretch of paved road interspersed. Figure 3 shows the distribution of the villages surveyed over this region.

Figure 3: Map of communities visited for language survey



For transportation, all the villages have *colectivos*, group taxis by which residents can travel to Juxtlahuaca, the market town. Otherwise, people walk or catch a ride with somebody they know who owns a car. Most people never travel to Oaxaca City, a trip that is made by *suburban*, large passenger vans that make the trip multiple times a day. Few people own their own car. The roads that connect each community are shown in Figure 4, which shows the general nature of these connections but not the actual shape and direction of the roads, as official maps of these dirt and gravel roads are not available.

Figure 4: Roads connecting communities surveyed



Most people seem to only travel to Juxtlahuaca, the market town, when they leave their community, and do so once a week for market day on Fridays, when they buy and sell produce. The weekly market is also where most people buy clothing and housewares.

Among the communities surveyed, cell phone service is only available in Santa Rosa Caxtlahuaca; the other communities are outside the range of cell coverage. Sometimes residents have sporadic cell service using large, long-range antennas that attempt to capture the cell service from Juxtlahuaca, where it is available.

The terrain in the region is quite rugged and difficult to traverse. This has contributed to the high level of isolation these communities have experienced over the centuries, which has likely contributed to the preservation of the language and culture. Only in the last few decades have infrastructure improvements made travel and interaction more common.

1.2 Demographic information

There are a number of demographic factors that are important to consider for a linguistic analysis of the area. Table 1 gives the population and location information for the towns surveyed as well as for the market town, Santiago Juxtlahuaca.

Table 1: Population and geographic information for villages surveyed
(Instituto Nacional de Estadística y Geografía 2010)

Community name	Population	Latitude	Longitude	Approx. Elevation
Santiago Juxtlahuaca (Market town)	9565	17.3353	-98.0124	1700m
San Martín Durazos (SMD) total:	967	17.3046	-98.1019	2000m
San Martín Durazos	460			
Yucuyi	231			
Río de Hielo	276			
Guadalupe Nundaca (GNA)	449	17.2987	-98.1343	2300m
Santa Rosa Caxtlahuaca (SRC)	1028	17.3035	-98.0163	1700m
Yucunicoco (YUC) total:	2954	17.2382	-97.9202	2300m
Santa María Yucunicoco	1511			
Lázaro Cárdenas Yucunicoco	484			
Cuauhtémoc Yucunicoco	137			
Zaragosa Yucunicoco	216			
Benito Juárez Yucunicoco	278			
La Laguna de Guadalupe Yucunicoco	189			
Buena Vista Yucunicoco	77			
Santa Cruz Yucunicoco	62			
Santa Catarina Noltepec (SCA)	585	17.2892	-98.0620	2300m
San Miguel Cuevas (SMC)	522	17.2441	-98.0507	2500m
San Pedro Chayuco (SPC)	500	17.2152	-97.9944	2000m

The SMD and YUC designations include multiple communities because of the close geographic proximity (within five kilometers) of those communities, and because local people reported in conversations that they consider them as single units during my prior six months of living in the area. This is perhaps due to a shared historical origin from a single community, or possibly because the smaller government councils of the communities subsumed under the SMD and YUC designations report, at least informally, to the government councils of the largest villages in each group.

The ethnicity of the communities surveyed is homogenously Mixtec, while the market town population also includes *mestizos*, or people with mixed Spanish and indigenous heritage. There is a high birth rate of about 3.5 (Instituto Nacional de Estadística y Geografía 2010), and generally, people have large families with multiple generations living together. Additionally, many people have some sort of handicap from accidents and lack of access to medical care.⁶

Most families live in one room homes (adobe or cement) with a separate kitchen house outdoors, usually made of wood. Most of the day is spent outdoors and the house is generally only used for sleeping at night. Homes may have dirt floors, though more and more people have conditioned floors (bare cement or tiled cement) since so many people work in the United States and send back money. Television programming is available in most communities, but not many families own their own TV set. Nearly all families have access to running water, though indoor plumbing is still quite rare. Water is pumped down the mountain from a nearby river, stream, or spring by the town governments.

⁶ The number of disabled people living in the area may appear abnormally high, since most of the able-bodied men have migrated elsewhere for work and are no longer living in the area.

Most families own at least a few acres of land and practice subsistence agriculture. Beans, squash, and corn are the most common crops. Many species of fruit grow naturally in the area, too, such as peaches, apples, and guava. I found the climate temperate (around 20°C on average) and most rainfall occurs during June and October.

There is a very high rate of movement to the United States. Based on my experience in SMD, I would estimate that roughly 80% of the men in most of these communities between the ages of 15 and 60 are either currently in the United States, have just returned from there, or are planning their next trip. It is not clear what effects movement has had on language use and language vitality.

1.3 Government information

Governmental policies and organization have a significant bearing on the sociolinguistic situation of the area. Each of the villages surveyed (including each of the three towns subsumed under SMD) is an *agencia*, the smallest Mexican governmental designation. Village leaders are elected every year or every two years by the community. Positions cover everything from being the *agente* (agent or primary leader), deputy *agente*, to overseers of education, construction, government food and welfare programs, health, and more. It seems that the municipal, state, and federal governments do not interfere much with individual *agencias* and allow communities to effectively rule themselves.

Government-run preschool and primary schooling is available in each village. For secondary school, villages have *telesecundarias*, where students receive schooling via satellite, video, and other distance education strategies. Many students in these communities do not complete their secondary education. On the rare occasion that students are able to pursue

higher education, they attend secondary school in Juxtlahuaca and then *preparatoria* in Oaxaca City, a preparatory school that precedes a bachelor's degree.

Bilingual education in Spanish and the local language formally began in Mexico in the 1970s (Hamel 2016), though it was only with the passage of the *Ley de derechos lingüísticos de los pueblos indígenas* (General Law of Linguistic Rights of Indigenous Peoples)⁷ in 2003 that speakers of indigenous languages began to see many fruits of this policy (Terborg & García Landa 2006:457). There is a federal government mandate for bilingual education in Spanish and the local language, but in the communities surveyed, teachers are almost never from the communities where they teach, which means that they do not speak the language. With the exception of a few attempts by teachers in Guadalupe Nundaca trying to develop teaching materials in the local Mixtec, none of the communities surveyed or their schools have any written materials in Mixtec. Sometimes teachers will attempt to use Mixtec materials from another region, but the variety is so different that it just causes confusion for students. Therefore, the reality in these communities, much as in the rest of Mexico (Hamel 2016:446), is that education is only provided in Spanish.

This brief sociolinguistic sketch forms the backdrop for my language survey work. Chapter 2 presents the methodology of the language survey, including my three main research tools; a General Wordlist, a Tone Wordlist, and a Sociolinguistic Questionnaire. Chapter 3, Findings, details the analyses and findings resulting from applying these tools. An analysis of similar phonological forms provides a straightforward basis for comparison of linguistic relatedness in combination with three new strategies. The first of these new strategies is a new application of

⁷ (Cámara de Diputados del Congreso de la Unión Mexicana 2003: Article 11)

an old tool, that of using isoglosses to map phonological correspondences. I use isoglosses to measure language relatedness in previously unstudied languages, rather than the traditional use of confirming existing classifications of well-studied languages. The second and third new tools are also of particular interest, as they present preliminary strategies for conducting language survey work in tonal languages, a previously unaddressed area of language survey work in unstudied languages. The tone analysis of individual forms uses Levenshtein Distance, a common method for evaluating relatedness, to measure the similarity of tones, rather than the typical comparison of segments. Based on these four analyses (similar forms, isoglosses, and two tone analyses), I find two linguistic groups of the Mixtec varieties in question, with two subgroups of two pairs of more closely related varieties that fall within the first group. Finally, these relatedness findings are compared with and corroborated by the reported intelligibility findings in the final analysis that I present. In Chapter 4, I discuss reservations about these new strategies for language assessment and areas for future study.

CHAPTER 2

METHODOLOGY

The survey discussed in this thesis was largely in line with traditional Rapid Appraisal (RA) language surveys (see Section 4.4 for a fuller description of RA surveys, including a potential objection to the sampling methods employed by RA.) For RA field trips, surveyors generally use questionnaires and wordlists in group settings, using individual interviews for clarifying areas of confusion. One of the primary developers of RA, Jürg Stalder, recommends meeting with at least two groups in each community (1993). In RA, the responsibility is on the researcher to decide when an area has been sufficiently surveyed, rather than applying each research tool a required number of times in each location. Sampling may not be representative but should, as much as possible, be from a variety of demographics (Stalder 1993). RA testing usually only requires a few hours in each village (Brye 2004).

RA stands in contrast to traditional language surveys, which are more in-depth and use a wider range of tools, such as Sentence Repetition Tests (SRT), in which participants listen to a sentence in the target language and respond with what they understood from the utterance (Radloff 1991), or Recorded Text Tests (RTT), in which participants listen to a brief story in the target language and respond, either by answering questions (Casad 1980) or by retelling it (Kluge 2007). Generally, these surveys are more time-consuming to develop and administer, and they most often use representative sampling methods.

For the purposes of this Rapid Appraisal survey, I used three main tools to gather data as I interacted with local residents during visits to Mixtec communities: a General Wordlist, a Tone

Wordlist, and a Sociolinguistic Questionnaire. For each tool, participants were native speakers of Mixtec who lived in the community and were willing to speak with us. They also had to speak enough Spanish to participate fully. We found participants that were available to work with us through a mutual acquaintance that served as an intermediary, or by approaching strangers on the street, or visiting a storekeeper, or knocking on doors. Men were nearly always out of the home during the daylight hours when we visited communities, and children were usually in school. In addition to cultural restrictions on interactions between women and men, these factors meant that my female teammate and I generally met with women. If men did participate, it was in the context of a group. Tools were administered to a mix of groups and individuals.

This overall approach to sampling certainly did not result in a representative sample of the population, but it did allow for the collection of data during just a few hours of a visiting a community while still respecting cultural norms.

While it is beyond the scope of this thesis to provide a thorough presentation of Mixtec phonology, I have found a basic understanding of the topic to be useful in developing tools for this research. For a description of Proto-Mixtec phonology, see McKendry (2013). Zylstra (1980) and Bradley (1970) provide sketches of the phonology of individual Mixtec varieties. Where appropriate, I have included basic phonological information to help the reader understand the development of research instruments and my analysis.

2.1 General Wordlist

Wordlists have long been a staple of dialectology and language assessment, and my General Wordlist was the basis for a significant portion of my analyses. For cross-linguistic

comparisons, standardized wordlists are one of the easiest tools to develop and administer. Wordlists make it possible to quantify inherently qualitative data, like phonological differences. Typically, they have been utilized for three purposes (Casad 1980):

- 1) analysis of the percentage of cognates,
- 2) analysis of percentage of similar forms, and
- 3) analysis of phonological correspondences displayed with isoglosses.

For this language survey, I used the General Wordlist for the latter two purposes.

2.1.1 A brief history of wordlists

In the mid-twentieth century, interest in historical linguistics and glottochronology motivated the development wordlists to use in language assessment. Swadesh published his first wordlist in 1950, which contained 225 items. This was later revised, through a series of iterations, to a 200 item wordlist and a 100 item wordlist, the final version of which was published in 1972. Swadesh developed his first list for the purpose of statistical comparison of lexical items (1950). This is generally understood as the beginning of the field of glottochronology, or the statistical study of the chronological development of languages and their divergences (Lees 1953:113). This field is connected to lexicostatistics and phonostatistics, or the statistical analysis of lexical and phonological similarities across time and languages (Lees 1953:113). I applied lexicostatistics in my analysis of similar forms (see Section 3.1), which is based on the framework laid out by Blair (1990). Blair's system operates on the basis of comparing parallel segments to determine their degree of similarity, and also provides criteria for the threshold that must be met in order for the words to be considered similar. Since Swadesh's initial list, which was largely based on his own fieldwork (Swadesh 1950),

many other wordlists have since been developed, some of which have been created based on cross-linguistic data, such as the Dolgopolsky list (1964) or the Leipzig-Jakarta list (2009).

Mapping has also been a long-time primary application for wordlist data. The first known attempt at “dialect geography” was made by Georg Wenker in Germany in 1876 (Chambers & Trudgill 1980:18). His study involved mailing a list of forty basic sentences in standard German to school teachers in the north of Germany so that they could rewrite the sentences into the local dialect (Wenker 1877). By the end of this project, he had received responses from a staggering 45,000 of the 50,000 school teachers that he contacted. The sheer scope of this project meant that Wenker eventually had to focus his attentions on mapping the variations of just a few select words from the data, but he later continued to gather other questionnaires (Chambers & Trudgill 1980:19). Another early attempt at dialect mapping was made in Denmark in 1898 by Marius Kristensen (Chambers & Trudgill 1980:19).

One of the most significant early advances in dialectology and language survey came from a Frenchman, Jules Gilliéron. In 1896, he was the first to use a trained fieldworker for a dialect survey. He used a 1,500 item wordlist of isolated, specific items to be elicited by a fieldworker, who transcribed each participant’s responses using phonetic notation worked out by Gilliéron. This fieldworker, Edmond Fremont, cycled through the French countryside for years, ultimately conducting 700 interviews in 639 different places (Jaberg, Jud & Scheuermeier 1928). Though the demographic variety of participants is questionable (only 60 were women and only 200 were educated beyond the norms of rural education for the time), this survey was quite influential for the development of later surveys (Chambers & Trudgill 1980:20). Gilliéron later consulted on and contributed to many other surveys of languages spoken in Europe.

Two of Gilliéron’s protégés, Karl Jagberg and Jakob Jud, went on to conduct another project of their own on Italian dialects spoken in Italy and southern Switzerland (Jaberg, Jud &

Scheuermeier 1928). Jud and another fieldworker from the Italian project, Paul Scheuermeier, then went on to the United States in 1931 to train other fieldworkers for the *Linguistic Atlas of the United States and Canada*, a massive undertaking (Chambers & Trudgill 1980). Another contemporaneous and influential undertaking was the Survey of English Dialects, begun in 1948. Out of financial necessity, one of the project's leaders, Harold Orton, decided to publish the survey with only a compendium of each participant's responses to each question of the interview (Orton, Sanderson & Widdowson 1978). This was done in lieu of the usual maps with responses overlaid, and it proved very useful for other researchers interested in the data and further investigation of linguistic variation (Chambers & Trudgill 1980:22). This allowed for side-by-side comparison of data, perhaps a precursor to today's phonostatistics and lexicostatistics.

Many people think about the work of Swadesh when they think about wordlists, but Antoine Meillet (1925) was another early user of wordlists. Meillet employed the three standard sources of evidence for establishing genetic relationships between dialects and languages—morphological similarities, phonological correspondences, and basic vocabulary. He favored studying morphological similarities in particular (Meillet 1925:36), though his discussions of regular phonological correspondences and “phonetic laws” are also well known (Campbell & Poser 2008:181). These three areas of comparison remain the basis for assessing relationships between languages, and I have used phonological correspondences as a major basis for my own analysis of language relatedness.

2.1.2 Using the General Wordlist

My primary tool for this survey was the General Wordlist, which consisted of 98 words and was specifically developed to target the phonological correspondences according to which

individual Mixtec varieties typically vary, and that are indicated by isoglosses, which are the geographic boundaries between phonological correspondences. They have been used by other linguists to differentiate language dialects and varieties, such as for Japanese (Shibatani 1990) and Dutch (Heeringa & Nerbonne 2001). However, their use in this thesis is for the initial establishment of the boundaries among understudied language varieties, rather than delineating differences among known language varieties.

Mixtec languages vary phonologically a great deal, but do so largely with respect to a few parameters, such as nasalization, palatalization, and tone. With the help of Inga McKendry, a linguist specializing in Mixtecan languages (2001, 2011, 2013), I created a list that aimed to capture the known phonological correspondences that coincide in their geographic distribution to isoglosses. The list consisted mostly of nouns and verbs, while words that would be difficult to elicit, such as pronouns, were not included in the list. Many varieties of Mixtec express verbal aspect with tonal and morphological changes, so verbs were elicited in a particular frame (third person masculine singular subject, perfective aspect, included the word *ayer* ‘yesterday’ in each phrase to clarify when the event occurred) to ensure consistency across participants and varieties. The General Wordlist was used to identify lexical and phonological differences among Mixtec varieties.

In each town, I elicited and audio-recorded at least one native speaker saying each of the items on the wordlist.⁸ I provided the target word in Spanish (or, in the case of verbs, the whole phrase including the frame) and the participant(s) said the word in their variety of Mixtec. I audio-recorded throughout this process of elicitation. In some cases, it was necessary

⁸ I used a Zoom H2 Handy Recorder for audio recordings.

to provide clarification about the meaning of a word for the participant, or ask the participant to repeat a word when it seemed likely that background noise obscured the recording.

In many cases, I elicited and recorded the wordlist with multiple participants to check for consistency or to verify a word that the first participant felt unsure about. The presence of more than one participant also served to verify data as members of the group corrected each other or worked together to figure out an unfamiliar word. I made audio recordings of each elicitation of the wordlist, whether with individuals or groups, which I then listened to at home and transcribed, checking my transcriptions multiple times against the recordings and against each of the other elicitations. For the sake of time, I did not transcribe data on the spot while eliciting. However, this approach also meant that if I was unsure of a word or sound, the participant was no longer present to clarify. On the other hand, I felt confident in my ability to listen and transcribe language data accurately, given my prior six months of language learning in SMD.

Table 2 shows the words that were elicited. Transitive verbs are listed with their object in parentheses.

Table 2: General Wordlist

	Spanish gloss	English gloss		Spanish gloss	English gloss
1	abeja	bee	50	harina	flour
2	abono	fertilizer	51	hervir (el agua)	to boil (water)
3	ala	wing	52	hilo	string
4	amarrar (la leña)	to tie together (firewood)	53	hombre	man
5	amarillo	yellow	54	jabón	soap
6	ampolla	blister	55	jícara	pitcher
7	ancho	wide	56	largo	long
8	anillo	ring	57	lejos	far away
9	árbol	tree	58	llegar	to arrive
10	blanco	white	59	lluvia	rain
11	borrego	sheep	60	luna	moon
12	bravo	bold	61	madera	wood
13	bueno	good	62	manteca	lard
14	caballo	horse	63	México, D.F.	Mexico City
15	cabello	hair	64	miel	honey
16	su (masc) cabeza	his head	65	mojado	wet
17	calabaza	squash	66	mucho	a lot
18	caminar	to walk	67	negro	black
19	camote	yam	68	noche	night
20	canasta	basket	69	nueve	nine
21	cantar (una canción)	to sing (a song)	70	nuevo	new
22	carne	meat	71	oscuro	dark
23	casa	house	72	pasto	grass
24	cerca	nearby	73	palabra	word
25	cerro	mountain	74	pecado	sin
26	cerdo	pig	75	peine	comb
27	chayote	chayote	76	pelear	to fight
28	chile	chili pepper	77	perro	dog
29	ciego	blind	78	plato	plate
30	colibrí	hummingbird	79	pobre	poor
31	colgar (la ropa)	to hang (laundry/clothing)	80	polvo	dust

32	comal	comal (<i>tortilla griddle</i>)	81	raíz	root
33	comprar (tortillas)	to buy (tortillas)	82	reír	to laugh
34	cortar (un árbol)	to cut (a tree)	83	río	river
35	coyote	coyote	84	rico	rich
36	cuatro	four	85	rojo	red
37	dos	two	86	sal	salt
38	dulce	sweet	87	serpiente	snake
39	elote	ear of corn	88	semilla de chile	chili seed
40	enfermedad	illness	89	siete	seven
41	escarbar (la tierra)	to dig (soil)	90	tenate	tenate (basket)
42	escuchar (una canción)	to listen (song)	91	tía	aunt
43	flor	flower	92	tlacuache	opossum
44	flor de calabaza	squash blossom	93	tortilla	tortilla
45	frijol	bean	94	vapor	vapor/steam
46	gallina	hen	95	vender (tortillas)	to sell (tortillas)
47	gallo	rooster	96	ver, mirar (una película)	to see/watch (movie)
48	grueso	thick	97	verde	green
49	hablar	to speak	98	zorrillo	skunk

2.2 Tone Wordlist

One of the primary ways in which Mixtec languages vary is in their tone patterns (Hollenbach 2003:2-3), as shown in example 1, so it is worthwhile to include a study of tone in any survey of Mixtec varieties when possible.

(1) 'bee' (General Wordlist #1)

ñùñù̀	L L	GNA
ñūñú́	M H	YUC
ñūñù̀	M L	SMD

For these three forms, the segments are nearly identical; SMD differs from the other two only in the palatalization of the second nasal. However, there are three distinct tone patterns present in the three forms. This is just one example of how tone can vary greatly from variety to variety in Mixtec.

Until now, assessing tones has been a significant and unmet challenge for language surveys of tonal languages. In my case, it was only possible to include tone in my survey thanks to many years of others' work on Mixtec tone. McKendry, whose dissertation on Mixtec tone provides one of the clearest understandings of the tone system of Mixtec languages to date (McKendry 2013), provided assistance with designing the tone portion of the survey. On the basis of thirty years of Mixtec study and analysis, she developed a Tone Wordlist of 50 items aimed at identifying the six most common tone patterns in Proto-Mixtec, which can appear in modern Mixtec varieties as patterns involving L (low), M (mid), and H (high) tones (McKendry 2013:157).⁹ Her work builds on the work of other linguists on the historical development of Proto-Mixtec into modern Mixtec varieties, including Longacre (1957) and Dürr (1987). Table 3 shows the words used to elicit these tone patterns. Tone Wordlist items that also occur in the General Wordlist are listed with their item number from the General Wordlist. Proto-Mixtec patterns are indicated by an asterisk preceding the pattern. The superscript glottal stop on the

⁹ It is notable that these six most common tone patterns do not include H tones. Dürr's reconstruction of the Proto-Mixtec tone system does include patterns that include H tones (1987:24), but the most common tone patterns in modern Mixtec varieties are derived from the six included in Table 3.

Proto-Mixtec patterns refers to the word-final glottal stops (usually analyzed as a feature of the vowel or morpheme, hence the superscript) believed to occur in Proto-Mixtec. These usually occur as word-final floating H tones in modern Mixtec varieties (McKendry 2013:21).

Table 3: Tone Wordlist

Pattern 1 *L		Gen. WL #	Pattern 4 *M ²		Gen. WL #
Spanish	English		Spanish	English	
dos	two	37	chile	chili pepper	28
nueve	nine	69	elote	ear of corn	39
peine	comb	75	espina	spine or	
sacerdote	priest		fiesta	party	
siete	seven	89	frijoles	beans	45
tres	three		masa	dough	
Pattern 2 *L ²		Gen. WL #	mecate	woven rope	
Spanish	English		petate	woven mat	
cuatro	four	36	tierra	soil/earth	
culebra	snake	87	Pattern 5 *ML		Gen. WL #
Spanish	English	jabón	Spanish	English	
lluvia	rain	59	flor	flower	43
maíz	corn		fuego	fire	
metate	grinding		papel	paper	
nube	cloud		perro	dog	
piedra	rock		pueblo	village	
plato	plate	78	venado	deer	
tortilla	tortilla	93	Pattern 6 *LM		Gen. WL #
Pattern 3 *M		Gen. WL #	Spanish	English	
Spanish	English		casa	adobe	adobe
uno	one	23	chapulín	grasshopper	
			ejote	green bean	
			hombre	man	53
			hormiga	ant	
			manteca	lard	62
			campana	bell	
			olla	pot	
			río	river	83

I elicited and recorded the Tone Wordlist at least once in each community, and often with more than one speaker, either in a group of participants or with individual participants, one

after the other. In most cases, the participant that provided the Tone Wordlist was the same person that provided the General Wordlist. Words were elicited by providing the Spanish meaning of each word. As with the General Wordlist, it was sometimes necessary to provide clarification about the meaning of a word for a participant. It would have been useful to elicit each word three times to improve accuracy, but words were only elicited once.

When working with tone, there is always a concern about the accuracy of the pitch transcriptions, but my six months of language learning greatly assisting in hearing and accurately transcribing tone data. Furthermore, I checked my transcriptions of tones multiple times and also by measuring the frequencies (in Hertz) using the computer program Praat (Boersma & Weenink 2012). While Praat is not recommended for the initial establishment of tones, it was useful for confirming my transcriptions and especially aided in clarifying less obvious pitch differences by identifying the pitch range for each tone for each speaker. Figure 5 gives an example of several LL word forms and Figure 6 gives an example of several MM word forms, shown in Praat. Both sections of recording are from the same female native speaker on the same occasion.

Figure 5: LL forms in Praat

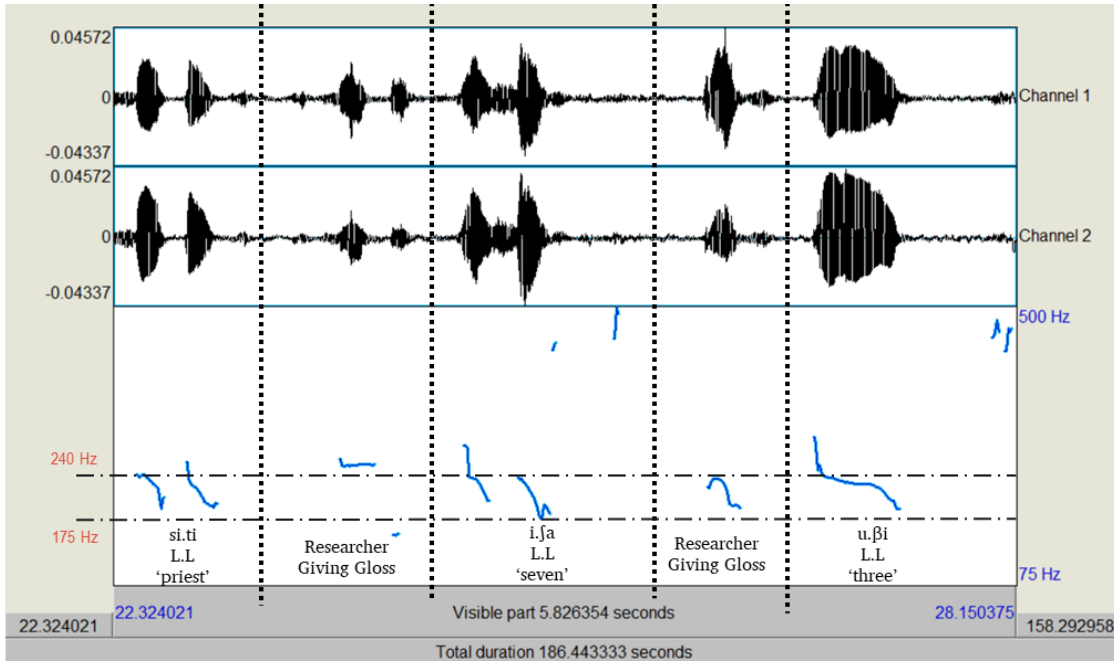
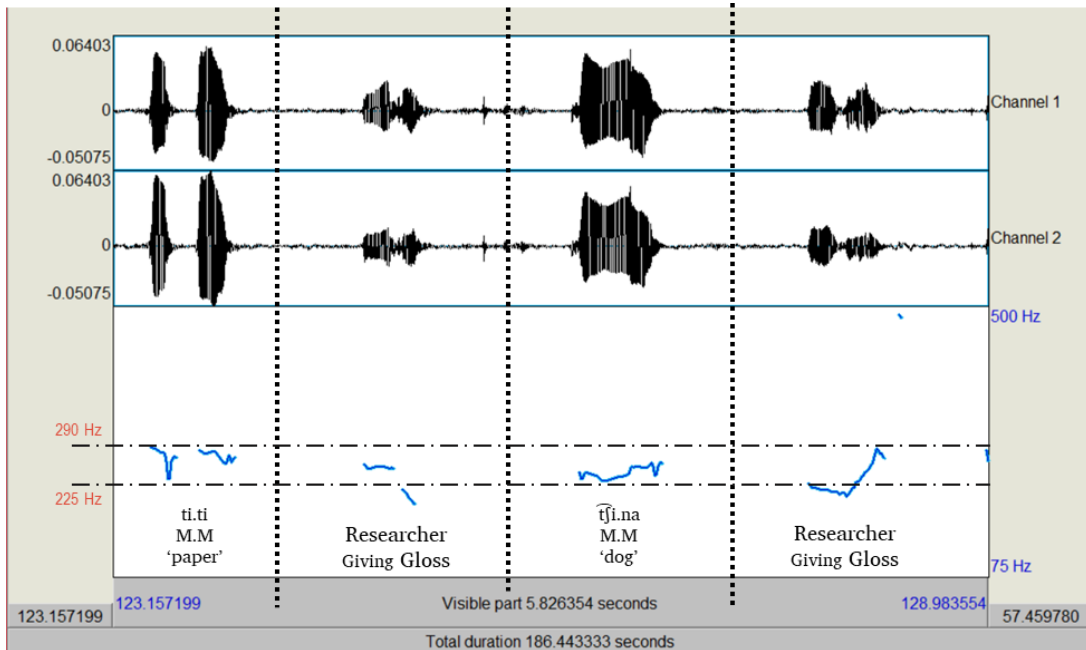


Figure 6: MM forms in Praat



In both cases, there is a clear range of frequencies for the pitches. Though these ranges overlap slightly, L tones are clearly distinct from M tones because they have a falling pitch on the second mora. Excluding syllable onsets, the average pitch for the L tones shown above is approximately 207.5 Hz, while the average pitch for the M tones shown is 257.5 Hz.

2.3 Sociolinguistic Questionnaire

My third tool was a Sociolinguistic Questionnaire, which I used to assess language use patterns and language attitudes, as well as reported intelligibility among Mixtec varieties.¹⁰ I made particularly heavy use of this tool in the villages that make up SMD (Río de Hielo, San Martín Durazos, Yucuyi), as one of the main goals of the survey was to determine the vitality of the Mixtec spoken in these communities (see Table 1). I did not administer the Sociolinguistic Questionnaire in Yucunicoco. For the purposes of this paper, I will focus on the use of the questionnaire for measuring reported intelligibility.

2.3.1 *A brief history of questionnaires*

Questionnaires, much like wordlists, have been an essential tool for language assessment since the mid to late twentieth century, though they were occasionally in use before then (Karan & Stalder 2000:198). As the study of intelligibility developed, researchers realized the importance of continuing to include questionnaires in their research in addition to more

¹⁰ In this thesis, I refer to “intelligibility” rather than “mutual intelligibility” in order to better address all the factors that influence comprehension of a language by a native speaker of a different language. Typically, “mutual intelligibility” is understood to refer to intelligibility that is a result of genetic relatedness of two languages, whereas “intelligibility” refers to comprehension that may be due to genetic relatedness and/or sociolinguistic factors, such as contact and language attitudes (Simons 1979:12).

quantitative tools, such as wordlists, sentence repetition tests (SRT), and recorded text tests (RTT) (Casad 1980:56).

A primary concern came from Voeglin and Harris' realization that intelligibility does not always go in both directions; people from town A might understand people from town B well without people from town B being able to understand much of what people from town A say (Voeglin & Harris 1951:329). Voeglin and Harris also distinguished between *neighbor intelligibility* (intelligibility that is due to extensive contact between two language groups) and *inherent or mutual intelligibility* (due to close genetic relationship) (Casad 1980:52). This distinction highlights the need not only for methods of testing mutual intelligibility and language relatedness, which affects inherent intelligibility, but also the sociolinguistic factors that affect neighbor intelligibility.

Furthermore, linguistic relatedness does not always predict intelligibility adequately. This is because intelligibility is not only impacted by linguistic variables, such as morphology, phonology, syntax, and lexicology, but also sociolinguistic variables, such as language contact, language attitudes, speaker aptitudes, and cultural values (Simons 1979:69). This further supports the need to use a variety of tools to measure the sociolinguistic factors of intelligibility.

Wolff (1964:441) argued that tests that primarily measure a subject's ability to translate between two languages (which involves more than just questions of intelligibility) are not a good measure of dialect distance. He also raised concerns about the fact that most participants were being asked to give responses in their third language, further limiting the accuracy of data, and that the conditions for testing were not standardized, allowing for bias and variation (Wolff 1964:76). Wolff's critiques strengthen the case for using both quantitative and opinion testing (i.e., questionnaires).

Though primarily used for measuring language use, attitudes, and vitality, sociolinguistic questionnaires can be useful for measuring intelligibility, keeping in mind that they measure only what participants *say* and *think* they understand, rather than objective evaluations of how much they comprehend. This is especially a concern in a shame-based culture, as I found the *Mixteca* (Mixtec-speaking region) to be, where speakers overreport their level of understanding. Attitudes toward other language communities may also skew participants' answers. To mitigate these effects, it is possible to use matched guise techniques to verify self-reports. In matched guise tests, such as those used by Lambert (Lambert 1967), participants are asked to evaluate the speakers of sections of recorded speech for kindness, intelligence, ambition, sincerity, etc. More than one of the sections is recorded by the same person, but in a different language/dialect (guise). Differences in evaluations of the sections with the same speaker reveal attitudes about the varieties in question (Karan & Stalder 2000:192). Surveyors could alternatively include matched guise questions in the questionnaire by requesting participants to pick between two imaginary but described people as the future spouse for a sibling, an elected government official, or somebody that is most likely to succeed (Karan & Stalder 2000:192).

Systematic sampling of households is ideal for questionnaires, such as in the case of Fowler's (2002) area probability sampling, in which the total target area is exhaustively subdivided into areas that are then sampled. However, inferior sampling methods, such as simply aiming to include a wide variety of demographics in the sample, can still provide data of sufficient quality (Karan & Stalder 2000:194). When using questionnaires, it is helpful to use a five-value scale (i.e. strongly disagree, disagree, uncertain, agree, strongly agree.) This makes it easy to analyze data and quantify motivations (Karan & Stalder 2000:195).

A well-designed questionnaire can touch on language vitality, language use, and language attitudes, provided that participants have the necessary self-awareness to recognize and

articulate their own beliefs and feelings, and are able to understand questions (Henerson, Morris & Fitz-Gibbon 1978:12). Whole studies have been based on simply using questionnaires, without the addition of RRT, SRT, or wordlists. For example, Haugen's 1966 study of Danish, Norwegian, and Swedish, one of the first uses of a questionnaire for measuring intelligibility, consisted of a questionnaires asking 300 participants about their experiences regarding each of the three languages in question. More recently, Chan (2007) used just a questionnaire to conduct a survey on intelligibility and language vitality of the Gan Yi and San Yi in China, which showed that these speech communities are undergoing language shift.

Questionnaires are easy to develop and administer and allow researchers to cover many topics in little time. They also allow for flexibility with respect to administration. The Sociolinguistic Questionnaire used for my survey proved to be a useful tool for precisely these reasons.

2.3.2 Using the Sociolinguistic Questionnaire

Questionnaires were administered to at least two participants per community. Sometimes the questionnaire was administered to individuals and sometimes to groups. Many cultural issues affected the administration of the questionnaire.

A primary concern with this tool is that, culturally, it is common for a person to tell another what they think the other wants to hear, particularly when there is an imbalance of power between the two people, as when there is a light-skinned, educated outsider asks to interview an indigenous speaker who may not speak Spanish well or have much formal education. Questions were designed to ask about specific, concrete experiences of participants so as to avoid skewing of answers that would be caused by this imbalance of power.

Another concern is the cultural value of saving face and avoiding shame, which might bias participants' answers. Native speakers of Spanish, who form the cultural majority of Mexico, consider Mixtec inferior to Spanish, so many Mixtec speakers feel shame due to their heritage and language. To some extent, shame is an unavoidable bias because it is a central part of the local culture. However, to minimize this bias, I intentionally discussed with the participants the value of Mixtec and the local culture prior to administering the questionnaire. This introduced a new possible bias in the opposite direction, whereby participants might tell me what they think I want to hear about how valuable their language is. To guard against this bias, I held long discussions with participants in which I probed their motivations further.

Yet another concern is that asking direct questions is often offensive in a Mixtec context, so each question was re-phrased as a statement when administering the questionnaire. For example, Question 3 from the Language Use and Vitality Section was written as, "When you speak to your neighbors, what language do you use?" But when administering the questionnaire, this was phrased as a statement, "When you talk with your neighbors, you use Mixtec or...." where the statement trailed off in inflection and volume at the end to indicate the interrogative purpose of the statement. This is a more culturally appropriate way of finding out information, as I learned during my prior months of language and culture learning.

Table 4 shows the questionnaire, which focused on three major topics: language use and vitality, intelligibility, and language attitudes.

Table 4: Sociolinguistic Questionnaire

LANGUAGE USE & VITALITY		
1	Cuándo usted habla con sus hijos, ¿qué idioma usa? ¿Cómo le responden—en Mixteco o español?	When you speak to your children, what language do you use? How do your children respond--in Mixtec or in Spanish?
2	Cuándo usted habla con su esposo/a, ¿qué idioma usa?	When you speak to your spouse, what language do you use?
3	Cuándo usted habla con sus vecinos, ¿qué idioma usa?	When you speak to your neighbors, what language do you use?
4	Cuándo usted hace sus compras aquí en el pueblo, ¿qué idioma usa?	When you buy things here in the village, what language do you use?
5	¿Usted llega al mercado de Juxtlahuaca? ¿Qué idioma usa para hacer sus compras allá?	Do you go to the market in Juxtlahuaca? What language do you use when shopping there?
6	Cuándo sus niños juegan juntos entre ellos, ¿qué idioma hablan?	When your children are playing together, what language do they speak?
7	¿Qué idioma usan los niños de la comunidad cuando juegan juntos?	What language do children in the community use when they are playing?
8	¿Sus niños tienen clases o actividades en mixteco en la escuela? Las entienden?	Do your children have classes or activities in Mixtec at school? Do they understand them?
9	¿Alguna vez ha escrito en su idioma?	Have you ever written in your language?
10	¿Alguna vez ha visto su lengua (el mixteco que habla) en forma escrita? ¿O algún idioma mixteco en general?	Have you ever seen your language written down? Or any other Mixtec?
11	¿Todas las generaciones hablan la lengua?	Do all generations speak the language?
12	Cuando usted va a la agencia/la presidencia y habla con las autoridades, ¿qué idioma usa?	When you go to the agencia/presidencia and speak to the authorities, what language do you use?
13	Si las autoridades visitan su casa, ¿qué idioma usan?	If the authorities visit your home, what language do they use?
14	¿Usted va a la iglesia? ¿Qué idioma usan para las misas?	Do you go to church? What language is used for mass?

Table 4 cont'd

INTELLIGIBILITY		
1	¿Usted ha viajado alguna vez a otra comunidad? ¿Cómo se comunicó con las personas allá?	Have you ever traveled to another community? How did you communicate with the people there?
2	¿Alguna vez ha venido alguien de otra comunidad a su pueblo? ¿Qué idioma usó para hablar con personas del pueblo?	Has somebody from another community ever come to your village? What language did they use to talk with people from your village?
3	¿Es más fácil para usted entender el mixteco de Yucunicoco o el de Yucuyi?	Is it easier for you to understand the Mixtec from Yucunicoco or Yucuyi?
4	¿Es más fácil para usted entender el mixteco de San Miguel Cuevas o el de Yucuyi?	Is it easier for you to understand the Mixtec from San Miguel Cuevas or Yucuyi?
5	¿Es más fácil para usted entender el mixteco de San Pedro Chayuco o el de Yucuyi?	Is it easier for you to understand the Mixtec from San Pedro Chayuco or Yucuyi?
6	Aparte de la lengua de su pueblo, ¿cuál es el mixteco que entiende mejor?	Apart from the Mixtec of your village, which Mixtec do you understand best?
7	¿Cuál es el mixteco más difícil de entender?	Which Mixtec is the hardest to understand?
LANGUAGE ATTITUDES		
1	¿Qué son unos de los pueblos más cercanos a su pueblo? ¿En estos pueblos, hablan la misma lengua como usted?	What are some of the villages that are closest to yours? In those villages, do they speak the same language as you?
2	¿Cuándo es útil en la vida diaria usar el mixteco?	When is it useful to use Mixtec in daily life?
3	¿Cuándo es útil en la vida diaria usar el español?	When is it useful to use Spanish in daily life?
4	¿Usted quiere que sus niños aprendan mixteco?	Do you want your children to learn Mixtec?
5	¿Cuándo es útil que sus niños sepan mixteco?	When is it useful for your kids to know Mixtec?

Responses to the questionnaire were compiled and used to create a map indicating the general level of understanding reported by each participant for each village discussed: understand well, more or less understand, and hard to understand.

2.4 Observation

In each community, I also wrote down observations about language use made during each visit as a way to supplement and verify what other tools showed. I mainly used observations to clarify points of confusion or conflicting participant answers.

2.5 Summary of methodology

There were three main tools used for this language survey: a General Wordlist, a Tone Wordlist, and a Sociolinguistic Questionnaire. Observations from visits to each village were occasionally used to verify findings. (See Appendix A for a calendar of survey trips.) Table 5 is a summary of the tools used and the number of survey events in each village. I sometimes interviewed individuals and sometimes small groups, so the number of times each tool was administered, as well as the number of participants that were involved in the administration of the tool, is recorded, as well as the total number (rightmost column) of participants that were involved in the administration of the tools in each village.

Table 5: Summary of tools and village visits

	Gen. WL	Gen. WL Participants	Tone WL	Tone WL Participants	Socioling. Quest.	Socioling. Quest. Participants	Total Participants
SMD	1	1	1	1	5	15	15
SMC	2	6	2	6	3	7	7
SPC	2	6	1	2	1	2	6
SRC	1	2	1	1	1	2	2
GNA	3	6	2	5	2	5	6
YUC	2	9	1	1	0	0	9
SCA	2	4	2	3	3	4	4
TOTALS	13	34	10	19	15	35	49

General Wordlist and Tone Wordlist data are also included in analyses from the nearby communities of Coicoyán de las Flores (COI) and San Juan Mixtepec (SJM), nearby Mixtec communities, when they provide useful reference points for comparison. These data are from personal correspondence with colleagues from SIL who have worked in those of Oaxaca areas for over a decade. The Mixtec varieties spoken in these communities have previously been identified by SIL as clearly distinct from each other and from the Mixtec variety spoken in YUC.

CHAPTER 3

FINDINGS

I present five analyses of the data: percentage of phonologically similar forms, phonological correspondences that can be displayed by isoglosses, two analyses of tone patterns, and reported intelligibility. The data used for these analyses can be found in Appendix B (General Wordlist Data) and Appendix C (Tone Wordlist Data). My analysis of isoglosses shows that this tool, typically used for corroborating existing boundaries between well-studied languages, is also useful for studying the relatedness of lesser-studied languages for which the boundaries are unclear. This is particularly the case when an analysis of isoglosses is used to supplement a traditional analysis of similar forms.

I also present two tone analyses. The first tone analysis focuses on the six main tone patterns found in present day Mixtec varieties, based on reconstructions of Proto-Mixtec. The second tone analysis focuses on the dissimilarity of tone patterns of individual words and syllables, which utilizes a new application of a traditional tool, Levenshtein Distance. Both tone analyses provide new, albeit preliminary, strategies for studying tonal languages, an issue typically left unaddressed by early language assessment research.

Taken together, these four analyses provide a clear picture of the linguistic relations of the Mixtec varieties studied. This picture of linguistic relatedness is then compared to the levels of intelligibility reported by native speakers from the Questionnaire. With minor exceptions, the proposed relatedness correlates closely with intelligibility reported by survey participants, supporting the validity of using these tools to measure language relatedness.

3.1 Analysis of similar forms

Using data from the General Wordlist (see Appendix B), it is possible to gain an idea of the phonological similarity of the languages surveyed, and also their phonological similarity to COI and SJM. Following the guidelines laid out by Blair (1990), the similarity of corresponding segments was used for a traditional evaluation of similar forms. For each word elicited in each variety, segments were compared and then grouped into one of three categories (Blair 1990:31):¹¹

Category One: Segments that are an exact match, vowels that only differ by one phonological feature, and phonetically similar sounds that consistently occur in the same position in multiple words. (e.g. [e] and [i], [ʰd] and [ʰd͡ʒ])

Category Two: Phonetically similar consonants that do not consistently occur in the same word position, and vowels that differ by two or more features. (e.g. [ʰd] and [t], [a] and [i])

Category Three: Corresponding sounds that are not phonetically similar, additions, and deletions. (e.g. [n] and [β])

The more segments that fall into Category One and the fewer that fall into Category Three, the more similar two words will be. Pairs of words are compared and then binarily judged as similar or not similar. Additional considerations for classifying segments are found in Appendix D. Tone was excluded from this analysis.

¹¹ Examples are from Mixtec data that I collected.

Table 6 provides the guidelines for the acceptable number of segments in each category in order for a word to be considered similar to another. For example, the first row shows that in order for a word with two segments to be considered similar, both of its segments must fall into Category One and none can fall into Categories Two or Three when compared to the other word. For words with six segments, there can be up to one Category Three segment, up to two Category Two segments, and at least three Category One segments in the word that is being compared.

Table 6: Guidelines for determining similar forms (Blair 1990)

<u>Number of segments</u>	<u>Category One</u>	<u>Category Two</u>	<u>Category Three</u>
2	2	0	0
3	2	1	0
4	2	1	1
5	3	1	1
6	3	2	1
7	4	2	1
8	4	2	2
9	5	2	2
10	6	3	1
11	6	3	2
12	6	3	3

Category One includes any sounds that have correspondences which may be mapped using the isoglosses presented in Section 3.2. This includes correspondences between a segment and null, such as when [ʒ] and [n] are added or deleted. For example, the word for ‘wood’ is [ĩtù] in SMD, [ĩtù] in GNA, and [ʒũtù] in SRC. Since the words are identical in GNA and SMD, three out of three segments fall into Category One. Upon first glance, the word in SRC might seem quite dissimilar from those in SMD and GNA.

However, lining up the words as in example (2) shows that they are, in fact, similar.

(2) Ø i t ũ SMD
 ʒ u t ũ SRC

The third and fourth segments are identical, so they are Category One. Furthermore, [ʒ] and null form a regular correspondence, as do [i] and [u], making them Category One sounds, too. Therefore, four out of the four sounds are Category One, thus exceeding the threshold of two Category One sounds, one Category Three sound, and one Category Four sound that is required for the two four-segment words to qualify as similar.

Looking at the word for ‘sheep’, [lā^ad̥ʒī] in SMD and [bōrégō] in SCA, all of the segments in the SCA form are Category Three when compared to SMD, meaning that these two words are not similar forms.¹² There are four segments in the SMD form that the SCA form is being compared to, so based on Table 6, there would need to be a maximum of one Category Three segment, a maximum of one Category Two segment, and at least two Category One segments to qualify the two forms as similar.

In cases of metathesis, differences in glottalization,¹³ and differences in nasalization,¹⁴ the segment being compared is downgraded one category. For example, ‘word’ is [t̃ũ?ũ] in SMC and [t̃ũũ] in SPC. My convention is to consider long vowels as a single segment, so the segments for both forms are the same, /t̃ũ/. However, the missing glottal stop in the SPC form,

¹² [bōrégō] is a loanword from the Spanish word for ‘sheep’ *borrego* and [lā^ad̥ʒī] may also be a loanword from the Spanish word for ‘wool’ *lana*, so it is already clear that these two words are dissimilar.

¹³ Glottalization can be analyzed as a feature of the root in most Mixtec languages. See Macaulay & Salmons (1995). It may also be analyzed as a feature of the vowel. See Josserand (1983) and McKendry (2013).

¹⁴ Nasalization can be understood as a feature of the morpheme in Mixtec languages. See Marlett (1992) and McKendry (2013).

which can be understood as a feature of the morpheme (Macaulay & Salmons 1995), means that one of its Category One segments is downgraded to a Category Two. Table 6 shows that for words with two segments, both must fall into Category One in order to qualify two words as similar. Therefore, according to the guidelines set forth by Blair, these are not similar forms, since the form in question has one Category One segment and one segment downgraded to a Category Two segment. For other considerations, exceptions, and a more detailed accounting of how segments were categorized, see Appendix D.

The total number of similar forms was then added up for each variety and divided by the total number of pairs. Table 7 shows the percentage of similar forms for each variety.

Table 7: Percentage of similar forms

SJM									
66	COI								
70	75	SMD							
70	77	90	GNA						
69	74	88	89	SRC					
68	73	88	90	96	SCA				
65	71	81	84	86	88	SPC			
66	70	79	82	83	84	84	SMC		
68	68	67	70	71	70	68	70	YUC	

COI and SJM had already been identified previously by SIL-Mexico as distinct languages that could not share the same literacy materials or language development materials. Unsurprisingly, they have the lowest percent of similar forms to each other (66%). YUC’s percentage of similar forms is not much higher (67% to 71%), signaling it is quite distinct from all the other varieties, too. The closest any two varieties are to each other are SCA and SRA, which have similar forms 96% of the time.

Based on these data, SMD, GNA, SCA, SRC, SPC, and SMC could all be grouped loosely together, while within that core group, SRC and SCA are significantly closer to each other than to the others (96%). Furthermore, all of these members of the core group can be tentatively classified as a variety of Mixtec separate from YUC. Usually, any percentage below 70% is taken to mean that the two varieties in question are distinct languages, while a percentage above 70% does not necessarily indicate close similarity, but rather that more information is needed to clarify whether or not the languages are the same (Summer Institute of Linguistics 1989). I will later use the analyses of phonological correspondences displayed by isoglosses and the analysis of tone to clarify the status of YUC.

A suggested grouping (in terms of similar forms) of the surveyed communities is as follows:

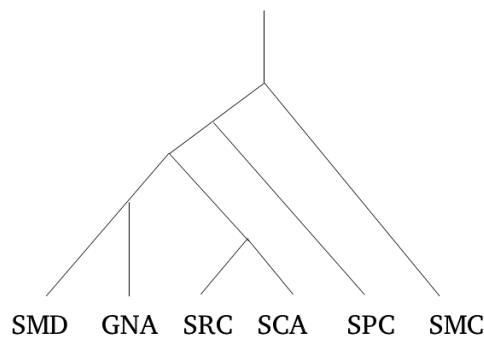
(3) Group 1: SMD, GNA, SCA, SRC, SPC, SMC

Subgroup 1A: SRC, SCA

Group 2: YUC

Figure 7 uses a tree to illustrate these relatedness groups for Group 1. Note that, though trees of this sort are typically used to demonstrate genetic relationships, this particular tree is merely showing proposed relatedness groups, which may be due to genetic relationships or other factors.

Figure 7: Tree of relatedness for Group 1



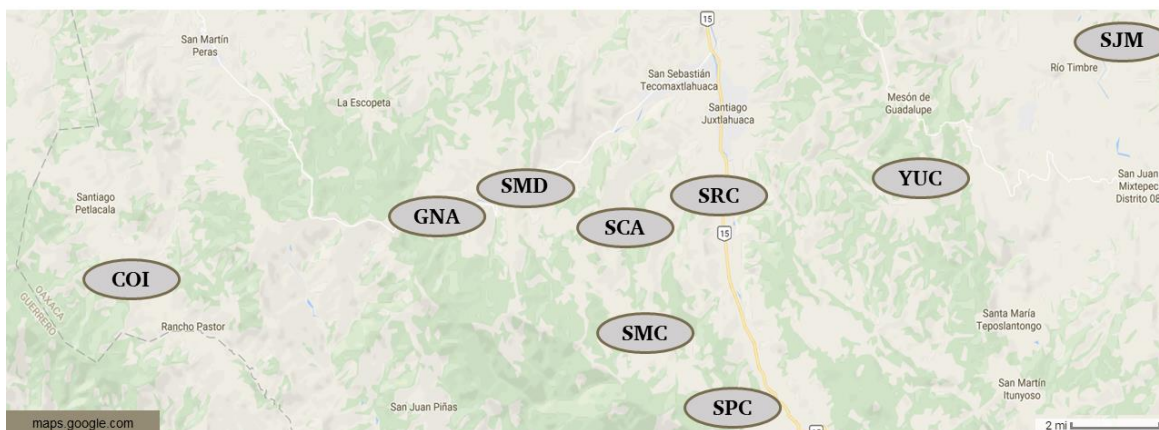
While it usually makes sense to base an analysis of relatedness on an understanding of the historical progression from the proto-language to the modern variety, the history and genetic relationships of Mixtec languages in this region are not well understood, but an in-depth understanding of such relationships is not necessary to make sense of today's realities. In fact, geographic distribution provides a feasible explanation for many of these patterns of similar forms. Geography is generally understood to affect borrowings, but it may have also played a role in the historical development of genetic relationships between languages in this region. For example, until a few decades ago, nearly all travel between communities in the area was on foot over very rugged terrain, mountains, and rivers, which increased the isolation of these language communities. When geography is discussed below, it is with the understanding that it includes both genetic relationships and borrowing.

Within Group 1, SMC is the least similar to the other varieties. This is to be expected, as the SMC community is more remote from the other communities and does not share a road with any other community. SRC and SCA, which are quite similar, are both located quite near to Juxtlahuaca, the market town, and near to each other geographically. Due to this close proximity, and in spite of never having shared a road, these two communities would likely have found it markedly easier to visit each other and interact over the course of history, compared to other communities. SRC is located on the same major highway that passes by SPC and historically within the same river valley, perhaps contributing to SPC's higher degree of similarity to SCA and SRC than to other communities. GNA and SMD, which have a slightly higher percentage of similar forms than other varieties, are located on the same road, which is the only major road in the region, and are fairly close together (see Figure 4). Finally, YUC, which has the most dissimilar forms, is presently only accessible by a two-hour drive up switchbacks on a gravel road. Historically, this community would have been very hard to get

to, as one must climb several thousand feet and traverse several mountains to reach it.¹⁵ This geographic separation provides a likely explanation for YUC's low degree of phonological similarity for the following reasons.

First of all, SJM and COI, which are provided for reference, are located many mountains and several hours away (by car) from the seven communities researched, and are correspondingly less similar to those communities. The relative location of COI and SJM is shown in Figure 8, although many of the roads connecting these communities and the ones surveyed are not shown since they are quite remote and usually unpaved. See Figure 4 for how roads connect these communities.

Figure 8: Relative location of COI and SJM



COI is most similar to Group 1 (70% to 77% similar, see Table 7) and is more similar than even YUC (70-67%). This is a strong case for placing YUC in its own linguistic group, as COI is considered a distinct language from the varieties in Group 1 since the experience of SIL colleagues working in COI was that COI is distinct (Bruce & Candice Beatham, p.c.). Therefore,

¹⁵ Today, YUC is 30km from Juxtla huaca by gravel road. The road is mostly switchbacks and usually takes between one and two hours to traverse, depending on conditions.

if YUC is less similar to Group 1 than even COI, it is likely a distinct variety from those in Group 1.

Furthermore, SJM is about as similar to Group 1 as YUC (65% to 70%), the difference being that SJM is closest to the SMD/GNA end of the tree (just as COI), while YUC is closest to the SMC end of the tree. Because of geography, one would expect COI to be more similar to the SMD/GNA end of the chain, as they are the nearest by road of the six communities.

However, out of the six communities in Group 1, SJM is the furthest geographically to SMD/GNA but closest phonologically to SMD/GNA. Additionally, YUC is quite far from SMC (about three hours by car) and closest to SRC by road, so it is unexpected that it would be roughly as similar to SMC as to SRC, which is much closer to YUC. In both of these cases, it is unclear the extent to which other factors, such as migration or contact, may contribute to this linguistic similarity despite geographic distance.

In most language surveys, and especially RA surveys, the principal measures of relatedness are the analysis of similar forms (lexicostatistic analysis), as I have just presented, and the analysis of a questionnaire (Section 3.4). However, I am proposing three additional tools that help to fine-tune the results of these more traditional measures:

- 1) analysis of sound correspondences using isoglosses (Section 3.2),
- 2) analysis of tone patterns (Section 3.3.1), and
- 3) using Levenshtein Distance to analyze tone differences (Section 3.3.2).

I conclude with an analysis of reported intelligibility (Section 3.4) to corroborate these findings.

3.2 Grouping language varieties according to phonological correspondences using isoglosses

Analyzing sound correspondences through the use of isoglosses constitutes the second part of my analysis. While typically used to confirm the boundaries between well-studied languages, isoglosses can be applied to an investigation of linguistic relatedness for largely unstudied language varieties, provided there are either prior research, historical reconstructions of the language, or the varieties in question vary in systematic ways. As mentioned above in Section 2.1.2, isoglosses are the geographic boundaries between sound correspondences. In data collected for the General Wordlist (see Appendix B), many phonological correspondences varied systematically, the resulting patterns of which can be displayed using isoglosses, as was expected. The data show eleven correspondences, which pattern in ways that are displayed by isoglosses. (These are listed as rows in Table 8 below.) These resulting isoglosses display boundaries among language varieties.

Table 8 presents the eleven correspondences, along with examples for each of the two sounds in each correspondence pair, labelled Zone A and Zone B. Note that the zones change for each isogloss depending on how the sound correspondences are geographically distributed; the zones do not correspond to the linguistic groups based on relatedness, proposed above in Section 3.1 or below in this section. Tones are excluded from the examples, as they vary across varieties, even when segments are the same. The sounds in question appear to occur without restrictions on word position (word-initial, word-medial, word-final). See Appendix E for a list of which words from the General Wordlist were used to identify these correspondences.

Table 8: Isoglosses

<u>Isogloss</u> #	<u>Zone A</u>	<u>Zone B</u>	<u>Gen.</u> <u>Wordlist</u> #	<u>Gloss</u>	<u>Zone A</u> <u>example</u>	<u>Zone B</u> <u>example</u>
1	k ^w	β	66	a lot, much	k ^w a?a	βa?a
2	ʃ	s/ʒ/ks	55	sing	ʃita	ʒita
			21	buy	ʃi	iksi
			33	gourd	ⁿ dʒaʃi	ʒasi
3	ʃi	iʃ	93	tortilla	ʃita	iʃta
4	t̃ʃ	t	24	near	ʒat̃ʃi	ʒat̃i
5	ʒ	∅ (null)	44	squash	ʒiki	iki
6	ʒ	ʃ	63	Mexico City	nukoʔʒo	nukoʔʃo
7	ⁿ dʒ	ⁿ d	48	thick	ⁿ dʒika	ⁿ dika
8	ⁿ d	l	79	poor	ⁿ daʔaβi	laʔaβi
9	n	∅ (null)	63	Mexico City	nukoʔʒo	ikoʔʒo
10	i	u	37	two	iβi	uβi
11	a	e	14	horse	k ^w ai	k ^w ei

In the case of the first correspondence ([k^w] vs. [β]), the varieties in Zone B do not always exhibit the sound [β] on the same words, but they do both realize the sound some of the time, while the varieties in Zone A do not. In the case of the [ʃ] vs. [s]/[ʒ]/[ks] correspondence (#2), the sounds [s]/[ʒ]/[ks] in Zone B are all realized as [ʃ] in Zone A. For the correspondences that include a null, the sound in Zone A corresponds to the absence of the sound in Zone B.

Figures 9 through 19 are maps that show the distribution for each of the correspondences and the resulting isoglosses.

Figure 9: Isogloss 1 – [k^w] vs. [β]

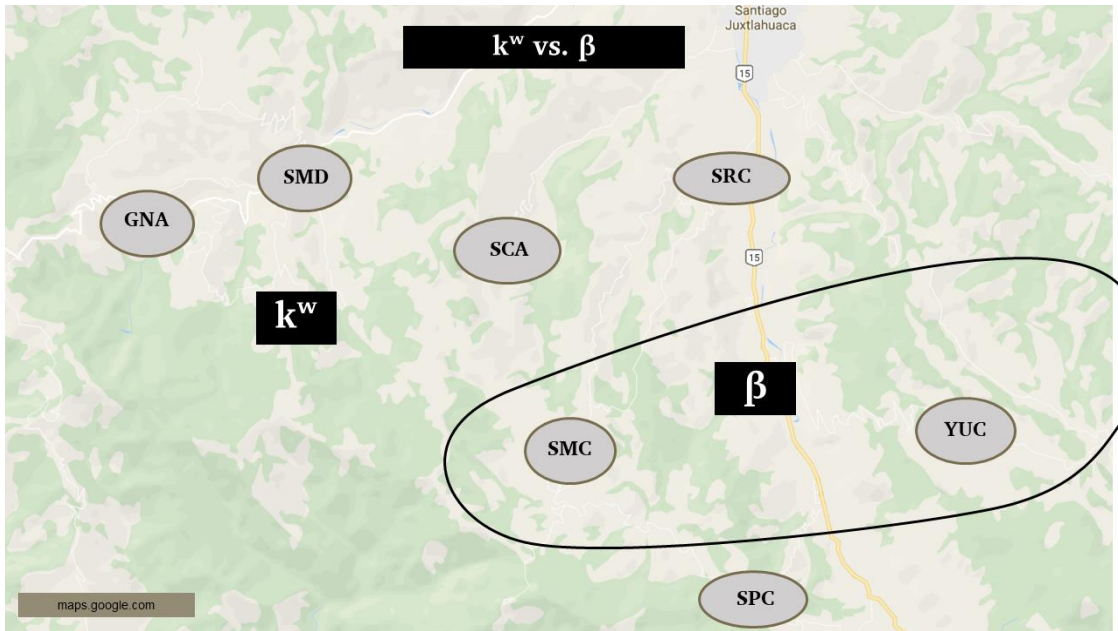


Figure 10: Isogloss 2 – [ʃ] vs. [s]/[z]/[ks]

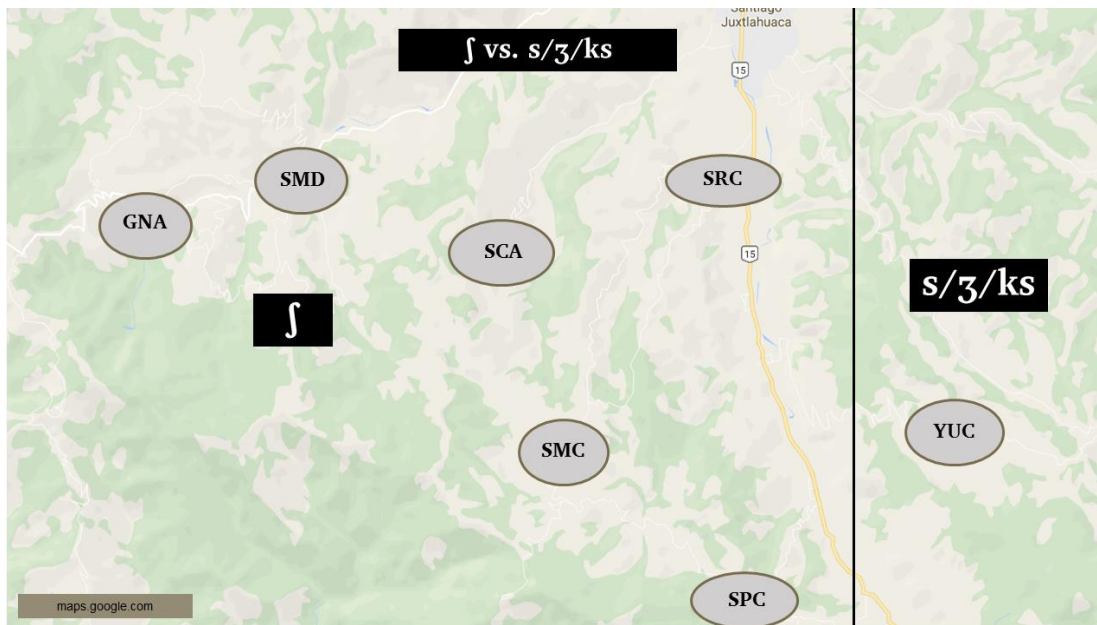


Figure 11: Isogloss 3 – [ʃi] vs. [iʃ]

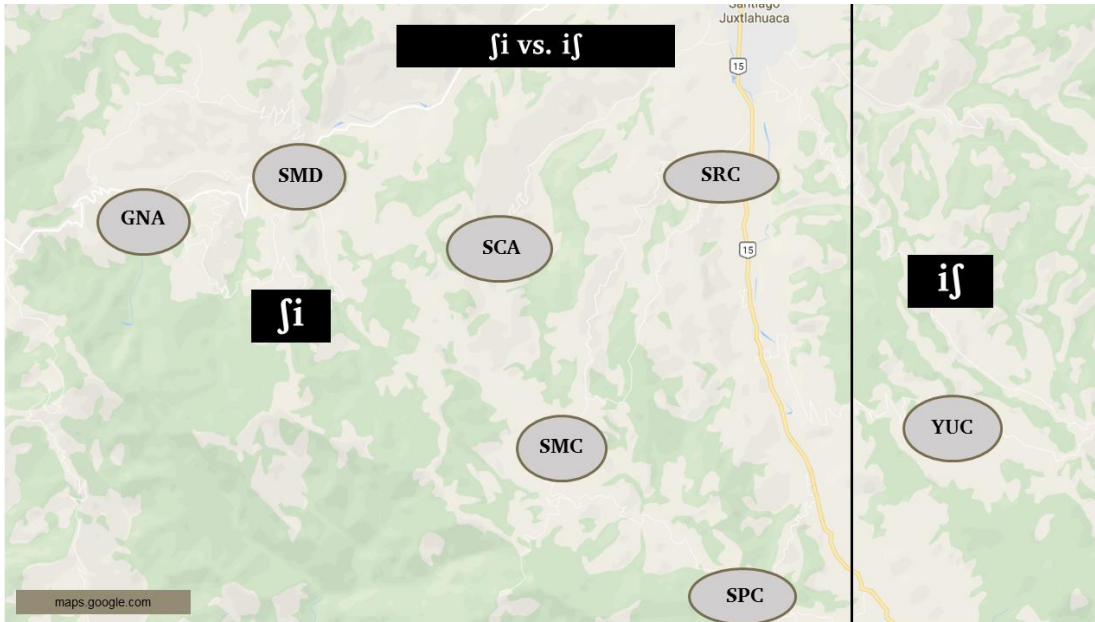


Figure 12: Isogloss 4 – [tʃ̃] vs. [t]

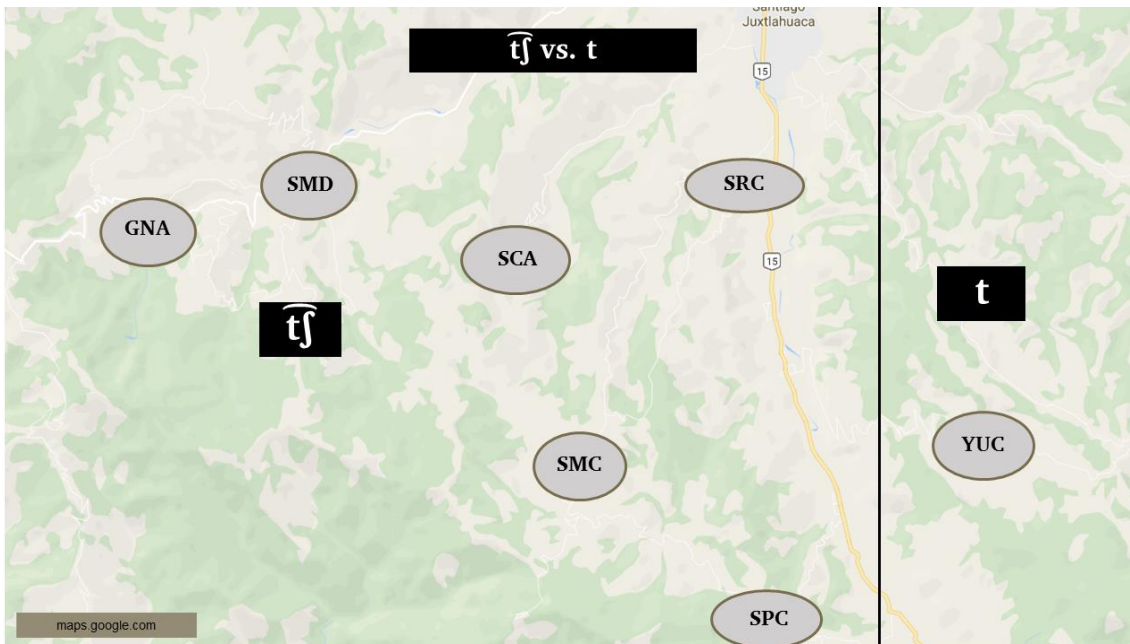


Figure 13: Isogloss 5 – [ʒ] vs. Ø

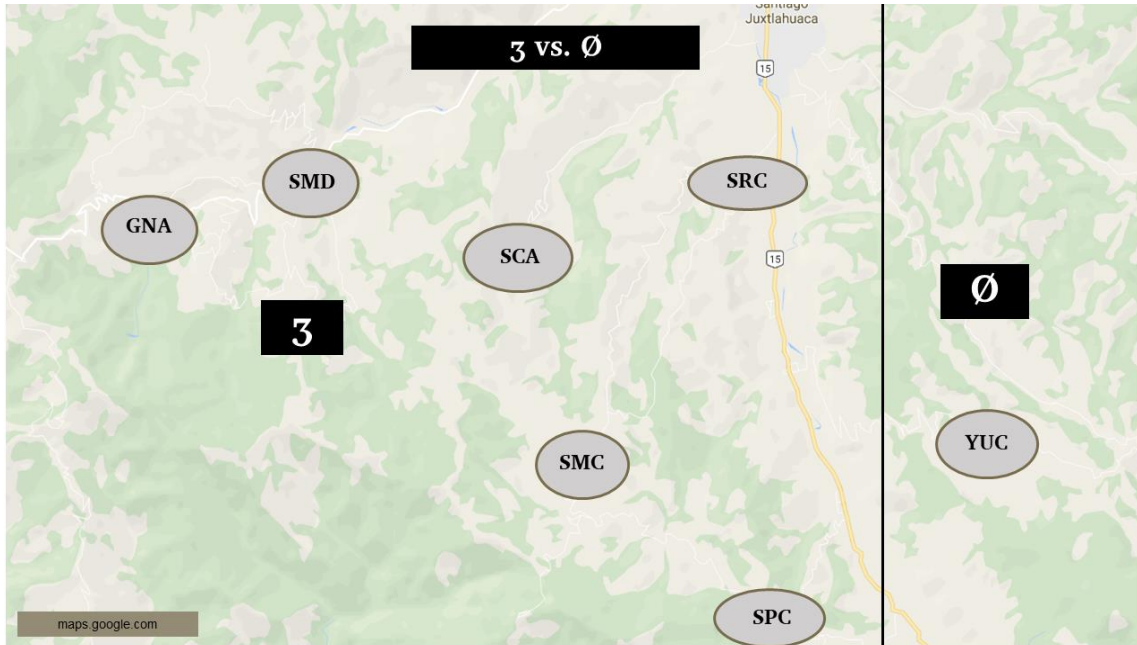


Figure 14: Isogloss 6 – [ʒ] vs. [ʃ]

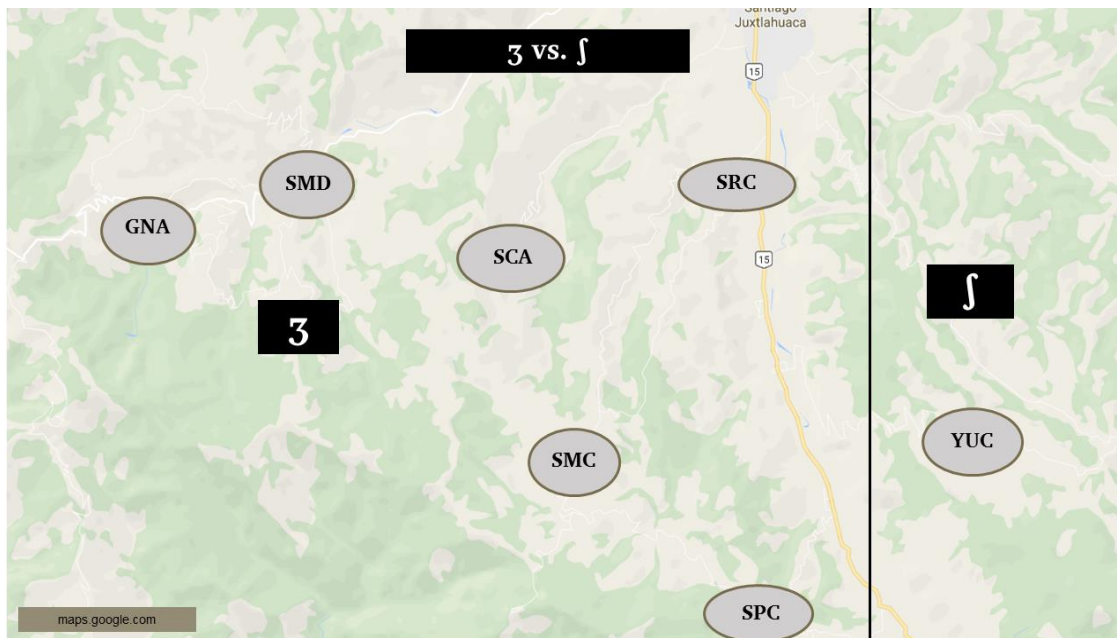


Figure 15: Isogloss 7 – [ˈd̥] vs. [ˈd]

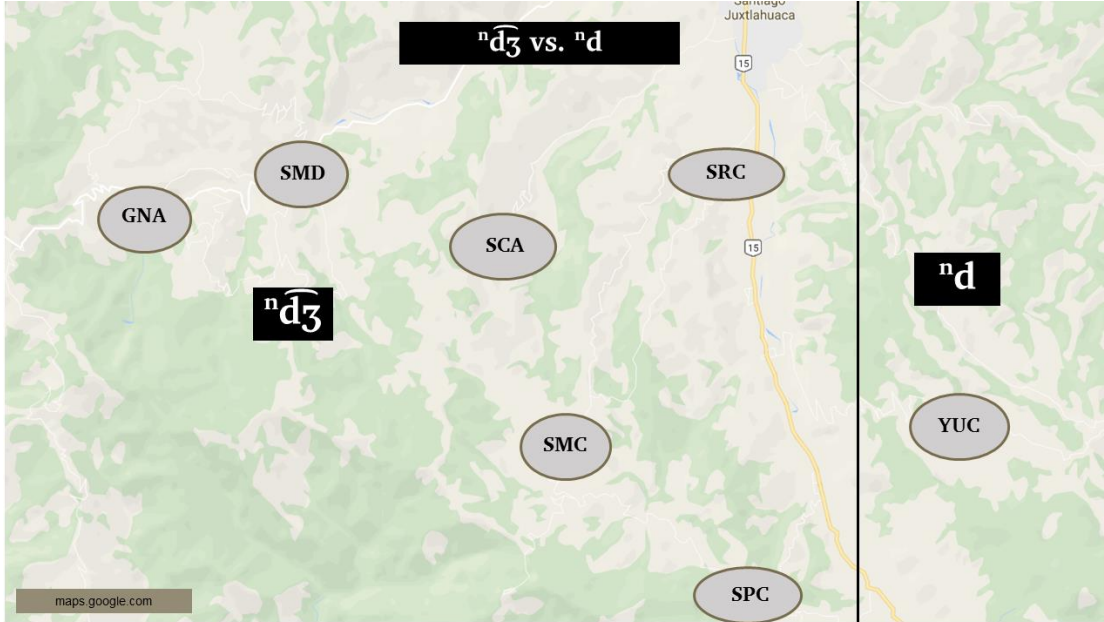


Figure 16: Isogloss 8 – [ˈd] vs. [l]

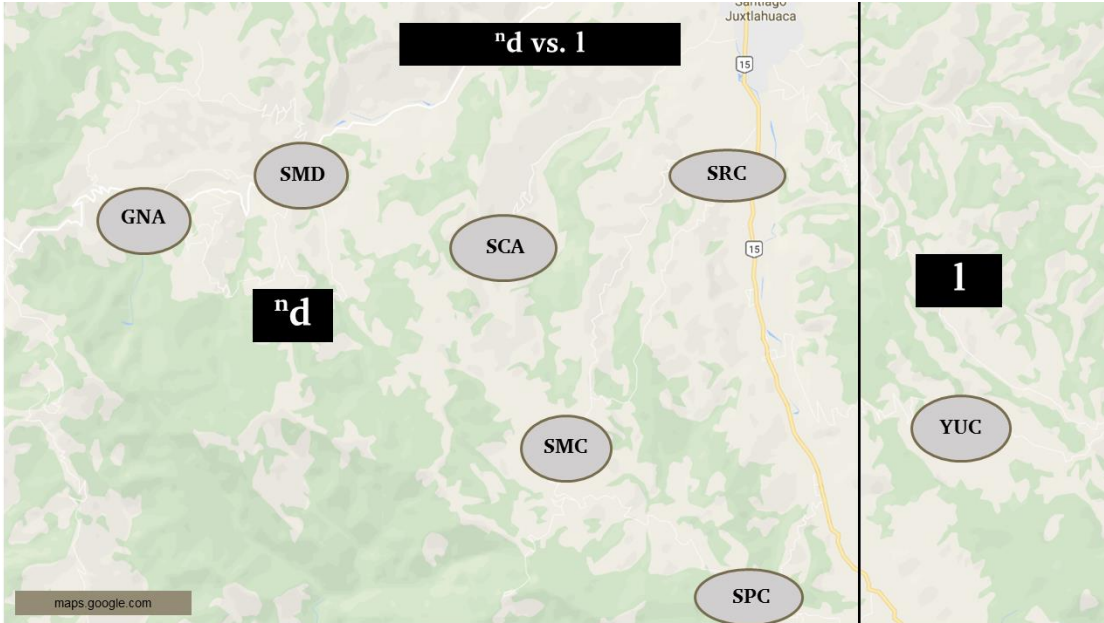


Figure 17: Isogloss 9 –[n] vs. Ø

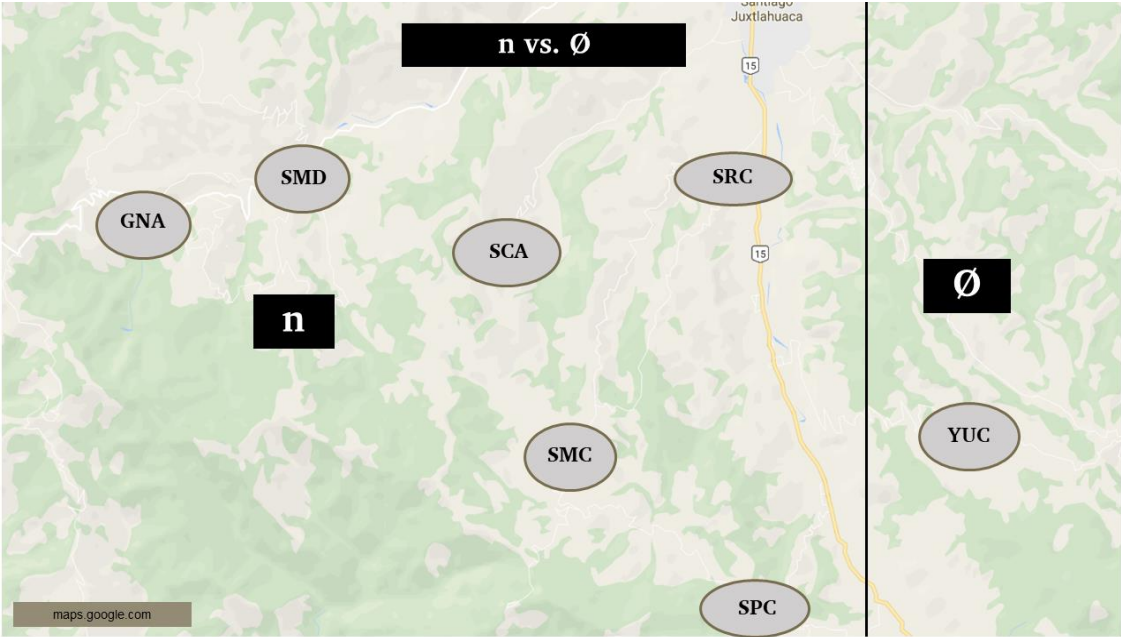


Figure 18: Isogloss 10 – [i] vs. [u]

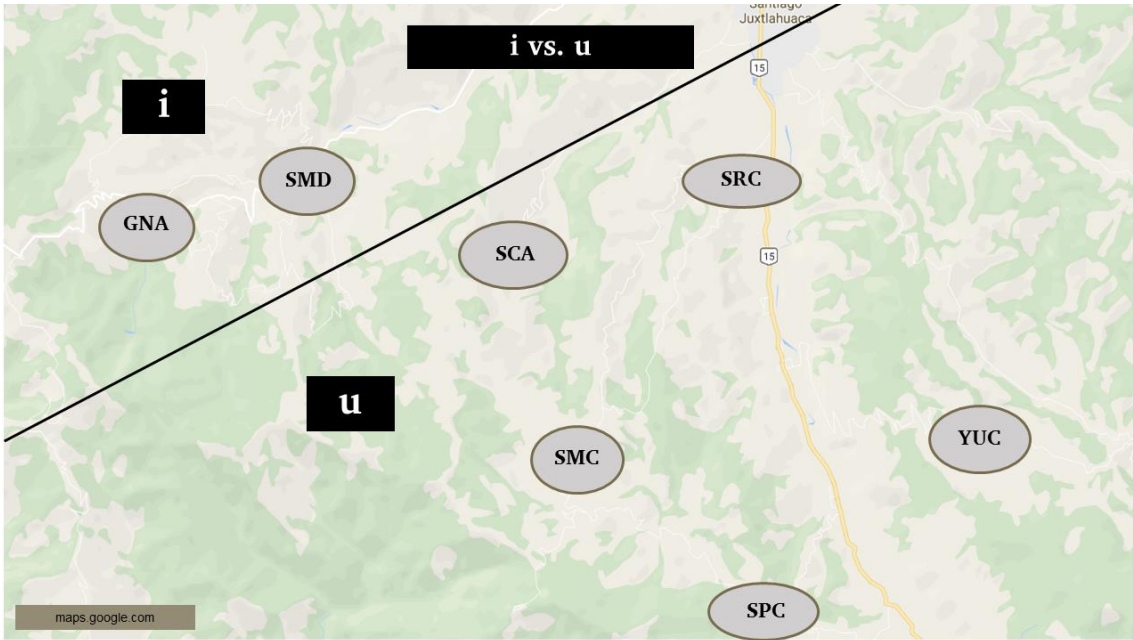


Figure 19: Isogloss 11 – [a] vs. [e]

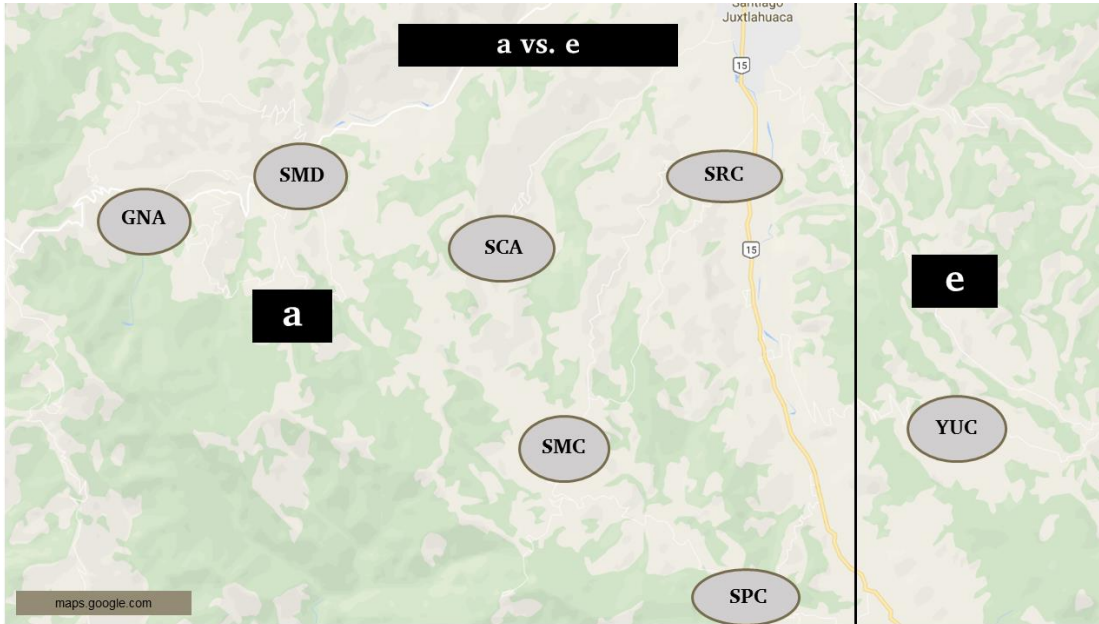
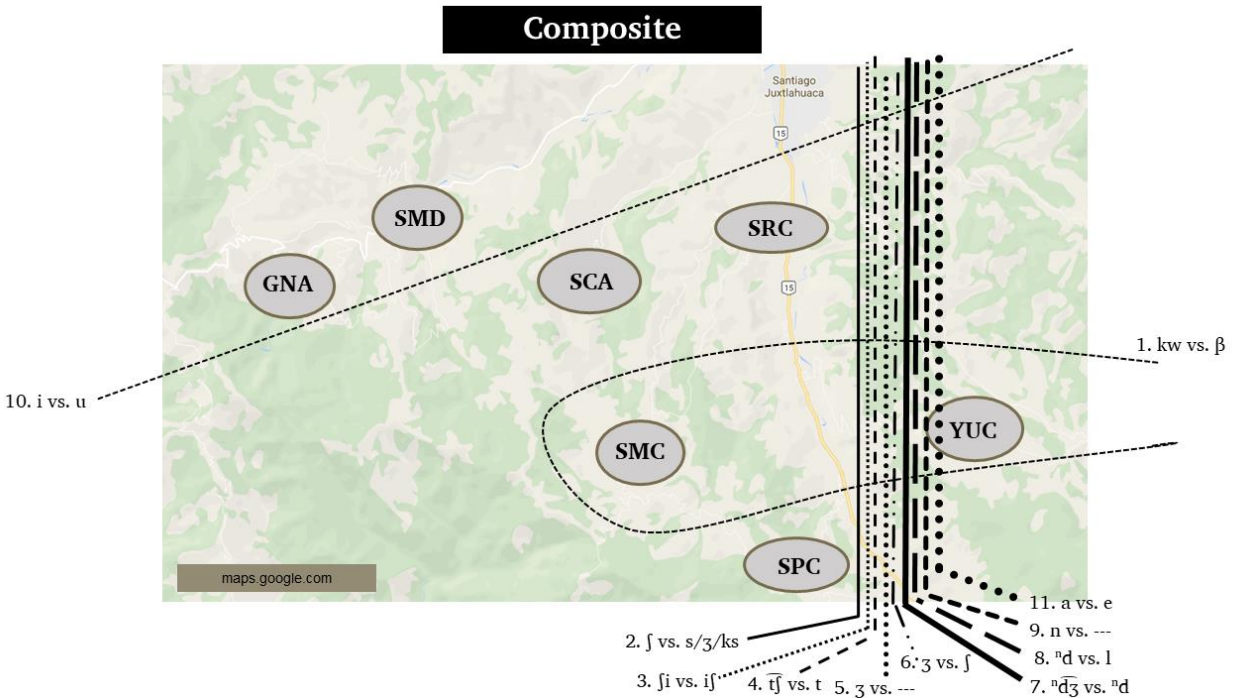


Figure 20 is a composite map of all of the isoglosses. By far the most common pattern is that of YUC in contrast to all of the other varieties.

Figure 20: Composite Map of Isoglosses



The isoglosses converge into three clear patterns, the first of which consists of isogloss 1. For this isogloss, SMC and YUC fall on one side of the isogloss, while SMD, GNA, SCA, SRC, and SPC all group together on the other side of the isogloss. This isogloss indicates slightly higher relatedness between SMC and YUC.

The second and by far the most common distribution consists of YUC falling into a separate group from the rest of the communities surveyed, which occurs for isoglosses 2-9, and 11.

Finally, the third pattern consists of isogloss 10. GNA and SMD group together apart from the rest of the communities, which supports slightly higher (compared to the other communities) relatedness between these two communities.

The revised groups, based on the analysis of similar forms and the analysis of phonological correspondences displayed using isoglosses, are summarized in (4):

- (4) Group 1: SMD, GNA, SCA, SRC, SPC, SMC
 - Subgroup 1A: SMD, GNA
 - Subgroup 1B: SRC, SCA
- Group 2: YUC (SMC?)

The Group 1/Group 2 split initially proposed in the analysis of similar forms (see Section 3.1) is corroborated by the analysis of phonological correspondences displayed using isoglosses. Based on the convergence of nine out of eleven isoglosses, it appears that there is significant relatedness among SMD, GNA, SRC, SCA, SMC, and SPC, but not YUC. The six communities in Group 1 (from the analysis of similar forms) are displayed in the same way by nine out of eleven isoglosses, while YUC is separated from the six varieties of Group 1 for those nine occasions.

Given that SMD and GNA are separated from the rest of Group 1 by an additional isogloss, they appear to consist of their own subgroup within Group 1, indicating slightly higher

relatedness to each other than the other varieties in Group 1. These two varieties shared 90% similar forms in the first analysis (see Table 7), so this is a feasible conclusion.

YUC demonstrates a low degree of relatedness with the other varieties, given that it is separated from the varieties in Group 1 by nine out of eleven isoglosses. However, it appears to have a slightly higher degree of relatedness with SMC, given that they are separated from the other varieties by the first isogloss ([k^w] vs. [β]). However, while the additional isogloss that affects SMD and GNA gave a clear indication of relatedness, we cannot immediately conclude that SMC should be grouped with YUC instead of Group 1. SMC has 70% similar forms to YUC, which is the threshold for considering varieties to be distinct from one another (see Section 3.1), so it seems unlikely that SMC should be designated under Group 2 with YUC. A possible explanation for this anomalous correspondence is that the varieties split at some point in the past, with phonological change for [k^w]/[β] affecting YUC/SMC differently from the other varieties. However, this does not account for how SMC patterns with the other communities and not YUC for nine out of eleven isoglosses. The genetic relationship between YUC and SMC remains cloudy.

Finally, the especially high degree of relatedness found in the analysis of similar forms between SRC and SCA is not supported by these isoglosses, but neither is it disconfirmed. I now present two analyses of tone that help to clarify this issue and general relationships among the varieties in question.

3.3 Tone

Even among Mixtec varieties that have similar segments, tone patterns can vary significantly. As such, the relatedness of varieties will be realized in their tones in addition to

the phonology of their segments. However, while tone is common in many of the world's languages, it is not generally considered in early analyses of relatedness of understudied languages. In my analysis, there are two ways in which I use tone to gauge relatedness of the seven understudied Mixtec varieties. The first compares overall tone patterns in words that are grouped on the basis of reconstructed Proto-Mixtec tone patterns. The second analysis compares the similarity of tones associated with corresponding syllables without regard to historical origin.

It is generally accepted that in present-day Mixtec varieties, there are at least three underlying tone levels, L, M, and H (McKendry 2013:157).¹⁶ For the purposes of this analysis, which largely relies on the comparison of surface pitch levels, I adhere to this general understanding.

3.3.1 Analysis of tone patterns

Table 9 shows the tone patterns found in each variety using the Tone Wordlist (see Table 3 and Appendix C), which is based on Proto-Mixtec tone patterns.¹⁷ The glottal stop in Proto-Mixtec patterns, as in Patterns 2 and 4 (column 1), is generally realized as a floating high or floating low tone in present day Mixtec varieties (McKendry, p.c.). I did not attempt to assess floating tones in this survey. I use IPA tone notation: H (high): á M (mid): ā L (low): à R (rising): ǎ F (falling): â.

¹⁶ However Tranel (2012) and Daly & Hyman (2007) analyze Mixtec tone systems as having two underlying tones.

¹⁷ The Proto-Mixtec tone patterns used do not include floating tones themselves.

Table 9: Tone patterns

	<u>Proto-Mixtec</u>	<u>YUC</u>	<u>SMD</u>	<u>GNA</u>	<u>SRC</u>	<u>SCA</u>	<u>SMC</u>	<u>SPC</u>	<u>COI</u>	<u>SJM</u>
Pattern 1	*L	LL	LL	LL	LL	LL	LL	LL	LL	LL
	‘two’	ùβì	ùβì	ìβì	ùβì	ùβì	ùβì	ùβì	ùβì	ùβì
Pattern 2	*L [?]	LM/MM	LL	LL	LL	LL	LL	LL	LH	LR
	‘four’	kùmī	kòmì	kòmì	kùmì	kòmì	kùmì	kùmì	kòmí	kùmǐ
Pattern 3	*M	MM	MM	MM	MM	MM	MM	MM	MM	MM
	‘house’	βē?ē	βē?ē	βē?ē	βē?ē	βē?ē	βē?ē	βē?ē	βē?ē	βē?ē
Pattern 4	*M [?]	MH	ML	ML	ML	ML	ML	ML	MH	FR
	‘corn’	tījí	ⁿ d̄zīfì	ⁿ d̄zīfì	ⁿ d̄zīfì	ⁿ d̄zīfì	ⁿ dīfì	ⁿ d̄zīfì	ⁿ dīfì	ⁿ dīfì
Pattern 5	*ML	ML	MM	MM	MM	MM	MM	MM	MM	ML
	‘flower’	ītà	ītā	ītā	ītā	ītā	ītā	ītā	ītā	ītà
Pattern 6	*LM	LM/ML	LM	LM	LM	MM	MM	MM	LM/ML/ MM	LM
	‘adobe’	ⁿ dò?ò	ⁿ dò?ò	ⁿ dò?ò	ⁿ dò?ò	ⁿ dō?ò	ⁿ dō?ò	ⁿ dō?ò	ⁿ dò?ò	ⁿ dò?ò

There are less common patterns in the varieties surveyed, such as MH or H, which do not show up in these common patterns but do appear in words elicited in the General Wordlist. This bears further study, but is perhaps due to the high level of tone perturbation and floating tones involved in Mixtec tone systems, or because these patterns are simply less common.

There are three additional tone patterns found in COI and SJM, neither of which was surveyed but are included as points of comparison with the communities that were surveyed. These three patterns are rising, falling, and semi-low.

The loanwords group (“Other Group”) in the original Tone Wordlist from McKendry was recorded but excluded from the intelligibility analysis since the tones and lexical items in that group varied widely, which made comparison unhelpful. In a few cases, noted in Table 9, there is no clear tone pattern for words in the Tone Wordlist. For example, looking at the realization of the Pattern 2 pattern in YUC, while the most common pattern is LM, there are several words in the group with MM tones, such as ‘tortilla’ [iʔtā] and ‘plate’ [kōō]. For Pattern 6, several words in YUC have a LM pattern (see ‘adobe’ above), while others have ML, such as ‘grasshopper’ [tikà] and ‘green bean’ [ˈdītʃi]. COI also has multiple patterns for words in Pattern 6. Many words in the group have the pattern LM, while the word for ‘grasshopper’ [tīàà] is ML and the word for ‘grasshopper’ [tikā] is MM.

Most of the varieties share tone patterns most of the time, with the exception of YUC which only shares the same pattern as the other varieties for Pattern 1 and Pattern 3. This further supports the claim that YUC is less related to the other varieties. SMD, GNA, SCA, SRC, SMC, and SPC all have identical tone patterns, with the exception of Pattern 6, where SMD, GNA, and SRC group together and SCA, SMC, and SPC group together. This supports the slightly greater relatedness between SMD and GNA, as seen in the isogloss and phonological analysis. It brings into question whether or not SCA and SRC share a greater relatedness, as

seen in 3.2.1, as they fall into different groups for Pattern 6. SMC patterns uniformly with SCA and SPC, rather than YUC, which supports the claim that it likely belongs to Group 1, rather than Group 2.

3.3.2 Tone analysis of individual words

3.3.2.1 Rationale for a second tone analysis

While a comparison of tone patterns is useful, it does not examine the full extent of variation in the reflexes of tone patterns, nor does it evaluate all of the tone patterns that occur in the data. Despite the fact that most of these varieties share the same major tone patterns, there are many differences in the tones for individual words on the General Wordlist (see Appendix B). For example, though SMD and GNA share the same tones for all the groups in section 3.3.1, the form for ‘wing’ is [kìfì] (LL) in SMD and [ˈd̪ɛ́ɪfì] (ML) in GNA. Though these varieties are quite similar phonologically (90% similar forms, nine out of 11 isoglosses), they differ in the first tone for this word, which is L in SMD and M in GNA. This is why it is important to analyze not only the distribution of tone patterns, but also the tones of individual words in each variety; similarity of tones is not tied to phonological similarity of segments. In order to have a fuller picture of the relatedness of the varieties in question, it is important to include an analysis of the distribution of tones in individual words, this time using the 98 items from the General Wordlist.

Another reason for the second tone analysis is that there are tone patterns that do not show up in the six main tone patterns in the analysis presented in 3.3.1. For example, none of the patterns in Table 9 include H for SCA, SMC, and SPC, but H does occur in these varieties, as in ‘meat’ [kūpú] in SCA. In addition to the H tone, this word’s MH pattern occurs with other varieties, like YUC, but not for SCA in Table 9. The analysis of tone patterns given in Table 9

therefore does not capture the entire range of tones for these varieties, so an additional analysis is needed.

Table 10 shows two words, ‘mountain’ and ‘meat’, which show the MH pattern not shown in the analysis of most common tone patterns.

Table 10: Additional tone correspondences

Gen. Wordlist #25: 'mountain'			
		<u>Tone pattern</u>	<u>Word form</u>
Base variety:	SMD	M L	ʒĩkì
	GNA	M L	ʒĩkì
	SPC	M L	ʒũkù
	SRC	M L	ʒĩkì
	SCA	M L	ʒĩkì
	SMC	M L	ʒĩkì
	YUC	M H	ʒũkú
Gen. Wordlist #22: 'meat'			
		<u>Tone pattern</u>	<u>Word form</u>
Base variety:	YUC	L L	kũɲ̀
	SPC	L M	kũɲ̄
	SMD	M H	kũɲ́
	SRC	M H	kũɲ́
	SCA	M H	kũɲ́
	GNA	M M	kũɲ̄
	SMC	M M	kũɲ̄

In the case of ‘mountain’, ML in SMD, GNA, SPC, SRC, SCA, and SMC corresponds to MH in YUC. In the case of ‘meat’, MH in SMD, SRC, and SCA corresponds to MM in GNA and LL in YUC. These correspondences were not seen in the analysis of major tone patterns, necessitating an additional analysis of tone.

Further investigation is also needed as to why the same patterns are playing out in different distributions, though this is beyond the scope of this thesis.¹⁸ There is a clear need for a second tone analysis that not only looks at the broad strokes of the main tone patterns, but evaluates the distribution of tones in individual words.

3.3.2.2 Modified Levenshtein Distance: A second tone analysis

I now utilize Levenshtein Distance (LD), a common tool for assessing relatedness, in a new way to examine the distribution of tone in individual words. Traditionally, LD is used to calculate the phonological distance between two words by counting the number of operations (deletion, addition, substitution of segments) needed to relate words in two varieties (Heeringa & Nerbonne 2001:11). For example, ‘mountain’ is [ʒíkì] in GNA and [ʒükù] in SPC. It would require two substitutions to arrive at the SPC form from the GNA form, giving a LD of 0.5.

LD was first introduced by Kessler (1995) in an analysis of Irish Gaelic dialects, and has been applied extensively in European linguistics since its introduction for the purposes of quantifying partial matches between segments across dialects. For example, it was used to corroborate the traditional delineations of Dutch dialects by Nerbonne in 1996. For a phone string comparison, the basic and most common use of LD, every operation has an equal cost of 1, and the sum total of all the operations needed for all of the words in the data set is then

¹⁸ One can theorize about some explanations. This is possibly due to how phonological segments may be changed or deleted in Mixtec languages, while still preserving the tones; tones and segments operate independently of each other, and phonological similarity for segments does not necessarily correlate with similarity of tones. There is also a great deal of tone perturbation and there are many floating tones in Mixtec, which could cause surface pitches to differ, even when varieties seem to adhere to the same underlying patterns.

divided by the total number of segments of all the words in the variety being used for comparison (Kessler 1995). LD has also been modified for less traditional purposes, such as the assessment of sign language wordlist items by Parks (Parks 2011).

For the purposes of this analysis, percentages of dissimilar tones were calculated with a simplified LD by comparing the tones for each TBU for each word of the General Wordlist, totaling the number of different tones for each variety compared, and then dividing by the total number of tone-bearing units (TBUs) in the base variety. Tones were compared for forms considered similar in the analysis of similar forms, excluding words found to be dissimilar (usually because one or both are loanwords), such as the word for ‘fertilizer’, which is [ābónó] in SMD and [jā?ā] in GNA. Words are also excluded when a form is missing for one of the varieties, such as the word for ‘to walk’, which is missing in the data from SPC.

LD does not differentiate between substitutions of similar sounds; a substitution of [b] for [x] is counted the same as a substitution of [b] for [p]. Therefore, in my analysis, tones that are further apart in pitch (i.e. H and L are further apart than H and M) are not counted as “more” different. Rather, all tone differences are counted the same.

Table 11 is an example of how this analysis was conducted, using the same examples as shown in Table 10.

Table 11: Examples of tone distance calculations

Gen. Wordlist #25: 'mountain'				
		<u>Tone pattern</u>	<u>Word form</u>	<u># Different tones</u>
Base variety:	SMD	M L	ʒĩkì	
	GNA	M L	ʒĩkì	0
	SPC	M L	ʒũkù	0
	SRC	M L	ʒĩkì	0
	SCA	M L	ʒĩkì	0
	SMC	M L	ʒĩkì	0
	YUC	M H	ʒũkù	1
Gen. Wordlist #22: 'meat'				
		<u>Tone pattern</u>	<u>Word form</u>	<u># Different tones</u>
Base variety:	YUC	L L	kũɲ̀	
	SPC	L M	kũɲ̃	1
	SMD	M H	kũɲ́	2
	SRC	M H	kũɲ́	2
	SCA	M H	kũɲ́	2
	GNA	M M	kũɲ̄	2
	SMC	M M	kũɲ̄	2

The other words in the General Wordlist were likewise analyzed. Their differences were then added together and divided by the total number of TBUs in the base variety to calculate the percentage of dissimilar tones.

Table 12 shows the percentage of tones that are different across the seven Mixtec varieties surveyed. Note that this is the percentage of *different* tones, rather than identical tones, so a lower score indicates a higher degree of relatedness. COI and SJM are included the analysis of common tone patterns above (see Section 3.3.1) but not included in this analysis

because their tonal inventories have already been shown to be extremely different from the varieties surveyed.

Table 12: Percentage of dissimilar tones

YUC						
53	GNA					
53	28	SMD				
58	23	21	SRC			
57	31	23	20	SCA		
61	31	31	23	29	SPC	
64	32	32	32	32	27	SMC

SRC and SCA are the most similar of any of the varieties; their tones differ only 20% of the time. This helps to confirm their greater relatedness in the face of their differing tone patterns for Group 6, as discussed above. It also confirms the closer relatedness seen in the analysis of similar forms. SRC, in general, has the lowest percentages of dissimilar tones with all the other varieties. This is perhaps due to its central geographic location.

SMD and GNA (28% dissimilar) do not have a much lower percentage of dissimilar tones, compared to SRC, SRC, SPC, and SMC, thereby casting doubt on their hypothesized greater relatedness and subgrouping. Instead, it appears that there is a subgroup of SMD, GNA, SRC, and SCA within the core Group 1. This is corroborated by the tone pattern analysis in 3.2.1, where these four varieties pattern together.

The most dissimilar varieties (with respect to tone) are YUC and SMC, whose tones differ 64% of the time. YUC is dissimilar not only from SMD, but from all the varieties surveyed (53% to 64% dissimilar). These data support grouping YUC separately from the other varieties, as discussed in the analyses of phonological correspondences displayed using isoglosses, phonological similarity, and tone patterns. This higher degree of tonal dissimilarity further

clarifies the question from Section 3.2 about whether SMC should be grouped with YUC.

Although SMC and YUC are displayed in the same way by an isogloss in that analysis, the high degree of tone dissimilarity shown in this analysis supports the claim that SMC should remain grouped separately from YUC.

The other varieties differ between 21% to 35% among each other, in contrast to differing from 52% to 63% of the time from YUC. This is a much lower degree of tonal dissimilarity, confirming grouping the other six varieties together as posited in 3.1 and 3.2, with the addition of SMD to subgroup 1A, summarized in (5):

- (5) Group 1: SMD, GNA, SCA, SRC, SPC, SMC
- Subgroup 1A: SMD, GNA, SRC, SCA
- Group 2: YUC

Further study of tone would better illuminate the nature of the subgroup and whether SMD and GNA should be part of the subgroup with SRC and SCA or form their own. SMD and GNA seem to pattern with SRC and SCA for tone, but not for segments.

3.4 Reported intelligibility

Intelligibility depends on a wide range of factors, including genetic relationships between language varieties, geography, contact, cultural practices, and politics. However, all other things being equal, intelligibility increases as language relatedness increases. I will now compare the intelligibility reported by survey participants toward other Mixtec varieties with the intelligibility suggested by the above findings on relatedness (reported in Sections 3.1-3.3).

To investigate intelligibility, participants were asked about how well they understand other varieties of Mixtec. These questions are presented in Table 13, and are a subset of the questions

presented in Chapter 2, Table 4. Questions referring to Yucuyi are referring to SMD Mixtec, since Yucuyi is a part of the SMD designation.

Table 13: Sociolinguistic Questionnaire questions about intelligibility

INTELLIGIBILITY		
1	¿Usted ha viajado alguna vez a otra comunidad? ¿Cómo se comunicó con las personas allá?	Have you ever traveled to another community? How did you communicate with the people there?
2	¿Alguna vez ha venido alguien de otra comunidad a su pueblo? ¿Qué idioma usó para hablar con personas del pueblo?	Has somebody from another community ever come to your village? What language did they use to talk with people from your village?
3	¿Es más fácil para usted entender el mixteco de Yucunicoco o el de Yucuyi?	Is it easier for you to understand the Mixtec from Yucunicoco or Yucuyi?
4	¿Es más fácil para usted entender el mixteco de San Miguel Cuevas o el de Yucuyi?	Is it easier for you to understand the Mixtec from San Miguel Cuevas or Yucuyi?
5	¿Es más fácil para usted entender el mixteco de San Pedro Chayuco o el de Yucuyi?	Is it easier for you to understand the Mixtec from San Pedro Chayuco or Yucuyi?
6	Aparte de la lengua de su pueblo, ¿cuál es el mixteco que entiende mejor?	Apart from the Mixtec of your village, which Mixtec do you understand best?
7	¿Cuál es el mixteco más difícil de entender?	Which Mixtec is the hardest to understand?
LANGUAGE ATTITUDES		
1	¿Qué son unos de los pueblos más cercanos a su pueblo? En estos pueblos, ¿hablan la misma lengua que usted?	What are some of the villages that are closest to yours? In those villages, do they speak the same language as you?

Participants were first asked about their comprehension of Yucuyi/SMD in the course of conversation. Questions 3-5 in the Intelligibility section of the Questionnaire ask participants to rank intelligibility of several varieties in comparison to SMD. Using a baseline of comparison to Yucuyi/SMD, survey participants' rankings provided a secondary comparison. For example, if a participant reported understanding SMC better than SMD (question 4) but understand SMD better than SPC (question 5), and they report understanding SMD well (question 3), it is likely

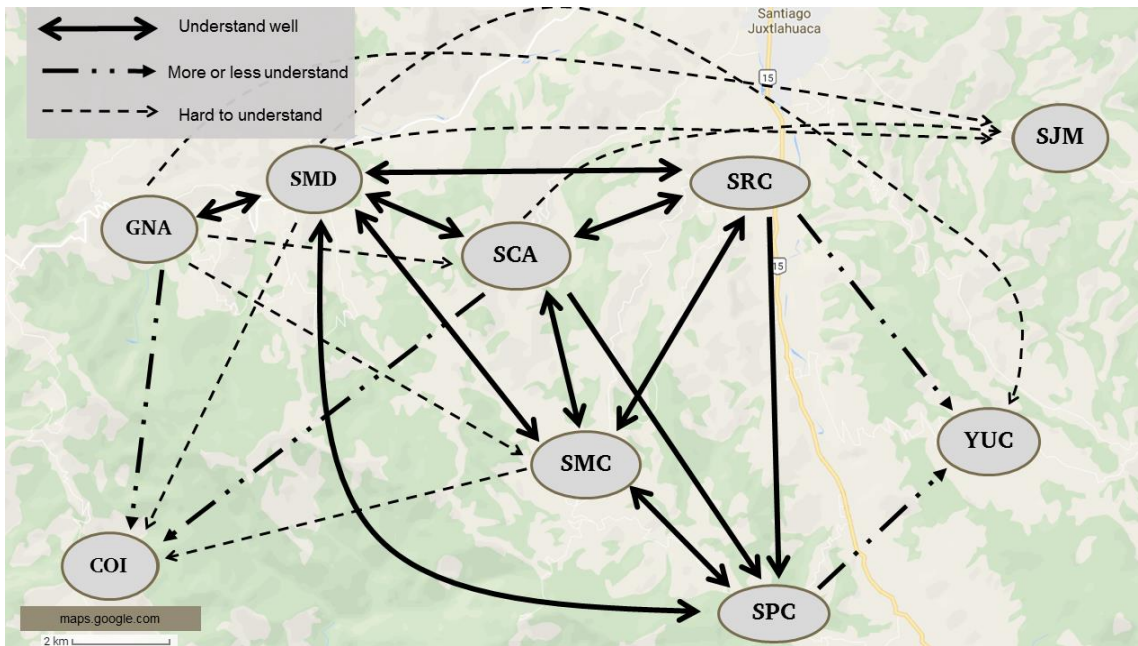
that they also understand SMC well and that their comprehension of SPC is less than that of SMC.

Prolonged discussion usually arose out of the questions above during which I asked participants about how well they understood the Mixtec varieties spoken in the other communities included in the survey. These responses are not quantifiable, given the impressionistic nature of questions and responses, but commonalities did arise in responses that made clear the general levels of intelligibility among communities.

Individuals from most of the villages surveyed reported understanding each other well, at least in daily conversations, with the exception of YUC. Most of the towns surveyed send large numbers of people to the weekly market day (Friday) in Juxtlahuaca, the market town for the region. This may be a factor contributing to the reported understanding between villages and that reinforces the inherent intelligibility of the varieties, with the exception of YUC. The fact that contact at the weekly market is not sufficient to generate comprehension of YUC points to the fact that it is less related to the other six varieties surveyed.

Figure 21 is a composite map of the feedback given in the Sociolinguistic Questionnaire regarding how well individuals from each village understand people of other villages, as well as the Mixtec varieties spoken in SJM and COI. This map was developed by compiling the responses to questions 1-7 in Section 2 and question 1 from Section 3 of the Sociolinguistic Questionnaire, shown above in Table 13, along with guided discussion with participants.

Figure 21: Map of reported intelligibility¹⁹



In many cases, participants from two villages, such as GNA and SMD, reported understanding each other. In such a case, there is an arrow in both directions. In other cases, such as SRC and YUC, participants in only one of the communities reported about the other, in which case the arrow only points in one direction. In some cases, no participant reported experiences with another community, such as SPC and GNA. In these situations, participants were prompted following the initial question to reflect on intelligibility with communities they had not mentioned in their preliminary response, but many participants had never been to, interacted with, or even heard of the community in question. This was especially common among women, as they are less likely to travel outside of their community in this cultural context.

¹⁹ The locations of SJM and COI are not to scale. See Figure 8 for a to-scale map.

Furthermore, as visits to COI and SJM were not conducted, arrows connecting these communities only point in one direction.

The significant degree of interaction that occurs for speakers of most of the varieties surveyed means that, without in-depth statistical analysis, it is difficult to separate inherent intelligibility (intelligibility resulting from genetic similarities between languages) and acquired intelligibility (intelligibility gained through repeated interactions with another language). Speakers from GNA reported less understanding of other Mixtec varieties, but this may be because they are located further away and fewer GNA residents attend the weekly market day. Speakers from GNA did report high intelligibility with SMD, however. GNA and SMD are high in similarity for phonological forms, phonological correspondences displayed by isoglosses, and tone patterns, suggesting that intelligibility between SMD and SMC (and maybe SCA) is at least partially inherent.

In general, the high intelligibility reported by participants in the Questionnaire among all the surveyed villages, but not with YUC, supports the findings from the linguistic analyses of relatedness.

Additionally, most survey participants stated that they find it difficult or even impossible to understand the Mixtec varieties spoken in COI and SJM. COI and SJM differ significantly with respect to tone and phonology from the varieties surveyed,²⁰ and COI and SJM speakers also attend market days in Juxtlahuaca in fewer numbers, so the low degree of intelligibility reported is not surprising.

²⁰ Information from personal communication with Bruce and Candice Beatham and also Millie Nieves and Gisela Beckmann, SIL colleagues working in these areas.

Reported intelligibility provides a useful measure for confirming the two groups of languages based on the analyses of relatedness. However, the subgroup (SRC and SCA) cannot be confirmed with the intelligibility analysis, since high degrees of intelligibility were reported for Subgroup 1A, just as for with other varieties in Group 1. A revised set of relatedness groups, based on reported intelligibility findings from the Questionnaire, is shown in (6):

(6) Group 1: SMD, GNA, SRC, SCA, SPC, SMC

Subgroup 1A: SMD, GNA

Subgroup 1B: SRC, SCA

Group 2: YUC

The intelligibility findings confirm the two major groupings demonstrated in previous analyses. The measures of relatedness (analysis of similar forms, displaying phonological correspondences using isoglosses, and two analyses of tone) provide a comprehensive picture of the overall relatedness of the language varieties in question, and the intelligibility proposed by these patterns of relatedness correlates positively with the intelligibility reported by survey participants. The small differences in the relatedness findings and reported intelligibility findings can be accounted for by low-level linguistic differences or by differences in language contact.

3.5 Summary of findings

The five analyses of similar forms, displaying phonological correspondences using isoglosses, both analyses of tone, and reported intelligibility all point to a high degree of relatedness among SMD, GNA, SRC, SCA, SMC, and SPC (Group 1), with a low degree of relatedness between those six communities and YUC (Group 2). Within Group 1, there are two subgroupings that have a higher degree of relatedness: SMD and GNA (Subgroup 1A) and SRC

and SCA (Subgroup 1B). The patterns of relatedness reflected in the first four analyses correlate closely with the patterns of reported intelligibility from the fifth analysis, suggesting that the analyses of relatedness provide a reasonable understanding of the relations among the seven varieties surveyed. It would also suggest that much of intelligibility for these Mixtec varieties is predicted by, if not dependent on, their inherent, genetic relationships, though issues of interactions and geography also play a role.

The final summary of these relationships, based on the five analyses above, is shown in (7):

(7) Group 1: SMD, GNA, SRC, SCA, SPC, SMC

Subgroup 1A: SMD, GNA

Subgroup 1B: SRC, SCA

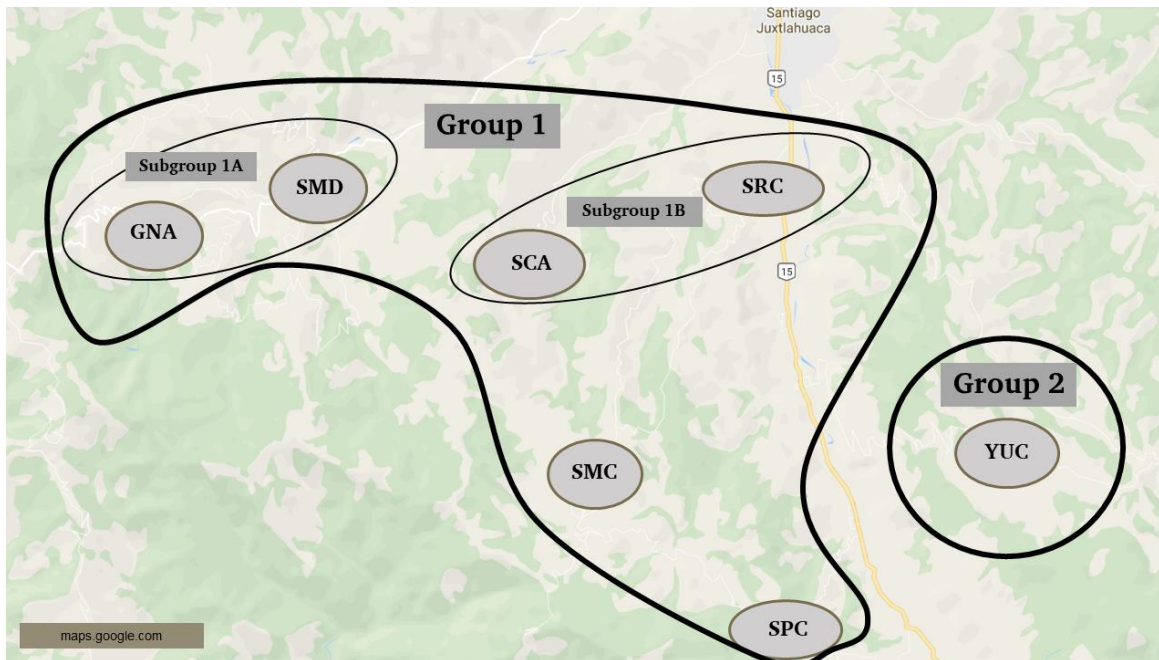
Group 2: YUC

Each of the five analyses supports grouping YUC separately from the other six varieties surveyed. YUC has only 67% to 71% similar forms, is characterized separately from Group 1 varieties for nine out of eleven isoglosses, shares only two common patterns with the varieties in Group 1, and was reported to have low intelligibility with the Group 1 varieties.

There is strong evidence for both of the subgroupings. For Subgroup 1A, SMD and GNA have 90% similar forms, are displayed in the same way by ten out of eleven isoglosses, share six out of six common tone patterns, have only 28% dissimilar tones, and report strong intelligibility. For Subgroup 1B, SRC and SCA have 96% similar forms, are displayed in the same way by nine out of eleven isoglosses, share five out of six common tone patterns, have only 20% dissimilar tones, and report strong intelligibility.

Figure 22 summarizes the relatedness groupings, which match up closely with reported intelligibility. This map shows that SMD and GNA form Subgroup 1A, while SRC and SCA form Subgroup 1B. It also groups YUC separately from the other varieties, as it is much less related to the other six varieties and also has much lower rates of intelligibility with those varieties.

Figure 22: Map of relatedness groups



The previously established understandings of this understudied region do not adequately reflect the linguistic relationships of the speech communities in question. The Ethnologue lists YUC together with the varieties in Group 1 under a single designation, Juchtlahuaca Mixtec [vmc] (Simons & Fennig 2017). My analysis shows the necessity of listing YUC separately from the other six varieties surveyed under a new designation. INALI lists four of the six varieties in Group 1 (SMC, SPC, SCA, SRC) with YUC under the *mixteco del oeste alto* (High Western Mixtec) designation (Instituto Nacional de Lenguas Indígenas:143), while SMD and GNA, also in Group 1, are listed separately under the *mixteco del oeste* (Western Mixtec) designation (Instituto

Nacional de Lenguas Indígenas:140). On the basis of my findings, YUC should be left under its current designation, the *mixteco del oeste alto* (High Western Mixtec),²¹ while all of the six varieties that I classify as Group 1 should be under the same designation, *mixteco del oeste* (Western Mixtec). In both cases, my recommendations, based on my RA language survey, better reflect the linguistic situation of the region than the current designations.

²¹ YUC is located at higher elevation than the other communities (see Table 1), hence the “High” Western designation.

CHAPTER 4

REFLECTION ON METHODOLOGIES AND DIRECTIONS FOR FURTHER STUDY

Now that I have presented my findings, I discuss some of the objections that might be raised in response to the analysis of isoglosses, the analyses of tone, and the analysis of reported intelligibility findings. I also give a more thorough discussion of the basis and history of Rapid-Appraisal surveys, which strengthens the validity of the data collected for this analysis. I conclude with a discussion of areas for further study.

4.1 Use of isoglosses to determine relatedness

Although isoglosses are an established part of dialectology and cross-linguistic comparison, I also found them very useful for organizing and focusing my study of relatedness. However, it should be clear that this tool is useful for identifying relatedness *only* in contexts where significant prior linguistic analysis of related language varieties has been done. My analysis of isoglosses was greatly influenced and facilitated by Josserand's (1983) work on Proto-Mixtec (which includes many maps and studies of isoglosses across many Mixtec languages) and McKendry's thesis (2001), which identified many of the words that exhibit correspondences that can be mapped with isoglosses. These works made it possible to identify likely areas of systematic phonological differentiation before beginning data collection.

It would be most profitable to undertake a similar analysis of isoglosses in languages that a) vary phonologically in systematic ways, such as Mixtec varieties, and b) have undergone prior linguistic study that has identified ways in which varieties tend to vary. If these conditions are met, an analysis of isoglosses can provide valuable insights into language relatedness.

Furthermore, isoglosses in and of themselves do not provide a sufficient basis for determining relatedness of languages, but they can provide organizing structure and insights when used in conjunction with other methods, as I have done here. For language families and groups that vary phonologically in very systematic ways, such as Mixtec, isoglosses are an organized way to understand just how and where they vary.

Isoglosses are especially useful for an RA analysis because they are efficient and easy to use. With the right wordlist, it is fairly simple to identify the phonological correspondences that are distributed according to isoglosses, and then map the isoglosses in question. Another advantage to using isoglosses for studying relatedness is that they allow for the inclusion of historical linguistics in conducting cross-linguistic comparisons. I have also shown that it is possible to group previously unstudied language varieties on the basis of isoglosses. Isoglosses can also demonstrate why language variety A understands language variety B if they are consistently in the same isogloss zone, but neither understands language variety C if language variety C does not consistently appear in the same isogloss zone. This strategy is especially useful for determining language development needs and which communities to include in a given language development project.

4.2 Use of tone to determine relatedness

Given that the majority of the world's languages are tonal (Yip 2002:1), there is a surprising dearth of surveys that include tone, especially with regard to its effects on the ability of two language communities to understand each other. The issue of tone is particularly relevant in the Mixtec context, as one of the primary ways that Mixtec varieties vary is in their tones. Additionally, as I saw in my own language learning, tone carries a high functional load in most Mixtec varieties, as it is used for marking aspect, mood, person, and many other grammatical functions. This is why tone is a central issue for measuring intelligibility among Mixtec varieties.

I found the inclusion of tone in my survey to be a rich complement to the other tools, but it was only possible because of tone analysis research available in related Mixtec varieties. This established a basic understanding of tone in Mixtec varieties. McKendry developed the Tone Wordlist for her Mixtec research and graciously shared it with me for my own survey work; she did the work of identifying the groups of words that share same tone patterns across Mixtec varieties.

Including tone in language survey will be most possible in situations similar to my experience, where there is already a fundamental understanding of the phonology and tone of the languages involved, as this groundwork provides the necessary basis for understanding collected data.²² The Tone Wordlist, or a similar tool, would be most useful for comparing languages that are closely related and have identifiable tone patterns, either because of a fairly

²² Tone survey has, of course, been included in language surveys before, but not for unstudied languages as in this language survey.

straightforward tone system (i.e. without many floating tones or perturbation), or because of prior study that gives at least a basic analysis of tone. The Tone Wordlist is also most useful when used in combination with other analyses of relatedness. In my case, I primarily used this tool for corroborating other findings on relatedness. For example, YUC and SMD only share two out of six tone patterns; this corroborated the analysis of similar forms and affirmed the low level of intelligibility as found in the results from the Sociolinguistic Questionnaire.

Moreover, surveying for tone should only be undertaken with the understanding that it is only possible to gather a superficial, impressionistic sketch of the tones in a given language at such an early stage of research. It is not possible to understand floating tones or tone perturbation, for example, with just a short wordlist and a visit to a community. The very nature of tone means that it is difficult to gain worthwhile information about it in a short survey, much less in an RA. Linguists often spend decades studying the tone system of a given language and still do not feel confident that they understand it well.

Furthermore, the analyses of tone proposed in Chapter 3 are for the purposes of comparison and understanding relatedness between language varieties, rather than outright phonological analysis of the individual tone systems in isolation. These analyses are measures of *relative* relatedness, rather than objective measures of each variety's tone system. Comparing tones in one variety relative to those of another variety provides a helpful perspective that, when taken with other analyses, gives a more thorough understanding of language relatedness.

A potential objection to this tool is that my calculations of tone dissimilarity were based solely on my hearing and transcriptions of tone for General Wordlist items, albeit with many rounds of checking. However, Mixtec tones are incredibly close together in their pitches. (The difference between a H and L is often a half-step on an octave scale or less.) It is difficult for any outsider to hear them correctly, and as an outsider, I likely made some mistakes in

transcribing pitches, despite listening to the recordings over and over and over again. To mitigate these errors, I relied on my prior months of language learning to identify the tones and used Praat (Boersma & Weenink 2012) to corroborate my transcriptions of tone. The tones transcribed are furthermore reflections of surface pitches, rather than the underlying tones that may be determined only after extensive tone analysis.

Related to this issue of underlying tones, another concern is that these two tools do nothing to capture floating tones and tone perturbation, which are so prevalent in Mixtec languages. A thorough study of this would require an in-depth and much longer survey.

4.3 Reservations about reported intelligibility findings

Regarding the qualitative method of using questionnaires to gauge intelligibility, I attest that they can still provide valuable information when used in conjunction with other, more quantitative measures, such as the tone and phonological analyses.

Questions of intelligibility are complicated by the frequency and depth of interactions of individual speakers of different varieties of Mixtec, which can cloud the distinction between inherent and acquired intelligibility. The more often speakers of different Mixtec varieties interact, the more likely they are to understand each other.

This issue is further confused in that questions about intelligibility are especially likely to provoke a desire for participants to “save face”. It is desirable from an economic and intellectual standpoint to understand other Mixtec varieties, since it means you can conduct commerce with speakers of them more easily and might appear smarter. As such, it is likely that many participants over-reported their ability to understand other Mixtec varieties, though

it would be impossible to measure the extent that this affected answers to the Sociolinguistic Questionnaire.

Finally, this data is incomplete; participants often were not familiar with the other communities or, if they did know of them, had no interactions with speakers from those communities that they could report on. A more in-depth study would ideally include data from more participants and require a certain number of responses reflecting intelligibility in both directions for each pair of communities, and would include considerations of gender and age to balance the data further. It would also be helpful in the future to research the historical trading relationships among these communities, as the language contact that results from trading interactions would have also affected intelligibility.

Despite these concerns, the Sociolinguistic Questionnaire provides a useful way to corroborate my findings on relatedness through the findings on reported intelligibility. The findings from this tool coincide closely with those of the other four tools employed, supporting their validity. For future use, the value of this tool could be enhanced by adding matched guise questions and including more participants to help compensate for when another participant does not know or recognize a community that they are being asked about.

4.4 Rapid Appraisal surveys

The style of the survey was in line with “Rapid Appraisal” (RA) language assessments, introduced and developed by SIL in Cameroon in order to gain a broad overview of a linguistic situation in very little time (Stalder & Starr 1990, Bergman 1991). While I have not found any specific objections to RA surveys, someone might argue against RA surveys on the grounds that they are not based on representative sampling and only use short visits to communities. RAs

also do not include RTTs or SRTs, more objective and quantitative tools. It is important to remember that the goal of RA is not to produce statistically valid findings, but rather to provide a brief, efficient snapshot of the linguistic situation of a given area. In my case, I used multiple measures of relatedness, which helps to address this concern about the reliability of results. Furthermore, the findings from each of the five tools that I employed were consistent with each other, strengthening the validity of my findings. Though RA may need to be followed with further research, it can determine a wide range of relevant facts when administered by a researcher who is familiar with the local situation. As a result, RA is useful for determining the language development needs of a community and in preparing for future, more in-depth survey work (Stalder 1993). RA surveys have successfully provided the basis for deciding whether or not a community needs language development (Stalder 1993).

Background research is crucial to RA, as in normal survey methodologies, including developing contacts and relationships in the language community so as to have a better idea of who to work with and how to approach the project (Stalder 1993). This was a major motivation for the six months of language and culture learning that I spent prior to beginning the survey. The many years of linguistic work in other Mixtec varieties by other researchers were invaluable for establishing a foundation of background information about the language and culture, which informed my work.

RA can be a very efficient way to ascertain language community needs while spending less time or money on the project. In a context where there are dozens or even hundreds of varieties and languages to be surveyed, as in the case of Cameroon, this means there is a greater likelihood of actually being able to visit all of the communities that need to be surveyed. It means these communities can make decisions about their language sooner because

they have information sooner, which means that they can begin a language development project more quickly.

RA surveys are not only more efficient than more in-depth surveys, but also allow for accommodation of cultural constraints. In the Mixtec context in which I was working, town leaders were more willing to allow an outsider into their town when it meant a short visit of just a few hours of talking with just a handful of residents, rather than a lengthy visit that involved many community members. RA was a way to gain enough information to make necessary organizational decisions, while still respecting the cultural norms of local communities.

4.5 Directions for further study

In a more in-depth study, there are several specific ways to improve the value and breadth of findings on relatedness. It would be of great value to measure how closely reported intelligibility correlates with more quantitative measures of intelligibility, such as SRT and RTT. Furthermore, as mentioned in Section 3.3, languages that seemingly coincided for all or nearly all of their basic tone patterns still exhibited a fair amount of tone disparities for General Wordlist items. It bears further investigation into why this is the case. Tone represents a yet untapped area of language survey and the study of language relatedness, and the linguistics community must do more to include tone in its language assessment undertakings.

A general improvement that should be made for future surveys in this geographic area in western Oaxaca is a more rigorous and representative sampling method than was used in this survey, which would enhance the validity and accuracy of the survey results. This is particularly true for the eight communities under the YUC designation. In this RA survey, I

only visited and collected data from residents of the largest village, Santa María Yucunicoco, which leaves uncertainty about the linguistic uniformity of the Mixtec spoken in the other seven communities compared to that of Santa María Yucunicoco. In the seven communities that I visited, a more rigorous, systematic method of sampling would have greatly improved the usefulness and reliability of my research.

The results of this survey provide a strong basis for future in-depth survey efforts in the region. The new strategies proposed for including tone in language survey, even RA-style surveys, open the door for a better understanding of the relatedness of tonal languages, both within the Mixtec language family and beyond. Mapping and analyzing isoglosses has been shown to be a useful tool for studying relatedness, further expanding the repertoire of tools at the disposal of language surveyors. Until such further research can be conducted, there is at least a much clearer understanding of the relationships among varieties spoken in the seven Mixtec communities surveyed, and there are new tools available for the study of other understudied linguistic situations, particularly those involving tonal languages.

REFERENCES

- Bergman, Ted G. 1991. Rapid Appraisal of languages. *Notes on Scripture in Use and Language Programs* 28:3–11.
- Blair, Frank. 1990. *Survey on a shoestring: A manual for small-scale language surveys*. Dallas: SIL International.
- Boersma, Paul & David Weenink. 2012. *Praat: Doing phonetics by computer*. Amsterdam: Phonetic Sciences, University of Amsterdam.
- Bradley, Henry. 1970. A linguistic sketch of Jicaltepec Mixtec. Norman, OK: Summer Institute of Linguistics. <https://www.sil.org/resources/archives/8627> (March 15, 2017.)
- Brye, Edward. 2004. *Rapid Appraisal of Beezen*. ms. Cameroon: SIL International.
- Cámara de Diputados del Congreso de la Unión Mexicana. 2003. *Ley general de derechos lingüísticos de los pueblos indígenas*. http://www.wipo.int/wipolex/es/text.jsp?file_id=220918 (March 16, 2017.)
- Campbell, Lyle & William J. Poser. 2008. *Language classification: History and method*. Cambridge: Cambridge University Press.
- Casad, Eugene (ed.). 1980. *Dialect intelligibility testing*. (Summer Institute of Linguistics Publications in Linguistics and Related Fields 38). Norman, OK: Summer Institute of Linguistics of the University of Oklahoma at Norman.
- Chambers, J.K. & Peter Trudgill. 1980. *Dialectology*. (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.
- Chan, Kenneth. 2007. *A bilingualism and language vitality survey of the Gan Yi in the Yi Mei Duo village of the Shilin county of the Yunnan province of China*. Survey Report. (SIL Electronic Survey Reports 7). Dallas: SIL International. <http://www-01.sil.org/silesr/2008/silesr2008-007.pdf> (September 20, 2016.)
- Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. Genus Mixtecan. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/languoid/genus/mixtecan#8/16.605/262.614> (May 12, 2017.)
- Dürr, Michael. 1987. A preliminary reconstruction of the Proto-Mixtec tonal system. *INDIANA* 11:19–61.
- Egland, Steven T. (ed.). 1978. *La inteligibilidad interdialectal en México: Resultados de algunos sondeos*. Mexico: Instituto Lingüístico de Verano.
- Fowler, Floyd J. 2002. *Survey research methods*. 3rd ed. (Applied Social Research Methods Series 1). Thousand Oaks, CA: Sage Publications.
- Hamel, Rainer. 2016. Bilingual education for indigeneous peoples in Mexico. In Nancy Hornberger (ed.), *Encyclopedia of Language and Education*. New York: Springer International Publishing.
- Heeringa, Wilbert & John Nerbonne. 2001. Dialect areas and dialect continua. *Language Variation and Change* (13). 375–400.
- Henerson, Maryelene E., Lynn L. Morris & Carol T. Fitz-Gibbon. 1978. *How to measure attitudes*. Newbury Park, CA: Sage.

- Hollenbach, Barbara E. 2003. The historical source of an irregular Mixtec tone-sandhi pattern. In Mary Ruth Wise, Thomas N. Headland, & Ruth M. Brend (eds.), *Language and life: essays in memory of Kenneth L. Pike*, 533-550. Dallas: SIL International and University of Texas at Arlington.
- Instituto Nacional de Estadística y Geografía. 2010. *Censo de población y vivienda*. Mexico: Instituto Nacional de Estadística y Geografía. <http://www.beta.inegi.org.mx/proyectos/ccpv/2010/> (June 5, 2017.)
- Instituto Nacional de Lenguas Indígenas. *Proyecto de indicadores sociolingüísticos de las lenguas indígenas nacionales*. http://site.inali.gob.mx/Micrositios/estadistica_basica/estadisticas2015/estadisticas2015.html (June 20, 2017.)
- Jaberg, Karl, Jakob Jud & Paul Scheuermeier. 1928. *Sprach- und Sachatlas des Italiens und der Sudschweiz*. Ringier: Zofinger.
- Josserand, Judy Kathryn. 1983. Mixtec dialect history: Proto Mixtec and modern Mixtec text. New Orleans: Tulane University PhD dissertation.
- Kessler, Brett. 1995. Computational dialectology in Irish Gaelic. *Proceedings of the seventh conference on European chapter of the Association for Computational Linguistics*, 60–66. Dublin, Ireland: Morgan Kaufman Publishers.
- Karan, Mark E. & Jürg Stalder. 2000. Assessing motivations: Techniques for researching the motivations behind language choice. In M. Paul Lewis & Gloria Kindell (eds.), *Assessing ethnolinguistic vitality: Theory and practice*. 189–205. Dallas: SIL International.
- Kluge, Angela. 2007. RTT retelling method: An alternative approach to intelligibility testing. <http://www-01.sil.org/silewp/abstract.asp?ref=2007-006> (June 21, 2017.)
- Lambert, Wallace E. 1967. A social psychology of bilingualism. *Journal of Social Issues* 23:91–109.
- Lees, Robert. 1953. The basis of glottochronology. *Language* 29:113–127.
- Longacre, Robert. 1957. Proto-Mixtecan. *International Journal of American Linguistics* 23:1-195.
- Macaulay, Monica, & Joseph C. Salmons. 1995. The Phonology of Glottalization in Mixtec. *International Journal of American Linguistics* 61:38–61.
- McKendry, Inga. 2001. Two studies of Mixtec languages. Grand Forks, ND: University of North Dakota MA Thesis.
- McKendry, Inga. 2011. Some features of hortatory texts in Southeastern Nochixtlán Mixtec. ms. <https://www.sil.org/resources/publications/entry/35846> (June 20, 2017.)
- McKendry, Inga. 2013. Tonal association, prominence and prosodic structure in South-eastern Noxchixtlán Mixtec. Edinburgh: University of Edinburgh PhD Dissertation.
- Meillet, Antoine. 1925. *La méthode comparative en linguistique historique*. English Translation (1967). Paris: Champion.
- Orton, Harold, Steward Sanderson & John Widdowson (eds.). 1978. *The linguistic atlas of England*. London: Croom Helm.
- Parks, Jason. 2011. Sign language word list comparisons: Toward a replicable coding and scoring methodology. Grand Forks, ND: University of North Dakota MA Thesis.
- Radloff, Carla. 1991. *Sentence repetition testing for studies of community bilingualism*. Dallas: SIL International Publications in Linguistics.
- Shibatani, Masayoshi. 1990. *The languages of Japan*. Cambridge: Cambridge University Press.

- Simons, Gary. 1979. *Language variation and limits to communication*. Ithaca, NY: Department of Modern Languages and Linguistics, Cornell University.
- Simons, Gary F. & Charles D. Fennig (eds.). 2017. *Ethnologue: Languages of the world*. 20th edn. Dallas: SIL International. ethnologue.com/language/jmx (June 13 2017.)
- Stalder, Jurg. 1993. *Rapid Appraisal: Strategy and Methodology as Applied in Cameroon*. Yaoundé, Cameroon: Ministry of Scientific and Technical Research.
- Stalder, Jurg & Alan Starr. 1990. *Rapid Appraisal: Strategy and methodology as applied in Cameroon*. Yaoundé, Cameroon: Ministry of Scientific and Technical Research.
- Summer Institute of Linguistics. 1989. Language assessment criteria: Conference recommendations. In Gloria Kindell (ed.), *Proceedings of the Summer Institute of Linguistics International language assessment conference*, 27–29. Horsleys Green, U.K.: SIL International.
- Swadesh, Morris. 1950. Salish internal relationships. *International Journal of American Linguistics* 16:157–167.
- Swadesh, Morris. 1972. *The origin and diversification of language*. Joel Sherzer (ed.). London: Routledge & Kegan Paul.
- Terborg, Roland & Laura García Landa. 2006. The language situation in Mexico. *Current Issues in Language Planning* 7:415–518.
- Voeglin, C.F. & Zellig S. Harris. 1951. Methods for determining intelligibility among dialects of natural languages. *Proceedings of the American Philosophical Society* 95:322–329.
- Wenker, Georg. 1877. *Das rheinische Platt*. Dusseldorf, Germany.
- Wolff, Hans. 1964. Intelligibility and inter-ethnic attitudes. In Dell Hymes (ed.), *Language in culture and society: A reader in linguistics and anthropology*, 440–445. New York: Harper and Row.
- Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.
- Zylstra, Carol. 1980. Phonology and morphophonemics of the Mixtec of Alacatlazala, Guerrero. SIL Mexico Working Papers 4. <http://www.mexico.sil.org/resources/archives/10878> (June 11, 2017.)

APPENDICES

APPENDIX A

CALENDAR OF SURVEY TRIPS

October – November 2016

<u>Sunday</u>	<u>Monday</u>	<u>Tuesday</u>	<u>Wednesday</u>	<u>Thursday</u>	<u>Friday</u>	<u>Saturday</u>
<u>Oct 23</u>	<u>Oct 24</u> YUC Gen WL	<u>Oct 25</u> SMD Gen WL, Tone WL, Socio Q	<u>Oct 26</u> YUC Tone WL, Gen WL	<u>Oct 27</u>	<u>Oct 28</u>	<u>Oct 29</u>
<u>Oct 30</u>	<u>Oct 31</u> SPC Gen WL, Tone WL, Socio Q	<u>Nov 1</u>	<u>Nov 2</u>	<u>Nov 3</u>	<u>Nov 4</u> SPC Gen WL	<u>Nov 5</u> SMD Socio Q (2)
<u>Nov 6</u>	<u>Nov 7</u>	<u>Nov 8</u>	<u>Nov 9</u>	<u>Nov 10</u>	<u>Nov 11</u> SRC General observation	<u>Nov 12</u> SMD Socio Q (2)

<u>Nov 13</u>	<u>Nov 14</u> Zaragosa Alacranes (Unsuccessful attempt)	<u>Nov 15</u> GNA Gen WL (2) , Tone WL, Socio Q	<u>Nov 16</u>	<u>Nov 17</u> SRC Met with town authorities	<u>Nov 18</u>	<u>Nov 19</u> GNA Gen. WL, Tone W, Socio Q
<u>Nov 20</u>	<u>Nov 21</u> SRC Gen WL, Tone WL, Socio Q SCA Met with town authorities	<u>Nov 22</u> SCA Gen WL (2), Tone WL (2), Socio Q (3)	<u>Nov 23</u> SMC Gen WL (2), Tone WL (2), Socio Q (3)	<u>Nov 24</u>	<u>Nov 25</u>	<u>Nov 26</u>

APPENDIX B

GENERAL WORDLIST DATA

Transitive verbs are listed with their object in parentheses.

	<u>Spanish</u>	<u>English</u>	<u>SMD</u>	<u>GNA</u>	<u>SRC</u>	<u>SCA</u>	<u>SMC</u>	<u>SPC</u>	<u>YUC</u>	<u>SJM</u>	<u>COI</u>
1	abeja	bee	ᵐᵁᵐᵁ	ᵐᵁᵐᵁ	ᵐᵁᵐᵁ	ᵐᵁᵐᵁ	ᵐᵁᵐᵁ	ᵐᵁᵐᵁ	ᵐᵁᵐᵁ	ᵐᵁᵐᵁ	ᵐᵁᵐᵁ
2	abono	fertilizer	ābónó	ᶑāᶑā	ābōnō	ābōnō	ābōō	ābónó	kāʔβā	tàtā	bónō
3	ala	wing	kīᶑī	ᵐᵁᶑīᶑī	ᵐᵁᶑīᶑī	ᵐᵁᶑīᶑī	ᵐᵁᶑīᶑī	ᵐᵁᶑīᶑī	dīsí	ᵐᵁᶑitsi	ᵐᵁᶑíᶑí
4	amarrar (la leña)	to tie together (firewood)	kātú	kātū	kātú	kātú	káátúᶑī	ᶑā	ᵐᵁᶑikù	ᵐᵁᶑikātú	kātú
5	amarillo	yellow	kʷāā	kʷāā	kʷāā	kʷāā	kʷāā	kʷāā	kʷāā	ᵐᵁᶑikʷáá	kʷáá
6	ampolla	blister	ᶑᶑīkī	ᶑᶑīkī	ᶑᶑīkī	ᶑᶑīkī	ᶑᶑīkī	ᶑᶑīkī	ᶑᶑīkī	ᶑᶑīkī	ᶑᶑīkī
7	ancho	wide	ᵐᵁᶑīkā	ᵐᵁᶑīkā	ᵐᵁᶑīkā	ᵐᵁᶑīkā	ᵐᵁᶑīkā	ᵐᵁᶑīkā	ᵐᵁᶑīkā	ᵐᵁᶑīkā	ᵐᵁᶑīkā
8	anillo	ring	ᶑēʔē	ᶑēʔē	ᶑēʔē	ᶑēʔē	ᶑēʔē	ᶑēʔē	ᶑēʔē	ᶑēʔē	ᶑēʔē
9	árbol	tree	ītú	ītú	zūtú	zūtú	zūtú	zūtú	tíkú	zūtú	ītú
10	blanco	white	kʷíᶑí	kʷíᶑí	kúᶑí	kúᶑí	zāā	kúᶑí	zāā	kʷíᶑí	jāā
11	borrego	sheep	lāᵐᵁᶑī	lāᵐᵁᶑī	bōrégó	bōrégó	bòrègò	bōrégó	tīkātᶑī	ᵐᵁᶑīkātᶑī	lāᵐᵁᶑī
12	bravo	bold	kʷèʔè	kʷèʔè	kʷèʔè	kʷèʔè	kʷèʔè	kʷèʔè	kʷèʔè	ᶑéé	kʷèʔè

	<u>Spanish</u>	<u>English</u>	<u>SMD</u>	<u>GNA</u>	<u>SRC</u>	<u>SCA</u>	<u>SMC</u>	<u>SPC</u>	<u>YUC</u>	<u>SJM</u>	<u>COI</u>
13	bueno	good	βà?à	βà?ā	βà?à	βà?ā	βà?à	βà?ā	βà?ā	βà?ā	βà?á
14	caballo	horse	kʷáī	kʷáī	kʷèi	kʷèi	kʷèi	kʷáī	kʷájū	kítī	kʷájù
15	cabello	hair	íʃí ʃɪnī	íʃí ʃɪnī	íʃí ʃɪnī	íʃí ʃɪnī	íʃí ʃɪnī	íʃí ʃɪnī	íʃí ʃɪnī	íʃí	íʃí
16	cabeza	head	ʃɪnī	ʃɪnī	ʃɪnī	ʃɪnī	ʃɪnī	ʃɪnī	ʃɪnī	ʃɪnī	ʃɪnī
17	calabaza	squash	ʒíkì	ʒíkì	ʒíkì	ʒíkì	ʒíkì	ʒíkì	íkí ɪnú	íkí	íkí
18	caminar	to walk	ʃíkā	ʃíkā	ʃíkā	ʃíkā	ʃíkā		síkā	ⁿdzìkà	ʃíkā
19	camote	yam	ɲá?mí	ɲā?mī	ɲāmī	ɲā?mī	ɲā?mī	ɲā?mī	ɲā?àmì	ɲā?mì	ɲā?mī
20	canasta	basket	tíkā	tìʒìkā	tùʒìkā	tùʒìkā	tùʒìkā	tùʒìkā	tíkā	t̄ʃìkā	tòkā
21	cantar (una canción)	to sing (a song)	ʃítā	ʃítā	ʃítā	ʃítā	ʃítā	ʃítā	ʒítā	ⁿdzìtā	ʃítā
22	carne	meat	kūɲú	kūɲũ	kūɲú	kūɲú	kūɲũ	kūɲũ	kùɲù	kũɲù	kũɲu
23	casa	house	βē?ē	βē?ē	βē?ē	βē?ē	βē?ē	βē?ē	βē?ē	βē?ē	βē?ē
24	cerca	nearby	ʒātʃī	ʒātʃī	ʒātʃī	ʒātʃī	ʒātʃī	ʒātʃī	ʒātī	jātʃī	jātʃī
25	cerro	mountain	ʒíkì	ʒíkì	ʒūkù	ʒūkù	ʒūkù	ʒūkū	ʒúkú	ʒúkú	ʃíkí
26	cerdo (puerco)	pig	kótʃí	kótʃí	kótʃí	kótʃí	kótʃí	kótʃí	kínì	kínì	kínì
27	chayote	chayote	ɲáɲā	nāɲā	nāɲā	nāɲā	nāɲā	nāɲā	nàɲā	nāɲǎ	nāɲá
28	chile	chili pepper	ʒā?à	ʒā?à	ʒā?à	ʒā?à	ʒā?à	ʒā?ā	ʒà?ā	jā?á	ja?á

	<u>Spanish</u>	<u>English</u>	<u>SMD</u>	<u>GNA</u>	<u>SRC</u>	<u>SCA</u>	<u>SMC</u>	<u>SPC</u>	<u>YUC</u>	<u>SJM</u>	<u>COI</u>
29	ciego	blind	k ^w āā	k ^w āā	k ^w āā	k ^w āā	k ^w āā	k ^w āā	k ^w āā	k ^w āá	k ^w āá
30	colibrí	hummingbird	ʒōʔlī	ⁿ d̄ʒōʔlí	ⁿ d̄ʒóʔlī	ⁿ d̄ʒóʔlō	ⁿ d̄ʒī ʒōʔō	ⁿ d̄ʒí ʒōʔō	ⁿ dīʒōō	ⁿ d̄ʒōʔō	ⁿ dīōʔō
31	colgar (la ropa)	to hang (laundry/clothes)	kātākā	kātāzā	kātākā	kātākā	kātākàʃi	kātāzī	kātākà	ⁿ d̄ʒīnāā	kātīkāā
32	comal	comal (tortilla griddle)	ʃòò	ʃōō	ʃòò	ʃòò		ʃīōō	sìʒō	tsìó	ʃiyó
33	comprar (tortillas)	to buy (tortillas)	ʃī	ʃī	ʃī	ʃī		ʃī	īksí	ⁿ dāk ^w āā	ʃīī
34	cortar (un árbol)	to cut (a tree)	ʃā ⁿ d̄ʒā	ʃá ⁿ d̄ʒā	ʃā ⁿ d̄ʒā	ʃá ⁿ d̄ʒā	ʃ ⁿ dātū	ʃā ⁿ d̄ʒā	ʃà	ⁿ d̄ʒaʔnt̄ʃā	ʃà ⁿ dīā
35	coyote	coyote	ⁿ d̄ʒī βāʔí	ⁿ d̄ʒīβāʔī	ⁿ d̄ʒī βāʔū	ⁿ d̄ʒī βōʔō	ⁿ d̄ʒī βōʔō	ⁿ d̄ʒī βāʔí	ⁿ dībōʔō	ⁿ dīβàʔū	ⁿ dīβáʔjū
36	cuatro	four	kòmì	kùmì	kùmì	kòmì	kòmì	kùmī	kùmī	kùmí	kòmí
37	dos	two	ìβì	ìβì	ùβì	ùβì	ūβī	ūβī	ùβì	ùβì	ùβì
38	dulce	sweet	βìʃì	βìʃì	βìʃì	βìʃì	βìʃì	βīʃī	βìʃì	βìʃì	βìʃì
39	elote	ear of corn	ⁿ d̄ʒīʃī	ⁿ d̄ʒīʃī	ⁿ d̄ʒīʃī	ⁿ d̄ʒīʃī	ⁿ d̄ʒīʃī	ⁿ d̄ʒīʃī	tìʃí	ⁿ dīʃí	ⁿ dīʃí
40	enfermedad	illness	kíʔì	k ^w èʔè	k ^w ēʔē	k ^w èʔè	k ^w èʔè	k ^w ēʔē	k ^w èʔè	k ^w èʔè	k ^w ēʔē
41	escarbar (la tierra)	to dig (soil)	ʃátʃā	ʃátʃà	t̄ʃátʃā	ʃátʃá	ʃátʃāzī	ʃiātʃā	t̄ʃītù	ⁿ d̄ʒàt̄ʃā	kāan
42	escuchar (una canción)	to listen (a song)	ⁿ d̄ʒā sōʔō	ʒōsō	sòʔò	tōsò	sòʔōʃī	tāsō	t̄ʃītà	ⁿ d̄ʒāā sōʔō	ʃīnī sòʔō

	<u>Spanish</u>	<u>English</u>	<u>SMD</u>	<u>GNA</u>	<u>SRC</u>	<u>SCA</u>	<u>SMC</u>	<u>SPC</u>	<u>YUC</u>	<u>SJM</u>	<u>COI</u>
43	flor	flower	ītā	ītā	ītā	ītā	ītā	ītā	ītā	ītā	ītā
44	flor de calabaza	squash blossom	ītā ʒíkī	ītā ʒíkī	βāʒá	ītā ʒíkī	ītā ʒíkī	ītā ʒíkī	ītā íkí	ītā tīβajà	ītā
45	frijol	bean	ⁿ dūtʃī	ⁿ dītʃī	ⁿ dūtʃī	ⁿ dūtʃī	ⁿ dūtʃī	ⁿ dūtʃī	ⁿ dūtʃī	ⁿ dútʃī	ⁿ dūtʃī
46	gallina	hen	ⁿ dūjí	ⁿ dījí	ⁿ dūjí	ⁿ dūjí	ⁿ dūjí	ⁿ dūjí	ⁿ dīsí	ʃjùú	ⁿ dīūjí
47	gallo	rooster	ʃjēlē	ʃjēlē	ʃjēlē	ʃjēlē	ʃjēlē	ʃjēlē	tōʔolō	tóʔlō	ʃjēlē
48	grueso	thick	ⁿ dʒēʔē	ⁿ dʒēé	ⁿ dēē	ⁿ dēē	ⁿ dēé	ⁿ dēē	ⁿ dēé	ⁿ dʒēé	ⁿ dūʔú
49	hablar	to speak	kāʔǎ	kàʔà	kāʔǎ	káʔǎ	kāʔà	kāʔǎ	kaʔà	kàʔǎ	kātʃī
50	harina	flour	āríná	āríná	āríná	āríná	āríná	āríná	ītʃiā	ārínà	ārínā
51	hervir (agua)	to boil (water)	sākiti	sītʃíkī	ʃíkítʃī	sákítʃī	ʃákítʃāʃī	ʃákítʃī	sàsā	skiti	sàk ^w isò
52	hilo	string	ʒòʔβà	ʒòʔβà	ʒūʔβà	ʒùʔβà	ʒūʔβā	ʒūʔβā	ʒūʔβà	jùʔβà	yùʔβà
53	hombre	man	ʃjāà	ʃjāā	ʃjāā	ʃjāā	ʃjāā	ʃjīāā	tāā	ʃjāá	tīāā
54	jabón	soap	nàmà	nàmà	nàmà	nàmà	nàmà	nāmā	nāmá	nāmă	námá
55	jícara	gourd	ⁿ dʒāʃī	ʒāʃī	ʒāʃī	ʒāʃī	ʒāʃī	ʒāʃī	ʒāsí	jātsí	jāʃí
56	largo	long	kāpnī	kāānī		kāāpnī	kāāpnī	kāpnā	kāānī	kānī	káʔnū
57	lejos	far away	ʃíká	ʃíká	ʃíkā	ʃíkā	ʃíkā	ʃíkā	síká	tsíkā	ʃíkā
58	llegar	to arrive	ʃà	kìʃà	kīʃā	kīʃā	ā ⁿ dʒá	kīʃā	ʃī	kìtsàà	ʃàà

	<u>Spanish</u>	<u>English</u>	<u>SMD</u>	<u>GNA</u>	<u>SRC</u>	<u>SCA</u>	<u>SMC</u>	<u>SPC</u>	<u>YUC</u>	<u>SJM</u>	<u>COI</u>
75	peine	comb	k ^w ɪkà	k ^w ɪkà	k ^w ɪkà	k ^w ɪkà	k ^w ɪkà	k ^w ɪkà	βikà	k ^w ɪkà	k ^w ɪkà
76	pelear	to fight	kāpī tá?ā	kà?à	kāpī tá?ā	kāpī tá?ā	kā ⁿ dʒá?ā	kāpī tá?ā	kànā	kākā	
77	perro	dog	tʃinā	tʃinā	tʃinā	tʃinā	tʃinā	tʃinā	tìnà	tǐnà	tǐnā
78	plato	plate	kò?ò	kò?ò	kòò	kòò	kò?ò	kò?ò	kò?ò	kò?ò	kò?ó
79	pobre	poor	ⁿ dā?í	ⁿ dā?ī	ⁿ dā?βī	ⁿ dá?āβí	ⁿ dā?āβī	ⁿ dā?āβī	lā?āβí	ⁿ dā?βí	ⁿ dá?βī
80	polvo	dust	mātʃī	mātʃī	ⁿ dʒikātʃī	mātʃū	mātʃū	mātʃī	záāká	jākā	jāká
81	raíz	root	ʒò?ò	ʒò?ò	ʒò?ò	ʒò?ò	ʒò?ò	ʒò?ò	ʒò?ò	jò?ó	tīò?ò
82	reír	to laugh	ʃàki	ʃàki	ʃàkū	ʃàkū	ʃàkū	ʃàkū	ʃiàkù	ⁿ dzàkù	ʃàkù
83	río	river	ʒitʃà	ʒitʃà	ʒūtʃà	ʒūtʃà	ʒūtʃà	ʒūtʃià	itē	jūtʃà	itīà
84	rico	rich	k ^w ɪkà	k ^w ɪkà	k ^w ɪkà	k ^w ɪkà	k ^w ɪkà	k ^w ɪkà	βiká	k ^w ɪkà	k ^w ɪkà
85	rojo	red	k ^w ā?ā	k ^w ā?ā	k ^w ā?ā	k ^w ā?ā	k ^w ā?ā	k ^w ā?ā	k ^w ā?à	ⁿ dik ^w á?à	k ^w á?á
86	sal	salt	nìi	nìi	nìi	nìi	nìi	nìi	ĩĩ	ĩĩ	ĩĩ
87	serpiente	snake	kòò	kòò	kòò	kòò	kòò	kòò	kòó	kòó	kòó
88	semilla de chile	chili seed	ⁿ dʒikī ʒā?ā	ⁿ dʒikī ʒā?ā	ⁿ dʒikī ʒā?ā	ⁿ dʒikī ʒā?ā	ⁿ dʒikī ʒā?ā	ⁿ dʒikī ʒā?ā	díkí ʒá?á	ⁿ díkí	ⁿ díkí
89	siete	seven	ìʃà	ìʃà	ùʃà	ùʃà	ùʃà	ùʃà	ùsìà	ùtsà	ùʃà
90	tenate	tenate (basket)	ⁿ dò?ò	ⁿ dò?ò	ⁿ dò?ò	ⁿ dò?ò	ⁿ dò?ò	ⁿ dò?ò	ⁿ dò?ó	ⁿ dò?ó	ⁿ dò?ó

	<u>Spanish</u>	<u>English</u>	<u>SMD</u>	<u>GNA</u>	<u>SRC</u>	<u>SCA</u>	<u>SMC</u>	<u>SPC</u>	<u>YUC</u>	<u>SJM</u>	<u>COI</u>
91	tía	aunt	ʃiʃi	ʃiʃi	ʃiʃi	ʃiʃi	ʃiʃi	ʃiʃi	íʃi	ʃiʃi	ʃiʃi
92	tlacuache	opossum	lángō	xākō	ʃákō	ʃákō	ʃákō	ʃākō	tʃōkò	tsókō	ʃākō
93	tortilla	tortilla	ʃità	ʃità	ʃità	ʃità	ʃità	ʃitā	īʃtá	stàá	ʃitá
94	vapor	vapor/steam	ʒòkò	ʒòkò	ʒòkò	ʒūkù	tòkò	ʒòkō	ʒòkó	jòkǒ	jòkó
95	vender (tortillas)	to sell (tortillas)	ʃikó	ʃikō	ʃikó	ʃikū	ʃikō	ʃikó	ʃikō	ʃikō	ʃikō
96	ver, mirar (una película)	to see/watch (a movie)	ⁿ d̄ʒĩʔé	ⁿ d̄ʒēʔē	ⁿ dēʔē	ⁿ déé	ⁿ déʔéʃi	ⁿ d̄ʒē	ⁿ dēʔē	ⁿ d̄ʒèʔē	ʃitōʔnī
97	verde	green	k ^w ĩ	k ^w ĩ	k ^w ĩ	k ^w ĩ	k ^w ĩ	k ^w ĩĩ	k ^w íĩ	k ^w íĩ	k ^w íĩ
98	zorrillo	skunk	ⁿ d̄ʒĩnĩʔí	tʃĩnĩʔĩ	tʃĩnĩʔí	tʃĩnĩʔí	ríló	tʃĩʔĩ	tĩʔĩtā	tíʔĩ itā	lírú

APPENDIX C

TONE WORDLIST DATA

TONE WORDLIST ABBREVIATIONS

L Low tone
 M Mid tone
 H High tone

F Falling (contour) tone
 R Rising (contour) tone

M1 Mora 1
 M2 Mora 2

GROUP 1 *L																			
<u>Spanish</u>	<u>English</u>	<u>SMD</u>		<u>YUC</u>		<u>SPC</u>		<u>GNA</u>		<u>SRC</u>		<u>SCA</u>		<u>SMC</u>		<u>COI</u>		<u>SJM</u>	
		M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2
dos	two	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
nueve	nine	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
peine	comb	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
sacerdote	priest	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
siete	seven	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
tres	three	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L

GROUP 2 *L²

<u>Spanish</u>	<u>English</u>	<u>SMD</u>		<u>YUC</u>		<u>SPC</u>		<u>GNA</u>		<u>SRC</u>		<u>SCA</u>		<u>SMC</u>		<u>COI</u>		<u>SJM</u>	
		M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2
cuatro	four	L	L	L	M	L	L	L	L	L	L	L	L	L	L	L	H	L	R
serpiente	snake	L	L	L	M	L	L	L	L	L	L	L	L	L	L	L	H	L	H
jabón	soap	L	L	L	M	L	L	L	L	L	L	L	L	L	L	L	H	L	R
lluvia	rain	L	L	L	H	L	L	L	L	L	L	L	L	L	L	L	H	L	R
maíz	corn	L	L	L	M	L	L	L	L	L	L	L	L	L	L	L	H	L	R
metate	woven belt	L	M	L	L	L	M	L	M	L	M	L	M	L	M	R	M	R	M
nube	cloud	L	L	M	H	L	L	L	L	L	L	L	L	L	L	L	H	L	R
piedra	rock	L	L	L	M	L	L	L	L	L	L	L	L	L	L	L	H	L	H
plato	plate	L	L	M	M	L	L	L	L	L	L	L	L	L	L	L	H	L	R
tortilla	tortilla	L	L	M	M	L	L	L	L	L	L	L	L	L	L	L	H	L	H

GROUP 3 *M

<u>Spanish</u>	<u>English</u>	<u>SMD</u>		<u>YUC</u>		<u>SPC</u>		<u>GNA</u>		<u>SRC</u>		<u>SCA</u>		<u>SMC</u>		<u>COI</u>		<u>SJM</u>	
		M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2
casa	house	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
uno	one	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M

GROUP 4*M²

<u>Spanish</u>	<u>English</u>	<u>SMD</u>		<u>YUC</u>		<u>SPC</u>		<u>GNA</u>		<u>SRC</u>		<u>SCA</u>		<u>SMC</u>		<u>COI</u>		<u>SJM</u>	
		M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2
chile	chili pepper	M	L	M	H	M	L	M	L	M	L	M	L	M	L	H	M	M	H
elote	ear of corn	M	L	M	H	M	L	M	L	M	L	M	L	M	L	M	H	F	R
espina	spine/thorn	M	L	M	H	M	L	M	L	M	L	M	L	M	L	M	H	F	R
fiesta	party	M	L	M	H	M	L	M	L	M	L	M	L	M	L	M	M	F	R
frijoles	beans	M	L	M	H	M	L	M	L	M	L	M	L	M	L	M	H	F	R
masa	dough	M	L	M	H	M	L	M	L	M	L	M	L	M	L	M	H	F	R
mecate	grindstone	M	L	M	H	M	L	M	L	M	L	M	L	M	L	M	H	M	H
petate	woven mat	M	L	M	H	M	L	M	L	M	L	M	L	M	L	M	H	F	R
tierra	soil	M	L	M	H	M	L	M	L	M	L	M	L	M	L	M	H	M	H

GROUP 5 *ML

<u>Spanish</u>	<u>English</u>	<u>SMD</u>		<u>YUC</u>		<u>SPC</u>		<u>GNA</u>		<u>SRC</u>		<u>SCA</u>		<u>SMC</u>		<u>COI</u>		<u>SJM</u>	
		M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2
flor	flower	M	M	M	L	M	M	M	M	M	M	M	M	M	M	M	M	M	L
fuego	fire	M	M	M	L	M	M	M	M	M	M	M	M	M	M	M	M	M	L
papel	paper	M	M	M	L	M	M	M	M	M	M	M	M	M	M	M	M	M	L
perro	dog	M	M	M	L	L	L	L	M	L	L	L	M	L	L	R	M	R	L
pueblo	village	M	M	M	L	M	M	M	M	M	M	M	M	M	M	M	M	M	L
venado	deer	M	M	M	L	M	M	M	M	M	M	M	M	M	M	M	M	M	L

GROUP 6*LM

<u>Spanish</u>	<u>English</u>	<u>SMD</u>		<u>YUC</u>		<u>SPC</u>		<u>GNA</u>		<u>SRC</u>		<u>SCA</u>		<u>SMC</u>		<u>COI</u>		<u>SJM</u>	
		M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2
adobe	adobe	L	M	L	M	M	M	M	M	L	M	M	M	M	M	L	M	L	M
chapulín	grasshopper	L	M	M	M	M	M	M	M	L	M	M	M	M	M	M	M	L	M
ejote	green bean	M	H	M	L	M	M	M	H	L	M	M	M	M	M	M	M	R	L
hombre	man	M	M	L	L	M	M	M	M	M	M	M	M	M	M	L	L	L	H
hormiga	ant	M	H	M	H	M	H	M	H	M	H	M	H	M	H	M	H	F	R
manteca	lard	L	L	M	L	M	M	M	M	L	M	M	M	M	M	M	M	L	L
campana	bell	L	M	L	H	M	M	M	H	L	M	M	M	M	M	L	L	L	M
olla	pot	L	M	L	M	M	M	M	M	L	M	M	M	M	M	L	M	L	M
río	river	L	M	L	M	M	M	M	M	L	M	M	M	M	M	L	M	L	M

APPENDIX D

CONSIDERATIONS AND EXCLUSIONS FOR ANALYSIS OF SIMILAR FORMS

The following criteria were used in the analysis of similar forms for categorizing corresponding segments.

- Lack of nasalization or glottalization and metathesis resulted in a downgrade of one category. Nasalization and glottalization are contrastive and features of the morpheme in Mixtec languages. (See Macaulay 1995, Marlett 1992, and McKendry 2013:67.) For example, ‘word’ is produced as [t̃ũʔũ̃] in SMC and [tũũ̃] in SPC. The missing glottal stop in SPC means that one of its Category One segments is downgraded to a Category two. My convention is to treat long vowels as single segments (see below), so this is considered as a word with two segments, /tũ̃/, with one Category One and one Category Two segment. Therefore, these are not similar forms.
- Long vowels were counted as a single segment, due to the fact that underlyingly, they are single length vowels. Words that are underlying CV are lengthened to CVV or CV?V to satisfy the minimal word constraint of two mora. (See McKendry 2013.)
- [kʷ] and [ᵑg] are considered to be single segments, as they are individual phonemes.
- There is a systematic correspondence between [ts] in SJM and [ʃ], [s], and [ʔd] in the seven varieties surveyed. (Category 1)
- There is a systematic correspondence between [ᵑd͡z] in SJM and [ᵑd͡ʒ] and [ᵑd] in the seven varieties surveyed. (Category 1)
- There is a systematic correspondence between [j] in COI and SJM and [ʒ] in the seven varieties surveyed. (Category 1)
- General Wordlist items #18 ‘to walk’ (missing in SPC), #32 ‘comal’ (missing in SMC), #33 ‘to buy’ (missing in SMC), and #56 ‘long’ (missing in GNA) were

excluded from the analysis of similar forms because of missing data for at least one of the varieties surveyed.

- The following distinctive features were used for determining the number of different features between vowels:

	a	e	i	o	u
[back]	+	—	—	+	+
[low]	+	—	—	—	—
[high]	—	—	+	—	+

The features [round] and [ATR] are not needed to distinguish the vowels of the Mixtec varieties surveyed and thus excluded from this table. While /i/ occurs in some Mixtec varieties, it is unattested in the ones surveyed.

APPENDIX E

GENERAL WORDLIST ITEMS AND SYSTEMATIC CORRESPONDENCES

Some correspondences have only one example from the General Wordlist, since McKendry (2001) has already identified many systematic correspondences and their glosses for Mixtec varieties.

<u>Isogloss</u>	<u>Zone A</u>	<u>Zone B</u>	<u>General Wordlist Items (Appendix B)</u>
1	k ^w	β	66, 75, 84
2	ʃ	s/ʒ/ks	3, 18, 21, 32, 33, 46, 55, 57, 70
3	ʃi	iʃ	91, 93
4	tʃ	t	24, 54, 72, 77
5	ʒ	∅ (null)	9, 44, 61, 83
6	ʒ	ʃ	63
7	ⁿ dʒ	ⁿ d	7, 35, 48, 88, 96
8	ⁿ d	l	79
9	n	∅ (null)	63
10	i	u	9, 25, 35, 63, 64, 80, 89
11	a	e	14