

# Colonial Valley Zapotec Effects on Bilingual Spanish

Jaime Metzger

A thesis submitted in partial fulfillment of the requirements for  
the degree of Bachelor of Arts in Linguistics

Haverford College

December 2018

## Abstract

Most linguistic research on Zapotec-Spanish language contact has centered on Spanish influences such as loanwords borrowed into Zapotec. The current state of the field fails to acknowledge the structural effects that Zapotec has had on modern Oaxacan Spanish. This thesis analyzes variation and innovation found in a corpus of Zapotec and Spanish bilingual manuscripts. The goal in variationist language contact studies such as this one is to tell the stories of the sociolinguistic identities whose language use informed the local variant. The patterns that emerged because of Zapotec grammar's influence can be captured dynamically across time by way of data visualization and complementary present-day studies, both of which this thesis will begin and suggest.

Tri-College Linguistics Department  
Under the supervision of Brook Lillehaugen

# Contents

Abstract

Acknowledgments	1
<b>1 Zapotec in Contact with Spanish</b>	<b>2</b>
1.1 Colonial Valley Zapotec . . . . .	3
1.2 Spanish as an Imperial Language . . . . .	6
1.3 Bilingualism . . . . .	8
1.4 Language Contact Studies . . . . .	9
1.4.1 Thomason & Kaufman 1992 . . . . .	9
1.4.2 Labov 1972 . . . . .	11
1.4.3 Muysken 2005 . . . . .	11
<b>2 Corpus</b>	<b>12</b>
2.1 A Closer Look . . . . .	12
2.2 Evidence of Bilingualism . . . . .	13
2.3 Data Collection Methods . . . . .	15
<b>3 Variation</b>	<b>15</b>
3.1 Agreement . . . . .	15
3.1.1 Number . . . . .	16
3.1.2 Gender . . . . .	16
3.2 Simplification . . . . .	16
<b>4 Innovation</b>	<b>19</b>
4.1 Quantifier and Dual . . . . .	20
4.2 Transfer . . . . .	22
4.3 San Lucas Quiavini Zapotec . . . . .	23
4.4 Modern Oaxacan Spanish . . . . .	25
<b>5 Diachronic analysis</b>	<b>26</b>
5.1 Proposed Diagnostic Tools . . . . .	29
5.1.1 Interference Suggests Bilingualism . . . . .	29
5.1.2 Transfer spectrum application . . . . .	30
5.1.3 Data visualization impact on findings . . . . .	30

### List of Figures

1	The Depth of Transfer spectrum mentioned above. . . . .	2
2	The Ticha Project’s map of manuscript origins. . . . .	4
3	Zapotec language text in the Archivo General del Estado de Oaxaca [12] . . .	13
4	The corpus shift in distribution according to conjoined elements’ roles . . . .	27
5	Timeline showing which towns contributed to quantifier dual tokens . . . . .	28
6	The above timeline zoomed in . . . . .	28
7	Individual translators’ contribution to quantifier dual tokens in corpus . . . .	29

### List of Tables

1	Potential directions of language contact interference . . . . .	10
2	What language background did these translators have? . . . . .	13
3	Spanish conjunctions . . . . .	22

## Acknowledgments

I had the wonderful opportunity to work with this corpus thanks to the advice of Brook Danielle Lillehaugen- and although that first spark of inspiration towards the ongoing project was transformative for me, I can say without a doubt that the most impactful help I received from her was her reliable and encouraging support as my academic advisor.

Both Brook and Emily Gasser provided many astute comments towards the improvement of this thesis for which I am extremely thankful.

I am also incredibly grateful for the consultation of Dr. Felipe H. Lopez and Elena Hernandez. Thank you for sharing your intuitions with me!

Since starting the analysis of this corpus, I have had the fantastic opportunity to travel and present at conferences as well as spend the summer occupying a digital scholarship research fellowship; without those opportunities I would not have achieved as much progress as I have. With that in mind I'd like to thank the Haverford College Louis Green Fund and the Haverford College Libraries team for the Slocum Fellowship experience, especially Mike Zarafonetis.

Finally, my peers have made my writing endeavors more fruitful by reviewing pieces of the thesis and checking the accessibility; an emphatic thank you is owed to Juliana Gonzales, Emily Lin, Grace Metzger, and the other members of my family.

# 1 Zapotec in Contact with Spanish

This thesis examines two Zapotec-Spanish language contact effects that will be analyzed at opposite ends of a spectrum: **Depth of Transfer**. The spectrum presents a common mystery to sociolinguists who study language contact situations all over the world. The shallow end of the spectrum hosts many transient, surface-level effects representing Zapotec grammar's interference in Spanish language formation. Examples of these effects are not expected to be uniform across all idiolects, but they may be useful for identifying bilingualism. They provide snapshots of the ongoing process of a speaker's second language acquisition.



Figure 1: The Depth of Transfer spectrum mentioned above.

At the other end, we place the effects that are more persistent. Their increased presence characterizes the affected language itself- in this case, Spanish. On a structural level, any construction that has grown roots in Spanish syntax would belong at this deep end. No feature of Zapotec could reach the deep end of **transfer**, without having first been **interference**. So why did some features of Zapotec that interfered with Spanish grammar remain relegated to the shallow end of this spectrum when other features actually became mapped onto the Spanish language as transferred pieces of Zapotec grammar?

The course of my analysis begins with uncovering variation in Spanish bilingual texts. Any Spanish constructions that would be ungrammatical in the Castilian Spanish grammar constitute candidates for contact effects if they reflect features from the Colonial Valley Zapotec grammar. With a diachronic analysis of Zapotec-influenced variation, the goal is to achieve a greater understanding of Oaxacan Spanish dialect development. Luckily, I have access to modern day speaker intuitions which allow for strong hypotheses on which transferred constructions are present in Spanish today, but examples of constructions from the 17th-18th centuries receive the most analysis in this thesis. These examples, which come from a corpus of legal manuscripts, demonstrate that bilingual translators produced a range of linguistic variation types which can be tracked back to Zapotec influence.

Spanish influence on Zapotec has been much more widely studied than the converse: Zapotec influence on Oaxacan Spanish. The instances of variation that I look at in this thesis, whether the result of language learning or dialectical innovation, all stem from language con-

tact between Zapotec and Spanish. Representations of this contact beyond glossed examples will be explored through data visualization, a necessary technique for historical linguists that can answer questions about the nature of linguistic interference and the spread of innovation.

My thesis puts forth an in-depth analysis of one Oaxacan Spanish construction which may have been formed based on Zapotec quantifier-building conventions: *los dos con* as an ambiguous conjunction used in naming groups. Given an increase in scope that this feature underwent, a few sketches of potential change dynamics will be offered with the caveat that modern Oaxacan Spanish data still needs to be developed on the topic. The classification of this quantifier strategy for groups relies upon on the depth of transfer level, and confirmation from modern Oaxacan Spanish speakers leads me to conclude that it qualifies as a deep transfer from Zapotec into Spanish. What’s more, it constitutes a dialectical innovation which exemplifies the long-term impact that Zapotec languages had on Spanish in Colonial Oaxaca.

When it comes to framing this study in the context of other linguistic projects, it applies research from the subfields of both bilingualism and dialect studies. For the descriptive analysis of contact-induced innovation, a review of seminal language contact works will be referenced. A contextualized comparison of these frameworks’ goals allows me the advantage of drawing on all three to explain the deponent structures found in the manuscripts. Non-canonical transfers, or borrowings, and structural transfers both belong at the deep end of the spectrum in question, but I will draw out the implications of each classification for the stories of these pueblos. In the end, examples of data visualization used to analyze my findings illuminate how Zapotec grammar persisted under the duress of imperialism. It is my aspiration that this thesis contributes to the ever-expanding field of analyzing and disseminating the features found in historical corpora.

## 1.1 Colonial Valley Zapotec

The Zapotec language family originated in Mesoamerica, branching off the larger Otomanguean language family. Today it consists of at least 57 distinctive Zapotec languages spoken by 441,182 people (Simons and Fennig 2018) [21]. The Zapotec language represented in the corpus studied for this thesis may have functioned as a standardized lingua franca variety that allowed myriad towns with different Zapotec dialects to communicate more easily. This variety, called Colonial Valley Zapotec, has been studied as “a historical form” from which the many diverse modern Valley Zapotec cognates developed (Heher et al 2014) [8]. The corpus of archival documents that provides Colonial Valley Zapotec data includes both handwritten

legal manuscripts and printed religious texts from a range of about 250 years between the 16th century and the late 18th century.

The linguistic features of CVZ have been sketched in previous works: just like its modern descendants, it follows VSO word order with some flexibility, and its templatic verbal morphology has been classified as polysynthetic. While modern Zapotec languages are tonal, the analyzed data for CVZ consists of solely written materials, so its tonal distinctions can only be estimated.



Figure 2: The Ticha Project's map of manuscript origins.

In Figure 2, a modern map of Mexico is marked by pointers showing the approximate locations of the towns whose legal institutions conserved the manuscripts which I study. All of the locations are based upon the geolocation of colonial churches in each town, as the layout of Oaxacan pueblos usually follows the pattern of the traditional Mexican centro with the Catholic church in a place of central importance. Some of the locations are indeterminate because of the names of places changing, but historical records of name changes have allowed me to guess the modern town corresponding to a renamed colonial town. Of course, due to potential overlap for named places, these are just hypotheses and not conclusions. Regardless, a discrete area appears in the map: the Valley of Oaxaca. Only four points stand apart from the conglomeration, and the four outliers only contribute one text each to the examined 54 texts. The complete list of towns and documents can be found online at

the Ticha project [12].

When it comes to language contact between Spanish and indigenous languages of Mesoamerica like Zapotec, linguists have studied the influence of Spanish on the substratum languages far more often than they have studied the opposite. Because Spanish was and continues to be an imperial language, and because it was later declared the national language of Mexico during the Mexican independence movement, it became ubiquitous in Mexico, even in Zapotec-speaking Oaxacan communities. Learning Spanish has usually been prioritized over native languages' preservation since the 20th century, which means that each generation has had fewer Zapotec speakers. Today, Zapotec languages are endangered, and there is a commonly held misconception that while the Spanish language affected Zapotec, Zapotec had no effects on Spanish whatsoever. This false idea paints the Zapotec languages as weak, which is damaging to language revitalization efforts and to linguistic identity of Zapotec-identifying people. The failure of academic institutions to study the impact of Zapotec languages on the imperial and nationalist language, Spanish, produces a perceivable absence of Zapotec voice that could be wrongly interpreted as a real lack of impact. This interpretation would not only be false, but would compound the damaging narrative of Spanish linguistic superiority which continues to harm many modern day Zapotec speakers by way of discrimination.

A few linguists have not missed the chance to study Spanish language change as it relates to indigenous language influence; for instance, the work of Hardman de Bautista (1982) touches on sound change as well as syntactic variation [7]. In regards to Zapotec linguistic studies, there is a concurrent paper by Rosemary Beam de Azcona that studies several effects of Zapotec on Spanish by investigating the Sierra Sur region's Spanish and finding that postverbal subjects and some other originally Zapotec features have transferred to the bilingual dialect [2]. Her work is critically important not just because it contributes to the field but because, as she puts it, "la forma en que se habla el español en la Sierra Sur cuenta una historia zapoteca" or "the form in which Spanish is spoken in the Sierra Sur tells a Zapotec history" (Azcona in preparation) [2, p. 3]. These works are increasingly relevant as speakers of endangered languages become activists seeking to discover more about the histories of their languages. The accessibility of these projects is more critical than ever. For the linguistic field, the clarity and organization of such works are also an issue. The frameworks accessible to sociolinguists are numerous and complex, but when it comes to diachronic studies of languages living in resistance, no clear style has emerged as coherent. For interested non-linguists and invested theoretical linguists alike to get something out of sociolinguistic writings, a consolidated model would help. With this in mind, my analysis

will attempt to touch on each of the the typological categories most important to influential authors like Thomason & Kaufman, Muysken, and Labov.

## 1.2 Spanish as an Imperial Language

In order to make clear the sociolinguistic context in which language contact took place, I will preface the highly political history of Spanish being spoken in Mesoamerica. The history of Spanish in Mexico is, by nature, imperial history. Before the sixteenth century, Spanish was not a language spoken in Mesoamerica; in fact, there are myriad substratum languages in the region, languages that developed there. The Mesoamerican language group is one of the most diverse groups worldwide, and within it "the area of greatest diversity is located in the State of Oaxaca (36,820 sq. miles)" (Suarez 1983) [23, p. 16]. In the communities that this thesis investigates, Zapotec was the substratum basis for language and communication.

Since 1520, when the Nahuatl empire was displaced by the Spanish conquest and empire, religious and political officials asserted the goal of disseminating Spanish speaking and writing among native peoples. Farriss (2018) [6] describes the unique obsession with documentation as a characteristic of the Spanish empire. This directly incited the rapid pace of Spanish language acquisition as well as the creation of the materials and practices that provided models for those acquiring Spanish.

Economic aims of the new empire contributed to the impetus to learn Spanish as well, because they included the exploitation of indigenous labor and lands. How could a family defend the resources they possessed when a new empire had changed the rules by which one claimed to own that land? Not only that, but the language in which these claims were made had changed in accordance with the new guard.

Finally, the sociopolitical goals of imperial agents figured into this language contact environment. The language choices made by bilingual speakers were highly stratified. While under the duress of colonial exploitation, indigenous people often married Spaniards or aiding the Spaniards in their imperial bureaucracy. Social movement was a matter of survival; it is estimated that by 1650, the indigenous population had suffered such losses from foreign disease that only 150,000 indigenous people remained in Oaxaca, although the 1520 estimate is 1.5 million. This immense loss necessitated drastic change in the functions of every community, and facilitated further dependence on the Spanish empire.

Those who bore mestizo children, or mixed children, brought about the familial structure that was the first certain catalyst for bilingualism. Other catalysts included efforts by Spanish friars to teach Spanish, which were initially unsuccessful until they had studied

Zapotec as well and cultivated bilingual texts; even then, Spanish monolinguals were far outnumbered in Oaxacan towns, so anyone who did not grow up in a bilingual home would have very few spoken sources of Spanish language input. These factors imply that, while some indigenous priests or *cassiques* may have had to acquire Spanish quickly by learning from monolinguals, that was not a popular case for language acquisition, and a town would likely not be bilingual during the first generations of language contact. Those generations would have been notably earlier than any of the documents in the corpus: 1528 was the year Dominicans settled in Oaxaca, so that is the point we will refer to as the start of language contact. Unlike in the Nahuatl empire, however, the elite class did not remain the only group to become bilingual. The cultural diffusion, religious fervor, and professional impetus surrounding it propelled Spanish to be coveted in some way by all kinds of people.

Therefore the key difference between the precolonial imperial structure and the colonial one was the intensity of political pressures exerted on peripheral colonies. This pressure has a causal relationship with contexts for furthering cultural diffusion. For instance, Catholic religion in Mesoamerica spread so widely precisely because it made use of linguistic studies. While the Nahuatl empire did not study and record the substrate languages of its subject ethnic groups, the Spanish empire did, and did so meticulously, because the goal of expressing religious concepts in these new tongues would require expert translation. While the translations studied in this thesis are Spanish representations of Zapotec texts, their conception was predicated on emphasis by the Spanish legal tradition on creating and curating detailed records. Their execution was assisted by the larger ongoing dialogue of bridging the gaps between Zapotec and Spanish, if only indirectly.

It was a realistic understanding of the city-state structured empire that might have led to the Spanish religious officials' decision to encode their texts in indigenous languages; the empire was too vast for Spanish priests to divide up the territory and go about teaching Spanish before converting anybody in a given town. Rather, at the outset of the imperial plan, common people spoke the substratum languages, so those native languages were indeed the languages in which the imperial religious texts should be made available. Spanish policy on language for the national religion fluctuated between anti- and pro-Mesoamerican languages, but the broad effects of the continued debate were that Spanish and Zapotec would coexist in many religious circles (Farriss 2018) [6].

By the late 1600s, bilingualism had spread to the Valley of Oaxaca. This did not mean that there were a majority of common people who were bilingual speakers of both Spanish and Zapotec, but it does mean that such bilingualism existed in at least the elite network of

the towns and was coveted for its power. Elites were required to translate official bureaucratic documents to Spanish. Whether or not they would be punished directly for solely overseeing the creation of Zapotec documents is unsure, but the Zapotec documents would not be useful to a higher court, so it served both local and central imperial interests for important documents to be recorded in a standard language, Spanish. The cultural perception of the day would have upheld Zapotec as an important community language, for daily business and use among friends and family, but at the same time Spanish would possess loftier importance because of the conversations that were held in Spanish: matters of life and death, wealth and land. These conversations are represented in the corpus of last wills and testaments, bills of sale, and legal complaints that this thesis examines.

In Mexico, the term *dialecto* is considered derogatory, so I choose to use the term ‘variant’ or ‘variation’ to describe the same concept as ‘dialect’ does: a language genetically related to a language that has been standardized which has developed differently from that standard language in certain aspects.

### 1.3 Bilingualism

Bilingualism is the phenomena of people acquiring more than one language. The first language, or native language of any speaker, is the one that they grew up speaking naturally with their community. If a speaker acquires a second language, they are considered bilingual. Second languages may be acquired later in life or may be acquired just as early as the first language. For this study, Zapotec will be considered the L1 even if bilingual Zapotec speakers learned Spanish and could be classified as “simultaneous bilinguals,” because the substratum language of the region remains Zapotec and therefore an individual’s exposure to Zapotec in this cultural context outweighed exposure to Spanish (Bhatia 2008:59) [3]. When someone acquires two or more languages, the term ‘second language acquisition’ can be used to describe the process that they underwent when learning their second language (Butler 2012) [4, p. 15]. Phenomena such as code switching can highlight bilingual speech, but in a translation, alternation between the target language and the substrate language would not be expected. How, then, can we know that the translations were done by bilingual speakers of both Spanish and Zapotec? How can we know what their first language might have been?

Bilinguals make use of different grammars, or linguistic frames of reference, for each of their languages. A grammar is, at its most basic, a network of lexical items, grammatical categories, and lexical connections to semantic domains. Instantaneous alternation between the two grammars is called code switching, while grammatical influence from one grammar

to the other is called interference. The documents tell nothing about code switching in the colonial era, because the translators did not switch back to a Zapotec grammar at any time while translating. However, they are filled with evidence of interference, which is exactly what I look for in any language contact effect. The effect, a describable linguistic feature, is what provides the example that I analyze, but the impetus for this effect is always the interference from a Zapotec grammar that has naturally arisen in the bilingual mind.

## 1.4 Language Contact Studies

Research on linguistic contact “stud[ies] the outcomes... when speakers of different languages or dialects share a common geographical space” (Wesley Raymond 2018) [26, p. 163]. Along this logic, a sociocultural linguistic study with syncretic religious lexical items as its data could be a language contact study, too, although its approach would look different from mine. A feature-based search returns data that can be analyzed for a better understanding of the processes in language contact.

### 1.4.1 Thomason & Kaufman 1992

According to Thomason & Kaufman’s seminal work on language contact, the outcomes of language contact can be described as “borrowing” or “substratum influence” (1992) [24, p. 20]. In the case of a borrowing, a foreign lexical item or construction is used in another language, following all the grammatical rules of the language grammar it has entered. This is a clear and sometimes marked phenomenon, but it does not constitute deep structure transfer between the two languages. Deep structure transfers occur under “substratum influence,” which is the second type of language contact effect Thomason & Kaufman describe. While the superstrate, or newly imposed prestige language, may have more political power, the substratum language provides the base language in the local speakers’ developing bilingualism. For instance, Zapotec is the substrate language in this study, because the area’s first language was Zapotec; the people translating the manuscripts spoke Zapotec as their mother tongue.

Language contact effects classified as “substratum influence” are features of the area’s original language that permeate the superstrate language’s structure and change its grammar in more ways than by just adding a new lexical term. In this study, the effects of language contact range from borrowing to potential substratum influence. These categories are important to the research goal of describing the language contact, but they are not the only categories that will be discussed. Because language acquisition is also a result of language

contact, and it propagates transient interference in the studied documents, I will include it in my study by putting forth another binary dividing the potential grammar gaps which result for surface representation variation.

Variation in constructions that exist in Spanish grammar but not in Zapotec grammar often represent idiolect-level bilingualism effects. When the underlying representation calling for Spanish morphology is transformed to a surface representation, the result is either the expected grammatical Spanish construction or a simplified version that passes the writer’s grammaticality judgment due to interference from Zapotec grammatical rules. The transience of this ‘simplification’ variation is evident in its failure to produce the same form consistently or follow any set of rules for when it occurs; it certainly does not persist and trigger language change. Simplification does not fall within Thomason Kaufman’s criteria for language contact, because there is no ‘borrowing’ or feature to transfer; there is a syntactic gap (1992) [24, p. 20].

On the other hand, some variation represents evidence of dialect evolution in the Zapotec bilingual community. Constructions that exist in Zapotec, but not in Spanish, or constructions that are semantically more complex in Zapotec, often require innovation to be expressed in Spanish. This complicating effect works in opposition to that of variation based on simplification. In this case bilingual innovation serves a purpose beyond the idiolect, as the community of bilinguals understands the Zapotec construction instigating the new Spanish construction. Variation created innovatively to aid translation efforts is likely to persist in a predictable pattern, following Spanish rules as a ‘borrowing’ or instigating structural changes in the target Spanish as a ”substratum influence” (Thomason Kaufman 1992) [24].

SIMPLIFICATION VARIATION	INNOVATION VARIATION
Given a Spanish construction	Given a Zapotec construction
a lack in Zapotec grammar	a lack in Spanish grammar
manifests as grammatical incompleteness in Spanish	manifests as an innovation in Spanish

Table 1: Potential directions of language contact interference

Given a Zapotec construction, a comparative lack of representation in semantically or syntactically related Spanish grammar manifests as an innovation in Spanish writing.

The grammatical incompleteness cited as a result of simplification variation notably differs from any narrative stating that bilingual variation is an error, but to a native Castilian Spanish speaker from the Colonial era, the variation would appear to be erroneous. Looking deeper than that, given the data that was culled from these documents, I will examine whether the simplification variation ever followed grammatical rules or appeared more frequently in certain semantic domains. That would bring it into the realm of the categories Thomason Kaufman described (1992) [24].

#### **1.4.2 Labov 1972**

In any language contact study, sociolinguistic consideration is key: the genetic makeup of the languages is never changed simply due to contact; time depth does not go far enough to make internally motivated change a viable explanation, and yet change occurs. All language contact is socially stratified, and the range of its resulting phenomena are a testament to stratification's strength (Labov 1972) [9]. Dialogism, which Goffman, other sociologists, and the philosopher Bakhtin furthered, reminds any scholar to consider the social context surrounding their data, as every utterance takes place in an ongoing social dialogue. In my study, the power dynamics inherent in imperial cultural contact are taken into account, as I have begun to elucidate in the above historical background section.

As his career in sociolinguistics evolved, Labov worked more toward the elucidation of different types of sociolinguistic effects, defining the difference between transmission and diffusion as recently as 2007. It would not be sufficient, in the view of modern sociolinguistics, to state that language change is totally unpredictable due to the social stratification of individual situations. While that would be true to an extent, the creation of categories can help us to dissemble histories of linguistic change. The binary described by transmission (naturally formed variation in an area marked by diverse dialects or languages) and diffusion (borrowing that does not go through natural language learning processes) actually lines up quite well with the spectrum for deep transfer: transmission is evidence of deep transfer, because it demonstrates a speaker's knowledge of the innerworkings of a linguistic system.

#### **1.4.3 Muysken 2005**

[16] One of the most unique concerns in the field brought up by Muysken in his 2005 data-driven, modular analysis of language contact acknowledges Cheshire's elucidation of the boundaries between pragmatic variation (communicative intent) and syntactic variation (equivalent constructions). This work takes a step further to ensure that the discursive value

of the innovation is discussed as evidence for its likely diachronic path, which is an often unfulfilled logic according to Muysken (Muysken 2005) [16, p. 33]

## 2 Corpus

### 2.1 A Closer Look

Imagine that it is 1741 in Oaxaca, Mexico- the pueblo of Santo Domingo del Valle, to be exact. Don Pedro de Zarate finishes translating a bill of sale, an official record of families exchanging land and money. In doing so, he legitimizes a legal act and galvanizes it for the future. In effect: “Yes, this property was and is owned by Zapotec people.” Signed, sealed, and put away for the potential resolution of future disputes.

There’s a discomfort in the need for translation. Zapotec speakers wrote the bill of sale, and it pertains to their Zapotec speaking community. But Zarate knows that not all of the town’s land disputes would be resolved in Zapotec. Translating a Zapotec-language document into Spanish was a necessity if that document was to be used as evidence in a court case.

In fact, Oaxaca’s current linguistic landscape would look strikingly different to Zarate, but it would not shock him, given the imperial dynamic of power that belonged to Spanish even in 1741.

Translating Zapotec documents into Spanish falls within the scope of Zarate’s oversight as cacique, or regional indigenous leader, and in that job a transfer of power and autonomy is implicit. While Spanish elites were in charge of large territories, they were also vastly outnumbered, so the strategy of granting colonial towns a local chief became a common one for furthering colonial expansion (Deagan 2003) [5]. While the translations of Zapotec documents studied in this thesis clearly differ from examples of natural speech used in other language contact studies, I agree with the argument of Hidalgo Mouton in their book about Mexican Spanish’s diversification: “the literary language coexists with the common language while the changes that are generalized in the latter are eventually accepted in the former” (Sánchez 2016) [20]. In other words, and more specifically related to the texts I have studied, it is clear that the conspicuous constructions analyzed as the results of language contact would have been too marked for professional writing if they not had been first ‘naturalized’ in the local speech patterns first.



Figure 3: Zapotec language text in the Archivo General del Estado de Oaxaca [12]

## 2.2 Evidence of Bilingualism

There are 23 translators represented in this corpus from 20 different pueblos, spanning 55 works of translation from Zapotec into Spanish. The following table predicts all of the possible linguistic identities which could have undertaken this job in Colonial Oaxaca.

Translator identities		
TYPE OF TRANSLATOR	L1	ROLE
Type A- Zapotec Bilingual	Zapotec	Town official
Type B- Spanish Bilingual	Spanish	Town official or priest
Type C- Spanish Monolingual	Spanish	Gov't official using dictionary resources to translate

Table 2: What language background did these translators have?

I argue that the majority of the translators in this corpus were bilingual L1 Zapotec speakers, corresponding with Type A above. Type C can be ruled out as a potential majority simply because the translation work of a Spanish monolingual would certainly not be comprehensive. I have found a couple of examples of Type C work in the corpus, but they are

distinctive in that they are incomplete; their translator did not speak Zapotec, so they had to leave parts blank if a dictionary resource did not supply them with the necessary words. In short, we can tell when a Spanish monolingual is translating, and in these texts, it is not the case.

Distinguishing between Type A and Type B requires a variationist perspective. “Sociolinguists tend to see multilingual speakers as actors of social life who draw on complex sets of communicative resources which are unevenly distributed and unevenly valued” (Wei and Moyer 2008, p 14) [25]. While Zapotec and Spanish have different grammars and speakers are aware of many differences between them, natural speech patterns such as code-switching have shown that bilingual people need not swap completely from one to the other. Given lexical decision tasks, bilinguals can often access one language’s data more easily than the other—and not necessarily across the board. Beam de Azcona generalizes that usually substratum influence is structural while prestige languages’ lexical items are taken up by the substratum, and the case is the same for the modern language Sierra Sur Zapotec in contact with Spanish: “Aunque haya influencia léxica y gramatical tanto en español como en zapoteco, las lenguas zapotecas han adquirido vocabulario de español y el español ha cambiado su estructura para converger con la estructura gramatical del zapoteco” or, “Although there has been lexical and grammatical influence in Spanish like in Zapotec, the Zapotec languages have acquired vocabulary from Spanish and Spanish has changed its structure to converge with the grammatical structure of Zapotec” (Aczona In Review p 12). This mutual influence follows patterns based on the social relationship that speakers have to each language and to the expression of identity that comes with each language. Therefore, in the investigation of these documents, I expect to find Type A linguistic identities where a fluent understanding of Zapotec is represented in the completeness of the Spanish translation and where Zapotec grammar arises in Spanish form.

The Spanish conquest of Nueva España is almost two centuries old in 1741. Zarate has already translated several legal manuscripts in his lifetime. Today, we’re lucky to have access to these texts. A linguistic analysis of Zarate’s Spanish translations confirms the historically sound hypothesis that as a cacique he was also a fluent speaker of both Zapotec and Spanish. Zapotec-Spanish bilingualism, as it came about in towns like Santo Domingo del Valle, is first visible among the religious and political elites who had reason to use Spanish professionally. In this piece we will first review the linguistic evidence in Zarate’s work that identifies him as bilingual.

This evidence includes variation, which we will define as alternation between multiple

grammatical forms. The linguistic innovation evident in his work fits with certain features of the Oaxacan Spanish dialect, and I hypothesize that the feature exists due to Zapotec language influence on the development of this dialect.

There are several markers of Zapotec influence in Zarate's works that we can analyze as linguistic variation data. Let's break down the following examples to identify Zarate's use of Spanish as Zapotec-influenced.

## 2.3 Data Collection Methods

I transcribed the documents by viewing high-resolution images of them on the Ticha project website [12]. Examples of variation or confusing constructions were recorded, and I chose to analyze most prominent groupings of variation by grammatical category. This analysis included finding a count for how many times a type of variation did or did not occur out of total potential occurrences. For different constructions that pertain to different semantic domains, the method for counting total potential occurrences changes. For instance, if a lack of noun and article/adjective agreement is the type of variation encountered in a text, I must conversely take note of every single time that noun and article/adjective agreement is done in the expected, standard Spanish manner. However, for the quantifier variation that will later be discussed as a type of conjunction in the bilingual Spanish texts, its semantic scope is understood to be quite specific. So I need not count every conjugation and quantifier, but rather every conjugation of two proper nouns. In the end, while counting tokens did not prove to form the most effective analysis of the variation, it was a good tool for progress because in counting the examples I also recorded all the examples of variation. From there "I compared the use of ... the closest semantic equivalents" to one another and found the most interesting diversity within the variation, not just by looking at the variation as an entire quantitative body (Muysken 2005)[16, p. 45].

## 3 Variation

### 3.1 Agreement

In Spanish, nouns are coded for number and gender. Determiners and adjectives used with nouns must "agree with" or match the number and gender marking of that noun. In Zarate's Spanish texts, though, we often find a surprising lack of agreement (marked with a ):

### 3.1.1 Number

- (1) los cinco sacramento  
los cinco sacramentos  
'the five sacraments' *Spanish* (SDE758T)
- (2) assi son los lindero  
asi son los linderos  
'this way are the boundaries' *Spanish* (SDE746T)

### 3.1.2 Gender

- (3) Las dos pedasos de los dos tierras  
ART.F.PL two piece.M.PL of ART.M.PL two land.F.PL  
'the two pieces of the two lands' *Spanish* (SDE704T)

## 3.2 Simplification

Why would Zarate, a professional Spanish translator, not always follow the expected patterns of number and gender agreement in noun phrases? The intended meaning may have even been well achieved, because one part of the sentence still has plural marking. Agreement variation is found in these manuscripts in the following distributions: between the noun and the adjective, and between the article and the noun, and between an antecedent noun and a pronoun (Heher et al. 2014) [8].

Consider this: in CVZ, number agreement is not required. A number preceding a noun is one way to pluralize something, but in general the form of the noun would not change.

- (4) tobi cueyoo  
'one field' *CVZ* (Munro y Sonnenschein 2007) [15]
- (5) topa cueyoo  
'two fields' *CVZ* (Munro y Sonnenschein 2007) [15]
- (6) teyona cueyoo  
'all three of the fields' *CVZ* (SDE758T)

Another way to construct a plural in CVZ is to put the plural *ra* in front of the noun phrase. Again, the form of the noun would not change in agreement; the added plural *ra* is enough.

(7) *wi*

'guava' *San Lucas Quiavini Zapotec* (Munro et al 1999, p. 27) [14]

(8) *ra wi*

'guavas' *San Lucas Quiavini Zapotec* (Munro et al 1999, p. 27) [14]

As for gender agreement, there are no pronouns in Zapotec that distinguish between masculine and feminine classes as Spanish does (Operstein 2003) and, more importantly, nouns are not marked for gender as they are in Spanish or German [17].

Zarate sometimes uses expected grammatical Spanish forms that show number and gender agreement in his translations, but other times, as in the above examples, he does not. The influence of Zapotec as his first language explains why. His Spanish grammar shows interference from Zapotec grammar—a language that does not utilize number and gender agreement in noun phrases. This variation is a window into his linguistic identity as a bilingual speaker.

A critical question for the simplification process regards the translator's processing the features to which the alternating morphology of Spanish responds. When reading the Zapotec statement, the translator does not necessarily process the features [+masculine] or [+feminine] on any nouns. However, they certainly process the features [+plural] and [+singular] in the original CVZ writing, because these features are not absent:

(9) *Un pedaso de tierra comprada suyo*  
 One.M piece.M of land.F bought.F POSS.M  
 'one piece of his purchased land'

*Spanish* (SDE746T)

While there is plural marking in Spanish and Zapotec plurals do not get marked, a greater conceptual distance exists between the basic existence of [masculine/feminine] gender feature in Spanish, its alternating gender marking on nouns and agreement with articles and adjectives, and the lack of feature existence and marking in Zapotec.

This crucial difference leads us to expect that, if the translator was not a fluent bilingual, we would find more variation in bilingual translators' Spanish gender marking than in their plural marking. We expect non-fluent bilingual translators to form bare representations of

gendered nouns, which in most lexical representations of Spanish would be the ‘default’ of a word. The instances of such bare representations of gendered nouns, however, are not notably fewer than on numbered nouns. This is because, often, the grammatical feature of gender is considered in the formation of a noun phrase, and lexical feature confusion leads to a mismatch rather than complete ignorance of the feature. When there is an error in number agreement, it necessarily overlooks part of the set of rules for pluralizing in Spanish, but it does not often overlook all of them and leave the noun phrase bare.

Just looking at a single document unveils an extremely diverse set of agreement puzzles:

(10) este dho escritura  
este dha escritura

(11) tradusido escritura  
tradusida escritura

(12) plata festivo  
plata festiva

(13) este escritura... tradusido  
esta escritura... tradusida

(14) la compro el mismo  
la compro la misma

(15) otro medida  
otra medida  
(An740T)

There are two types of gender marking variation: 1) feature gap and 2) lexical gap. I would call the mismatch in example (14) feature-sensitive and simply lacking in lexical realization is that those examples exhibit a clear understanding of certain gender features at play, whether the masculine or feminine agreement with *la* or *el mismo* is expected- the correct form does partially appear.

Note that in (12) and (15) there is an apparent gap in the concept of any gender agreement whatsoever, but in (9) above the formation is full of the correct morphology types despite

its exact features being misrepresented. It represents a complete knowledge of productive pluralization practices with partial interference from Zapotec evident.

Because I find both types of variation and not just feature gap variation in the corpus, I operate on the basis that the naturalness of forming Spanish sentences was NOT at question for these L1 Zapotec speakers; their L2 was indeed Spanish. The interference present does actually describe the bilingualism of this community as L1 Zapotec L2 Spanish speakers. The first step of this investigation, then, is fulfilled in confirming the bilingual nature of these texts.

This analysis draws from Lightfoot's 1979 concept of "competing grammars" (Muysken 2005) [16, p. 32]. Rather than code-switching between two completely separate grammars, the speaker is always partially tuned into each grammar and cannot completely separate the naturalness of any given grammar from an utterance. These types of variation do not appear to follow any predictable rules for when agreement should and shouldn't occur, so there is no potential for substratum influence. There is no system being passed onto the Spanish of the region, as the Zapotec effect on the bilingual data is clearly one of simplification. In this study, the increased syntactic and semantic complexity of certain Zapotec constructions gives reason for innovation, but in a case where the substratum grammar is actually sparser than the target language's, there is no such impetus. These language contact effects did not form any lasting trend in the variant development of the region.

Next, we will look at more variation... this time, variation that constitutes innovation and has repercussions for the future development of Spanish spoken in Oaxaca.

## 4 Innovation

Innovation can be understood as a "novel linguistic creation," purposefully or inadvertently generated in speech or writing (Paradowski Jonak 2012, p. 132) [18]. The following variation constitutes innovation because I have not found a previous model for its structure in Spanish, but it occurs as a useful turn of phrase in several Zapotec translators' manuscripts- not just Zarate's. Take a look at this Spanish sentence from the 1741 bill of sale, or its English translation:

- (16) comparesieron      Don Sebastian de Gusman los      dos con su      hijo  
Appeared.in.court Don Sebastian de Gusman ART.PL two with POSS son  
lxitimo.  
legitimate.

‘Don Sebastian de Gusman the two with his legitimate son appeared in court.’

-

*Spanish* (SDE741T)

How many people appeared in court? “The two with” evokes a calculation,  $2 + 1$ , for me. In English and Spanish syntactic rules, *with/con* can introduce a prepositional phrase to modify the entities participating in verb phrases or belonging to nominal phrases. Therefore the lexical item *with/con* leads me to believe that something is being added. However, only two parties are elucidated: Don Sebastian de Gusman and his legitimate son. Looking at the original Zapotec text can provide a better idea of what this means.

## 4.1 Quantifier and Dual

To provide the morphological background needed for an analysis of the Zapotec, this section briefly explains the construction of a definitive quantifier in Zapotec. It meets the “cline of borrowability” for innovation because it offers something unique into the Spanish grammar system [19].

The verbal irrealis prefix *qui* gives aspect to verbs by attaching before the root; all Zapotec verbs require tense/mood/aspect prefix marking (Smith Stark 2008) [22].

Deponency is defined as a “mismatch between morphosyntactic values and morphological form” (Baerman 2007) [1, p. 3]. The verbal morphology *qui-* has a morphosyntactic value, as the irrealis aspect applying to a verb, that it often appears without. For instance, in the following CVZ example from our corpus, the *qui-* prefix attached to a number creates a quantifier, and this construction will later be shown to be extendable into conjugation-like structures.

(17) Ni lao quixopa Pr. Alcald ordin-s...  
that before the.two principal mayor ordinary-pl

‘Here before the two Principal Mayors...’

Zapotec (1715)

Typology for deponent structures includes the diachronic possibility that either an ancient genus developed by way of “homophonous interference” into all of a language’s demonstrated uses unaccounted for grammatical rules, or there were originally multiple lexemes so there is “no issue of syncretism” (Baerman 2007) [1, p. 5]. So where did *qui-* the quantifier prefix come from? It is a difficult question to answer, and this ambiguity makes it likely that the speakers of CVZ had no easy way to translate the particle into Spanish except by their understanding of its distribution.

The distribution of *qui-* in Zapotec aligns with nothing in Spanish, but its particular use as a quantifier that simultaneously conjoins its arguments brought about an innovation in Spanish that modifies a Spanish quantifying strategy. The Zapotec basis for the innovation is exhibited in 12:

Now, creating a full translation of 12 into English will be nearly impossible, but we can see that the conjunction between the two individuals is asyndetic but supported by the semantic value of *quiropa*. By putting *qui-*, the irrealis prefix, on a number, you can get a quantifier in Zapotec that defines the size of the group that is going to be discussed; in this example, *quiropa* constitutes a third person dual. ‘Dual’ will refer to a Zarate’s translations include several examples of *los dos con* as a *quiropa*-based construction. The *con* appears optional, as the construction functions without it in example d.

(18) gue guetao Ant.o de Silva quiropa xinigananij  
 ART deceased Antonio de Silva NOM.two POSS.son  
 ‘(of) the deceased Antonio de Silva the two his son’ *Zapotec* (SDE736)

(19) del difunto Ant.o de Silva los dos su hijo  
 POSS.ART deceased Antonio de Silva ART.PL two POSS son  
 ‘of the deceased Antonio de Silva the two his son’ *Spanish* (SDE736T)

(20) Cuye=nij lao Don Sebastian de Gusman qui-ropa xini-gana=nij  
 ?=3 face Don Sebastian de Gusman IRR-two child-oldest=3  
 ‘Don Sebastian de Gusman the two with his legitimate son appeared in court.’  
 - *Zapotec* (SDE741)

Is this construction a calque? According to Farriss (2018), calques can be defined “broadly” as “any word or phrase translated literally” [6, p. 223]. The *los dos con* construction could not have been a literal translation per se due to its highly flexible, analytic semantic role that deals with so few real-world objects. Beyond its syntactic credibility as an independent construction representing bilingual speech, related translations of *qui-* quantifiers in these texts show that forming the construction was the result of productive processes. “framework where. . . styles and registers are defined as repositories for usage patterns, constructions, etc” One admirable point of Muysken’s framework is the outright assentation that a problem is finding frequent, outer representations of variation even though it should also be noted that there are inner forms of variation resulting from language contact (Muysken 2005) [16, p. 33].

There is some variation within this innovative construction, but it never differs from one of the three word orders described below:

I. ‘A los dos con B’

II. ‘A los dos () B’ -*meaning same as above*

III. ‘los dos A con B’ -*possibly denotes a patient role when naming two people*

The above options constitute innovation because they are Zapotec-influenced and well-formed according to Spanish rules; they constitute transmission rather than diffusion, following the rule that “dialect contact may influence transmission in a dialectally diverse community” but the results of transmission are complex and representative of a dialect’s underlying structure rather than a simplified lexical borrowing (diffusion). I did not include III. in my analysis of the transmission, however, because it is not marked in meaning for Spanish speakers; ie: no monolingual Spanish speaker outside of Oaxaca would find it as unnatural as they might find I. and II. [10, p.].

I. and II. differ from what Spanish already had available in terms of conjunctions:

TANBIEN/TAMBIEN/TAMVIEN	Y	CON
conjoins clauses or NPs	conjoins clauses or NPs	conjoins by adding NPs
tight or loose	tight	tight
additive relation		extends argument of predicate

Table 3: Spanish conjunctions

While the constructions based on *quiropa* are well-formed, the issue of ambiguity arises with the third person dual. It is possible that the open interpretation of this phrase could have incited further structural change in the Oaxacan Spanish variant. Potential starting points and paths for language change will be discussed in the next section.

## 4.2 Transfer

In this section, I will discuss potential reconstructions of the Zapotec towns’ histories regarding language change.

### Origin 1

If, contrary to what I argued above, the construction “*los dos con A*” as a signifier for two parties did not exist in bilingual Spanish at the time of the translation, then its origin would be as a tool for translation. While nothing in Spanish had the form to represent an active pronominal quantifier, it did have quantifiers and conjugation. The translator, likely Don Pedro de Zarate or one of his predecessors, could have calqued the expression together and

used it throughout his career. However, it would be a stretch to assume that one person’s innovation would occur in a vacuum; others likely innovated the same, and if they did not, then this path makes absolutely no sense.

### Origin 2

The construction was likely a part of natural speech before appearing in these testaments. This is supported by the fact that similar productive translations of *qui-* quantifiers are consistent, so the... and that the naturalness of speech innovation and circulation lines up with the high number of token examples we see even within a given town like Santo Domingo del Valle. “Since exact equivalence is usually impossible to achieve, the translation process involves an uneasy balancing act of compromises, trade-offs, and other kinds of accommodation” (Farriss 2018) [6, p. 213]. If the variation we found for these *qui-* quantifiers were less concentrated toward a certain form, Farris’ description of translation techniques would apply to *los dos con*.

I cannot possibly interpret the standard PP pattern with its semantic distribution of conjoining agent-role nouns and possessor nouns as an in-process translator’s calque or a ‘compromise.’ If this had been a calque solely based on *quiropa*, its distribution would have been wider. Its interaction with preexisting Spanish conjunctions in the speech of local bilinguals shaped its destiny as a more specific construction than just a quantifier. This is why I find Origin 2 so much more likely than Origin 1. In the narrative of Origin 1, this variation was already well on its way to becoming a normal part of the local variant in the Valley of Oaxaca.

What path did this construction take to end up disassociated from the Zapotec languages and associated with Oaxacan Spanish variants instead? Useful next steps would include getting data from non-Zapotec towns, or any Oaxacan towns later in the postcolonial period, to see if and when the construction decentralized from being a direct language contact effect. Without a study of that period, I move on to the discussion that according to modern Oaxacan Spanish grammaticality judgments, the construction underwent a change in scope from what the colonial texts include. This will be discussed alongside an updated variant of the CVZ *qui-* quantifier in a modern Zapotec language.

## 4.3 San Lucas Quiavini Zapotec

28,000 people speak a Western Tlacolula Valley Zapotec language, with about 2,000 of those people living in the town of San Lucas Quiavini and an indeterminate number of others

outside the town who acquired the language there (Simons and Fennig 2018) [21]. While the language is endangered, it is widely spoken and understood in the town (Lee 2006) [11]. Dr. Felipe H. Lopez, an award winning Zapotec poet and PhD in Urban Planning, worked with me to explain how the quantifier construction works productively in SLQZ, or San Lucas Quiavini Zapotec [13].

The particles of interest here are the number words and the suffix *-ën*. In another study on quantifiers, it was found that in the Tlacolula Valley Zapotec variant which SLQZ belongs to, “quantifiers can be used with added morphemes that look like regular verbal aspectual prefixes” and the descendant of the *qui-* morpheme was reduced to *y-* (Munro Sonnenschein 2007) [15]. In the examples that Dr. Felipe H. Lopez provided, the same holds true, except with the *y-* weakened in example 16 to the point where it’s barely audible before the rhotic phoneme [13].

(21) We went to the restaurant.

Byoën rrestauran.

(22) The two of us went to the restaurant.

(y)ropën byoën rrestauran.

y-rop-ën byo-ën rrestauran

IRR-two-3p.prox restaurant

(23) The three (of them).

ygyontën.

y-gyon-tën

IRR-three-3p.prox

(24) The four (of them).

Ydyeaptën.

y-dyeap-tën

IRR-four-3p.prox

(Dr. Felipe H. Lopez, 2018) [13]

The construction is not only productive in the Zapotec speaking community, but also retains its grammatical structure. “There are some aspects of the syntax of numbers (and other quantifiers) in TVZ that make them seem slightly predicate-like. . . quantifiers, like verbs, are required to have an associated nominal ‘argument’” (Munro and Sonnenschein 2007) [15, p. 4]. The *-ën* suffix is required above to make those examples grammatical; that creates the parenthetical ‘of them’ which I did not attempt to elicit but was required

by Dr. Lopez for the construction's naturalness. Comparing this grammatical rule with the testaments, we find that they followed it both in Zapotec and Spanish; in the Spanish version, one name is always stated after the *los dos con* and it would not be grammatical to leave *con* without an argument.

#### 4.4 Modern Oaxacan Spanish

In Oaxaca, the *los dos con* construction is still used today to express a dual.

- (25) Vimos la película los dos con Esme.  
We.saw the movie the two with Esme.  
We both saw the movie (Esme and I). (Dr. Felipe H. Lopez, 2018) [13]

The interpretation that you access when reading the sentence is determined by your lexicon, or your mental word inventory, and how the lexicon connects to the syntactic rules of your native language. I have surveyed some native speakers of modern Spanish, and the interpretations are split. Is this a grammatical sentence that means two people went to a movie, or is it a slightly confusing structure that means 2 + 1 people went to a movie?

Without having conducted a formal study, I can't be certain, but it seems that the people who interpret 2 people are actually speakers of a Spanish variant, who may on a structural level see the *los dos con* construction in its totality as a second person quantifier.

In Oaxaca, the *los dos con* construction is still used today to express a dual:

- (26) Fuimos al cine con Juan.  
We.went to.the movies with Juan.

This sentence can have two potential interpretations:

- (i) We went to the movies with Juan (3+ total people).
- (ii) We went to the movies, Juan and I (2 total people).

The interpretation that you access when reading the sentence is determined by your lexicon, or mental word inventory, and how it connects to the syntactic rules of your native language. I have informally posed this interpretative question to some native speakers of Spanish while sharing this research in Philadelphia, Los Angeles, and Oaxaca; their intuitions are not the same. Some people can access both meanings, and some people can access only the first.

Those who can access both interpretations appear to be speakers of a Zapotec influenced Spanish variant. As this *los dos con* phrase became common in the local dialect, more speakers would have grown accustomed to its role as a dual and incorporated it into their

structural understanding of the language. A long term impact may be that, today, adding *con* to a sentence modifies the subject by specifying or naming one of the parties included, not necessarily by adding anything.

In Bermúdez-Otero’s conclusion in Baerman’s 2007 collection of articles on deponency, he discusses language change creating productive morphology that has an irregular, difficult to explain usage [1]. Perhaps in the course of language change, *los dos con* underwent “further detachment” from *quiropa* simply because the group of speakers enlarged to include non-speakers of Zapotec. Bermúdez-Otero describes reasons for deponency as “developments” that can include a suite of related constructions in the language causing a semantic shift, or the original syntactic system gradually weakening in its rigidity [1, p.2]. Future studies investigating Oaxacan Spanish’s development can respond to this open question further. Under the hypothesis that the structure broadened over time to include first person as well as third person, I determine that the quantifier construction had been deep transferred to Spanish as substratum influence (Thomason & Kaufman 1992) [24]. The construction could also be a non-canonical transfer that does not signify deep transfer or syntactic influence, but only if it was used in a purely lexical fashion; both the Colonial variation in form and the modern day state of the dual proffer a deeper structural transfer. As far as I have seen, it is not accessible to speakers as a variation option to take the place of other lexical conjunction words in Spanish, like *y* or *también*, in domains outside of specifically describing people groups. It carries the syntax of quantifier embedding from Zapotec, but its semantic role consolidation is owed not to Zapotec but to the social context of its origin in bilingual communities. The legal texts only used the construction in sentences where the conjoined group acted as agent or possessor; these semantic roles are coded for people to fulfill easily, and so the construction continued to develop in the direction of describing people groups.

Bilingual Zapotec-Spanish speakers like Don Pedro de Zarate mapped *quiropa* onto Spanish and developed a novel part of Oaxacan Spanish grammar, an innovation, that persevered in the region over several centuries.

## 5 Diachronic analysis

Of course, the data from the Colonial era alone are not sufficient to reach a conclusion about the development of any one feature in the Oaxacan Spanish dialect. A relevant research question to philologists is, How can I represent this data in a way that reveals trends, or the processes and spaces in between what is known? The only way to make these findings

more salient is to clarify which parameters we are using and look at them in relation to one another, which is greatly assisted by data visualization tools.

The Changing Role of Innovative Quantifier Conjunctions



Figure 4: The corpus shift in distribution according to conjoined elements' roles

The quantifier dual in bilingual Oaxacan Spanish does not belong in every semantic domain; this is not only suggested by the data, but by the very discursive motivation to innovate such a construction based on Zapotec quantifier conjunction; if the intended meaning is different enough to be innovated separately from what Spanish already had in the way of conjunction strategies, then of course it should have a discrete semantic role when mapped onto Spanish. Interestingly, the semantic role evoked by the *los dos con* construction appears to have changed within the short time depth represented by our texts, and this supports my hypothesis that it broadened over time. In the beginning, the legal texts only used the construction in situations of possession, to describe groups of possessors whose land or money was being discussed. Then, new meaning arose and became more widely used than the first: the bright blue segments represent examples where the quantifier dual did not describe a group of possessors in the sentence, but did describe the party occupying the sentence's agent role.

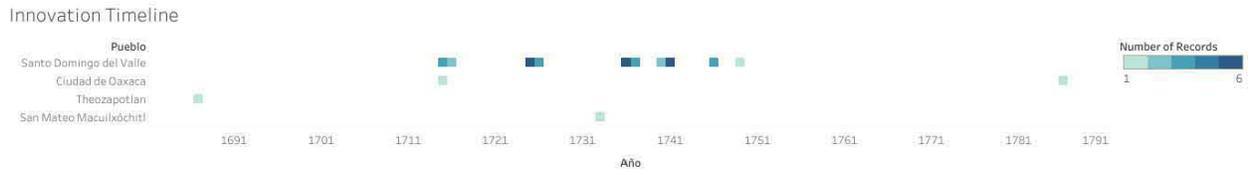


Figure 5: Timeline showing which towns contributed to quantifier dual tokens

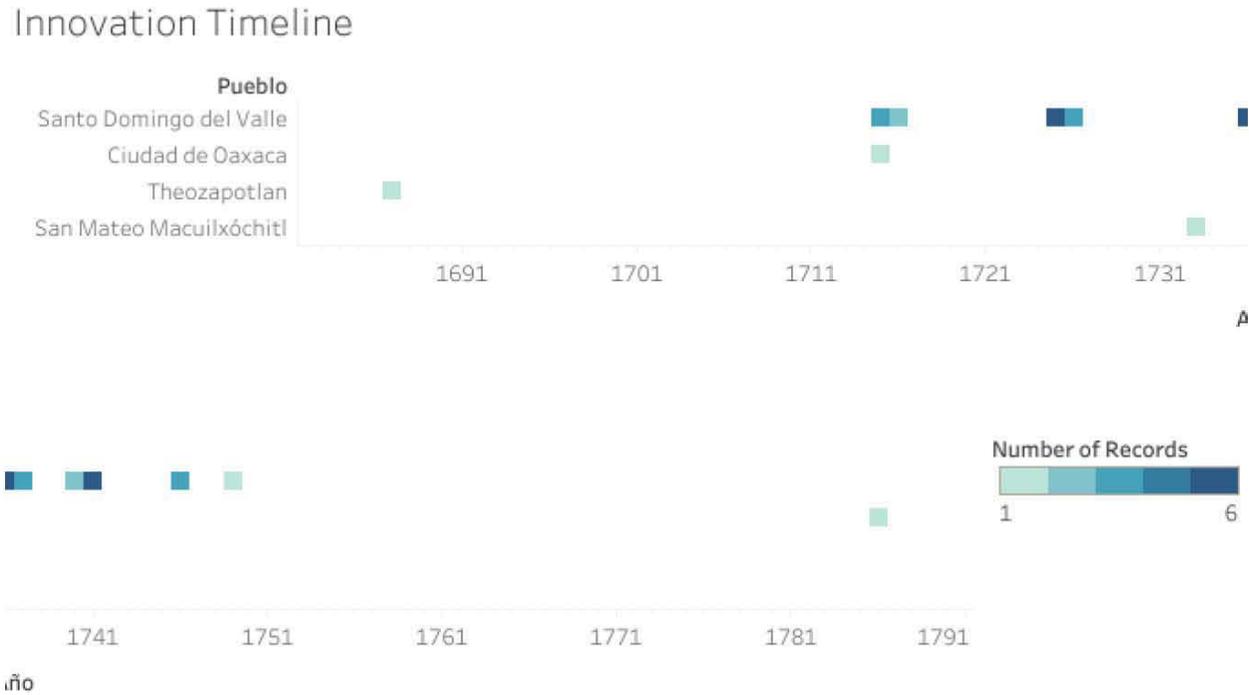


Figure 6: The above timeline zoomed in

As the timeline shows, the most impactful region for the *los dos con* construction was Santo Domingo del Valle, and the peripheral effects are visible as well in towns like San Mateo Macuilxochitl. What is most mysterious is the fact that Theozapotlan had such an early token of the construction. That Theozapotlan document represented by the square seems isolated, but it could have been accompanied by more tokens in other documents- our corpus does not have more than one document from Theozapotlan.

## Translators' Examples of Innovation

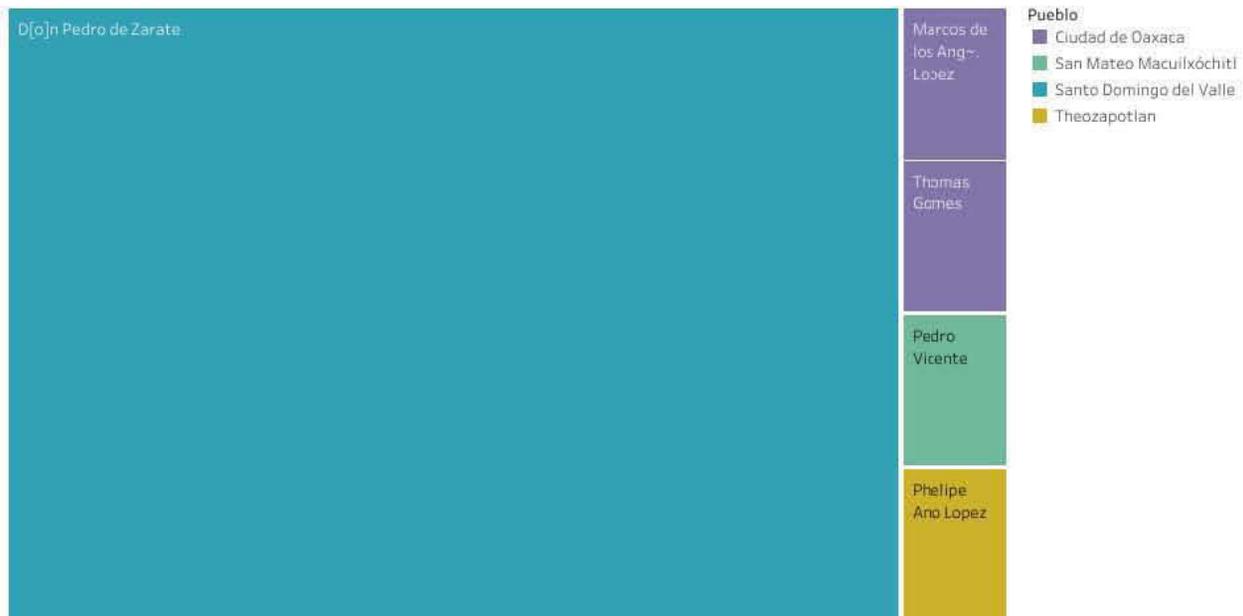


Figure 7: Individual translators' contribution to quantifier dual tokens in corpus

The individual translators who created the written models for the quantifier dual variation are visualized here. While Don Pedro de Zarate had by far the most examples of this construction in his works, it does not mean that the construction was necessarily more natural for him than it was for those in neighboring towns; we only have access to the legal documents, and in those the construction only appeared in contexts discussing two named persons who had identical semantic roles in a sentence. Therefore, if the sentence was about someone selling their house to someone else, *los dos con* would never appear unless two people were on the selling or receiving end, and both were named to a certain degree. I believe that Don Pedro de Zarate's writing style was simply the best vessel for these types of sentences to arise, and the fact that the construction appeared at all in other towns is just as striking to me, because it was likely a part of natural speech first no matter where it originated.

## 5.1 Proposed Diagnostic Tools

### 5.1.1 Interference Suggests Bilingualism

The diversity of data shown in this thesis as interference from Zapotec is fantastic to behold, but difficult to analyze meaningfully. One useful analysis of this variation is to use it as

a diagnostic for bilingualism in reconstructions of history that aim to amplify the Zapotec voice. If these were documents written by Spanish friars, they would not have such constellations of agreement puzzles caused by bilingual interference. While I could not arrive at a recommended quantitative threshold for variation as a diagnostic for bilingualism (because a lack of interference variation examples in one text does not necessarily mean an author will also lack variation in another one of their texts), the linkage between bilingualism and agreement variation has been well established in this corpus.

### **5.1.2 Transfer spectrum application**

By drawing one focus from each of the seminal works I studied (Thomason & Kaufman: lexical versus syntactic impact of interference; Labov: naturalness of learning any given type of variation in social context; and Muysken: discursive motive for innovating variation), I was able to build a nuanced analysis of the variation exemplified in the corpus of bilingual Zapotec-influenced documents. For these documents, nothing less would have been appropriate, because the prestige associated with the legal realm and Spanish contrasts so boldly with the Zapotec features present in the data, and those thematic comments alone bring up modular concerns outside of the features themselves. I think that the next step in this framework would be to stretch the transfer spectrum across other dimensions to include all of the categories brought up in this thesis, from transmission and diffusion to borrowing and substratum influence. Locating different types of attested variation from diverse communities on this type of a conceptual graph would build good grounds for linguistic variation comparison.

### **5.1.3 Data visualization impact on findings**

The impact that applying data visualizations had on my research was not only consequential but massively so. The patterns that emerge when one factor of a complex construction is highlighted (for instance, the alternation of distribution in possessor nouns or agent nouns) are made far more accessible with tools such as the one I used, Tableau. Tableau has a bit of a learning curve, but is accessible to students for free and has the useful ability to export its creations online for distribution. I predict that timelines concerning specific towns and graphs describing the identities of translators would be the most accessible and popular for the interested non-linguist community online. For the theoretical linguists, I would anticipate that these data visualizations would only be more useful for projects with bigger corpora.

## References

- [1] Matthew Baerman, Greville G Corbett, Dunstan Brown, and Andrew Hippisley. Dependency and morphological mismatches. 2007.
- [2] Rosemary G. Beam de Azcona. El sustrato zapoteco en el español de la sierra sur. In preparation.
- [3] Tej K Bhatia and William C Ritchie. *The handbook of bilingualism*. John Wiley & Sons, 2008.
- [4] Yuko G Butler and Kenji Hakuta. Bilingualism and second language acquisition. *The handbook of bilingualism*, pages 114–144, 2004.
- [5] Kathleen Deagan. Colonial origins and colonial transformations in spanish america. *Historical Archaeology*, 37(4):3–13, 2003.
- [6] Nancy Farriss. *Tongues of Fire: Language and Evangelization in Colonial Mexico*. Oxford University Press, 2018.
- [7] Martha James Hardman-de Bautista. The mutual influence of spanish and the andean languages. *Word*, 33(1-2):143–157, 1982.
- [8] Brook Danielle Lillehaugen Michael Galant Heher, Anneke. El bilingüismo zapoteco-español temprano: evidencia de documentos coloniales. In *Coloquio sobre Lenguas Otomangues y Vecinas VI: Mario Molina Cruz, Oaxaca*, 2014.
- [9] William Labov. *Language in the inner city: Studies in the Black English vernacular*, volume 3. University of Pennsylvania Press, 1972.
- [10] William Labov. Transmission and diffusion. *Language*, 83(2):344–387, 2007.
- [11] Felicia Lee. *Remnant raising and VSO clausal architecture*, volume 66. Springer, 2006.
- [12] George Aaron Broadwell Michel R. Oudijk Laurie Allen May Plumb Lillehaugen, Brook Danielle and Mike Zarafonetis. Ticha: a digital text explorer for colonial zapotec, first edition, 2016.
- [13] Dr. Felipe H. Lopez, 2018.
- [14] Pamela Munro and Felipe H Lopez. *Di'csyonaary X: tè'e'n Dii'zh Sah Sann Lu'uc: San Lucas Quiavini Zapotec Dictionary*. Chicano Studies Research Center, 1999.

- [15] Pamela Munro and Aaron Huey Sonnenschein. Four zapotec number systems. 2007.
- [16] Pieter Muysken. A modular approach to sociolinguistic variation in syntax. *Syntax and variation: Reconciling the biological and the social*, pages 31–53, 2005.
- [17] Natalie Operstein. Personal pronouns in zapotec and zapotecan. *International journal of American linguistics*, 69(2):154–185, 2003.
- [18] Michał B Paradowski and Łukasz Jonak. Diffusion of linguistic innovation as social coordination. *Psychology of Language and Communication*, 16(2):131–142, 2012.
- [19] Gillian Sankoff. 25 linguistic outcomes of language contact. *The handbook of language variation and change*, page 638, 2002.
- [20] ISRAEL SANZ-SÁNCHEZ. Chapter seven a diachronic perspective on latin american spanish verbal morphology: Reassessing the role of koineization israel sanz-sánchez. *Diachronic Applications in Hispanic Linguistics*, page 239, 2016.
- [21] Gary F. Simons and Charles D. Fennig. Ethnologue: Languages of the world, twenty-first edition, 2018.
- [22] Thomas C Smith Stark. La flexión de tiempo, aspecto y modo en el verbo del zapoteco colonial del valle de oaxaca. *Memorias del Coloquio Francisco Belmar*, pages 377–419, 2008.
- [23] Jorge A Suárez. *The mesoamerican indian languages*. Cambridge University Press, 1983.
- [24] Sarah Grey Thomason and Terrence Kaufman. *Language contact, creolization, and genetic linguistics*. Univ of California Press, 1992.
- [25] Li Wei and Melissa G Moyer. *The Blackwell guide to research methods in bilingualism and multilingualism*. John Wiley & Sons, 2009.
- [26] Chase Wesley Ramond. On the relevance and accountability of dialect: Conversation analysis and dialect contact. *Journal of Sociolinguistics*, 22:161–189, 2018.