CONSIDERATIONS IN THE CREATION OF AN ELECTRONIC DATABASE FOR COLONIAL VALLEY ZAPOTEC

GEORGE AARON BROADWELL
University of Florida

BROOK DANIELLE LILLEHAUGEN
Haverford College

ABSTRACT. There is a rich corpus of texts written in Zapotec during the Mexican colonial period that remains relatively understudied. The nature of the corpus poses significant challenges to would-be readers; for example, the texts were written using the Roman alphabet with few standardized spelling conventions, resulting in a large number of homographs and a wide range of variation in the spelling of any particular lexical item. To facilitate access to the corpus we developed an interrelated database of Colonial Valley Zapotec texts and morphemes using Fieldworks Language Explorer (FLEx). Here we describe the issues involved in creating this database and summarize the current results, evaluating the benefits and challenges of using FLEx for this type of corpus in order to contribute to a growing conversation in methodologies in digital philology.¹

Keywords: Zapotec, philology, databases, lexicography

¹ We thank the editors of the journal and the anonymous reviewers for their feedback in the preparation of this article. Preliminary versions of this work were presented at the Quinto Coloquio Sobre Lenguas Otomangues y Vecinas (COLOV) Antonio de los Reyes (Broadwell & Lillehaugen 2012) and the first International Conference on Mesoamerican Linguistics (Broadwell & Lillehaugen 2013).

We are grateful to Michel R. Oudijk (UNAM) for sharing his time and resources, including the Smith Stark database. We are especially grateful to Pamela Munro, Kevin Terraciano, Michael Galant, John Foreman, Aaron Sonnenschein, and the late Thomas Smith Stark for helping us understand the grammar of Colonial Valley Zapotec. We owe our thanks to the other members of the UCLA Zapotexts research group, led by Munro and Terraciano, and the Seminario de zapoteco colonial at the UNAM in Mexico City led by Smith Stark: Bety Cruz, Uliana Cruz, Christina Esposito, Xochitl Flores-Marcial, Marita Martin, Olivia Martinez, Mercedes Montes de Oca Vega, Julie Morgenlender, Maria Ornelas, Rosa Maria Rojas Torres, Diana Schwartz,
1. Background and the Colonial Valley Zapotec Language. Linguists, anthropologists, and historians are fortunate to have access to printed and handwritten texts from the Mexican colonial era written in Valley Zapotec. These materials include administrative documents written by native speakers as well as religious and linguistic texts described further in §2. However, there are challenges to using and interpreting these materials, which we present in §3. For example, the main dictionary of Colonial Valley Zapotec, Cordova’s Vocabulario (1578b), has Spanish headwords with Zapotec translations. There is no Zapotec to Spanish section of the dictionary or any other way for a user to look up an unknown Zapotec word. There are modern materials based on the Vocabulario (e.g. Whitecotton & Whitecotton 1993, Smith Stark et al. 1993, and Oudijk & Miceli 2015) that, in contrast to the Vocabulario, allow a user to look up a Zapotec word, but while these are useful tools, they underrepresent the full range of lexical spelling variation found in the corpus, thus being of limited use to researchers wishing to consult or translate texts in the Colonial Valley Zapotec corpus.

To facilitate access to and translation of the corpus, we developed a Colonial Valley Zapotec database using the freely available Fieldworks Language Explorer, version 8 (hereafter ‘FLEx’).¹ We describe the benefits of using FLEx for this type of database in §4 and some challenges we faced in using this software for this corpus of documents in §5. In §6 we give an example of how the database can be used to inform linguistic analysis, showing some sample lexical results. Finally, we offer our conclusions in §7, hoping that this discussion will be useful not only to Zapotecanists, but also to others considering using FLEx for corpora of historical manuscripts.

Zapotec (Otomanguean) is a large family of related languages, spoken in Oaxaca, Mexico (ISO 639-3 code [zap], Lewis et al. 2014). There are texts written in Zapotec from the Mexican colonial period in several Zapotec languages; our work focuses on those written in Valley Zapotec varieties. We refer to the Valley Zapotec language as attested in this colonial corpus as Colonial Valley Zapotec (CVZ).

The modern Zapotec speaking areas of the Valley of Oaxaca are quite linguistically complex. Almost every pueblo has a distinct variant of Zapotec. It is likely that there is dialect diversity within the CVZ corpus. The work that we describe here is the beginning of a corpus which is detailed enough to identify systematic dialect differences in CVZ and to distinguish such dialect differences from other kinds of idiosyncratic and orthographic differences.

and Lisa Sousa. We also owe our thanks to Michael Galant for discussion of the Colonial Spanish and to Flora Berklein, Alex Delgado, Emily Drummond, Rachel Elizabeth Weissler, and especially May Plumb for their research and editorial support.

As always, we are thankful to our Zapotec language teachers who generously share their language with us, especially Roberto Antonio Ruiz and Josefin Antonia Ruiz (from Tlacolula de Matamoros), Felipe López Hernández (from San Lucas Quiaviní), and Luisa Martínez (from San Dionisio Ocotepec). ¡Xtyozën yuad! All errors are, of course, our own.

¹More information on FLEx and software downloads can be found at http://fieldworks.sil.org/.
The exact relationship between Colonial Valley Zapotec and modern Valley Zapotec varieties is currently unclear; however, it is likely that CVZ is the direct ancestor of the Zapotec languages in the Tlacolula Valley, including San Lucas Quiaviní Zapotec (SLQZ, Munro & Lopez et al. 1999; Munro et al. 2007; ISO 639-3 code [zab], Lewis et al. 2014).

2. **The Colonial Valley Zapotec Corpus.** The materials in the CVZ corpus share many overall properties with Spanish colonial texts from other parts of the Americas. Priests and colonial administrators chose to translate similar materials (e.g. catechisms, confessional) into the various indigenous languages of the empire. Thus the genres of texts in Colonial Zapotec often match those found in other Mexican languages, such as Nahuatl and Mixtec. Nonetheless, we take the time to describe the Colonial Valley Zapotec corpus as many Zapotecanists, not to mention linguists in general, are unfamiliar with it. Moreover, some of features of the texts are language specific, or have language specific ramifications.

The corpus of materials available in Colonial Valley Zapotec can be roughly divided into two sub-parts: (i) texts produced under the auspices of the Catholic Church and (ii) texts written by native speaker scribes for the Zapotec community. We are following Lockhart here, who makes the distinction between ‘Nahuatl writing [that] had the purpose of communication among indigenous people’ and writing ‘produced for Spaniards’ (1990:95). While both sets of texts have a plethora of cultural, historic, and linguistic data to offer, they were created under different conditions, which is relevant to interpreting this data. The texts produced through the Church, including Feria’s doctrine of Catholic faith (1567) and Cordova’s dictionary (1578b), are works that translate Spanish material into Zapotec and were produced for Spaniards. Both non-native and native Zapotec speakers participated in the creation of these texts. For the most part, these are long, typeset texts, some of which are bilingual. The shorter, hand-written manuscripts created by native speaker scribes form the second sub-part of the corpus. In contrast to the first part of the corpus, these materials are not translations: they are monolingual texts, primarily administrative in nature, written by native speakers for use within the Zapotec community.

The context and linguistic methodology involved in the creation of the texts is important in understanding how to interpret them. For this reason, we do not necessarily take data from any of these texts as authoritative exempla of CVZ and we proceed with special caution with the texts produced through translation or elicitation. Rather, we take the full range of these texts as the data necessary for a better understanding of the grammar and lexicon of CVZ. Each of these sub-parts of the corpus is described more fully below.

There are a variety of colonial documents written in CVZ through institutions within the Catholic Church—e.g. a dictionary, a grammar, doctrines of faith, collections of parables, and confessional; see Smith Stark 2007 for more information.
concerning these types of documents. The most well known and widely used of the CVZ texts is Cordova’s 1578 Vocabulario en lengva çapoteca, an extensive dictionary containing approximately 30,000 entries, organized from Spanish to Zapotec (Smith Stark 2009). The Vocabulario is a great resource for those who wish to learn more about the language and culture of colonial Valley Zapotec communities. However, there are limitations to the usefulness of the Vocabulario as a tool for reading and understanding other texts written in CVZ. As the Vocabulario consists of Spanish headwords with Zapotec translation equivalents, it cannot be used to find the Spanish meaning of an unknown Zapotec word. In addition, the morphological characteristics of Zapotec complicate the presentation of Zapotec citation forms, especially of the verbs, in any dictionary—even modern ones (Munro 2005:97). Because most verbs in Zapotec languages occur with an obligatory aspect prefix, it is generally necessary to choose some aspectual form to serve as a citation form, and users of the Vocabulario must learn how to deduce the root and other aspectual forms of a verb from the citation form, which we discuss further in §5.2.

To further complicate the situation, although the Vocabulario is generally organized alphabetically, there are sections that are not strictly alphabetical, with certain conventions superseding alphabetical order. This can be observed in Figure 1, where different forms are organized by roots and not by strict alphabetization; for example, towards the bottom of the first column, dexar como el que muere que dexa la hazienda is followed by dexada ser assi; likewise, halfway through the second column, dezir o hablar no con nadie is followed by di, di. digo a uno, and further down ‘dezir’ appears again in dezir o hablar algo a otro. This makes the location of any particular lexical item more difficult to predict.

Cordova also directed the creation of a grammatical description of CVZ, his 1578 Arte en lengva zapoteca. The Arte contains valuable linguistic data but is challenging to use effectively, even more so than the Vocabulario. One major obstacle is the organization and presentation of information: the Arte is organized in the classical style of grammar, and so describes Zapotec in terms of Latin grammatical categories (Smith Stark 2009). For example, Cordova describes the ‘declinations’ of nouns in Zapotec in terms of nominative, genitive, dative, accusative, vocative, and ablative cases (1578a:1v). While Latin has morphological case, these categories are not very useful for understanding Valley Zapotec, which does not have morphological case. Obviously, Zapotec grammar is different from that of Latin. While overall it is impressive how well Cordova dealt with a language that was so typologically different from any language he had ever been exposed to, the Latin framework obscures the details of the grammar of Zapotec itself. Users are left wondering where they might look for data on any particular aspect of Zapotec grammar.

---

2 Thanks to Robert Germany for thoughtful suggestions on this topic.

3 This section of the Arte can also be consulted online through the Ticha project (Lillehaugen et al.). It can be found in §2.2 Declination, and can be viewed as a transcription of the original or as a version with modernized Spanish spelling and word boundaries from the Arte page: http://ticha.haverford.edu/en/arte_original.
Moreover, while there were certain writing conventions, there was no standardized way to write Zapotec in the colonial period (Broadwell 2010, 2015b); Smith Stark 2003 describes the spelling conventions in Cordova’s Vocabulario. Cordova often writes the same word in two (or more) different ways, even within the same text. These words may be written in additional different ways in other materials, such as the Doctrina and the handwritten administrative manuscripts described below. Part of this variability in spelling is likely due to the fact that certain phonemic contrasts in

FIGURE 1. An example of the use of alphabetical order in the Vocabulario (Cordova 1578b:138r)
Zapotec are difficult to represent using the Latin alphabet, such as contrasts in tone and phona- tion as well as fortis and lenis contrasts in consonants. This is similar to Mixtec colonial texts, where there was ‘not a consistent designation’ for the high back unrounded vowel, which is not present in Spanish (Terraciano 2001:91). Readers familiar with Mesoamerican philology may know that many of the variant spellings and scribal traditions for Nahuatl have been established (e.g. Sousa, et. al., 1998:46; Lockhart 1999:188–189, Lacadena 2008), but these results do not provide much guidance for interpreting colonial Zapotec texts, given the very different phonological and morphological properties of the language. As Nahuatl philologist Canger states, ‘Nahuatl is phonetically a simple language, so the Spanish had only a few problems in writing it, and there are not many disagreements in the orthographies that they used’ (1990:108). Zapotec, on the other hand, is not a phonetically simple language and there are many unresolved issues in terms of the representation of the phonology of Zapotec in the colonial texts.

Spanish orthography of this period can be extremely variable (e.g. Catalan 1989, Lapesa 1988), and as a result the Spanish in these texts is not always written in a completely predictable manner, especially in the case of certain phonemes, such as /u/ and /s/ or letters and letter combinations including < b, c, ç, g, gu, h, i, j, qu, ss, s, u, v, x, y, z >. Lockhart notes, for example, that ‘in early modern times alternate [Spanish spelling] conventions existed and might be used by the same writer… as in ‘iglesia, yglesia’ or ‘saber, sauer’ (1992:580); for a fuller discussion of the Spanish used in Cordova we refer the reader to Smith Stark 2003. Important for our purposes here is that the variability in orthography can make searching for a Spanish word more complicated; one might need to search for Spanish words in the Vocabulario under multiple spelling possibilities.

The final text created within the auspices of the Catholic Church that we discuss here is primarily a religious document, as opposed to the Vocabulario and Arte, which are linguistic descriptions. The Doctrina christiana en lengua castellana y capoteca by Feria (1567) is a doctrine of the Christian faith. Each page of the Doctrina is organized into two columns—the left column is written in Spanish and the right in CVZ. A full transcription of the Spanish in the Doctrina can be found in Resines 2002, which also contains an analysis and comparison of this with other contemporary Catholic texts. While primarily religious in content, the structure of the document lends itself to linguistic investigation. However, the user should be careful in assuming that the two columns of text are always entirely parallel: we have, on more than one occasion, found significant differences between the Zapotec and Spanish versions of the text, such as a passage on folio 64v in which God is compared to an abusive husband in the Spanish, but the Zapotec compares the abusive husband to those under God’s judgment (Broadwell 2013). Part of this passage is seen in example 3 in §4.3, where

4 See Lillehaugen et al. 2014 for discussion of using statistical methods to better understand orthographic tendencies in the CVZ corpus, especially the orthographic representation of stressed vowels and affricates.
pulling a wife’s hair is given as an example of normative behavior by Spanish husbands of the 16th century.

These printed materials are consulted and used by academics and non-academics alike. Cordova’s *Vocabulario*, for example, is cited by linguists (e.g. Smith Stark 2003, Rojas Torres 2009), historians (e.g. Tavárez 2010), and archaeologists (e.g. Marcus & Flannery 1994, Zulauf 2013). Following Lockhart’s model of ethnohistory (e.g. 1992, 2001) it is clear that our understanding of Mesoamerican history will benefit from the reading of texts in Mesoamerican languages.

In addition to these early printed materials, there are at least 400 handwritten documents created by native Zapotec speaking scribes in the Valley Zapotec language during the Mexican Colonial era that have survived to date (Michel Oudijk p.c.). The majority of these documents are wills, but there are also bills of sale and formal complaints. Many of these documents are held in the *Archivo General de la Nación* of Mexico, but there are also documents in other archives, including the *Archivo General del Poder Ejecutivo de Oaxaca*. Over 100 of these documents can be consulted online through the Ticha project (Lillehaugen et al.; http://ticha.haverford.edu/en/handwritten).

Unlike the more widely used (primarily bilingual) printed materials produced through institutions associated with the Catholic Church, these (monolingual) archival documents are used almost exclusively by specialists in the language, addressing a range of questions of interest to the fields of typology, theoretical linguistics, and Zapotecan linguistics including negation (Anderson & Lillehaugen 2016), conjunction (Broadwell 2002, Plumb 2015), language change (Broadwell 2015a), possession (Galant 2011a, 2011b), locatives (Lillehaugen 2006), grammaticalization (Lillehaugen 2014), the realization of subjects (Munro 2015a), relative clauses (Munro 2002b), and tense/aspect marking (Smith Stark 2004, 2008, Broadwell 2015a).5

When we refer to potential readers of the Colonial Zapotec texts we are thinking broadly of a range of potential users— from Zapotec specialists, to academics in other fields, to non-academics— who might want to understand these documents. A lack of understanding of the conventions of the text and of the structure of the language can result in a superficial use of the materials and an underutilization of the richness of information contained within them. Resines’ (2002) very useful analysis of Feria’s *Doctrina*, for example, would have been much more powerful if it had been able to take advantage of the Zapotec text as well as the Spanish. One benefit of the FLEx database is that it can be exported as XML, making it convenient to transform the data so it can be used for a variety of purposes, such as sharing the lexicon publicly on a website.6

---


6 The pop-up morphological analysis of the Zapotec words in Cordova’s *Arte* on the Ticha
3. CHALLENGES IN TRANSLATING AND UTILIZING THE CVZ MATERIALS. There are a variety of modern resources that exist for reading and translating the CVZ texts, which we describe in §3.1. However, the documents remain challenging to read and translate, despite these materials, partially because of the nature of the CVZ materials themselves. We give an overview of these challenges in §3.2 and provide a case study, detailing the process of translating a small passage in §3.3. We propose that a database of CVZ would address some of these challenges in utilizing CVZ texts and we describe our current database in §4.

3.1. MODERN RESOURCES. There are three important modern resources that are particularly helpful in the reading of CVZ. Whitecotton & Whitecotton 1993, a Zapotec-Spanish wordlist, is an index of the Junta Colombina (1893), written by an anonymous author circa 1793, though not published until much later. Originally published as a book, it is now available as a searchable PDF and thus can be very useful to a researcher looking for the meaning of an unknown CVZ word.

The other important modern tools are both based on Cordova’s Vocabulario. Smith Stark et al. 1993 is an electronic index of Cordova 1578b organized from Zapotec to Spanish. The file was first created in Corel Word Perfect then later converted to Microsoft Access by Oudijk and his colleagues. These electronic versions were key to our creation of the CVZ database described in §4. Oudijk then further refined that database, creating an online, searchable version (Oudijk & Miceli 2015; http://www.iifilologicas.unam.mx/cordova). As Smith Stark et al. 1993 is an electronic archive, it is, of course, also searchable. These are incredibly useful tools for the readers of CVZ materials and are a critical step forward in enabling the reading and translation of these documents.

Although Whitecotton & Whitecotton 1993, Smith Stark et al. 1993, and Oudijk & Miceli 2015 are all more useful than the original material on which they are based, they are limited by the fact that they only contain word forms such as they are written in Cordova 1578b (Smith Stark et al. 1993, Oudijk & Miceli 2015) or the Junta Colombina (Whitecotton and Whitecotton 1993). They do not reflect the full range of colonial orthographic practices. In CVZ material, we often come across words written in forms other than those that appear in either Cordova 1578b (and thus Smith Stark et al. 1993 and Oudijk & Miceli 2015) or the Junta Colombina (and thus Whitecotton & Whitecotton 1993). In addition, these resources neither include all the Zapotec words in the corpus (there are words in the texts that do not appear at all in the Vocabulario or Project’s website (http://ticha.haverford.edu/en/arte, Lillehaugen et al.) is an example of the way lexical information in a FLEX database can be connected to larger digital projects for public dissemination.

7 The exact relationship between the Junta Colombina (1893) and Cordova’s Vocabulario is complicated and contested (Oudijk p.c.); it may be an attempt by an eighteenth century author re-elicit Cordova and respell the words as they sounded at the time. See Broadwell (2010) for further discussion.
the *Junta Colombina*) nor all the morphological forms of all the words in the corpus. A would-be reader of CVZ would benefit from a more extensive, flexible, and powerful dictionary than these.

**3.2. Challenges based on Characteristics of the Documents Themselves.** While there are various resources, colonial and modern, to assist the would-be reader of the CVZ documents, none of the resources are ideal as translation tools. But perhaps an even greater challenge to the reader of CVZ lies in some of the characteristics of the documents themselves.

First, the orthography in these colonial materials is inconsistent, and many times is distinct from that (or those) used in the *Vocabulario*. This means that even if one finds a Zapotec word in a colonial document, and if this same word happens to be in the *Vocabulario*, the orthography may be so different that the reader will fail to find it there. Thus, the richness of the *Vocabulario* cannot fully be taken advantage of either in its original form nor its modern, digital forms.

Secondly, as is typical in handwritten manuscripts in any language of the same time period, word divisions (i.e. what is written as a single word or multiple words), vary greatly. What appears as a single word in one document may appear as two or more words in another. This variation complicates the problem presented in the previous paragraph even more.

Thirdly, the verbs of Zapotec appear in a variety of aspectual forms that are marked with prefixes and are morphologically distinct. In addition, nouns, such as body part words, can be incorporated into the verb (post-root) and various clitics can follow verb stems. Thus, any verb has many distinct forms with prefixes and/or clitics. The verbal lexical entries in the *Vocabulario* do not demonstrate the full range of possible verb forms.

As a final point, the semantic extensions of the lexical entries in the documents may be beyond those indicated in the *Vocabulario*. All of this represents a great amount of information for a reader to keep in mind while reading a colonial document. It makes translating CVZ documents time consuming, even for experts, and creates a particularly high barrier of entrance to the corpus for new users. We believe a modern database can not only facilitate translation of texts for experts, but can open the corpus to additional interested individuals. Before describing the database in §4, we use an excerpt from Feria’s *Doctrina* to give the reader a sense of how one could use currently available resources to try and read a CVZ passage.

**3.3. A Case study: an example from Feria 1567.** Here we examine the step-by-step process of searching for a word while attempting to read a passage in CVZ through a mini case study. Consider the paragraph in Figure 2, taken from the *Doctrina* (Feria 1567): *Chioayale beniati, yobi anima toninabani quitobi pelalatini, tonibiniquito-biçôbeni*, with the corresponding Spanish sentence: ‘*Nascido ya el hombre, el alma es*
la que da vida a todo el cuerpo: y mueve todas sus partes. [The human being, having been born, the soul is what gives life to the whole body: and moves all of its parts.]

FIGURE 2. Chioayale beniati... (Feria 1567:7v) (Image published courtesy of the John Carter Brown Library at Brown University)

Because of the accompanying Spanish translation, we know that the first part is likely associated with nacer ‘be born’, so we can look up the verb nacer in Cordova’s Vocabulario (1578b), which can be found on page 279r, as shown in Figure 3. The
first pertinent entry is ‘Nacer animal y toda cosa assi. tàlea.co.ca.Plu.tàlena.’ The first form shown in the Cordova entry is the verbal form in the habitual aspect. From what we know of modern forms of Valley Zapotec, we know that the t- is the prefix that marks habitual aspect and the verb root is álea. But Cordova does not give us the root directly—we must deduce this. The entry furthermore gives us prefixes for the perfective (co-) and potential (ca-) forms, though, again, it does not indicate what exact root these prefixes attach to. At the end it gives us the first person plural subject form of the same verb, preceded by the Latin Plu[ral].

If we look for nacer animal y toda cosa assi in Smith Stark et al. 1993, we find the same form that appears in Cordova (1578b), shown in example 1; this is expected, as it is a digital version of Cordova. The Smith Stark et al. database contains different fields for the different parts of the entries, and many entries, though not this one, contain morphological segmentation of the Zapotec word.

(1) nacer animal y toda cosa assi Talea prt> co fut> ca Plu prt> talena

Finally, we look for nacer in the PDF of Whitecotton and Whitecotton 1993. As is shown in Figure 4, one finds the form ràle ‘nacer hombre ó animal’. The <r-> in the Junta Columbina, and therefore in Whitecotton & Whitecotton, corresponds to a flap in modern varieties. Cordova writes these consistently as <t>. Regarding this, Smith Stark states:

---

8 Not all verb entries in Cordova contain extra information about other asceptual forms or additional subjects, as can be seen in the following entry for nacer hazer otra cosa, which lists only three verb forms, all in the habitual aspect (beginning with t-) and with a first person singular subject (ending with =a). It is not clear why the first person plural verb form is included in the entry for nacer animal y toda cosa assi, as it is regular.
Una de las características más desconcertantes de la ortografía de Cordova es su uso de la <t> en palabras donde muchas formas modernas del zapoteco tienen una /r/. Esto es especialmente notorio porque cita todos los verbos en su forma habitual... como <t(i)->....

‘One of the most disconcerting characteristics of Cordova’s orthography is his use of <t> in words where modern Zapotec varieties have a /r/. This is especially noticeable because he cites all of the verbs in their habitual form... as <t(i)->...’ (2003:219)

Again, without knowledge of Zapotec morphology, it is not obvious that the root of ràle is ale.

Let us now return to our example from Feria. Here the verb ‘to be born’ appears in the ‘perfect (prf)’ aspectual form, with the prefix oay-.9 This form does not appear in any of the three sources consulted. Nor does the verb root ale appear in them. In addition, while chioayale is written together without any noticeable spaces, we believe that chi ‘when’ is a separate morphological and phonological word, based on our knowledge of modern Valley Zapotec varieties. We present our morphological analysis of the first few words in the Feria passage below in 2.10

(2) Chioayale beniati yobi anima...

Chi oay-ale beniati yobi anima
when prf-be.born human.being same soul

“When a person is born, the same soul...’

4. A LEXICAL AND TEXTUAL DATABASE OF CVZ. It is clear that in order to read a document written in CVZ one needs substantial information that cannot be found in any available dictionary: for example, understanding where and how to separate strings of letters into words, recognizing prefixes that appear before the root, and identifying alternative forms of words found in documents. In order to address some of these challenges, we have created a database of texts and a lexicon using the

---

9 A modern cognate can be found in Isthmus Zapotec: ‘El aspecto perfecto se marca con el prefijo hua. El verbo en esta forma significa acción repetida sobre el tiempo. No se usa si la acción no se ha llevado a cabo antes. ‘The perfect aspect is marked with by the prefix hua-. The verb in this form represents an action repeated over time. It is not used unless the action has taken place before’. (Pickett et al. 2001:54), as illustrated in (i), with our glossing. Thanks to Steve Marlett, John Foreman, and Michael Galant for discussion on this.

(i) Ma’ hua-yeeda Betu chonna tiru (Isthmus; Pickett et al. 2001: ex 63a)
already prf-come Beto three times
‘Beto has already come three times’

10 The following abbreviations are used: 1s, first person singular; 3, third person; CAUS, causative; CVZ, Colonial Valley Zapotec; FLEX, Fieldworks Language Explorer; HAB, habitual; IZ, Isthmus Zapotec; MZ, Mitla Zapotec; PERF, perfective; PRF, perfect; SDOZ, San Dionisio Ocotepec Zapotec; SLQZ, San Lucas Quiavini Zapotec; SPGZ, San Pablo Gúilá Zapotec; and STA, stative.
FieldWorks Language Explorer program (FLEx). This program is created by and distributed free of charge through SIL International, and there are interfaces in English, Spanish, and other languages. In this section we present the benefits that using FLEx has provided for our database project. Overall, FLEx has served us well; nevertheless, in §5 we outline the challenges of using FLEx for this particular corpus.

FLEx is designed to integrate a lexicon of morphemes (i.e. a ‘dictionary’), morphological analysis constrained by rules, and analyzed texts. Figure 5 shows the ‘Texts & Words’ tab of the FLEx interface (bottom of far left column), viewing the list of interlinearized texts (top of far left column) with the text associated with pages 7v–8r of the Doctrina (Feria 1567) selected.

In addition to having the power to integrate the morphologically analyzed texts with a complex lexicon, the FLEx database has many advantages over the tools that already exist for analysis of CVZ. For example, it is possible to register all the orthographic variants of one lexical item, morphological analysis can indicate when an orthographic word should join with, or separate from, another word, and the lexicon recognizes a large variety of affixes and clitics. Additionally, the database allows for the inclusion of pure verb roots in the lexicon and can associate any number of meanings with each lexical item. Finally, the efficiency of the concordance feature of FLEx allows us to search for (almost) identical words or phrases in the corpus of already analyzed documents to aid the analysis of new documents.

11 A publicly accessible version of the resulting CVZ dictionary can be found online at https://ticha.haverford.edu/en/vocabulary/.

12 For an image of the ‘Lexicon’ tab of the FLEx interface and a sample entry, see Broadwell & Lillehaugen 2012: Figures 12 and 13.
4.1. The morphological analysis of texts. One of the most powerful aspects of FLEX is that it ‘learns’ the analysis of words entered in the lexicon or previously approved by a user in the morphological analysis of a text. When the identical string of segments is encountered again, for example in a new text, FLEX will offer the previously approved analysis as a possible parsing of these subsequent appearances. These suggestions help enormously in the analysis and translation of new texts. These suggested glosses can be accepted or rejected, as deemed appropriate by the user. Moreover, each part of an analysis can be changed in the future as necessary, and the changes can be applied either individually or to all the appearances of the word in the corpus. This feature of FLEX has been highly useful to our analysis of new texts. Furthermore, the process of analyzing new texts, including associating new spelling variants with a lexeme in the lexicon and approving the morphological analysis of yet-unattested words, only strengthens the power of this mechanism, making future parsing even quicker.

4.2. The concordance. FLEX has a powerful concordance feature that is very useful to the reader of CVZ documents. This concordance feature allows a user to search for patterns within any field (or combination of fields) in the database. Possible searches include, but are not limited to, patterns in the Zapotec data, the Spanish data, the analyses, or the notes. Figure 6 shows a search for all the instances of the morpheme hue ‘perfect aspect’ in the corpus, which happens to be hue-2 in our lexicon. Note that as it is concorded on the lexical entry, it returns all known spelling variants of that morpheme, including oay-. A concordance search for the string hue in the Zapotec field would also be possible, and would return all and only instances where the string of letters hue is found, regardless of morphological analysis.

The concordance function can be very helpful to the researcher on at least two fronts. First, the concordance can help a reader understand an unknown CVZ word. In these cases, the reader can perform a search for the word throughout the entire corpus. The larger the text corpus, the more results will be given. Ultimately, the investigator will likely be able to deduce the meaning of the unknown word based on the results of the search of the concordance. The concordance is also useful in understanding the grammar of CVZ as a researcher can search for morphemes or lexical items to facilitate an understanding of the language, such as the tense/aspect system (Broadwell 2015a) or positional verbs (Foreman & Lillehaugen 2017).

4.3. Filters and the lexicon. FLEX allows users to filter the data in various ways. We have found the option to filter for certain strings or properties in the lexicon particularly useful and well suited for our project. As of November 2014, our lexicon had over 48,000 entries, partly because we benefitted from importing all of the entries

---

13 A screenshot of a text with suggested analysis (those marked in blue) can be seen in Broadwell & Lillehaugen 2012: Figure 14.
FIGURE 6. The FLEX CVZ database: a concordance search for instances of the lexeme *hue* ‘perfect aspect’

from Cordova (1578b) via Smith Stark et al. 1993 (via Oudijk’s Microsoft Access database).\(^{14}\) When working with such a large lexicon, the most productive strategy for analyzing a text often involves using filters to narrow down the possible candidates that might match an unknown word in a text. In this section we illustrate how to use the filter option through an example we encountered in our database.

Consider the phrase in 3a, from the *Doctrina* (Feria 1567). In this case, we wanted to understand the two verbs ‘pull (hair)’ and ‘drag’. As we had a Spanish translation, we first attempted to find the Zapotec word in the database by filtering the Spanish column in the lexicon for the expected Spanish verb stem without the infinitival -ar/er/ir. Specifically, we were looking for a word that we anticipated would mean *messar*

\(^{14}\) The process of importing these lexical items is described in Broadwell & Lillehaugen 2013. For more details on the technical implementation (including all the specifications of import settings and necessary pre- and post-editing of the files) see Broadwell’s blog post at http://workinglinguist.blogspot.com/2012/05/more-on-importing-cordova-into-flex.html.
so we filtered for the string ‘mess’, shown at the top of the third column in Figure 7. In this case, we successfully located the entry àxe ‘messar pelando [pull hair]’, shown in the habitual form in 3, with a third person subject =ni.

(3) a. oalijca tàxeni quichaquiquieni, tigòba yoonini (Feria 1567:64v) ‘la messa los cabellos, y la arrastra [he pulls her hair and drags her]’

b. t-àxe=ni

HAB-pull.hair=3

FIGURE 7. The FLEX CVZ database: Filtering for the string <mess> in the Spanish gloss field of the lexicon

FLEX allows users to filter on many possible columns. This is well suited for our database project, as occasionally we will want to filter on the CVZ column (i.e. the lexeme form) with the intent of finding an unknown Zapotec word in the database. Recall, however, that CVZ orthography is very unreliable, especially in terms of the fortis/lenis distinction for consonants. It might seem, at first, that one would have to search several times, using multiple possible combinations of letters to account for the variety of ways in which the word might have been spelled. FLEX, however, provides at least a partial alternative to this.
Returning to our example in 3a, consider the second verb. We expect the habitual form of this verb to be \textit{ti-gobayoo}. Therefore, if we want to find this verb root in our lexicon, we might search for \textit{gobayoo}. But we know that /g/ is often spelled <c> in the Cordova dictionary and /b/ is often <p>. The vowel length is also unreliable. Instead of manually searching for multiple spelling alternatives (such as \textit{gobayoo}, \textit{goobayo}, \textit{gobayo}, \textit{cobayoo}, \textit{cobayo}, \textit{copayoo}, \textit{copayo}, etc.) we used the option of regular expressions in the filter field. We can use the filter \textit{o+[bp]a+}, which reads ‘one or more <o>s, followed by <b> or <p>, followed by one or more <a>s’. This yielded a number of hits including the correct verb, as shown in Figure 8.\footnote{\textit{Gobayoo} certainly merits its own entry distinct from the other items it is currently listed with. Many of the entries imported from Smith Stark et al. 1993 contain multiple roots that need separate entries; this problem will be addressed in future work with the database.}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure8.png}
\caption{The FLEX CVZ database: Filtering for the string “\textit{o+[bp]a+}” (using regular expressions) in the Zapotec lexeme form field of the lexicon}
\end{figure}

Given the unreliable nature of the orthography, it may be necessary to try a few different filters on the Zapotec and Spanish fields to find an appropriate entry. For example, the filter \textit{[gc]o+[bp]a+} would not have returned the correct verb to us, since the first <o> in our verb has a grave accent. Nevertheless, it is significantly quicker than trying all the possible spelling variants one by one.
4.4. Collaboration. One of the largest benefits that FLEX version 8 has provided our database project is the ability to collaborate with multiple users. Currently, the FLEX CVZ team consists of two linguists, the two authors, and several student research assistants. At times we have had an even larger team. In the future, we would like to include more collaborators; thus, we wanted a database that could be used within a collaborative framework.

Prior to FLEX 8, doing collaborative work was quite inconvenient, though possible. For example, we used the FLEX 7.3 ‘backup / restore project’ function together with Dropbox to facilitate collaborative work on a single database, as described in Broadwell & Lillehaugen 2012.

In April 2013 the FLEX development team released version 8 of FLEX. This version has significantly changed the way teams can collaborate on a database. It is now possible to have several team members working on a single database simultaneously. It is important to note, however, that FLEX does not work using cloud computing, such as Google Drive, which allowed multiple users to simultaneously access and edit a document in real time. Instead, it works through a process of occasionally syncing the database, through a ‘send and receive’ feature. One advantage to this is that a team member can work on the database without internet connection, and sync when they have a connection. There are several ways team members can store their projects for syncing. We use the Language Depot / FLEXBridge option.

5. Challenges and Concerns. We have found FLEX to be quite suitable for our database. However, FLEX was not designed for inconsistent archival corpus data such as this, and thus we have faced certain challenges in using FLEX for this type of corpus. We mention here two main challenges: one based on the nature of the CVZ documents, such as the paleography (§5.1), and the second related to the nature of the Zapotec language itself, in particular verbs and verbal morphology (§5.2).

5.1. Representing paleographic detail. In this section, we exemplify several types of paleographic detail that are difficult or impossible to represent in FLEX.

The main entry field for a text in FLEX is ‘the baseline’, which is fairly limited in formatting options. The baseline, quite reasonably, accepts text that can be broken into words and lines. The morphological parser, however, assumes that a word will be written together on the same line. It is often the case, though, that in CVZ documents a word is split across lines, as in Figure 2, where seven of the eleven lines end with

---

16 Our current and former student research assistants are Carolyn Jane Anderson, Flora Berklein, Brent Coulter, Anneke Heher, Jordan Lavender, Oanh Luc, May Helena Plumb, Cameron Rees, Zach Sedefian, Allyson Stronach, Mattie Wechsler, and Rachel Elizabeth Weissler.

17 For more details on our team’s experience, see Broadwell’s blogpost at http://working-linguist.blogspot.com/2013/04/the-sendreceive-feature-in-flex-8.html.
words continued on the next line. This type of situation is, of course, quite common in manuscripts. We considered how to approach it in our database: if we are faithful to the paleography, we lose the advantage of the morphological parser. On the other hand, if we change line boundaries, we lose information about the original document. We opted for something of a compromise: in order to take advantage of the morphological parser, we finish writing any word in the baseline on the same line on which that word begins in the original text. However, we mark such alternations with a ‘[‘ before the letters that originally were not on that line. For example, the first line of our baseline, which corresponds with the first line in Figure 2, ends with ‘ani[ma’, meaning that the ma originally was on the following line. In this way, we take advantage of the strength of the morphological parser while losing as little information about the original text as possible.

There are other aspects of the paleography that seem less easy to represent in the baseline, such as decorative images, like the elaborate leaf at the bottom of Figure 2. In other cases, there are more detailed pictures or elaborated letters. Such decorations are interesting parts of the original document, but have no place in our current database. While we are often conflicted about leaving out such detail, it does not represent linguistic information. There are other aspects of the documents that are equally difficult to represent in the baseline that do—or might—represent linguistic information, such as letters written above other letters and missing or illegible text. We are currently developing conventions for dealing with these aspects of the archival documents, but FLEX has no built-in options for representing such paleographic detail.

5.2. Listing Zapotec Verbs in a Dictionary. As mentioned in §4.3, we imported all of Cordova’s Vocabulario entries into our database, via Oudijk’s Microsoft Access version of Smith Stark et al. 1993. One immediate concern that arose was that CVZ verbs are typically listed in their first person habitual forms in Cordova, where <ti> or <t> (among others) are the orthographic representation of the habitual prefix and <a> or <ya> (among others) are typical orthographic representations of the first person clitic used to mark the subject in this case. Recall from Figure 3 that our verb ‘be born’, which has the root ale, appears in Cordova as tālea which can be analyzed as shown in 4.

(4) ‘Nacer animal y toda cosa assi [be born (of) an animal and all such things]
   tālea...’ (Cordova 1578b:279r)
   t-āle=a
   HAB-be.born=1s
   ‘I am born’
While the first person habitual may be a reasonable citation form of the verb for Cordova’s purposes, the roots of the verbs are needed in order to use the lexicon as a tool to parse text. Therefore, we wanted to create a lexeme form of the verb without the aspectual prefix, but at the same time, we wanted to not lose any information.

Smith Stark et al. (1993) identifies and indicates the morphological divisions for approximately 70% of the entries, such as the entry in 5a, facilitating our task greatly. Note that the CVZ form appears with morphological divisions: Smith Stark et al. use ‘+’ for pre-root morphology and ‘–’ for post-root morphology. Thus not only does 5a indicate some of the morphological complexity of the word, it also indicates that chiba ‘sit elevated’ is analyzed as the root. Example 5b shows our analysis of the same entry, which—while not identical—also analyzes chiba as the verb root.

(5)  a. to+chiba-ya ticha-pi=tao ‘bendezir algo o consagrar [bless something or consecrate]’
   b. t-o-chib=ya ticha pi=tao
      hab-caus-sit.elevated=1s word life.force=great
      put god/great spirit
‘I put god’s word [on someone or something]’

18 The habitual aspect in fact seems like a good choice for a citation form of verb, as most verbs have a habitual form and the habitual form is morphologically quite regular. Many authors of modern Valley Zapotec dictionaries make the same choice, including Munro and Lopez et al. (1999) in their dictionary for San Lucas Quiaviní Zapotec and Stubblefield & Stubblefield (1991) in their Mitla Zapotec dictionary. (See Munro 2002a and 2005 for discussion on choice of citation forms for Zapotec verbs.)

The first person subject form, however, is not an ideal citation form, as it often has morphophonological irregularities. Cordova was likely influenced by Latin grammar in his decision to use the first person form as the citation form.

19 We follow Marcus (1983:345), Marcus & Flannery (1994:57–58), and Starr (1987:374) in analyzing pitao as morphologically complex, consisting of pi meaning something like ‘wind’, ‘breath’, or ‘life force’ and =tao meaning ‘great’. Interestingly, the Zapotec word pitao ‘great spirit’ is used here and not the Spanish borrowing Dios ‘God’. For more on the contrast between the use of pitao and Dios in CVZ documents, see Marcus 1983 and Valdivia 2012.

Zapotec positional verbs, like chiba ‘sit (elevated)’, have causative forms with ‘put’-type meanings (Foreman & Lillehaugen 2013). This is also true in CVZ (Lillehaugen & Foreman 2013).

We analyze ticha pitao as a possessed noun phrase, though we would expect the possessive prefix xi- on ticha ‘word’ as in the entry in the Vocabulario: “Bendito ser asi [be blessed, thusly]. tichibaya xiticha pitao” (Cordova 1578b:53v). For more on possession in CVZ see Galant 2011a and 2011b.
In cases such as 5a, we used the divisions indicated in Smith Stark et al. 1993 to strip away the prefix and suffix and constructed the lexical entry in Figure 9. Notice that the headword appears without the aspectual (and causative) prefix to- and without the first person marker =ya. In order to not lose information contained in the original Cordova entry, we created several custom fields. We retain the entire original citation form in a field called Cordova original. We list just the habitual prefix used in Cordova in a field we call Cordova habitual. (Note that we include causative o- as part of the habitual prefix for this purpose).\textsuperscript{20} Many Cordova entries also contain abbreviations related to the prefixes used for the completive and potential aspects; when these are present, we list them in the custom fields Cordova completive and Cordova potential respectively. The page number in Cordova is listed in source, which is a regular field (i.e. not a custom field). We believe that by using this method we can extract morphological information that will enable analysis of the texts while not destroying any information from the original. The flexibility to add custom fields in FLEX has, thus, been extremely useful.

6. Results. The CVZ database provides access to a rich variety of information about the language. In this section, we report on our results in better understanding the morphophonology, syntax, and semantics of selected CVZ verbs.

In constructing the (Spanish to Zapotec) Vocabulario, Cordova often listed the same Zapotec word under several different Spanish entries. For example, consider the sixty-nine entries from the Vocabulario in (6), in which a Zapotec root like <chille> ~ <chilla> is present. It is clear that many of the Spanish terms are synonyms of each other; we also see that the Zapotec verbs may have a variety of prefixes, including <na-> ‘stative’, <ti-> ‘habitual (intransitive)’, <t-o-> ‘habitual (transitive)’, and <hue-> ‘perfect’.

\textsuperscript{20} For the sake of simplicity, we have followed Smith Stark’s morphological segmentation in isolating the root. We acknowledge, however, that the habitual we have listed as /to-/ is morphologically complex, and includes both the aspectual /t/ and a reflex of the proto-Zapotec causative */o/, though it is unclear if the o- causative was productive in CVZ (Lillehaugen 2012). (Valence alternations associated with the causative o- are lexicalized in modern Tlacolula Valley Zapotec languages (Munro 2015b)).
**FIGURE 9.** The FLEX CVZ database: the entry for *chiba ticha-pitao* ‘bless or consecrate (something)’
(6) Entries from Cordova containing <chille> ~ <chilla> (and other spelling variants)

<table>
<thead>
<tr>
<th>Spanish gloss</th>
<th>Cordova original</th>
<th>Folio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abrir carta cerrada</td>
<td>tochillea, tixàllea</td>
<td>f4</td>
</tr>
<tr>
<td>Abrir la mano vazio</td>
<td>Tochille ñaaya</td>
<td>f4</td>
</tr>
<tr>
<td>Abrir los braços este[n]die[n]dolos.</td>
<td>Tochilleñaaya, tocigãñaaya.</td>
<td>f4</td>
</tr>
<tr>
<td>Adjuinar por suertes o agueros</td>
<td>Tochilla, tibeea. {vel.} tebeea pijci. /cv/ col.</td>
<td>f10</td>
</tr>
<tr>
<td>Agudo de ingenio</td>
<td>nachilla</td>
<td>f15</td>
</tr>
<tr>
<td>Agudo ser assi</td>
<td>tichilla,achtachi</td>
<td>f15</td>
</tr>
<tr>
<td>Apartar casados o a dos que están juntos</td>
<td>Tochilléa, tochillaquisòa, til láaya. /prt/ co</td>
<td>f32</td>
</tr>
<tr>
<td>apartar casados o a dos que están juntos</td>
<td>Tochilléa, tochillaquisòa, til láaya. /prt/ co</td>
<td>f32</td>
</tr>
<tr>
<td>Bulicioso sin sosiego, o trauesso</td>
<td>nachilla</td>
<td>f62</td>
</tr>
<tr>
<td>Casados apartar</td>
<td>Tochilla</td>
<td>f74</td>
</tr>
<tr>
<td>Casamientos sortear antiguamente el sortilegio para ver si eran para en vno</td>
<td>tochilla</td>
<td>f74</td>
</tr>
<tr>
<td>Crucificar a otro</td>
<td>tochilleñaaya láni</td>
<td>f99</td>
</tr>
<tr>
<td>Cuydado tener o estar con el de lo q[ue] te[n]go de hacer</td>
<td>Tichillalâchia</td>
<td>f102</td>
</tr>
<tr>
<td>Derribar edificio o casa</td>
<td>tochillea, {{ve}l.} tochijlalla</td>
<td>f119</td>
</tr>
<tr>
<td>Desarrollar</td>
<td>Tochillea</td>
<td>f121</td>
</tr>
<tr>
<td>Desarrugado ser</td>
<td>Tichille</td>
<td>f121</td>
</tr>
<tr>
<td>Desarrugar lo arrugado</td>
<td>Tochillea</td>
<td>f121</td>
</tr>
<tr>
<td>Desassosegado assi</td>
<td>nachilla,achtachi</td>
<td>f121</td>
</tr>
<tr>
<td>Desassosegado ser o estar, o enfermo assi</td>
<td>tichilla</td>
<td>f121</td>
</tr>
<tr>
<td>Desassosegamiento</td>
<td>nachilla,achtachi</td>
<td>f121</td>
</tr>
<tr>
<td>Desbaratada ser ge[n]nte assi</td>
<td>tichilla</td>
<td>f121</td>
</tr>
<tr>
<td>Desbaratar algo como edificio o casa</td>
<td>tochilla</td>
<td>f121</td>
</tr>
<tr>
<td>Descasarse ellos [los casados]</td>
<td>Tichillea</td>
<td>f122</td>
</tr>
<tr>
<td>Spanish gloss</td>
<td>Cordova original</td>
<td>Folio</td>
</tr>
<tr>
<td>------------------------------------------------------------------------------</td>
<td>------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Desembuelto ser [serle quitada la envoltura]</td>
<td>Tichillea</td>
<td>f125</td>
</tr>
<tr>
<td>Desemparejado ser</td>
<td>tichillea</td>
<td>f125</td>
</tr>
<tr>
<td>Desemparejado ser</td>
<td>tichillaya</td>
<td>f125</td>
</tr>
<tr>
<td>Desemparejar dos cosas</td>
<td>tochillea</td>
<td>f125</td>
</tr>
<tr>
<td>Desempegar</td>
<td>tochillaya</td>
<td>f125</td>
</tr>
<tr>
<td>Desemperezar</td>
<td>tichillaryèchia</td>
<td>f125</td>
</tr>
<tr>
<td>Deshecho ser assi [casa o edificio]</td>
<td>tichilla</td>
<td>f127</td>
</tr>
<tr>
<td>Deslizarse assi delas manos</td>
<td>Tichille</td>
<td>f129</td>
</tr>
<tr>
<td>Desmandarse en hablar</td>
<td>chillaya</td>
<td>f129v</td>
</tr>
<tr>
<td>Despartido ser</td>
<td>tichillea</td>
<td>f131</td>
</tr>
<tr>
<td>Despartido ser assi alos que esta[n] juntos</td>
<td>tichillaya</td>
<td>f131</td>
</tr>
<tr>
<td>despartir o dividir</td>
<td>tochillaya</td>
<td>f131</td>
</tr>
<tr>
<td>Despedirse el que se parte</td>
<td>tochillaya ticha làoaça-ya. {{ve]}t. teòchillaya</td>
<td>f132</td>
</tr>
<tr>
<td>Despedirse el que se parte</td>
<td>Tochillaya ticha ça-ya</td>
<td>f132</td>
</tr>
<tr>
<td>Distinguir o apartar vno de otro</td>
<td>tochillea</td>
<td>f142</td>
</tr>
<tr>
<td>Enconada cosa assi</td>
<td>nachillaquij</td>
<td>f162</td>
</tr>
<tr>
<td>Enconado ser assi</td>
<td>tichillaquij</td>
<td>f162</td>
</tr>
<tr>
<td>Enconado ser assi [herida, llaga o hinchazon]</td>
<td>Tichillaquij, /prt/ pi</td>
<td>f162</td>
</tr>
<tr>
<td>Estenderme en luengo o lo encogido</td>
<td>Tichillea</td>
<td>f189</td>
</tr>
<tr>
<td>Despartir o diuidir</td>
<td>Tochijllea,</td>
<td>f131</td>
</tr>
<tr>
<td>Estender outra cosa o ami mesmo</td>
<td>Tochijllea</td>
<td>f189</td>
</tr>
<tr>
<td>Estendida cosa assi. s. el braço o pierna</td>
<td>Nachijleñaa</td>
<td>f189</td>
</tr>
<tr>
<td>Estender el braço. s. estendido ser</td>
<td>tichijleñaya, tigáanaaya, tilijñaya.</td>
<td>f189</td>
</tr>
<tr>
<td>Descasar los casados</td>
<td>tochijllaya</td>
<td>f122</td>
</tr>
<tr>
<td>Sortear echar suerte</td>
<td>Tochijllaya</td>
<td>f387</td>
</tr>
<tr>
<td>Spanish gloss</td>
<td>Cordova original</td>
<td>Folio</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>----------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Deshazer casa o edificio</td>
<td>Tochijllaya</td>
<td>f127</td>
</tr>
<tr>
<td>Descasarse ellos [los casados]</td>
<td>tichijllaya</td>
<td>f122</td>
</tr>
<tr>
<td>Suerte hechar</td>
<td>tochijllaya</td>
<td>f390</td>
</tr>
<tr>
<td>Despoblarle [al pueblo]</td>
<td>tochijllea</td>
<td>f133</td>
</tr>
<tr>
<td>Desdoblado ser</td>
<td>Tichijllea</td>
<td>f125</td>
</tr>
<tr>
<td>Despartir ruydo o renzilla</td>
<td>tochijllea</td>
<td>f131</td>
</tr>
<tr>
<td>Desplegados ser assi [quitados los pliegues]</td>
<td>Tichijllea</td>
<td>f133</td>
</tr>
<tr>
<td>Destrchada ser assi [casa].</td>
<td>Tichijllea</td>
<td>f134</td>
</tr>
<tr>
<td>Despartido ser assi alos que esta[n] juntos.</td>
<td>Tichijllea</td>
<td>f131</td>
</tr>
<tr>
<td>Desenhebrar como los cabellos.</td>
<td>Tochijllea</td>
<td>f126</td>
</tr>
<tr>
<td>Descasar los casados</td>
<td>Tochijlléa</td>
<td>f122</td>
</tr>
<tr>
<td>Entesar</td>
<td>Tochijllea</td>
<td>f173</td>
</tr>
<tr>
<td>Desdoblar</td>
<td>Tochijllea</td>
<td>f125</td>
</tr>
<tr>
<td>Tender lo encogido</td>
<td>Tochijllea, tochijllaya</td>
<td>f396</td>
</tr>
<tr>
<td>Estender la mano vazia.</td>
<td>Tochijlleláchinaaya. {[ve]l.} ñaaya {solum.}</td>
<td>f189</td>
</tr>
<tr>
<td>Agudamente</td>
<td>huachilla láchichi</td>
<td>f15</td>
</tr>
<tr>
<td>Despoblador</td>
<td>huechijlle</td>
<td>f113</td>
</tr>
<tr>
<td>Estendedor</td>
<td>Huechijlle</td>
<td>f189</td>
</tr>
<tr>
<td>Agudo o dilige[n]te.</td>
<td>náchilla</td>
<td>f15</td>
</tr>
<tr>
<td>Diligente</td>
<td>Nachijlla, naciña, natiti, naquèñe, nanij.</td>
<td>f140</td>
</tr>
<tr>
<td>Diuidir</td>
<td>Tochillea, tochij-çòoa</td>
<td>f143</td>
</tr>
</tbody>
</table>
To understand how many different verb roots are involved in this list, the range of Spanish translations for each, and the range of morphemes that are shown for each verb, we began with these sixty-nine entries and used the FLEX function which allows a user to merge entries. After combining synonyms and comparing with forms in modern Zapotec languages, we arrived at eight entries for forms, which are shown in (7). The intransitive and transitive forms with meanings related to separation (e.g. of spouses, parts of buildings) we combine into chilla₁ and chilla₂. The transitive and intransitive senses related to smoothing and the removal of wrinkles are shown in chilla₃ and chilla₄. Finally, some idiosyncratic and isolated senses of the root, with meanings related to talking, divination, restlessness, and cleverness are shown in chilla₅, chilla₆, chilla₇, and chilla₈.

(7) <chilla> entries in FLEX CVZ database

chilla₁ [tsilla], v 1) disunite; divide; split up; distinguish or separate | Desunir; Despar-tir o diuidir; Diuidir; Distinguir o apar-van de otro. syn: chij-çòo⁷; syn: tècha₂. 2) separate married people or two who are together; unmarry | Apartar casados o a dos que estan juntos; Casados apartar; Descasar los casados. syn: chillaquicò; syn: lláaya. 3) demolish (building or house); tear down a house or building; ruin (a building) | Derribar edificio o casa; Deshazer casa o edificio; Desbaratar algo como edifi-cio o casa.; Desbaratar gente o cosa assi. 4) make unequal | Desemparejar dos cos-sas. nipichilla a thing made unequal 5) Divide | Despartir o diuidir caus/non-cause: billa₃; caus/non-cause: chilla₃. [HAB to-; PERF co-] [Original Cordova to+chilla-ya, To+chilla-ya, To+chilla-ya, To+chilla-ya, tochijlla-ya, to+chille-a, To+chille-a, tochillea, Tochillea, Tochillëa, tochillea, To+chille-a, To+chille-a, To+chijlle-a, To+chillë-a, To+chijlle-a, To+chijlla-ya, totèchaya, tochillaquicòa, til láaya. /prt/ co., ti+quíxe-a, /prt/ co+ti, tixàllea ] {compounds chillaquicò separate married people or two who are together, chille quique yöho unroof a house, quelà hue+chille action of creating inequality}

chilla₂ [dziHlla] v 1) be parted; be separated; be split up | Despartido ser assi alos que esta[n] juntos; Despartido ser; Despartido ser assi alos que esta[n] juntos. 2) be made unequal | Desemparejado ser; Desemparejado ser. 3) have one’s married dissolved; be divorced | Descasarse ellos [los casados]; Casados apartarse. 4) be unmade (of a house or building); have the roof taken off | Deshecho ser assi [casa o edificio]; Destechada ser assi [casa]. caus/non-cause: chilla₂. [HAB ti-, Ti; PERF pi-] [Original Cordova ti+chilla-ya, ti+chilla-ya, Ti+chilla-ya, Ti+chilla, Ti+chilla-a, Ti+chille-a, Ti+chille-a, Ti+chille-a, Ti+chille, to+cijlle-a ] {IZ ribiá ~ ruchiá} {MZ ru-tzujL-i ‘destruir, descoser, desbaratar’; bihL ‘to be undone’ ~ tshuL ‘to undo’} {SLQZ rtse’ilhly, rtsë’ilhly ‘undoes, unwraps, destroys’} {SPGZ bìld ~ tsìld (LC)} {compounds chillañà crucify, quelà na+chilla action of making something unequal, Quela na+chille divorce}
chilla₃ [tsilla] v uncurl; unfold; untwist; (like horses’ manes); unfold; unroll; unwrap; open a closed letter | Desarrugar lo arrugado.; Desplegar quitar los pliegues; Desdoblar; Desempegar.; Desenheattr como los cabellos.; Desarrollar; Desembolear, o desdoblar ropa.; Abrir carta cerrada. 

**syn:** xàlle. **caus/non-cause:** chilla₄ [Original Cordova tochíllea, tochijllea, tochillea] {MZ r-tsuL ‘to smooth out’} {compound Quela na+chijlle unfolding}

chilla₄ [dzilla] v 1) be smoothed out; be unfolded | Desarruggado ser; Desdoblado ser; Desplegados ser assi [quitados los pliegues]. 2) be unwrapped | Desembuelto ser [serle quitada la envoltura]. **caus/non-cause:** chilla₅. [HAB tí-] [Original Cordova tichíllea] {MZ Mitla has a bi-replacive biL ‘to be smoothed’}

chilla₅ v get out of control in talking | Desmandarse en hablar. [Original Cordova chilla-ya]

chilla₆ v 1) decide marriages (in the old manner by chance) | Casamientos sortear antigüamente el sortilegio para ver si eran para en vno. 2) divine through omens | Adiuinar por suertes o agueros. **syn:** bee₆. [HAB to-; PERF col] [Original Cordova Tö+chilla-ya, Tochillaya, tibeea. {vel.} tebeea pijci. /cv/ col-] {Compounds Quela-hue+chijlla luck, chances, lots, Xòopa hue+chijlla. lots that are thrown (of corn, small sticks, beans)}

chilla₇ 1) clever, sharp | Agudo de ingenio 2) keen or diligent | Agudo o diligen[nte. [STA na-, nà-] [Original Cordova na+chilla, nà+chilla]

chilla₈ v 1) be restless or ill; be thrown into confusion | Desassosegado ser o estar, o enfermo assi; Desbaratada ser ge[n]te assi. nijaxtenni yaquichijllati lachitooni loochijllati, § so that you will not be restless with the work of the body | para que assi podais mejor y con mas quietud 2) bustling, busy | Bullicioso sin sosiego, o trauiesso. [HAB tí-; STA na-] [Original Cordova ti+chilla, Na+chilla] {compound chilla-láchi be uneasy, anxious}

While we think that it is very likely that some of these entries have further relationships to each other, we have tried to be conservative so far in combining the Cordova entries in our current database. It also seems clear to us some of the senses are far apart semantically; we think these may be different words, where the orthography of CVZ fails to show a number of features that distinguish words phonologically in modern Valley Zapotec languages (features such as fortis/lenis contrasts in consonants, and tone and phonation type of vowels).

After the headword of selected entries we give a probable pronunciation of the root, based on a comparison of modern Valley and Isthmus Zapotec cognates, when available, from San Lucas Quiavini Zapotec (SLQZ), Mitla Zapotec (MZ), Isthmus Zapotec (IZ), San Dionisio Ocotepex Zapotec (SDOZ), and San Pablo Güilá Zapotec
We also group the senses of each root into the four or five main areas represented by the Spanish glosses, retaining the original Spanish and providing an English translation that tries to capture the most essential semantic features. Each entry also shows the information for the aspectual prefixes found in the Cordova entries, the original Zapotec form of all the entries, the modern cognates, and the compounds that include the current root as a part. Thus the database allows us create entries for CVZ lexemes that give a far clearer understanding of the range of senses associated with a root, the various spellings found in Cordova and other sources, the morphology associated with the root, and modern cognates to the root.

7. Conclusions. While there is a valuable corpus of colonial documents written in CVZ, the translation of these documents has proven slow and labor intensive. We believe the development of a CVZ database will be a powerful tool for translating and analyzing colonial documents as well as for using and interpreting the CVZ linguistic data. Despite the challenges of using FLEx for paleographic materials, FLEx 8 has proven a powerful and useful platform for developing this type of database, especially given the linked lexicon and textual database, support for morphological analysis, and ability to work collaboratively through syncing. We believe it would be a beneficial tool for others working with similar corpora.

References


21 As the forms from modern languages shown, the transitive form of the root ‘destroy, undo’ typically begins with /ts/ and has a breathy vowel (shown as H in the pronunciation field). The intransitive form ‘be destroyed, undone’ in most modern languages, however, has an initial /bi-/ in place of the consonant, and thus it does not match the forms recorded in chilla. Based on the normal fortis-lenis alternation found in other such pairs in Zapotec (e.g. Munro 2015b), we would expect the intransitive form to have initial /dz/.


de la Cruz, Víctor; Marcus Winter; Ausencia López Cruz; José Leonardo López Zárate; Gonzalo Sánchez Santiago; Nimcy Arellanes Cancino; and Thomas C. Smith Stark. 2009. Una lectura del texto zapoteco del mapa de Santo Domingo Niaa Guehui. Cuadernos del Sur 28.9–46.


LILLEHAUGEN, BROOK DANIELLE; GEORGE AARON BROADWELL; MICHEL R. OUDUK; LAURIE ALLEN; MAY PLUMB; and MIKE ZARAFONETIS. 2016. TICHA: A digital text explorer for Colonial Zapotec, 1st edn. Online: http://ticha.haverford.edu/.

LILLEHAUGEN, BROOK DANIELLE; GEORGE AARON BROADWELL; MICHEL R. OUDUK; and VICTORIA KRAZ. 2013. Outline of Cordova’s 1578 *Arte en lengua zapoteca*. Online: https://ticha.haverford.edu/en/outline.


MUNRO, PAMELA; KEVIN TERRACIANO; XÓCHITL FLORES-MARCIAL; MICHAEL GALANT; BROOK DANIELLE LILLEHAUGEN; MARIA ORNELAS; AARON HUEY SONNENSCHEIN; and LISA SOUSA. 2018. The Zapotec-language testament of Sebastiana de Mendoza, c. 1675. *Tlalocan*, to appear.

MUNRO, PAMELA; KEVIN TERRACIANO; XÓCHITL M. FLORES MARCIAL; MICHAEL GALANT; AARON HUEY SONNENSCHEIN; and DIANA SCHWARTZ. 2013. Un testamento zapoteco del valle de Oaxaca, 1614. Los Angeles: University of California, MS.


Instituto Nacional de Antropología e Historia.


**Smith Stark, Thomas C.; Áurea López Cruz; Mercedes Montes de Oca Vega; Laura Rodríguez Cano; Adam Sellen; and Alfonso Torres Rodríguez, with Vicente Marcial Cerqueda and Rolando Rosas Camacho.** 2008. Tres documentos zapotecos coloniales de San Antonino Ocotlán, *Pictografía y escritura alfabetica en Oaxaca*, ed. by Sebastián van Doesburg, 287–350. Oaxaca, Mexico: Instituto Estatal de Educación Pública de Oaxaca, México.


**Stubblefield, Morris, and Carol Miller de Stubblefield.** 1991. *Diccionario zapoteco de Mitla, Oaxaca*. Mexico City: Instituto Lingüístico de Verano, A.C.


Broadwell
Department of Anthropology
University of Florida
Turlington Hall, Room 1112
PO Box 117305
Gainesville, FL 32611-7305
[broadwell@ufl.edu]

Lillehaugen
Haverford College
370 Lancaster Avenue
Haverford, PA 19041
[blilleha@haverford.edu]