Supplementary Material: Orthography conversion and sources

Mixtec Sound Change Database

Sandra Auderset and Eric W. Campbell

November 16, 2023

Contents

1 Converting orthography to IPA

The orthographies used in materials on Mixtecan languages vary widely. Often each author and each source uses a system differing from all others in certain aspects. This section contains an overview of how the orthographies of each data source were converted to (a standardized version of) the International Phonetic Alphabet (IPA). This step is necessary for adequately comparing and aligning entries to identify cognate sets and sound changes. We provide details for each doculect for purposes of transparency and re-usability of the Mixtec Sound Change Database. Broadly, one can identify three major groups of orthographic systems currently in use:

1. Orthographies based on Spanish: This set of orthographies tries to stay as close to the Spanish system as possible. The sound [k], for example, is represented by <qu> preceding front vowels and <c> preceding non-front vowels. Additions are made as necessary, usually with digraphs. Nasality on vowels, for example, is represented by an <n> following the vowel in question. The glottal stop is usually written as <h>. This system is most often found in pedagogical materials and/or materials developed by SIL missionaries.
2. Orthographies based on the Americanist Phonetic Notation, also referred to as American(ist) Phonetic Alphabet (APA), is a phonetic system originally developed for the transcription of Native American languages (cf. the summary by the Western Institute for Endangered Language Documentation). It relies heavily on diacritics and mixes Latin and Greek characters. The sound [ʃ], for example, is represented as <š>. The glottal stop is usually written with the IPA symbol. Nasality on vowels is usually represented by an ogonek underneath the vowel in question, e.g. <ą> for [ã]. This system is found in older sources and those predominantly geared to an academic audience.
3. Orthographies based on neither of the above: Some newer documentation efforts use a system that could be referred to as latin-based, i.e. it uses Latin characters with few diacritics but does not emulate Spanish orthography. The glottal stop is usually represented by an apostrophe or saltillo. Nasality on vowels is usually represented by an <n> following the vowel in question.

In all of these orthographic systems, some graphemes are easy to convert to IPA because there is no ambiguity as to what sound they represent. This is the case for the nasals *n*, *m*, and *ñ* and for most vowels. Other graphemes, however, are notoriously ambiguous also due to differences in the sound systems of these languages. These are:

- <y which sometimes represents [j], but other times [ʒ]
- <x> which often represents [ʃ], but is other times used for [x] or [ɕ]
- <h> which is used as the glottal stop in Spanish-based orthographies, but in others often represents [h]
- <j> which is sometimes used as [h] other times as [j]

Below, we elaborate on some general principles for conversion and then address the specifics of each doculect beyond those general principles, providing detailed discussion only for potentially ambiguous graphemes. The doculects are ordered alphabetically by the language code used throughout the supplementary materials. For practical reasons, graphemes will not be written within angle brackets but presented in italics. IPA correspondences are rendered in square brackets as is customary (e.g. the grapheme *i* represents IPA [i]).

2 General Conversion Principles

2.1 Prenasalized stops vs. nasal + stop sequences

In most analyses, Mixtec stops do not exhibit a voicing distinction, but are described as opposing a series of voiceless stops to a series of prenasalized stops. Some scholars have analyzed these as combinations of nasal and stop as sequences, rather than a single phoneme (see e.g., Pike & Oram (1976) and Macaulay (1996)). Iverson & Salmons (1996), on the other hand, analyze prenasalization as an effect of hypervoicing, positing a series of plain voiced stops at a phonological level. On empirical grounds, this issue can only be resolved by a detailed analysis of the distributional and phonetic evidence in each variety, which lies outside the scope of this work. For practical purposes, i.e. for sequence alignments, consistency is the most important consideration, but there are no obvious advantages to representing all as prenasalized stops or nasal + stop sequences. We have decided to adopt the former solution, because it simplifies syllable structure, which could be helpful for automatic processing of the data.

2.2 Long vowels

Sequences of identical vowels are represented as such and not as a single long vowel, e.g. aa is represented as [aa] and not [a:]. We do this for two reasons: 1) The tone bearing unit in Mixtec languages is the mora, so it is easier and less confusing to represent tone with each vowel written separately, and 2) There is variation between and within Mixtec varieties in words of the form $CV_iʔV_i$, such that these can contract to CV_iV_i. For comparative purposes, it is more straightforward to align across such variation if the vowels are not written as one segment.

2.3 Glottal stop and glottalization

The synchronic phonological representation of laryngealization/glottal stop in Mixtec languages has received considerable attention and divergent analyses. For example, Castillo García (2007) (on Yoloxóchitl), McKendry (2013) (on Nochixtlán) and Hinton et al. (1991) (on Chalcatongo) treat laryngealization as a vocalic feature, as does Gerfen (1996) for Coatzospam Mixtec, where it is automatically inserted word-medially. Macaulay & Salmons (1995) treat laryngealization as a contrastive floating feature of the root in Chalcatongo Mixtec, and Carroll (2015) and Mendoza Ruíz (2016) adopt similar analyses for Ixpantepec Nieves and Alcozauca Mixtec, respectively. North & Shields (1977) and Pike & Cowan (1967) consider it to be a glottal stop consonant in Silacayoapan and Huajuapan Mixtec, respectively. Josserand (1983) analyzes glottalization as a feature of the vowel, because it was treated this way by the majority of descriptive studies available at the time. We follow Kaufman (in press) in that we include the glottal stop as a consonant, but this has practical rather than theoretical motivations. The representation of the glottal stop in the database (either as a consonant or as glottalization) does not affect the reconstruction nor the characterization of sound changes in the database. In other words, the diachronic behavior of the glottal stop is such that the practical results would be no different if we analyzed it as a vowel feature. Since it is representationally simpler to write a full glottal stop, this is what is implemented in our database.[1]

[1] We note that this can easily be changed in a future/derived version of the database.

3 Language-specific details

This section briefly discusses language-specific issues in the conversion to IPA. We start by outlining conversion principles for the two comparative works that cover multiple varieties.

3.1 Josserand 1983

The data gathered in Josserand (1983) covers 120 Mixtec varieties. They are presented in an Americanist phonetic notation at a level between phonetics and phonology. The conventions are explained in prose in Josserand (1983:472-474) and summarized here for convenience.

- caron above a consonant = palato-alveolar, e.g. *š* = [ʃ]
- *s̆* = fronted, pronounced with the blade of the tongue, which is likely alveolo-palatal, i.e [ɕ]
- *ṣ̆* = retroflex, i.e. [ʂ]
- *x* = velar fricative, i.e. [x]
- *h* = almost frictionless velar fricative (from PMx *s and not *x), i.e. [h]
- *ʉ* = [ʉ]
- sometimes an initial glottal stop is noted, but this is never contrastive and was deleted for better comparison
- nasalization on vowels is marked by a cedilla underneath the respective vowel: e.g. *ą* =[ã]
- prenasalized consononants are marked with a preposed superscript *n*, e.g. *ⁿd* = [ⁿd]
- palatalized consononants are marked with a postposed supercript *y*, e.g. *tʸ* = [tʲ]
- capital letters represent voiceless or aspirated vowels or consonants
- specific to (San Martín) Peras Mixtec [pera11]: this variety has a voiceless fricative [h], which appears only between vowels and is represented as a capital letter of the same vowel, e.g. *oOo* = [oho] (cf. also 3.23 for a closely related variety with the same phenomenon)

In order to make the material more comparable with other sources, we have standardized the entries such that phonetic variation or allophony is not represented. A list of these adjustments is given here:

- ʉ as a variant of u (usually in the contect of [j]): in Guadalupe Villahermosa, San Jorge Nuchita, San Martin del Estado, San Luis Morelia, San Miguel Ahuehuetitlan
- ʈʂ(ʲ) as an allophone of tʃ: tʃ ➔ ʈʂ(ʲ)/_e,a in Guadalupe Villahermosa
- $t^n_{\circ}$ as a phonetic variant of t^n in San Agustin Tlacotepec, San Bartolome Yucuane, San Miguel Achiutla, San Pedro Molinos, Santa Maria Yolotepec
- ɲ̃ as a phonetic variant of ɲ in San Juan Diuxi Mixtec, San Juan Ñumí, San Miguel Chicahua, Santa Lucia Monteverde, Santiago Tilantongo
- xʲ as an allophone of x: x ➔ xʲ/_e in San Antonio Huitepec, Santa Ana Cuauhtemoc
- ʒ as an allophone of ʃ: ʃ ➔ ʒ/#_ in San Juan Tamazola,
- ʂ as phonetic variant of ʃ in San Juan Teita, San Juan Tamazola, Santiago Tilantongo
- ʂ and sʲ as phonetic variants of ɕ in Santiago Juxtlahuaca
- ʂ as allophone of ʃ: ʃ ➔ ʂ/_i in Santa Cruz Itundijia

- echo-vowels after glottal stop before consonant
- ⁿr as an allophone of ⁿd: ⁿd ➔ ⁿr/_a,o,u in San Miguel Chicahua
- devoiced vowels as phonetic variants before voiceless stops in Santa Ana Cuauhtemoc, Tepango
- tʲ as an allophone of t: t ➔ tʲ/_u in Santo Domingo Nuxaa, Zapotitlan Palmas
- ⁿdʲ as an allophone of ⁿd: ⁿd ➔ ⁿdʲ/_u in Santo Domingo Nuxaa
- tʃ is actually ts in Mixtepec, as shown by Durr and later scholarship

3.2 Dürr 1987

The data presented in Dürr (1987) covers 17 Mixtec varieties. All data have been aggregated from published sources (cf. Dürr 1987:36-37 for the complete list) and standardized in Americanist phonetic notation. The segmental notation is thus essentially the same as in Josserand (1983) and the reader is referred to Section 3.1 for more details. A few subphonemic details such as nasalization have been included for better comparison. The cognate sets are arranged in the same way as in Josserand (1983).

3.3 Abasolo del Valle Mixtec [abas]

This variety is spoken in a diaspora community in the state of Veracruz. The data come from Galindo Sánchez (2009). The author claims that the variety is identical to that spoken in San Juan Mixtepec. The orthography is latin-based without further explanation given (Galindo Sánchez 2009:12, 18-23). Unclear IPA correspondences are resolved based on closely related varieties (cf. 3.32).

- *x* = [ʃ] and y = [ʒ]
- *ku*V = [kʷ]

3.4 Alacatlatzala Mixtec [alac]

The data for this variety comes from five sources: Josserand (1983), Dürr (1987), Zylstra (1991), Anderson (2006) and Zylstra (2012). Tones are marked in all of these sources, but not for all entries in Josserand (1983). As detailed below, the orthographies used across the five sources are different and produce certain conflicts and overlaps (especially in tone notation), but can be grouped into three sets:

Set A = Anderson (2006) and Zylstra (2012), Set B = Zylstra (1991), and Set C = Josserand (1983) and Dürr (1987) (see Sections 3.1 and 3.2 for more details).

Set A Anderson (2006:viii) and Zylstra (2012:13) both use a latin-based system

- digraphs: ti*V* = [tʲ], ku*V* = [kʷ]
- *y* is a semivowel, i.e. [j]
- *x* = [ʃ]

Set B Zylstra (1991:8) is also a latin-based system but with different grapheme choices and tone notation

- but *x* = [x] (*sh* = [ʃ])
- glottal stop as *h*

3.5 Alcozáuca de Guerrero Mixtec [alco]

The data for this variety come from three sources: Josserand (1983), Mendoza Ruíz (2016), and Swanton & Mendoza Ruíz (2021). Tone is marked only for a few entries in Josserand (1983), with diacritics. For the segments, we follow Mendoza Ruíz's (2016) representation, as that of Swanton & Mendoza Ruíz (2021) is more abstract (or phonological) than other material we have, which would make comparison difficult.

- *c* = [c] (this consonant is unique to Alcozauca Mixtec and transcribed as [t] in Josserand (1983))
- *b* = [β]; as in other varieties, this is realized as [b] word-initially, but [β] word-medially, but since this kind of allophony is not represented in other sources and for better comparison, we use [β] as the only representation

3.6 San Esteban Atatlahuca Mixtec [atat]

The data for this variety comes from three sources Alexander (1980), Josserand (1983), and Dürr (1987). Tones are marked in all sources, but not on every entry in Josserand (1983). The orthography used in Alexander (1980:3-6) is Spanish-based:

- *jn* = [n̥]
- *x* = [ʃ], *y* = [ʒ] (not specified, but most likely comparing with other sources)
- *j* is not further explained apart from being pronounced like Spanish, but comparing with the other sources most likely it is [h]
- glottal stop is written as *h*

3.7 La Batea Mixtec [bate]

The data for this variety was shared with us by collaborator and native speaker Yésica Ramírez (Ramirez 2020). The data were collected between 2016-2020 in Oxnard and Santa Barbara, California, as part of a collaboration between the Linguistics department of the University of California Santa Barbara and the Mixteco/Indígena Community Organizing Project in Oxnard, California (Hernández Martínez et al. 2021). The orthography used is latin-based.

- there are palatalized consonants: *ti*V = [tʲ], *tsi*V = [tsʲ]
- *y* = [j]
- *x* = [ʃ]

3.8 Cahuatache Mixtec [cahu]

The Cahuatache Mixtec data originally come from Schultze-Jena (1938), a collection of early travel accounts. It is reproduced in Josserand (1983) and Dürr (1987). Although a valuable resource, the interpretation of the material is not easy, especially regarding the tone notation. Therefore, I only use the data as reproduced in the later collections, not from the original.

- *y* is resolved to [j], based on the information on other varieties spoken in Guerrero

3.9 Chalcatongo de Hidalgo Mixtec [chal]

The data for this variety come from three sources: Josserand (1983), Macaulay (1996), and Swanton & Mendoza Ruíz (2021).

Macaulay (1996:19) uses an Americanist notation, with a few idiosyncrasies:

- *b* is described as a voiced stop and written as such, although it is pronounced either [ᵐb] or [β] depending on the position
- nasalization is marked with a tilde

Swanton & Mendoza Ruíz (2021) write in IPA with small modifications:

- *y* is a vowel [ʉ] in one entry (LAZY)

3.10 San Agustín Chayuco Mixtec [chay]

The data for this variety come from three sources Pensinger et al. (1974), Josserand (1983), and Dürr (1987). Pensinger et al. (1974:137) use a Spanish-based orthography and tones are marked only for a few entries. The same is true for Josserand (1983), while those from Dürr (1987) are marked for tone throughout. The Josserand (1983) entries could be from a different, nearby village, given that there are slight but regular segmental differences, such as [tʃ] where the others have [s].

- *h* = [ʔ]
- *z* in one word only 'child' is 'pronounced like in Spanish'; based on the other sources, this must mean pronounced like Castillian Spanish, i.e. as [θ]; in other lexical items it seems to be [s]
- *x* = [ʃ] (only in Pensinger et al. 1974)

3.11 San Juan Coatzospam Mixtec [coat]

The data for this variety come from three sources: Josserand (1983), Dürr (1987), and Small (1990). Tones are not marked in Josserand (1983), but they are marked in the other two sources.

Small (1990:268) uses a latin-based system:

- *d* is the voiced fricative [ð], there's also a palatalized version *dy* = [ðʲ]
- *h* is the glottal stop

3.12 San Juan Colorado Mixtec [colo]

The data for this variety come from Josserand (1983) and Stark et al. (1986).

Stark et al. (1986:145) use a Spanish-based orthography:

- *x* = [ʃ], cf. *"x representa un sonido semejante al de la ch pero pronunciada sin pegar la lengua a los alveolos"*
- *j* = [h] and y = [j]; it is only mentioned that they are pronounced like Spanish but comparing to Josserand (1983) these values make the most sense

3.13 San Juan Diuxi Mixtec [diux]

The data for this variety come from three sources: Josserand (1983), Dürr (1987), and Kuiper & Oram (1991).

Kuiper & Oram (1991:186) use a latin-based system

- *y* is a voiced fricative [ʒ]
- *x* = [x], *sh* = [ʃ]
- laryngeal h (glottal stop)

3.14 San Martín Duraznos Mixtec [dura]

The data for this variety comes from two sources: Josserand (1983) and the ongoing documentation project of Sandra Auderset and native speaker Carmen Hernández Martínez (see the deposit on ELAR: Auderset & Hernández Martínez 2022). Tones are only marked on a few entries in Josserand (1983), but they are marked throughout in the newer source. In the current documentation project, we use a latin-based orthography in which the mid tone is unmarked. This variety exhibits a rare split between alveo-palatal and palato-alveolar sibilants and affricates, which is not reflected in Josserand (1983).

- palato-alveolar: *sh* = [ʃ], *ch* = [tʃ], and *nch* = [ⁿdʒ]
- alveolo-palatal: *x* = [ɕ], *tx* = [tɕ], and *ntx* = [ⁿdʑ]
- *y* = [ʒ]
- *ku*V = [kʷ]
- glottal stop is written as apostrophe

3.15 San Marcos de la Flor Mixtec [flor]

The data for this variety were shared with us by collaborator and native speaker Moisés. The data were collected between 2019-2020 in Oxnard and Santa Barbara, California as part of a collaboration between the Linguistics department of the University of California Santa Barbara and the Mixteco/Indígena Community Organizing Project in Oxnard, California (Hernández Martínez et al. 2021). This variety is closely related to Piedra Azul Mixtec and has the same phoneme inventory. It also uses the same writing system (refer to Section 3.23 for details).

3.16 San Miguel El Grande Mixtec [gran]

The data for this variety come from two sources: Josserand (1983) and Dürr (1987) and conversion details are covered in the respective sections (3.1 and 3.2).

3.17 Santo Domingo Huendío Mixtec [huen]

The data for this variety come from Becerra Roldán (2015). The author uses a latin-based system and tones are marked throughout with diacritics.

- *x* = [ʃ] and y = [ʒ]
- *j* = [x]
- *ku*V = [kʷ]

3.18 Santiago Jamiltepec Mixtec [jami]

The data for this variety come from two sources: Josserand (1983) and Johnson (1988).

Johnson (1988:18) uses a latin-based system:

- *y* is the semivowel [j]

- *x* = [x]
- there are palatalized consonants: ty = [tʲ] and ndy = [ⁿdʲ]

3.19 El Jicaral Mixtec [jicl]

The data for this variety was shared by JN Martin (Martin 2020). The orthography used is latin-based and tones are marked throughout.

- *x* = [ʃ]
- *y* = [j]
- *ku*V = [kʷ]
- there are palatalized consonants: *ti*V [tʲ] and *nti*V [ⁿdʲ]

3.20 San Sebastián del Monte Mixtec [mont]

The data for this variety come from two sources: collaborator and native speaker Juvenal Solano (Solano 2020) and Josserand (1983). Tones are not marked in Josserand (1983). The data from collaborator Juvenal Solano were collected between 2016-2020 in Oxnard and Santa Barbara, California as part of a collaboration between the Linguistics department of the University of California Santa Barbara and the Mixteco/Indígena Community Organizing Project in Oxnard, California (Hernández Martínez et al. 2021). A latin-based system is used:

- *x* = [ʃ] and *y* = [j], *j* = [h]
- there are palatalized consonants: *ty* = [tʲ] and *ndy* = [ⁿdʲ]

3.21 Magdalena Peñasco Mixtec [mpen]

The data for this variety come from an extensive, recently published dictionary by Hollenbach (2017) using a latin-based system. The graphemes are listed in Hollenbach (2017:xix-xxi). Information for the corresponding IPA sounds has also been taken from the grammar by the same author (Hollenbach 2013:9-14, 16-18).

- *d* = [ð]
- *j* = [h]
- *x* = [ʂ]
- *y* = [ʒ] ([ʃ] at the beginning of words)
- digraphs: *ku*V = [kʷ], *nd* = [ⁿd] and *tn* = [tⁿ]
- there are palatalized consonants: *ti*V = [tʲ] and *si*V = [sʲ] (the latter does not appear in the collected entries)

3.22 Santo Tomás Ocotepec Mixtec [ocot]

The data for this variety comes from three sources: Josserand (1983), Dürr (1987) and Alexander (1988).

Alexander (1988:170) uses a latin-based system:

- *y* is a voiced fricative, i.e. [ʒ], *x* = [x]
- glottal stop is written as *h*

3.23 Piedra Azul Mixtec [piaz]

The data for this variety was provided by linguist and native speaker Inî G. Mendoza with Simon L. Peters (Mendoza & Peters 2020). Our collaboration is part of the MILPA initiative, a collaboration between the Linguistics department of the University of California Santa Barbara and the Mixteco/Indígena Community Organizing Project in Oxnard, California (Hernández Martínez et al. 2021). The orthography used is latin-based and tones are marked throughout.

- *j* = [h]
- *x* = [ʃ]
- *y* = [j]
- *ku*V = [k^w]
- palatalized consonants: *ki*V = [k^j], *ti*V = [t^j], *tsi*V = [ts^j], *ntsi*V = [$^n ts^j$]

3.24 San Jerónimo Progreso Mixtec [prog]

The data for this variety come from three sources Josserand (1983), Dürr (1987), and Shields (1988). The IPA correspondences for Shields (1988) were identified based on the phonology outline provided in North & Shields (1977). Note that the sources refer to this variety as Silacayoapan Mixtec, using the name of the municipality. They all specify, though, that the data were gathered in the town of San Jerónimo Progreso.

Shields (1988:312) uses a latin-based orthography:

- *h* = [ʔ]
- *x* = [h]
- *y* is a voiced fricative, i.e. [ʒ]
- palatalized consonants: ky = [k^j], xy = [h^j], kwy = [k^{wj}]

In Josserand (1983) and Dürr (1987), some graphemes are used differently:

- *h* represents [h] and *x* = [x]

3.25 Tepango Mixtec [tepa]

The data for this variety come from three sources Josserand (1983) (see 3.1), Dürr (1987) (see 3.2), and Hills (1990). All three sources refer to the variety as 'Ayutla Mixtec', referencing the municipality that Tepango is part of. They all specify that the data were gathered with speakers in the village of Tepango. This is one of only two Mixtec varieties that retains final glottal stop, the other being Santa María Zacatepec.

Hills (1990:8) uses a latin-based system:

- palatalized consonsants: *ty* = [t^j], *ny* = [n^j]
- glottal stop is *h*
- *y* is semivowel [j]

In Josserand (1983) and Dürr (1987), some graphemes are used differently:

- *y* is semivowel [j], but it appears twice at the end of words in Josserand (1983) and in both cases Hills (1990) has [ũ]; because of this and because Mixtec words cannot end in a consonant, I convert it to [i] in this position
- *ɛ* is neutralized to [e]

3.26 San Pedro and San Pablo Teposcolula Mixtec [tepo]

For this variety there are historical records in the form of a colonial era vocabulary by de Alvarado (1962 [1593]). This source uses a Spanish-based writing system and tones are not marked. The variety is also represented in Josserand (1983), who standardized and interpreted the entries to some degree. More recently, Swanton (2021) provides an updated analysis and interpretation of this historical source and we follow him in almost all aspects of the conversion from orthography to IPA. For details, please see the original publication (Swanton 2021). A few more tricky equivalences are summarized here:

Phoneme	Grapheme(s)
ɨ	*e*
i	*i, y*
u	*u, #v*
ð	*dz*
ⁿd	*nd, #d, #nd*
ʃ	*s*
k	*c, q, qu*
kʷ	*qu, cu*
w	*vu, u, #hu*
j	*y*
ʔ	*ø, h*

3.27 Tlahuapa Mixtec [tlah]

The data for this variety come from collaborator and native speaker Griselda Reyes Basurto (Reyes Basurto 2020). The data were collected between 2016-2020 in Oxnard and Santa Barbara, California as part of a collaboration between the Linguistics department of the University of California Santa Barbara and the Mixteco/Indígena Community Organizing Project in Oxnard, California (Hernández Martínez et al. 2021). Reyes Basurto (2020) uses a latin-based system and tones are marked throughout.

- palatalized consonants: *ty* = [tʲ]
- *x* = [ʃ] and *y* = [j]

3.28 Xochapa Mixtec [xoch]

The data for this variety come from Stark et al. (2013), who use a latin-based orthography. Spanish loans, however, are written in Spanish orthography Stark et al. (2013:85-90).

- prenasalized: *nd* = [ⁿd]
- *ku*V = [kʷ]
- *x* = [ʃ] and *y* = [j]

3.29 Yoloxochitl Mixtec [yolo]

The data for this variety come from Josserand (1983) and Amith & Castillo García (n.d.). Amith uses a latin-based system:

- prenasalized: *nd* = [ⁿd]

- *x* = [ʃ] and *y* = [j]
- palatalized consonants: *ti*V = [tʲ]
- here *w* = [β] in Josserand (1983)

3.30 San Pedro Yosoñama Mixtec [yosn]

The data for this variety come from Gittlen (2016)'s grammar, which refers to the variety as 'Mixteco del norte de Tlaxiaco' (Mixtec of northern Tlaxiaco), but details that the data were collected in the town of San Pedro Yosoñama. A latin-based system is used, but tones are not marked. The orthography is explained only in relation to Spanish (Gittlen 2016:3-10).

- *x* is [ʃ]
- *y* is [ʒ]
- glottal stop is marked by apostrophe

3.31 Santiago Yosondúa Mixtec [yoso]

The data for this variety comes from two sources: Josserand (1983) and Farris (1992).

Farris (1992:8) uses a latin-based system:

- *x* = [x] and *y* = [j]
- palatalized consonants: *shy* = [ʃʲ]
- preaspirated nasals: *Nn* = [ʰn], *Ññ* = [ʰɲ]

3.32 Yucunani Mixtec [yucn]

The data for this variety was shared by collaborator and native speaker Jeremías Salazar in collaboration with Guillem Belmar, among others (Salazar et al. 2021). The orthography used is latin-based and tones are marked throughout.

- prenasalized: *nt* = [ⁿd], *nch* = [ⁿtʃ]
- *ku*V = [kʷ]
- *x* = [ʃ] and *y* = [j]

3.33 Yucuquimi de Ocampo Mixtec [yucq]

The data for this variety come from three sources: Josserand (1983), León Vázquez (2017) and Swanton & Mendoza Ruíz (2021).

- *b* = [β]
- *y* = [y]
- *kɥ* = [kʷʲ] (this conversion is done mostly to make comparison easier)

3.34 San Andrés Yutatío Mixtec [yuta]

The data for this variety come from an extensive recent dictionary Williams et al. (2017). The orthography used in this source is latin-based and tones are marked throughout. It is presented in a chart with little further explanation (Williams et al. 2017:x-xiv). However, a more detailed description is provided in the pedagogical grammar by the same author (Ferguson de Williams 2007:9-20) and we base our interpretation on this grammar.

- *d* = [ð]
- *j* = [h]
- *y* = [j] 'se pronuncia como suena la ll de *calle* al estilo oaxaqueño' (it is pronounced like the ll of 'street' in the Oxacan style)
- *x* = [ʃ]
- glottal stop is apostrophe

3.35 Santa María Zacatepec Mixtec [zaca]

The data for this variety comes from three sources: Josserand (1983), Swanton & Mendoza Ruíz (2021), and Towne (2011). This variety is very important for reconstruction because it preserves final glottal stop. However, this feature is missing from a lot of entries in Towne (2011).

- *y* = [j]
- *ku*V =[k^{w}]
- *x* = [ʃ] (x does not appear in the other two sources, so there is no overlap)
- there is one entry with *j* = [h], which had to be adjusted by hand (because it is used as [j] in the other sources)

References

Alexander, Ruth Mary. 1980. *Gramática mixteca de Atatlahuca.* México, D.F.: Instituto Lingüístico de Verano, A.C.

Alexander, Ruth Mary. 1988. A syntactic sketch of Ocotepec Mixtec. In C. Henry Bradley & Barbara E. Hollenbach (eds.), *Studies in the syntax of Mixtecan languages*, vol. 1, 151–304. Dallas & Texas: Summer Institute of Linguistics and the University of Texas at Arlington.

de Alvarado, Francisco. 1962 [1593]. *Vocabulario en lengua mixteca. Reproducción facsimilar con un estudio de Wigberto Jiménez Moreno.* México, D.F.: Instituto Nacional Indigenista e Instituto Nacional de Antropología e Historia.

Amith, Jonathan D. & Rey Castillo García. n.d. Recursos lexicosemánticos para el mixteco de Yoloxóchitl, municipio de San Luis Acatlán, Guerrero (Glottocode yolo1241; ISO 639-3 xty). unpublished.

Anderson, Lynn. 2006. *Vocabulario de palabras que se relacionan con el maíz en mixteco de Alacatlatzala, Guerrero.* México, D.F.: Instituto Lingüístico de Verano, A.C. 2nd edn.

Auderset, Sandra & Carmen Hernández Martínez. 2022. Documenting Tù'un Na Ñuu Sá Matxí Ntxè'è, a Mixtec language of Oaxaca, Mexico. Endangered Languages Archive. https://www.elararchive.org/dk0629/.

Becerra Roldán, Braulio. 2015. *Un estudio fonológico del mixteco de Santo Domingo Huendio, Oaxaca.* México, D.F. ENAH MA thesis.

Carroll, Lucien Serapio. 2015. *Ixpantepec Nieves Mixtec word prosody*: University of California San Diego PhD dissertation.

Castillo García, Rey. 2007. *Descripción fonológica segmental y tonal del mixteco de Yoloxóchitl, Guerrero.* México, D.F. CIESAS MA thesis.

Dürr, Michael. 1987. A preliminary reconstruction of the Proto-Mixtec tonal system. *Indiana* 11. 19–61.

Farris, Edwin R. 1992. A syntactic sketch of Yosondúa Mixtec. In C Henry Bradley & Barbara E Hollenbach (eds.), *Studies in the syntax of Mixtecan languages*, vol. 4, 1–171. Summer Institute of Linguistics and the University of Texas at Arlington.

Galindo Sánchez, Bernardo. 2009. *Vocabulario Básico Tu'un Savi - Castellano.* Xalapa, Veracruz: Academia Veracruzana de las Lenguas Indígenas 1st edn.

Gerfen, Henry James. 1996. *Topics in the phonology and phonetics of Coatzospan Mixtec*: University of Arizona PhD dissertation.

Gittlen, Laura. 2016. *Gramática popular Mixteco del norte de Tlaxiaco.* México, D.F.: Instituto Lingüístico de Verano.

Hernández Martínez, Carmen, Eric W. Campbell & Griselda Reyes Basurto. 2021. MILPA (Mexican Indigenous Language Promotion and Advocacy): A community-centered linguistic collaboration supporting Indigenous Mexican languages in California. In Justyna Olko & Julia Sallabank (eds.), *Revitalizing Endangered Languages: A Practical Guide*, 216–217. Cambridge University Press.

Hills, Robert A. 1990. A syntactic sketch of Ayutla Mixtec. In C. Henry Bradley & Barbara E. Hollenbach (eds.), *Studies in the syntax of Mixtecan languages*, vol. 2, Summer Institute of Linguistics and the University of Texas at Arlington.

Hinton, Leanne, Gene Buckley, Marv Kramer & Michael Meacham. 1991. Preliminary analysis of Chalcatongo Mixtec tone. In *Papers from the American Indian Languages Conference, University of California, Santa Cruz, July and August*, 147–155.

Hollenbach, Barbara Elena Erickson. 2013. *Gramática del mixteco de Magdalena Peñasco (Sa'an Ñuu Savi).* Tlalpan & Mexico: Instituto Lingüístico de Verano, A.C.

Hollenbach, Barbarba E. 2017. *Diccionario mixteco de Magdalena Peñasco.* Instituto Lingüístico de Verano, A.C.

Iverson, Gregory K & Joseph C Salmons. 1996. Mixtec prenasalization as hypervoicing. *International Journal of American Linguistics* 62(2). 165–175.

Johnson, Audrey F. 1988. A syntactic sketch of Jamiltepec Mixtec. In C Henry Bradley & Barbara E Hollenbach (eds.), *Studies in the Syntax of Mixtecan languages*, vol. 1, 11–150. Summer Institute of Linguistics and the University of Texas at Arlington.

Josserand, J. Kathryn. 1983. *Mixtec dialect history*: Tulane University PhD dissertation.

Kaufman, Terrence. in press. Comparative Oto-Mangean grammar research: Phonology, aspect-mood marking, valency changers, nominalizers on verbs, numerals, pronouns, deictics, interrogatives, adpositionoids, noun classifiers, noun inflexion, compounds, word order, and diversification model. In Søren Wichmann (ed.), *Languages and linguistics of Mexico and Northern Central America: a comprehensive guide*, De Gruyter Mouton.

Kuiper, Albertha & Joy Oram. 1991. A syntactic sketch of Diuxi-Tilantongo Mixtec. In C Henry Bradley & Barbara E Hollenbach (eds.), *Studies in the syntax of Mixtecan languages*, vol. 3, 185–408. Summer Institute of Linguistics and the University of Texas at Arlington.

León Vázquez, Octavio. 2017. *Sandhi tonal en el mixteco de Yucuquimi de Ocampo, Oaxaca.* México, D.F. CIESAS MA thesis.

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

Macaulay, Monica Ann. 1996. *A grammar of Chalcatongo Mixtec.* Berkeley: University of California Press.

Martin, J N. 2020. Diccionario del Mixteco de El Jicaral y Otros Pueblos Fronterizos de Guerrero y Oaxaca. unpublished.

McKendry, Inga. 2013. *Tonal association, prominence and prosodic structure in South-Eastern Nochixtlán Mixtec.* Edinburgh: University of Edinburgh PhD dissertation.

Mendoza, Inî G. & Simon L. Peters. 2020. Vocabulario del Tù'un Sàjvĭ (Mixteco) de Piedra Azul y Paredón, San Martín Peras. unpublished.

Mendoza Ruíz, Juana. 2016. *Fonología segmental y patrones tonales del Tu'un Savi de Alcozauca de Guerrero.* México, D.F. CIESAS MA thesis.

North, Joanne & Jäna Shields. 1977. Silacayoapan mixtec phonology. In William R. Merrifield (ed.), *Studies in Otomanguean Phonology*, 21–33. Summer Institute of Linguistics and the University of Texas at Arlington.

Pensinger, Brenda J et al. 1974. *Diccionario Mixteco: mixteco del este de Jamiltepec, pueblo de Chayuco.*

México, D.F.: Instituto Lingüístico de Verano, A.C.

Pike, Eunice V & John H Cowan. 1967. Huajuapan Mixtec phonology and morphophonemics. *Anthropological Linguistics* 9(5). 1–15.

Pike, Eunice V & Joy Oram. 1976. Stress and tone in the phonology of Diuxi Mixtec. *Phonetica* 33(5). 321–333.

Ramirez, Yesica. 2020. Vocabulario del Mixteco de San Juan Mixtepec. unpublished.

Reyes Basurto, Griselda. 2020. Diccionario del Mixteco de Tlahuapa, Guerrero, Mexico. unpublished.

Salazar, Jeremías, Saleem Alfaife, Guillem Belmar Viernes, Eric W. Campbell, Gabriel Mendoza, Olguín-Martínez Jesús, Griselda Reyes Basurto, Catherine Scanlon, Giorgia Troiani & Alonso Vásquez-Aguilar. 2021. Vocabulario de Yucunani Sà'án Sàvĭ (Mixteco). unpublished.

Schultze-Jena, Leonhard. 1938. *Bei den Azteken, Mixteken und Tlapaneken der Sierra Madre del Sur von Mexico*, vol. 3 Indiana. Jena: Gustav Fischer.

Shields, Jana. 1988. A syntactic sketch of Silacayoapan Mixtec. In C. Henry Bradley & Barbara E. Hollenbach (eds.), *Studies in the Syntax of Mixtecan Languages*, vol. 1, 305–449. Summer Institute of Linguistics and the University of Texas at Arlington.

Small, Priscilla C. 1990. A syntactic sketch of Coatzospan Mixtec. In *Studies in the Syntax of Mixtecan languages*, vol. 2, 261–479. Summer Institute of Linguistics and the University of Texas at Arlington.

Solano, Juvenal. 2020. Vocabulario del Mixteco de San Sebastián del Monte. unpublished.

Stark, Sharon C., Audrey P. Johnson & Filiberto Lorenzo Cruz. 1986. *Diccionario Mixteco de San Juan Colorado.* México, D.F.: Instituto Lingüístico de Verano, A.C.

Stark, Sharon C., Audrey P. Johnson & Benita González de Guzmán. 2013. *Diccionario básico del mixteco de Xochapa, Guerrero.* Instituto Lingüístico de Verano, A.C. 3rd edn.

Swanton, Michael. 2021. Un acercamiento a la ortografía dominica del mixteco de Teposcolula: un enfoque comparativo. In Michael Swanton (ed.), *Filología Mixteca: Estudios sobre textos virreinales*, 1–76. México, D.F.: Instituto de Investigaciones Filológicas UNAM.

Swanton, Michael & Juana Mendoza Ruíz. 2021. Observaciones sobre la diacronía del tono en el Tu'un Savi (mixteco) de Alcozauca de Guerrero. In Francisco Arellanes & Lilián Guerrero (eds.), *Estudios lingüísticos y filológicos en lenguas indígenas mexicanas: Celebración de los 30 años del Seminario de Lenguas Indígenas*, Ciudad de México: Universidad Nacional Autónoma de México.

Towne, Douglas. 2011. *Gramática popular del tacuate (mixteco) de Santa María Zacatepec, Oaxaca.* Instituto Lingüístico de Verano, A.C.

Ferguson de Williams, Judith. 2007. *Gramática popular del mixteco del municipio de Tezoatlán, San Andrés Yutatío, Oaxaca.* Instituto Lingüistico de Verano, AC.

Williams, Judith F., Gerardo Ojeda Morales & Liborio Torres Benavides. 2017. *Diccionario mixteco de San Andrés Yutatío, Tezoatlán, Oaxaca.* Ciudad de México: Instituto Lingüístico de Verano, A.C.

Zylstra, Carol F. 1991. A syntactic sketch of Alacatlatzala Mixtec. In C Henry Bradley & Barbara E Hollenbach (eds.), *Studies in the Syntax of Mixtecan languages*, vol. 3, 1–177. Summer Institute of Linguistics and the University of Texas at Arlington.

Zylstra, Carol F. 2012. *Gramática del Tu'un Savi (la lengua mixteca) de Alacatlatzala, Guerrero.* Instituto Lingüístico de Verano, A.C.