# *Subgrouping of Coahuitlán Totonac*[1]

DEVIN MOORE
*University of Alberta*
devin.moore@ualberta.ca

## Abstract

Coahuitlán Totonac is spoken in Veracruz, Mexico, and has been variously ascribed to two different branches of the Totonacan family tree. While recent work has begun to bring empirical evidence to the internal structure of this family tree, there remain several important areas of disagreement, in addition to the disputed affiliation of Coahuitlán. This article informs the family tree and demonstrates that Coahuitlán belongs to the Northern branch using shared innovations and two computational methods. The comparative method seeks sets of shared innovations for evidence of subgrouping. This article presents proposed shared innovations in phonology, morphology, and lexicon, which fall into two sets, one belonging to the Sierra and Lowland branches, and the other belonging to the Northern. Coahuitlán Totonac overwhelmingly shares innovations found in Northern languages and lacks innovations found in Sierra. Two quantitative methods are also used to show that Coahuitlán groups groups closely with other Northern languages.

**Keywords:** historical linguistics, subgrouping, Totonacan, Mesoamerican languages

## Résumé

Coahuitlán Totonac est parlé à Veracruz au Mexique et s'est vu assigné à deux branches différentes de l'arbre familial Totonacan. Malgré les travaux récents qui portent de nouveaux faits empiriques concernant la structure interne de cet arbre familial, plusieurs sujets inspirent encore la controverse. À l'aide d'innovations communes et de deux méthodes computationnelles, cet article éclaircit l'arbre familial et montre que Coahuitlán appartient à la branche du nord. La méthode comparative cherche des innovations communes pour établir des sous-groupes. Cet article présente des innovations communes aux niveaux phonologiques, morphologiques, et lexiques qui se divisent en deux groupes, l'un appartenant aux branches Sierra et Lowland, l'autre appartenant à la branche du nord. Coahuitlán Totonac présente surtout ces

innovations caractéristiques des langues du nord et manque des innovations attestées en Sierra. Deux méthodes quantitatives sont également employées afin de montrer que le groupe Coahuitlán est similaire aux autres langues du nord.

**Mots clés:** linguistique historique, sous-groupe, Totonac, langues méso-américaines

## 1.   INTRODUCTION

Coahuitlán Totonac (Ch) is a largely undescribed language spoken in the municipality of Coahuitlán, Veracruz, Mexico, by about 3,800 speakers (SEFIPLAN 2013). Coahuitlán is located on the border of two branches of the Totonac family, Northern and Sierra (see Figure 1), and the variety has been variously described as belonging to one or the other. It is first mentioned in the literature as one of the communities marking the northern boundary of Sierra Totonac, where Aschmann (cited in Ichon 1973) describes it as part of the Sierra branch. Brown et al. (2011) tentatively place Coahuitlán (there spelled 'Cohuahuitlán') in the Northern branch, based on reports of higher mutual intelligibility between speakers of Coahuitlán and speakers of Upper Necaxa Totonac. My own fieldwork in Coahuitlán confirms that speakers there consider Upper Necaxa Totonac more intelligible than the varieties spoken in Coyutla or Filomeno Mata. MacKay and Trechsel (2011) also place Coahuitlán in the Northern branch, on the basis of a few morphological patterns. Ethnologue (Lewis 2015) does not treat Coahuitlán Totonac as a variety per se, instead grouping it with nearby Filomeno Mata, and claiming that "Filomeno Mata-Coahuitlán" is "linguistically between" Northern and Sierra. McFarland (2009) claims that Filomeno Mata Totonac is a highlands, or Sierra, variety, although with characteristics of both Northern and Sierra branches; and that it is distinct from Coahuitlán Totonac. Speakers of Coahuitlán Totonac do not consider Filomeno Mata Totonac to be the same variety, and report low mutual intelligibility between the two varieties. This article will demonstrate that Coahuitlán belongs to the Northern branch by proposing shared innovations in phonology, morphology, and lexicon. Given the relative absence of systematic comparative work, the cognate sets and shared innovations presented here will also contribute to a better understanding of the larger Totonacan language family. Two quantitative analyses are also presented, which support the conclusions of the traditional method.

Section 2 provides a brief background of Totonacan languages and their traditional and more recent classification. Section 3 presents the primary evidence for my argument in the form of proposed shared innovations. The section begins with the theory of subgrouping by shared innovations, and examines phonological (3.1),
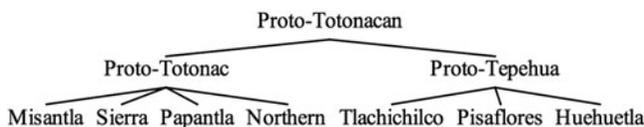


**Figure 1:** Traditional classification

morphological (3.2), and lexical (3.3) isoglosses to posit shared innovations relevant to subgrouping. Because of the relative homogeneity of phonological and morphological systems across Totonac languages, shared lexical innovations offer the most productive source of evidence. Section 4 presents additional evidence based on computational methods: first, lexical similarity evidence using ASJP algorithms (4.1), and second, a phylogenetic network of Totonacan languages (4.2). Conclusions are discussed in Section 5. Appendix A provides a complete list of the languages and their abbreviations as used in this article, and the primary data sources.

## 2.   BACKGROUND

Totonacan languages are spoken in eastern-central Mexico in the states of Puebla, Veracruz, and the eastern edge of Hidalgo, in the Sierra Madre Oriental and along the Gulf Coast. Totonacan is a well-established language family, but there has been little systematic work on its internal subgrouping. The traditional classification, represented in Figure 1, is based on various hypotheses put forward by early fieldworkers (McQuown 1940, Arana Osnaya 1953, García Rojas 1978, Ichon 1973, Levy 1987, and MacKay 1994).

   The traditional classification divides the family into two main branches, Tepehua and Totonac. There are three Tepehua languages – Huehuetla, Pisaflores, and Tlachichilco. The exact number of Totonac languages is unknown, though estimates range from three or four (MacKay 1999) to between 14 and 20 (Brown et al. 2011). These are grouped into four branches: Misantla, Northern, Lowland (or Papantla), and Sierra. Misantla Totonac is spoken in a few communities in the area south of the major urban centre of Misantla, Veracruz. Lowland is spoken by a number of communities around Papantla, Veracruz, and along the Gulf coast. The Lowland branch is sometimes called Papantla Totonac, which also refers to the only documented language in the Lowland Branch. Northern and Sierra are both spoken in many communities in the Sierra Madre, mostly in the state of Puebla. Figure 2 shows a map of some of the larger Totonacan communities.

   This classification was constructed with little empirical basis. While the general lack of documentary and comparative work in the family poses an ongoing challenge, there are recent attempts to study the classification of Totonacan languages empirically. MacKay and Trechsel (2011, 2015) have investigated morphological patterns of Totonacan languages. Brown et al. (2011), attempt to reconstruct proto-Totonacan roots for comparison with other language families. This work has begun to provide support for the basic structure of the family, but significant disagreements remain concerning the relationships within these higher level branches, and the assignment of individual languages to specific branches, including the disputed affiliation of Coahuitlán. Brown et al. (2011) present a tentative classification, summarised in Figure 3, that is based largely on lexicostatistical analysis carried out by the ASJP consortium (Wichmann et al. 2013) instead of on informed determinations of cognacy.
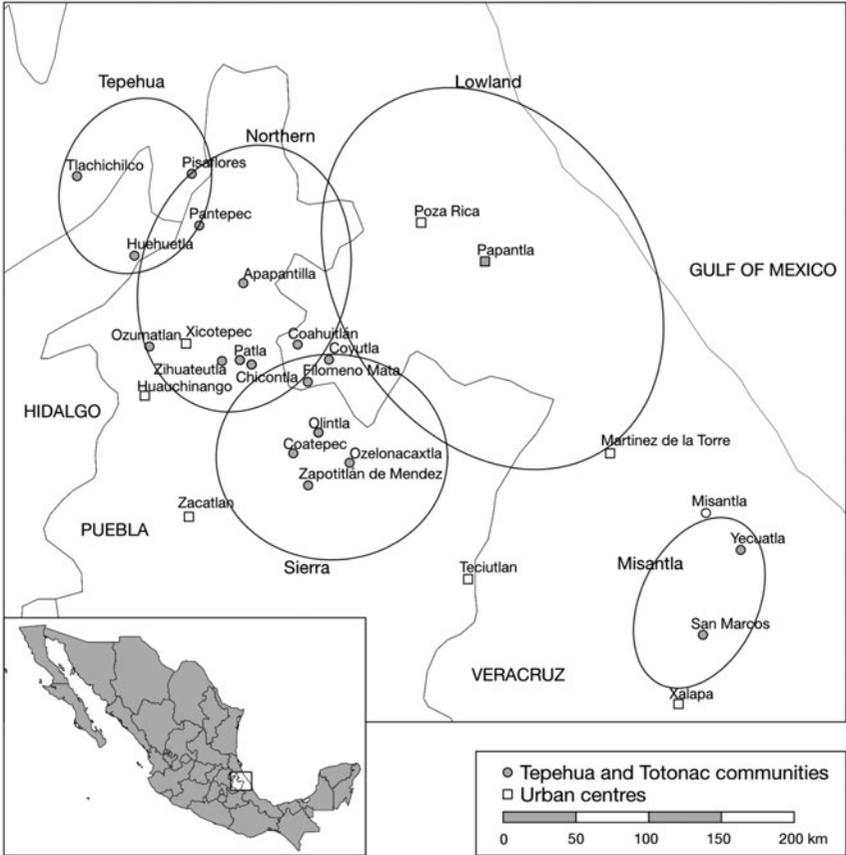
**Figure 2:** Map of Totonacan communities

The main features of this classification involve further subgrouping of the Totonac branches. Misantla has been set off from what they call Central Totonac, which includes Northern further set against Lowland-Sierra. They do not include
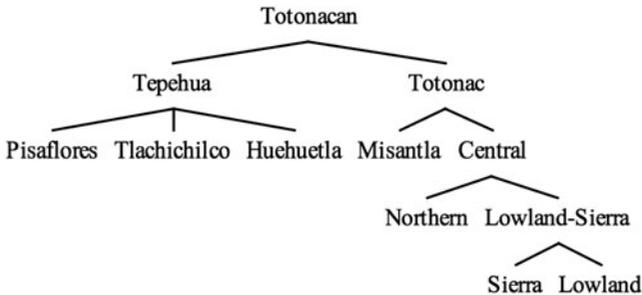


**Figure 3:** Brown et al.'s (2011) Classification

data from Ch, but have grouped it with the Northern branch based on reported high mutual intelligibility with Northern varieties. MacKay and Trechsel (2015) investigate phonological, morphological, and lexical data. They present three morphological features of Sierra languages, the presence of which they consider to be necessary and sufficient to determine Sierra affiliation. They conclude that there is some support for the traditional classification generally; however, they suggest a Sierra vs Northern-Lowland distinction at odds with Brown et al.'s proposed Northern vs. Sierra-Lowland, noting that although Sierra and Lowland share some lexical items, Lowland does not have the morphological features of Sierra languages. They include Coahuitlán in the Northern branch based on the fact that it lacks their three features of Sierra languages. While the present article tends to support a Northern vs. Sierra-Lowland split, it aligns with MacKay and Trechsel in placing Coahuitlán in the Northern branch, here by means of shared phonological, morphological, and lexical innovations; lexical similarity evidence from ASJP and a phylogenetic network created in SplitsTree4.

Focusing on the Sierra, Lowland, and Northern languages (Central), I present two sets of shared innovations, one shared by Northern languages, and another shared by Sierra-Lowland languages. Coahuitlán Totonac overwhelmingly shares innovations found in Northern languages and lacks innovations found in Sierra.

## 3. SHARED INNOVATIONS

The primary criterion for subgrouping is shared innovation (Fox 1995, Campbell 2013). A shared innovation is a feature belonging to a subset of daughter languages that set it off from other members of its family. The shared innovation is assumed to have occurred in an intermediate proto-language, which then diversified into the subset of languages which all share the innovated feature. Languages with the shared innovation inherited it from this intermediate proto-language, while languages without the shared innovation do not descend from that intermediate parent. However, there are reasons besides shared innovation that a subset of languages may have features in common. Three alternate scenarios[2] that may result in similarity but do not give evidence for shared subgrouping are 1) shared retention, 2) parallel innovation, and 3) contact.

Shared retention refers to a feature of the proto-language that has been retained in a subset of daughter languages. Retention alone does not give evidence for subgrouping because many scenarios may have lead to individual languages keeping or losing some particular feature – closely related languages may differ in retention of some features, and distantly related languages may have independently retained other features. A simple example of shared retention is the presence of *škaːn* in Tepehua, Misantla, and Northern branches (see section 3.3.1, below). The fact that this word

---

[2]Chance, sound-symbolism, and so-called "nursery forms" are other possible sources of similarity.

| H   |           | Ch  | ta:wan   |
| --- | --------- | --- | -------- |
| Pf  |           | Co  | ƛa:wan-  |
| T   |           | Ct  | ƛa̰:wán  |
| M   | tana:nan  | FM  | ƛa:wan   |
| A   | ƛa:wan    | Z   | ƛa̰:wán  |
| U   | ła̰:wán[4] | P   | ƛa̰:wán  |

**Figure 4:** 'to walk around'

has been retained in each branch does not support any subgrouping of these branches into a single unit.

Parallel innovation, also called language drift (Sapir 1921), is when innovations occur independently in different branches of a set of related languages. Often these are due to typologically common processes, like the devoicing of word-final stops, which may have existed in the proto-language, and undergo independent innovation in a subset of daughter languages. However, whether or not a daughter language undergoes this innovation is not dependent on other daughter languages, and does not give evidence for subgrouping. An example in Totonacan is the case of the alveolar lateral affricate *ƛ. This undergoes the same diachronic shift in two distant branches of the family, as seen in the cognate set for 'to walk around', shown in Figure 4.[3]

In Ch and M, *ƛ has changed to /t/. One might initially posit that Ch and M form a subgroup together, but this one piece of evidence is at odds with the geographic distance and great lexical dissimilarity between Ch and M. Further investigation reveals that in other languages, /ƛ/ is described as unstable. Levy's (1987) phonology of Papantla Totonac notes that, while prosodic evidence points to /ƛ/ as a phoneme, there is considerable variation by word and by utterance in Papantla, with a wide range of possible realizations including [ƛ], [ł], [t], [lʔ], and even an oddly metathesized [łt]. This variation by word and utterance is attested in other languages as well. While at first this may have been taken to be evidence of a relationship between two distant and dissimilar varieties, *ƛ was a good target for sound change, which occurred independently in M, and Ch.

---

[3]A full list of language abbreviations is found in Appendix A. This article uses an Americanist form of IPA commonly used by Totonacists, with the following notable differences from IPA: c = voiceless alveolar affricate, ƛ = voiceless lateral affricate, y = palatal approximant, ꞉ after a vowel indicates length, ˯ under a vowel indicates laryngealization, ' after a consonant indicates ejectivization, ´ above a vowel indicates stress. I have tried to make the transcriptions from different sources uniform, according to IPA and these differences.

[4]In U, *ƛ has undergone merger with /ł/.

|          | A              | U              | Ch      |
|----------|----------------|----------------|---------|
| 'boy'    | qawačU         | ʔawáča[5]      | qo:č    |
| 'make or do' | ƛawa    ƛo:    | ɬawá    ɬo:    | to:     |
| 'sit down' | tawi̧lá    to:ļá | tawilá    to:la | to:la   |
| 'go up, climb' | tawa̧ká́    to:ká̧ | tawaká    to:ka | to:ka   |
| 'be closed' | lakčo:        | lakčawá    lakčo: | lakčo: |
| 'dry'    | skawawa        | skawáwa̧        | sko:wa  |
| 'take'   | tayá           | tayá    te:     | te:     |

**Figure 5:** Glide reduction

The last scenario is horizontal transfer, or contact. A language of one branch that has borrowed a feature innovated by another branch might appear to have shared the innovation rather than just sharing the feature. A borrowing from an unrelated language appearing in languages from different branches could likewise appear to give evidence of an innovation shared by the languages possessing it. The influence of Spanish on Mesoamerican languages is an example where contact with an outside language has resulted in numerous shared forms in multiple languages that result not from genetic descent, but from borrowing. Another example of contact is what appears to be a dialect-chain phenomenon in Totonacan. Vowel-glide-vowel sequences in some languages are reduced to one long vowel. This is most often /awa/ to /o:/ and /aya/ to /e:/, as shown in Figure 5, but can also include other vowels, as seen here with /awi/ in the case of A 'sit down'.

The /awa/ form is well attested in other branches, but while this sound change is attested in the lexica for only these three Northern varieties, which are all geographically close, there are two reasons to believe that it does not represent a shared innovation exclusive to the Northern branch. First, the sound change is attested in both A and U, but is in fact restricted to just one of two dialects of U. While this sound change has run its course in Ch, which does not allow the unreduced forms, in A and U (Patla), the reduced and unreduced forms appear in variation conditioned largely by rate of speech, with the reduced form appearing in faster, more fluent speech and the unreduced form in slower, more emphatic, or elicited speech. However, in U (Chicontla) has only the unreduced form, and this sound change does not occur (Beck, p.c.). If this were an innovation shared across the Northern branch, one would not expect such a dialectal difference in U. Second, while reduced forms do not appear in the lexica for Sierra languages, they do occur in the speech of many Sierra languages, conditioned by rate of speech (Levy, p.c.). In addition to the spread of this change reported in Sierra, there is a similar phenomenon

---

[5]U has lost uvular stops and *q becomes /ʔ/.

| H | ʔušint'i | Ch | wiš |
|---|---|---|---|
| Pf | ušíndi | Co | wiš |
| T | ušínt'i | Ct | wiš |
| M | wiš | FM | wiši |
| A | wiš | Z | wi̩š |
| U | wiš | P | wiš |

**Figure 6:** 'you (sg)'

involving glide reduction which is found in Tepehua languages, as shown in the cognate set for 'you (sg)', in Figure 6.

Where Totonac *wiš* 'you (sg)' begins with a glide-vowel sequence, this seems to have been reduced to a single vowel in Tepehua languages. As in Totonac, the place features of the reduced form seem to have been affected by the glide. Glide reduction seems to have some features of a dialect chain, and may be conditioned by contact between neighbouring languages.

Because of these potentially confusing factors, a comparison of shared features in a family is likely to give conflicting evidence for groupings. A single piece of evidence does not give the full picture, and the best evidence for internal relations is *sets* of multiple shared innovations, just as the best evidence for genetic relatedness is sets of correspondences. Ultimately, these sets of innovations allow the reconstruction of a proto-form from which the development of each language's synchronic form may be traced.

The first step to identifying shared innovations is to assemble isoglosses. The term isogloss originally referred to a line that could be drawn on a map to represent the geographical boundaries of regional linguistic variants based on the distribution of particular dialectal features. It is used here by extension to refer to the dialect features themselves. The distribution of these features can provide evidence for shared innovations. Each of the following sections includes a grammatical or lexical feature that is shared by a subset of the varieties under examination, beginning with a single phonological isogloss (section 3.1), morphological isoglosses (section 3.2), and lexical isoglosses (section 3.3). Ch patterns closely with other Northern languages, sharing all but one of six Northern innovations, and sharing only two of 18 Sierra innovations (section 3.4).

### 3.1  Phonological isoglosses

One type of shared innovation that can be used to establish phylogenetic proximity between two languages is regular sound change. This type of innovation is especially useful because it is relatively salient and often affects large numbers of items. Totonac languages have relatively homogenous phonological systems, and nearly identical reflexes populate many cognate sets. This is perhaps due to the shallow time depth of the family, and a relative paucity of phonological innovation. It may

|      | 'road' | 'hail'  | 'good'        | 'cough'  | 'eight'  |
|------|--------|---------|---------------|----------|----------|
| H    | ti:    | čaxʔiit | qox           |          |          |
| Pf   | tih    | čaxʔíti | óši           | uxʔún    | =cajín   |
| T    | ti:n   | čahʔiti | o:ši          | uhʔún    | cahín    |
| M    | ti:ł   | čehet   |               | ʔɔhɔ     | ciyan    |
| A    | tex    | čexet   | cex / ceya    | o̲xonu̲:n | =ceyan   |
| U    | tex    | čexé:t  | cey / céya̲    | o̲xo:nún  | =cayán, =ceyén |
| Ch   | te:    | čexet   | cé:           | o̲xonun   | =ceyán   |
| Co   | tihi   | če:he:t | ƛa:n / la:n   |          |          |
| Ct   | tixiʔ  | čixí:t  | ƛa:n          | kuxu:nan | cayan    |
| FM   | tíxʔi  | čáxati  | ce:           | qoxo:nun | cayán    |
| Z    | tihi̲   | číhi:t  | ƛa:n          | ku̲hu:nán | =cayán   |
| P    | tixia  | číxi:t  | ƛa:n / ła:n   | ku̲xu:nán | =cayán   |

**Figure 7:** Cognates with unconditioned /e/ and /o/

also be due in part to obscuring factors, such as parallel development and horizontal transfer, as in the cases of *ƛ and /awa/, seen above.

There is one phonological isogloss that provides evidence for a shared Northern innovation, which is also shared by Ch. While Proto-Totonacan has been described as having a three-vowel system (/i/, /u/, and /a/), most varieties have mid vowels, /e/ and /o/, as allophones of /i/ and /u/, respectively (Brown et al. 2011). These vowels are conditioned primarily by proximity to /q/ or proto *q. Brown et al. (2011) claim that the emergence of phonemic /e/ and /o/ is a distinguishing feature of the Northern branch, supported by the appearance of /e/ and /o/ in U and A forms without a clearly identifiable conditioning environment, although they note that proximity to *x, and to a lesser extent to *y, may be a possible conditioning environment for the development of /e/ and /o/ in Northern. I do not exhaustively argue here for the phonemic status of /e/ and /o/ in Ch, but advance a number of cognate sets, shown in Figure 7, where /e/ and /o/ appear in this environment of proximity to *x and *y, but not to *q.

In these glosses, A, U, and Ch consistently pattern together with regard to /e/ and /o/, while the other languages typically have /i/ or /a/ instead of Northern /e/, and /u/ instead of Northern /o/. Assuming that the three-vowel system traditionally reconstructed for proto-Totonacan is correct, this supports the conclusion that Northern languages – including Ch – share the innovation of phonemic /e/ and /o/.

## 3.2   Morphological isoglosses

Like sound changes, changes in morphology may also provide evidence of shared innovations that can be used as examples of affinity and subgrouping. I first consider three isoglosses suggested by MacKay and Trechsel (2011, 2015): marking of 2nd-person subject and 1st-person object (section 3.2.1), unselective 3rd-person plural agreement marker (section 3.2.2), and word-final [y] (section 3.2.3). Following that, I look at three isoglosses from Beck (2012): the locative (section 3.2.4), the desiderative (section 3.2.5), and the negative (section 3.2.6).

### 3.2.1   Marking of 2nd-person subject and 1st-person object

All Totonacan languages have agreement for the person and number of both subject and object on transitive verbs. MacKay and Trechsel (2015) observe that a distinguishing morphological feature of Sierra languages is what they refer to as "unambiguous marking" of 2nd-person subjects with 1st-person objects. Tepehua, Misantla, Northern and Lowland languages use a non-compositional pattern of affixes to express second person subject and first person object agreement when either subject, object, or both is plural, as in (1):[6]

(1)   U
   kila:musuːyáːuw
   ki–laː–musuː–ya–w
   1OBJ–RECIP–kiss–IMPFV–1PL.SUBJ
   'you (sg.) kiss us', 'you (pl.) kiss me', 'you (pl.) kiss us'

(Beck 2004: 34)

This combination of morphemes expresses any of three scenarios: 2nd singular subject and 1st plural object, 2nd plural object and 1st singular subject, and 2nd plural subject and 1st plural object.

   Sierra languages have innovated distinct, unambiguous affixal sequences for these combinations of persons and numbers that specify the number of both subject and object, as in (2):

(2)      Ct
   a. kinkaːpaːškiːyaʔ
      kin–kaː–paːškiː–ya–ʔ
      1OBJ–PL.OBJ–love–IMPFV–2SG.SUBJ
      'you (sg.) love us'

   b. kimpaːškiːyatín
      kin–paːškiː–ya–tin
      1OBJ–love–IMPFV–2PL.SUBJ
      'you (pl.) love me'

---

[6]The abbreviations used in the interlinear glosses are as follows: 1, 2, 3 = first-, second, third-person; DCS = decausative; FUT = future; IMPFV = imperfective; INC = inchoative; IRR = irrealis; OBJ = object; PFV = perfective; PL = plural; RECIP = reciprocal; SG = singular; SUBJ = subject; TOT = totalitive.

    c.  kinka:pa:ški:yatín
        kin–ka:–pa:ški:–ya–tin
        1OBJ–PL.OBJ–love–IMPFV–2PL.SUBJ
        'you (pl.) love us'

<div align="right">(McQuown 1990: 166, 169; interlinear gloss added)</div>

These Sierra constructions are unambiguous and compositional. Ch has the non-compositional construction analogous to that in (1) above, as shown in (3).

(3)  Ch
    kila:pucayá:w
    ki–la:–puca–ya–w
    1OBJ–RECIP–search–IMPFV–1SUBJ.PL
    'you (sg.) look for us', 'you (pl.) look for me', 'you (pl.) look for us'

As in most Totonacan languages, Ch uses 1st-person object agreement, the reciprocal marker, and 1st-person plural subject agreement to mark agreement for the three scenarios with 2nd-person subject and 1st person object when one or both are plural. Ch thus lacks the Sierra innovation.

### 3.2.2    3rd unselective plural /qu:/

The second feature said by MacKay and Trechsel to be typical of Sierra languages is the 3rd-person plural agreement marker, /-qu:/ or /-qu̱:/. This agreement marker is unselective in that it does not select for subject or object, instead marking agreement with a subject or object that is 3rd-person plural, as in (4).

(4)  Ol
    lkapá:stákqó:h
    laka=pa:stak–qu:–ya
    remember.X–PL–IMPFV
    'she/he/it remembers them', 'they remember her/him/it', 'they remember them'

<div align="right">(MacKay and Trechsel 2015: 20)</div>

In this example, /qu:/ marks agreement with an unspecified 3rd-person plural argument, allowing the ambiguous meaning of either a 3rd-person plural subject, a 3rd-person plural object, or both.

    The source of this marker in Sierra varieties is the pan-Totonac terminative or totalitive suffix, which marks "the termination of an event, and/or that all participants in an event have been affected" (Beck 2012: 593), as in (5).

(5)  U
    ta̱knu:ʔo̱:łcá̱
    ta–a̱k–nu:–ʔo̱:–li=cá'
    DCS–head–in–TOT–PFV=now
    'he sank in completely'

<div align="right">(Beck 2011: 61)</div>

In this example, /ʔo̱:/ acts as the totalitive and conveys aspectual meaning, namely, that he sank all the way, completely.

Ch does not exhibit this Sierra innovation. Ch /qo:/ is not used as an agreement marker, but instead acts as a true totalitive (6a). Different patterns mark third-person plural agreement of subject (6b) and object (6c).

(6)     Ch

   a)  makaskakqo:ɬ
       maka–skak–qo:–li
       hand–dry–TOT–PFV
       'she finished drying her hands'

   b)  talaqcín
       ta–laqcin
       3PL.SUBJ–see
       'they see her/him/it'

   c)  ka:laqcín
       ka:–laqcin
       PL.OBJ–see
       'she/he/it sees them'

The pattern in (6b) and (6c) is the same as in Northern languages for analogous sentences. Ch lacks the Sierra innovation of /qo:/ as a person agreement marker; instead /qo:/ is used as a totalitive marker, and there are different affix patterns to mark 3rd-person plural agreement of subject and object.

### 3.2.3   Word-final sonorant [y]

MacKay and Trechsel's final Sierra isogloss is a palatal sonorant, [y], which occurs word-finally as the marker of the imperfective aspect for vowel-final stems, shown in (7).

(7) Ct
    taštúy
    ta–štu–ya
    INC–outside–IMPFV
    'she/he exits'

                                                              (MacKay and Trechsel 2015: 26)

In some varieties, this final /y/ is realized as an aspirated palatal fricative or glottal fricative, as in (8).

(8)     Oz

   a)  ɬtatáh
       ɬtata–ya
       sleep–IMPFV
       'she/he sleeps'

       Ol

   b)  ɬtatáç
       ɬtata–ya
       sleep–IMPFV
       'she/he sleeps'

                                                              (MacKay and Trechsel 2015: 26)

The source of this sound after vowels seems to be the pan-Totonacan imperfective morpheme /-ya/, which is realized in Sierra languages at the end of verbs as [y] if there are no other morphemes following it. In non-Sierra varieties, such as U, (9a), the imperfective morpheme is entirely unrealized in this position.

(9)     U

    a)  taštú
        taštu– yaː
        exit–IMPFV
        'she/he exits'

    b)  taštuyaːtít
        taštu–yaː–tít
        exit–IMPFV–2PL.SUBJ
        'you (pl.) exit'

<div align="right">(Beck 2004: 34; interlinear gloss added)</div>

In (9a), the verb is imperfective, but no form of /-yaː/ surfaces, while in (9b), the imperfective marker appears between the stem and the 2nd-person plural subject marker. Beck (2004) describes this as a morphophonemically conditioned syncope, where the imperfective morpheme is realized only in cases where it is "protected" by a following morpheme.

    Ch behaves like U, and does not realize the imperfective morpheme in unprotected position (10).

(10)    Ch

    a)  kaci
        kaci–ya
        exit–IMPFV
        'she/he knows'

    b)  kaciyaːtít
        kaci–yaː–tít
        exit–IMPFV–2PL.SUBJ
        'you (pl.) know'

Sierra languages share this feature, and Ch does not pattern accordingly. However, if /yaː/ is the source of this feature in Sierra, it is odd for an innovation to restore phonological material that has been lost in other branches. There seems to be a familial drift towards final syncope and devoicing, which may have been arrested in the Sierra branch for some reason. Alternatively, the word final [y] may have a different source, perhaps created by some kind of phonotactic process in final stressed vowels.

### 3.2.4   Locative

Totonacan languages have a locative prefix or clitic, which differs in form across the family, as shown in Figure 8.

    Focusing on the Northern, Sierra and Papantla varieties, there are two forms, *nak-* as a prefix or clitic, and the prefix *k-*, likely a phonologically reduced form.

| H | laqa:-, laka:- | Co | nak- |
|---|---|---|---|
| Pf | laka- | Ct | nak- |
| T | laqa:-, laka:- | CX | nak= |
| M | pu:- | FM | k- |
| A | nak- | Ol | k- |
| U | nak= | On | k- |
| Ch | nak= | Z | nak- |
| | | P | nak, k= |

**Figure 8:** Locative

| H | -putun | Co | -pu̠tun |
|---|---|---|---|
| Pf | -putun | Ct | -putun |
| T | -putun | CX | -pu̠tun |
| M | pan- | FM | -kutun |
| A | -ku̠tun | Ol | -putun |
| U | -ku̠tun | On | -pu̠tun |
| Ch | -ku̠tun | Z | -pu̠tun |
| | | P | -pu̠tún |

**Figure 9:** Desiderative

The reduced form seems to be limited to Sierra varieties (FM, Ol, On) and P, and may represent a Sierra-Lowland innovation. U has a different reduction, in some cases reducing *nak=* to *n(a)=* or even *ŋ=* (Beck, p.c.), *nak=* to *n(a)=* is also reported in Apapantilla (Reid 1991: 76). Ch does not have the Sierra-Lowland innovation, as it does not allow shortening to *k-*.

### 3.2.5 Desiderative

Totonacan languages have a desiderative marker, which attaches to verbs to indicate the subject's desire for the completion of the action of the verb (Beck 2004). There are three forms for this marker: *-pu̠tun/-putun, pan-*, and *-ku̠tun/-kutun*, as shown in Figure 9.

The form *pan-* is exclusive to M, and has not been etymologically connected to the other forms. Tepehua, Sierra, and Papantla (H, Pf, T, Co, Ct, Ol, On, Z, P) have *-putun*, while Northern varieties (A, U, Ch, FM) have *-ku̠tun.*[7] It seems likely that

---

[7]Totonac forms, such as *-pu̠tun* and *-ku̠tun*, which are given in the prose to refer to different glosses present in groupings of the varieties, are not intended to represent reconstructed forms.

| H | ha:ntu | Co | ni: |
|---|--------|-----|-----|
| Pf | ha:ntu | Ct | ni: |
| T | ha:ntu | CX | yi̧:ntú: |
| M | la: | FM | ła:= |
| A | tu̧: | Ol | ni: |
| U | xa: | On | ni: |
| Ch | la: | Z | ni: |
| Zh | ła: | P | ni: |

**Figure 10:** Negative

-*ku̧tun* represents a shared Northern innovation – the replacement of /p/ by /k/ – because the Tepehua and Sierra languages both share the form -*pu̧tun*. Ch shares this Northern innovation.

### 3.2.6   Negative

There is considerable formal variation across Totonacan varieties in the negative morpheme, as shown in Figure 10.

There seem to be three etyma, one belonging to Tepehua, one to Northern and Misantla, and one to Sierra. The reflex in CX is similar to Tepehua, as is that in A. If these forms are shared retentions, the Tepehua etyma likely represents the proto-Totonacan form, but these forms are also possible borrowings, as A and CX, a variety whose affiliation, like that of Ch, has not been definitively determined, are both somewhat adjacent to Tepehua. The *ła:/la:/xa:* forms in Misantla, Northern (U, Ch, Zh), and FM seem to represent a Totonac innovation, perhaps originally *la:, as it remains in M and Ch, changing to *ła:* in FM and Zh. U *xa:* is possibly derived from this, a hypothesis reinforced by the presence of an archaic form *ła:*. Another hypothesis is that these forms came from *xa: or *ha:, perhaps taken from the first part of *ha:ntu*, although the directionality of the sound change from /x/ to /ł/ and /l/ seems less plausible. The Sierra languages Co, Ct, Ol, On, Z, and Lowland P have the form *ni:*, an apparent Sierra-Lowland innovation that Ch does not share.

### 3.3   Lexical isoglosses

The lexical isoglosses were drawn from a manually compiled list of some 180 cognate sets from lexica for A, Ct, Co, FM, H, M, P, Pf, T, U, and Z, and my fieldwork in Ch. Of these sets, around one third were largely homogenous across Totonac and Tepehua. Another third provide strong evidence for a Tepehua-Totonac divide.

Rather, they should be read as convenient shorthand referring to groups with the same or very similar forms.

Of the remaining cognates, some are obscured by lateral transfer, like the word 'chicken' for which many languages use some form of Spanish *pollo*, and some simply don't present a clear picture, such as the 3rd-person singular pronoun 'he or she'. While the Tepehua forms are all similar to each other, the Totonac varieties have one of two forms, one that comes from a possessive form and another that seems to come from a demonstrative form. The distribution of these two forms does not seem to correspond to any *a priori* grouping. However, there are 16 isoglosses, which give clear and relevant evidence to the Central Totonac divisions of Lowland, Sierra, and Northern.

### 3.3.1    'water' and 'rain'

There are isoglosses for 'water' with two forms, *ška:n* and *čučut*, distributed as in Figure 11.

The form *čučut* 'water' is restricted to the Sierra-Lowland languages (Co, Ct, FM, Z, P), while *ška:n* is found in Tepehua, Misantla and Northern (H, Pf, T, M, A, U, Ch). This distribution suggests a Sierra-Papantla innovation, which is not shared by Ch.

The word *ška:n* has the further meaning of 'rain (n)' in those languages having this form. This meaning is seen most commonly in a periphrastic construction with *min* 'to come', which is found in Siera and Lowland, Northern, and Tepehua languages, as illustrated in (11).

(11)  Pf
      šká:n kaminá?
      ška:n   ka–min–ya–?
      water   IRR–come–IMPFV–FUT
      'it will rain'

<div align="right">(MacKay and Trechsel 2013: 209)</div>

Interestingly, in those languages with *čučut* 'water', the word does not have this second meaning of 'rain'. With the exception of FM, these languages use another form, *saʔin/se:n*, as in Figure 12.

A reflex of this form is found in M, A, and U; however, in these languages the etymon has the meaning of 'thunderstorm, downpour', a difference in meaning

| H  | ška:n | Ch | ška:n  |
|----|-------|----|--------|
| Pf | ška:n | Co | čúčut  |
| T  | ška:n | Ct | čučut  |
| M  | ška:n | FM | čúčuti |
| A  | ška:n | Z  | čúčut  |
| U  | ška:n | P  | čúčut  |

**Figure 11:** 'water'

| H   |         | Ch  |              |
|-----|---------|-----|--------------|
| Pf  |         | Co  | si:n         |
| T   |         | Ct  | si:n         |
| M   | *sę:n*  | FM  | sa?ín, sayín |
| A   | *ta:sąi̯:n* | Z | se:n, si:n   |
| U   | *tą:są?i̯:n* | P | sę:n         |

**Figure 12:** 'rain'

signalled by italics in Figure 12. Since Misantla and Northern languages share this meaning, this is possibly the meaning of the proto-Totonac form. The Sierra and Lowland languages would thus share a lexical innovation in changing the meaning to 'rain'.

Additionally, some of these languages have innovated a verb 'to rain' from the nominal form *si:n*, which is used instead of the *min* construction. Co, Ct, and Z all have a form *si:nan* or *se:nan*, 'to rain', derived from the nominal form by the suffix *-nan*, described as a detransitivizer that changes transitive verbs into intransitive, but also forms verbs "from nouns or adjectives denoting processes or activities strongly associated with the meaning of the root" (Beck 2004: 64). P uses the *min* construction with *sę:n*. FM also uses the *min* construction with *sayín* (McFarland, p.c.).

The forms *čučut* 'water' and *si:nan* 'to rain', and the shift in meaning of *si:n* from 'downpour' to 'rain' all represent Sierra-Lowland innovations. Ch has the conservative *min* construction, and uses *ška:n* to mean both 'water' and 'rain'. It lacks all three Sierra-Lowland innovations.

### 3.3.2   'sand'

There are two roots for 'sand', one of which is restricted to Sierra and Lowland. The other is found across every other branch. These are shown in Figure 13.

| H   | ku:ku:  | Ch  | kukú                 |
|-----|---------|-----|----------------------|
| Pf  | kúku    | Co  | muncaya              |
| T   | kúkuh   | Ct  | kuku                 |
| M   | kukɔ:h  | FM  | kúku, muncáy?a       |
| A   | kukux   | Z   | kuku:, kúkuh, muncaya̧ |
| U   | kukúx   | P   | muncaya              |

**Figure 13:** 'sand'

The form *kuku* is found in all four branches, including the most divergent divisions Tepehua and Misantla, which argues for it being the proto-form. The form *muncaya* is found only in the Sierra and Lowland languages, and is likely a Sierra-Lowland innovation not shared by Ch, and also not shared by Ct, which instead has the conservative form found in other Northern languages.

### 3.3.3   'see'

There are two roots for 'see', shown in Figure 14.

Tepehua, Misantla, and Northern share the form *laqci:n*, while the Sierra-Lowland languages, with the exception of FM, have the innovated form *akšila/a̰kšiła*, unique to that branch. Ch has the conservative form, not the Sierra-Lowland innovation.

### 3.3.4   'ear'

There are two forms for 'ear', one restricted to the Sierra-Lowland varieties, the other found across the rest of the family, as shown in Figure 15.

The distribution of (*q*)*aqašoł* across Tepehua, Misantla, and Northern suggest that it is the proto-form, while *ta:qé:n* is a Sierra-Lowland innovation, not shared by Ch.

| H | laqc'in | Ch | laqcín |
|---|---------|----|--------|
| Pf | la?c'ín | Co | ukšiła |
| T | laqc'ín | Ct | ?akšił- |
| M | la̰qḛn | FM | la:qcin |
| A | laqcḭ:n | Z | a̰kšiła |
| U | la̰?cín | P | akšila |

**Figure 14:** 'see'

| H | ?aqašoł | Ch | aqášqoł |
|---|---------|----|---------|
| Pf | ?a?aš?ół | Co | taqé:n |
| T | ?aqášqoł | Ct | taqa:n |
| M | qa̰qašqɔł | FM | taqé:n |
| A | a̰qašqoł | Z | taqá:n, taqé:n |
| U | a̰?aš'ół | P | taqá:n, taqé:n |

**Figure 15:** Nominal form of 'ear'

| H  | qay      | Ch | qaṯa          |
|----|----------|----|---------------|
| Pf | ʔay      | Co | langa, ƛanga  |
| T  | ʔay      | Ct | ƛankaʔ        |
| M  | qaṯ      | FM | lankʔa        |
| A  | qaƛa     | Z  | ƛanka         |
| U  | ʔáƚa     | P  | ƛanká, langa  |

**Figure 16:** 'big'

### 3.3.5  'big'

The words for 'big' are presented in Figure 16.

While the Tepehua and Misantla forms bear some resemblance to the Northern forms, there is clearly a new innovated form in the Sierra languages. Because the form *langa/ƛanka* is restricted to Sierra-Lowland languages, it is likely a shared innovation. Ch does not share this innovation.

### 3.3.6  'chest'

The isoglosses for 'chest' include two roots, shown in Figure 17.

There is a clear Tepehua-Totonac division, with two Totonac forms, both based on *kuš*. It is not clear which Totonac form is innovative, but in this case, Ch patterns with the Sierra and Lowland languages, not with Northern.

### 3.3.7  'liver'

There are three related forms for 'liver', shown in Figure 18.

Tepehua and Northern have one form, which includes the prefix *mak-* 'body', a combining bodypart prefix. Its presence in Tepehua and Northern suggests shared retention. This form is similar to the Misantla word, but M lacks the 'body' prefix. The Sierra languages Ct, FM, Z and Lowland P have another form, lacking the body prefix and with an extra final syllable. Ch does not share the Sierra and Lowland innovation, rather having the conservative form shared by Tepehua and Northern.

| H  | tankilhaka | Ch | kušmu:n   |
|----|------------|----|-----------|
| Pf | tankilakna | Co | kušmú:n   |
| T  | tankilákna | Ct | kušmú:n   |
| M  |            | FM | kušmú:n   |
| A  | kušan      | Z  | kušmú:n   |
| U  | kušan      | P  | kušmú:n   |

**Figure 17:** 'chest'

| H | makłp'aka | Ch | makłwák |
|---|---|---|---|
| Pf | makłwáka | Co | |
| T | makłúk'a | Ct | łwakaka? |
| M | huka̠:k | FM | łwakák?a |
| A | | Z | łwaka̠ka̠ |
| U | makłwáka̠ | P | łwaka̠ka̠ |

**Figure 18:** 'liver'

| H | šqoy | Ch | páłma |
|---|---|---|---|
| Pf | | Co | pałma, tuwá:n |
| T | šqoy | Ct | tawa:n, tuwa:n |
| M | paqala̠t | FM | pałma, tuwá:n |
| A | | Z | tu̠wá̠:n |
| U | pá?łma̠ | P | tu̠wá̠:n |

**Figure 19:** 'leaf'

### 3.3.8 'leaf'

There is a very clear Sierra-Lowland isogloss for 'leaf', Figure 19.

Tepehua and Misantla have unique forms, and U, Co, and FM have the form *pa̠?łma̠/pałma*.[8] The form *tuwa:n* is found only in Sierra and Lowland languages and seems to be an innovation proper to these branches. Although *tuwá:n* is present in FM, it is a specific type of leaf; *pałma* is the generic word 'leaf'. Ch does not share this Sierra-Lowland innovation, instead using *páłma*, as does U.

### 3.3.9 'fire'

The isoglosses for 'fire' include three distinct etyma, shown in Figure 20.

The form *łkúya:t* in Sierra and Lowland is likely the innovative form because Northern *makskut* is quite close to Misantla *mukskut*. Ch does not share the Sierra-Lowland innovation, though it has modified the form slightly and lost the /k/, with compensatory lengthening.

### 3.3.10 'girl'

The word for 'girl' has three forms, shown in Figure 21.

---

[8]The form in U, FM, and Ch *pa̠?łma̠/pałma* bears some resemblance to Spanish *palma* 'palm tree'. However, it seems to be derived from a Totonac verb *pa̠?ł-* 'to bloom, flower, sprout' with suffix *–ma* 'by-product' (Beck, p.c.).

| H   |          | Ch  | ma:skút    |
|-----|----------|-----|------------|
| Pf  | híkmi    | Co  | łkúya:t    |
| T   | híkmi    | Ct  | łkuya:t    |
| M   | mukskut  | FM  | łkúya:ti   |
| A   | makskut  | Z   | łkúya:t    |
| U   | makskút  | P   | łkúya:t    |

**Figure 20:** 'fire'

| H   |          | Ch  | cumaxá:t   |
|-----|----------|-----|------------|
| Pf  | hac'i?   | Co  | cuma:t     |
| T   | ha:cí?i  | Ct  | cuma:t     |
| M   |          | FM  | cumaxá:ti  |
| A   | cumąxa:t | Z   | cumą:t     |
| U   | cumąxá:t | P   | cuma:t     |

| H   | ?aci?    |

**Figure 21:** 'girl'

Setting aside the Tepehua *xac'i?*, the forms in Northern and Sierra-Lowland are closely related. It seems that Sierra (Co, Ct, Z) and Lowland (P) have lost the penult syllable. FM has the unreduced from, as does Ch.

### 3.3.11     'tomorrow'

Three forms of 'tomorrow' are observed, all of which seem to have a common root *li-* or *łi-*. These forms are shown in Figure 22.

The two different Totonac forms seem to consist of a body part prefix: *laqa-* 'eye' in M, A, U, and Ch; and *čạ:* 'shin' in Ct, FM, Z, and P. As it is shared by

| H   | łi?      | Ch  | laqalí:        |
|-----|----------|-----|----------------|
| Pf  | łi       | Co  | čạ:lí, čạ:líh  |
| T   | łiy      | Ct  | ča:lí          |
| M   | laqali:  | FM  | ča:lí          |
| A   | laqali:  | Z   | čạ:lí          |
| U   | lạ?alí:  | P   | čạ:lí          |

**Figure 22:** 'tomorrow'

| H | ʔaɬunut /ʔaɬunuʔ | Ch | nakú |
|---|---|---|---|
| Pf | haɬúnuti | Co | nakú / naku̯h |
| T | haɬánuti | Ct | nakú |
| M | la:kạci:n | FM | nakú |
| A | akɬkunuk | Z | nakú̦: / naku̯h |
| U | ạkɬku̦núk | P | nakú |

**Figure 23:** 'heart'

Northern and Misantla, *laqa-* may represent a Totonac innovation, with the further Sierra-Lowland innovation being to use *čạ·-* instead. Ch uses the same form as other Northern languages and does not have the Sierra innovation.

### 3.3.12    'heart'

The word 'heart' has four forms, shown in Figure 23.

The Tepehua and Northern etyma may be related, while Misantla has its own form. The form *nakú* clearly belongs to Sierra-Lowland. In this case, Ch has the Sierra-Lowland form, and not the Northern.

### 3.3.13    'nose'

There are two Totonac forms for the body part 'nose', shown in Figure 24.

The Northern languages U and A have a form distinct from that found in the Sierra-Lowland languages. Typically the nominal form of a body part is formed from the combining form with the morpheme *-ni̦* following consonants, or *-n* following vowels (Beck 2004). In the case of 'nose', the combining form is quite important. Tepehua, Northern, Sierra, and Lowland languages share the combining form *kinka- / kanka-*. This suggests that Sierra has retained the expected form *kinkan / kankan*, while Misantla and Northern have reduced this form. The alternative hypothesis would be that Misantla and Northern have reflexes of an irregular independent

| H | ɬi:šin | Ch | kinkán |
|---|---|---|---|
| Pf | ɬišín | Co | kinkán |
| T | líšin | Ct | kankan |
| M | ki̦ʔ | FM | kinkán |
| A | kini̦ | Z | kankan, kinkán |
| U | kíni̦ | P | kinkán |

**Figure 24:** 'nose'

| H   |                           | Ch  | škaya:wa              |
| --- | ------------------------- | --- | --------------------- |
| Pf  | st'a?a:w                  | Co  |                       |
| T   | štak'áu                   | Ct  |                       |
| M   |                           | FM  | łkayíw?a, škayíw?a    |
| A   | łka̠ya̠:wa̠, ska̠ya̠:wa̠    | Z   |                       |
| U   | łkayá:wa̠, skayá:wa̠       | P   |                       |

**Figure 25:** 'green'

nominal (cf. 'ear', section 3.3.4), which has been regularized in Sierra (and Ch). However, this hypothesis would also require the derivation of *kinka-* / *kanka-* from a much smaller nominal (unlike 'ear', where the combining form *qaqa-* / *aqa-* can easily be derived from the longer irregular nominal *qaqašqoł*/*?aqašqoł*). It seems more likely that Misantla and Northern independently reduced *kinkan* to M *ki̠?* and Northern *kín̠i*. In this case, the isogloss presents a Northern innovation, which Ch does not share.

### 3.3.14  'green'

The colour word 'green' is a clear Northern isogloss, shown in Figure 25.

Other branches have no separate word for the colour 'green' instead conflating green and yellow, as Ct *smukúku* 'light yellow, green' or green and blue Ct *spukuku* 'blue, green'. The Northern languages have reflexes cognate with Ct for yellow and blue, but the form *škaya:wa* 'green' seems to be a Northern innovation. Ch shares this Northern innovation, as does FM.

### 3.3.15  'finger- and toenail, claw'

The words for 'fingernail' and 'toenail' are given in Figure 26.

Many of these forms are created by the addition of a combining body-part prefix and a base. In Tepehua, these are the prefixes *č'an-* 'foot' and *mak-* or *max-* 'hand'. Totonac has the prefixes *tu:-* 'foot' and related prefixes *maq-* (Northern) and *maka-* (Sierra) 'hand', the form *maq-* likely being a syncopated form of *maqa-*. H, Z, and P allow the stem to occur on its own, without a combining prefix, and Ct and FM do not have any form attested with a combining prefix. The stem can be used independently in the Sierra languages and H. Northern languages do not allow the stem to occur independently, which may represent an innovation to restrict the distribution of the stem. In addition to this behaviour, the stem has different forms; Tepehua *qesi:t*, Misantla *-sɛ:h*, Northern *–si:n*, and Sierra-Lowland *siyín/sixán*. Sierra-Lowland has a longer form, which may have changed independently in the other branches. The Tepehua form is only roughly similar, Misantla appears to have lost the final syllable, and Northern has lost the middle and undergone compensatory lengthening. To

'toenail'

| H | č'anqesi:t, qesi:t | Ch | tu:sí:n |
| Pf | | Co | |
| T | c'an'ésiti | Ct | siyín |
| M | tu:sɛ:h | FM | sixán |
| A | tu:sị:n | Z | sẹxán, sịxán |
| U | tu:sí:n | P | tu:sịxan, siyạ́n |

'fingernail'

| H | makqesi:t, qesi:t | Ch | maqsí:n |
| Pf | mah?ésiti | Co | |
| T | mah?ési:ti | Ct | siyín |
| M | maqasɛ:h | FM | sixán |
| A | maqsị:n | Z | makasịyạ́n, sẹxán, sịxán |
| U | mạ?sí:n | P | makasịxán, siyạ́n |

**Figure 26:** 'toe- and fingernail'

| H | qoš | Ch | ce: |
| Pf | ?oši | Co | ƛa:n, la:n |
| T | o:ši | Ct | ƛa:n |
| M | qɔɬaná? | FM | ce: |
| A | cex, ceya | Z | ƛa:n |
| U | cey, céyạ | P | ƛa:n, ła:n |

**Figure 27:** 'good'

summarise, the Northern branch uses the prefixes, *tu:*- and *maq*-, a bound stem, and has compensatory lengthening of the stem. Ch shares each of these features.

### 3.3.16    'good'

There are three forms for 'good', Figure 27.

The form in M, *qɔɬaná?*, is particularly interesting because it appears to resemble both of the very different forms in Tepehua and Sierra-Lowland. It looks like the proto-form was something like the M form, from which Tepehua took the beginning, and Sierra-Lowland the end. Northern *cex* is unrelated and represents a clear Northern innovation, which is shared by Ch.

## 3.4  Summary

The innovations proposed fall into two groups: first, a large group of likely innovations shared by the Sierra and Lowland languages, and a smaller group of probable shared Northern innovations. Table 1 shows each innovation for the Central Totonac languages – A, U, Ch, Co, Ct, FM, Z, and P. The Tepehua languages have no innovations from either group, nor does M. As in the text above, the forms given do not represent reconstructed forms, and are instead used as shorthand for the innovation discussed above.

The Northern languages A and U have none of the Sierra innovations, while most Sierra languages have most of the Sierra innovations. There are two languages, FM and P, in addition to Ch, which are of special interest.

FM is unique in being quite mixed, having 11 of the 20 Sierra innovations and three of the six Northern innovations. MacKay and Trechsel (2015) classify FM as Northern, as it lacks the first three innovations in Table 1, which they identify as

| | A | U | Ch | FM | Co | Ct | Z | P | |
|---|---|---|---|---|---|---|---|---|---|
| 2 Subj (Pl) > 1 Obj (Pl) | – | – | – | – | + | + | + | – | Sierra Innovations |
| Unselective plural /quː/ | – | – | – | – | + | + | + | – | |
| Word-final [y] | – | – | – | – | + | + | + | – | |
| Locative /k-/ | – | – | – | + | – | – | – | + | |
| Negative /niː/ | – | – | – | – | + | + | + | + | |
| *čučut* 'water' | – | – | – | + | + | + | + | + | |
| *siːn* 'rain' | – | – | – | + | + | + | + | + | |
| *siːnan* 'to rain' | – | – | – | – | + | + | + | + | |
| *muncaya* 'sand' | – | – | – | + | + | – | + | + | |
| *akšila* 'to see' | – | – | – | – | + | + | + | + | |
| *taqeːn* 'ear' | – | – | – | + | + | + | + | + | |
| *ƛanka* 'big' | – | – | – | + | + | + | + | + | |
| *kušmun* 'chest' | – | – | + | + | + | + | + | + | |
| *nakú* 'heart' | – | – | + | – | + | + | + | + | |
| *łwakaka* 'liver' | – | – | – | + | – | + | + | + | |
| *tuwaːn* 'leaf' | – | – | – | + | + | + | + | + | |
| *łkuyaːt* 'fire' | – | – | – | + | + | + | + | + | |
| *ƛaːn* 'good' | – | – | – | – | + | + | + | + | |
| *cumaːt* 'girl' | – | – | – | – | + | + | + | + | |
| *čaːlí* 'tomorrow' | – | – | – | + | + | + | + | + | |
| Phonemic /e/ and /o/ | + | + | + | – | – | – | – | – | Northern Innovations |
| Desiderative /kutun/ | + | + | + | + | – | – | – | – | |
| *kini* 'nose' | + | + | – | – | – | – | – | – | |
| *škayaːwa* 'green' | + | + | + | + | – | – | – | – | |
| *siːn* '-nail' | + | + | + | – | – | – | – | – | |
| *ceː* 'good' | + | + | + | + | – | – | – | – | |

**Table 1:** Shared innovations in Central Totonac

morphological isoglosses and features of Sierra languages. While MacKay and Trechsel argue that these are necessary and sufficient criteria, I treat them as innovations shared by the Sierra branch, but not *a priori* more important than any of the other shared innovations identified. Considering the entire list of innovations, FM does not clearly fall into either the Sierra or the Northern branch. It is possible FM belongs to either Sierra or Northern, and the similarities to the other are due to contact. However, though FM is located close to the border between Northern and Sierra languages, other languages near this border, such as Ch and Co, clearly fall into one or the other camp. The status of FM remains uncertain, awaiting further study.

Lowland P has all of the Sierra innovations, except MacKay and Trechsel's (2011, 2015) three Sierra isoglosses. Considering only these three would lead to a clear differentiation between Sierra and Lowland, but the entire list shows that while Lowland P shares none of the Northern innovations, it has all of the other innovations shared by the Sierra languages, including other morphological innovations. While some other features differentiate Lowland, it appears to be very closely related to the Sierra branch. This suggests that Central Totonac can be further divided between Sierra-Lowland and Northern, as suggested in Brown et al. (2011).

Ch shares all but one Northern innovation, but only two of the 20 Sierra innovations. Interestingly, all of the three innovations where it does not align with Northern involve body parts, 'chest', 'heart', and 'nose', though 'heart' is one of the few body parts which does not have a combining prefix and thus does not participate in very common body-part constructions. These differences may be explained by borrowing, although the body has been identified as a semantic field resistant to borrowing (Tadmor et al. 2010). However, Totonac body parts are a formal class defined by co-occurrence with verbal roots in complex verb structures, structures that have a very high frequency (Levy 1999). This high frequency may explain why the two innovated Sierra forms may have been borrowed into Ch, but would not explain 'heart', which does not belong to this formal class. The Northern innovation may have been lost by borrowing back a conservative Sierra form, or may represent a further subdivision in the Northern branch.

## 4. COMPUTATIONAL METHODS

A number of computational tools borrowed from evolutionary biology are becoming increasingly popular in historical linguistics. While promising methodological tools, they are still often used only heuristically in linguistics (Schnoebelen 2009), offering a rapid and efficient method of finding patterns in the data and proposing relationships. In the context of historical linguistics, these tools have been adapted to use measures of phonological or lexical distance, by creating matrices of distance measurements from which to run their analyses, on the assumption that these distances correlate with genetic relatedness. These measures of similarity follow the assumption that high synchronic similarity correlates with genetic closeness. While similarity often does correlate, it does not provide such strong evidence for subgrouping,

because, as outlined in section 3, there are various pathways to synchronic similarity. The analyses and visualizations can nevertheless be informative and help find patterns in the data that may support the conclusions arrived at through more traditional methods.

This article presents two analyses using such distance-based measurements, calculated from different types of data. The first analysis, in section 2.1, is a tree calculated from the 100-word Swadesh lists of 12 Totonacan languages with the Neighbour-Joining algorithm used by the Automated Similarity Judgment Program (ASJP) consortium (Wichmann et al. 2013), which presents trees based on lexical similarity. The second analysis is a phylogenetic network created using SplitsTree4 and the Neighbour-Net algorithm based on binary characters coded from a list of cognate sets (section 2.2).

## 4.1   ASJP lexical distance

The ASJP World Tree is a massive collaborative project that calculates and graphically illustrates the lexical similarity between a large number of the world's languages (Müller et al. 2013). It currently has data, in the form of a shortened 40-word Swadesh list, for 4,401 of the 7557 languages recognized by ISO 639–3. The 40-word list is the result of a calculation to determine the most stable lexical items on the original 200-word Swadesh list (Holman et al. 2008). The World Tree is calculated on the basis of a neighbour-joining algorithm, which uses a distance matrix generated from the word lists to optimize a tree, beginning with equal distances between each node. Each stage of clustering will minimize branch length between nodes to create a cladistic tree (Saitou and Nei 1987), with each branching representing a historical event of splitting or "speciation". The distance matrix is based on the Levenshtein or edit distance, which is the smallest number of edits between two strings. In this article, the ASJP algorithms have been applied to 100-word Swadesh lists drawn from 12 Totonacan languages, including Ch. Figure 28 presents the resulting tree.[9]

The tree generated by ASJP shows important differences from the traditional classification. While there is only one Tepehua language in the ASJP database, T, it is divided from the other Totonac languages at the highest level. The Totonac languages are not split into four main groups; instead the analysis finds three: Misantla, as expected, stands apart, and the Northern languages are set off from Sierra. Lowland is not clearly split from Sierra, with P being grouped closely with Co, Z, and Ct. FM and CX are at the edge of this branch. Geographically, CX is distant from other Sierra languages, but FM is located close to Sierra communities Co and Z. Both are close to Northern languages, FM being one of the varieties located nearest to Ch. Importantly, Ch is grouped with the Northern languages. In this table, Coahuitlán is equally distant from A and U-Zh.

---

[9]The ASJP database has different labels for certain varieties, notably "Highland" for Z, and "Xicotepec" for A; the abbreviations used here are found to the right, beside the ASJP labels.
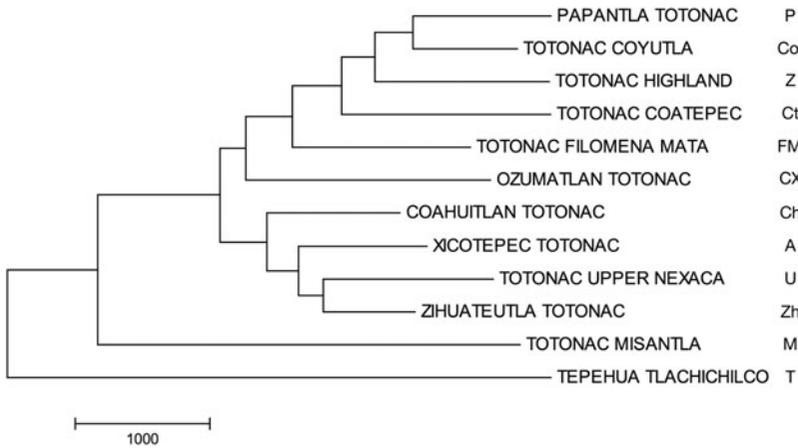
**Figure 28:** ASJP calculation of Totonacan lexical similarity

## 4.2   Phylogenetic network

One weakness in the tree model often used to represent genetic relations is that the analysis of a dataset can often lead to multiple possible trees for any set of data. This can be due to noise or uncertainty in the data, and/or because innovations, which could have been used to identify a branch splitting, may be borrowed between branches after a split. Another representation of phylogenetic relations is a network, which has the advantage of representing multiple possible hypotheses. Networks show something of an average over large, noisy datasets, offering a cumulative visualization of the possible trees, and visually represent the differing amount of support given to groupings in the data (L. Campbell 2013). Trees are derivable from networks, but networks give a richer representation of the data. By way of example, Figure 29 shows a network representation of Romance languages, created using cognates from the Comparative Indo-European Database (Dyen et al. 1997) and SplitsTree4.

Phylogenetic networks like this one transparently represent the distance of the relation between two nodes by the length of the line separating them, and the uncertainty by webbing displayed on the network (Schnoebelen 2009). The length of the lines between taxa indicates their likely closeness, so the tight cluster of Vlach and Rumanian suggest these two languages are closely related. Increased webbing is the result of multiple possible trees. Vlach and Rumanian have little webbing, while Sardinian and Italian have more, reflecting greater uncertainty in the data between the nodes of Sardinian languages and Italian.

While networks provide rich visualizations, they ultimately rely upon the assumption that similarity corresponds with genetic closeness. In this analysis, the data come from manually compiled cognate sets (cf. section 3.3), taken from lexica for A, Ct, Co, FM, H, M, P, Pf, T, U, and Z, and my fieldwork in Ch. This list of cognates is based first on the Swadesh 100 list, with the remaining items
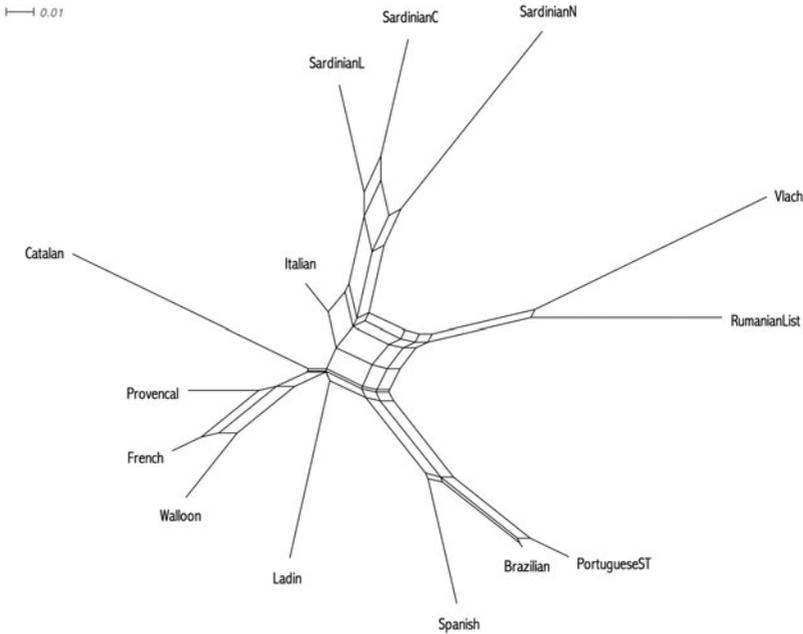
**Figure 29:** Phylogenetic network of romance languages

having been gathered somewhat opportunistically, although including many cultural-
ly central items (Kaufman et al. 2004). To calculate a distance matrix, these cognate
sets are first coded into characters – parameters for which languages may agree or
differ. Each cognate set consists of a gloss, and a number of Totonac stems, or
cognate classes. Basing a character on the gloss would result in a multistate character,
with each cognate class being a possible state. Because phylogenetic networks typ-
ically rely upon binary characters, the cognate sets are made binary by coding
each cognate class as a character, which has the added advantage of allowing the en-
coding of a language with two cognate classes for the same cognate set (Dunn 2014).
A list of 218 characters is encoded from the cognate sets, and a distance matrix is cal-
culated in SplitsTree4 (Huson and Bryant 2006) using the Neighbour Net algorithm
(Bryant and Moulton 2004). An example of this encoding is seen in Figure 30, with
two characters coded from the cognate set 'water' (cf. Figure 11 in section 3.3.1).

In the above characters, each language having a reflex of the cognate class or
stem, *ška:n* or *čučut*, is coded in that character with a 1, indicating a shared state,
so that in the first character, each 1 represents the stem *ška:n* in that language, and
in the second, each 1 represents the stem *čučut*. The languages with no reflex are
coded with unique states, indicating that there is no shared state between them.
Although termed "binary", the coding of these characters is not represented in 1s
and 0s, because if the absence of the reflex were coded as a 0, SplitsTree4 would
treat this as a feature shared by those languages. As noted by Ringe et al. (2002),
the absence of a feature does not give evidence for a relation, so each absence is

| Variety | Reflex | Coding | Variety | Reflex | Coding |
|---------|--------|--------|---------|--------|--------|
| A | ška:n | 1 | M | ška:n | 1 |
| Ct | | 2 | P | | 5 |
| Co | | 3 | Pf | škan | 1 |
| Ch | ška:n | 1 | T | ška:n | 1 |
| FM | | 4 | U | ška:n | 1 |
| H | ška:n | 1 | Z | | 6 |
| A | | 2 | M | | 5 |
| Ct | čučut | 1 | P | čųčut | 1 |
| Co | čų́čut | 1 | Pf | | 6 |
| Ch | | 3 | T | | 7 |
| FM | čúčuti | 1 | U | | 8 |
| H | | 4 | Z | čúčut | 1 |

**Figure 30:** Encoding of characters for 'water'

coded as a unique state. In section 3.3.1 we saw this same cognate set, and used the distribution of the two reflexes to argue that *čučut* represents a Sierra innovation. The presence of *ška:n* in each of these languages is then a shared retention, and does not provide evidence that the languages have this stem for any kind of group. However, the phylogenetic analysis will use these two character sets to group first the set of languages having *ška:n*, and second, the set having *čučut*, taking only into account these shared reflexes. Although this is counter to what we know, it is hoped that some sort of evening-out will occur across the greater number of data points the phylogenetic analysis is capable of computing. We rely upon the correlation between similarity and genetic closeness. With this caveat, the phylogenetic network is presented in Figure 31.
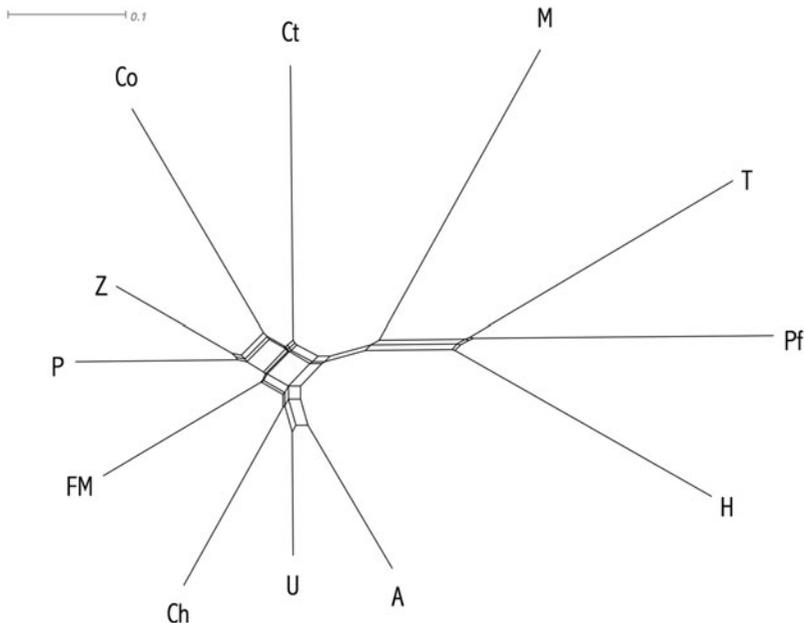


**Figure 31:** Phylogenetic network of Totonacan

The three Tepehua languages are represented on the right, quite distant from Totonac languages. Misantla is also quite distant from the other Totonac languages, in agreement with the ASJP results. On the top left, three Sierra varieties, Ct, Co, and Z, group quite closely with Lowland, P. While the ASJP tree puts P closest to Co, which is also the closest geographically, the phylogenetic network puts P closest to Z, which is much more distant. This would suggest that the branch spread widely before further innovation in the geographic centre, Co, and Ct. P is grouped more closely to Z than Z is to Co and Ct, giving little evidence for P as a distinct branch. FM appears somewhat between the other Sierra languages and the Northern languages seen at bottom left, U and A. Ch appears quite close to these Northern languages, here closer to U than to A, while it was equidistant to U and A on the ASJP tree. As in ASJP, Ch clearly groups closely with the Northern languages.

## 5. CONCLUSIONS

Although we have seen that quantitative measures provide less solid evidence, the two analyses do correlate with the evidence from shared innovations summarised in section 3.4. Each approach agrees with the traditional classification to separate Tepehua from Totonac, and Misantla from the other Totonac languages, in agreement with MacKay and Trechsel and with Brown et al. Three further points arising from the shared innovations are also represented in the quantitative analyses. First, while FM lacks the three features MacKay and Trechsel use to delimit the Sierra branch, it clearly shares a large number of both Sierra and Northern innovations. In the ASJP tree, FM appears at the periphery of the Sierra family in the ASJP tree, and appears roughly between the Sierra and Northern clusters in the phylogenetic network. The mixed distribution of shared innovations presents difficulties for both MacKay and Trechsel's Northern analysis, and for Brown et al.'s Sierra-Lowland hypothesis. FM certainly seems to be a promising area for further phylogenetic research. Second, Lowland P is shown to be very close to the Sierra languages, occurring very close to Co in the ASJP tree and to Z in the phylogenetic analysis. While Co is geographically proximate, Z is at the other side of the geographic range of Sierra languages. This close relation is affirmed in the large number of innovations shared by Lowland and Sierra. This supports Brown et al.'s grouping of Sierra and Lowland together opposed to Northern. Finally, the sharing of Northern, but not Sierra, innovations in Ch is matched by the quantitative data. In the ASJP tree, Ch is shown within the Northern branch, and in the phylogenetic network Ch appears close to A and U. Ch clearly belongs in the Northern branch, evidenced by both quantitative analysis and shared innovations.

### REFERENCES

Aranya, Osnaya, 1953. Reconstrucción del protototonaco: Huastecos, totonacos y sus vecinos, ed. Ignacio Bernal. *Revista Mexicana de Estudios Antropológicos* 23: 123–130.

Aschmann, Herman P. 1973a. *Diccionario totonaco de Papantla*. México, DF: Instituto Lingüístico de Verano.

Aschmann, Herman P. 1973b. *Vocabulario totonaco de la Sierra*. México, DF: Instituto Lingüístico de Verano.

Beck, David. 2004. *A Grammatical Sketch of Upper Necaxa Totonac*. München: LINCOM Europa.

Beck, David. 2011. *Upper Necaxa Totonac Dictionary*. Berlin: Mouton de Gruyter.

Beck, David. 2012. Apéndice: Tablas de morfología comparativa. In *Las lenguas totonacas y tepehuas: Materias para su estudio*, eds. Paulette Levy and David Beck, 587–596. Mexico City: UNAM Press.

Brown, Cecil H., David Beck, Grzegorz Kondrak, James K. Watters, and Søren Wichmann. 2011. Totozoquean. *International Journal of American Linguistics* 77: 323–372.

Bryant, David and Vincent Moulton. 2004. NeighborNet: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution* 21: 255–265.

Campbell, Eric. 2013. The internal diversification and subgrouping of Chatino. *International Journal of American Linguistics*, 79(3): 395–420.

Dunn, Michael. 2014. Language phylogenies. In *The Routledge Handbook of Historical Linguistics*, ed. Claire Bowern and Bethwyn Evans, 190–211. London: Routledge.

Dyen, Isidore, Joseph B. Kruskal, and Paul Black. 1997. Comparative IndoEuropean database collected by Isidore Dyen.

Fox, Anthony. 1995. *Linguistic Reconstruction: An Introduction to Theory and Method*. Oxford: Oxford University Press.

García Rojas, Blanca. 1978. Dialectología de la zona totonaco-tepehua. Honors thesis, Escuela Nacional de Antropología e Historia, México.

Holman, Eric W., Søren Wichmann, Cecil H. Brown, Viveka Velupillai, André Müller, Pamela Brown, and Dik Bakker. 2008. Explorations in automated language comparison. *Folia Linguistica*, 42: 331–354.

Huson, Daniel H. and David Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23(2): 254–267. Software available from www.splitstree.org.

Ichon, Alain. 1973. *La religion de los totonacas de la sierra*. México: Instituto Nacional Indigenista.

Kaufman, Terrance, Carolyn MacKay, and Frank Trechsel. 2004. Cuestionario lingüístico para la investigación de las variaciones dialectales de la lengua totonaca.

Levy, Paulette. 1987. *Fonología del totonaco de Papantla, Veracruz*. México, DF: Universidad Nacional Autónoma de México.

Levy, Paulette. 1990. *Totonaco de Papantla, Veracruz*. México, DF: Colegio de México.

Levy, Paulette. 1999. From 'Part' to 'Shape': Incorporation in Totonac and the issue of classification by verbs. *International Journal of American Linguistics* 65: 127–175.

Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig eds. 2015. *Ethnologue: Languages of the World*. 18th ed. Dallas, Texas: SIL International. Online version at http://www.ethnologue.com.

MacKay, Carolyn J. 1994. A sketch of Misantla Totonac phonology. *International Journal of American Linguistics* 60: 369–419.

MacKay, Carolyn J. 1999. *A Grammar of Misantla Totonac*. Salt Lake City: University of Utah Press.

MacKay, Carolyn J., and Frank R. Trechsel. 2005. *Totonaco de Misantla, Veracruz*. México, DF: Colegio de México.

MacKay, Carolyn J. and Frank R. Trechsel. 2010. *Tepehua de Pisaflores, Veracruz*. México, DF: El Colegio de México.

MacKay, Carolyn J. and Frank Trechsel. 2011. Relaciones internas de las lenguas totonaco-tepehuas. Memorias del V Congreso de Idiomas Indígenas de Latinoamérica.

MacKay, Carolyn J. and Frank Trechsel. 2013. A sketch of pisaflores tepehua phonology. *International Journal of American Linguistics* 79: 189–218.

MacKay, Carolyn J. and Frank R. Trechsel. 2015. Totonac-Tepehua genetic relationships. *Amerindia* 37(2): 121–158.

McFarland, Teresa. 2009. The phonology and morphology of Filomeno Mata Totonac. Doctoral dissertation, University of California, Berkeley.

McQuown, Norman, A. 1940. A Totonac Grammar. Doctoral Dissertation, Yale University.

McQuown, Norman A. 1990. *Gramática de la lengua totonaca* (Coatepec, Sierra Norte de Puebla). México, DF: Universidad Nacional Autónoma de México.

Müller, André, Viveka Velupillai, Søren Wichmann, Cecil H. Brown, Eric W. Holman, Sebastian Sauppe, Pamela Brown, Harald Hammarström, Oleg Belyaev, Johann-Mattis List, Dik Bakker, Dmitri Egorov, Matthias Urban, Robert Mailhammer, Matthew S. Dryer, Evgenia Korovina, David Beck, Helen Geyer, Pattie Epps, Anthony Grant, and Pilar Valenzuela. 2013. ASJP World Language Trees of Lexical Similarity: Version 4 (October 2013), available at http://asjp.clld.org.

Reid, Aileen A. 1991. *Gramática totonaca de Xicotepec de Juárez, Puebla*. México, DF: Instituto Lingüístico de Verano.

Reid, Aileen A. and Ruth G. Bishop. 1974. *Diccionarío totonaco de Xicotepec de Juárez, Puebla*. México, DF: Instituto Lingüístico de Verano.

Ringe, Don, Tandy Warnow, and Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100(1): 59–129.

Saitou, Naruya and Masatoshi Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4(4): 406–425.

Sapir, Edward. 1921. *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace.

Schnoebelen, T. 2009. A how-to guide for using phylogenetic tools on linguistic data (SplitsTree, MrBayes). Ms., Stanford University. Revised on April 23, 2009.

Secretaría de finanzas y planeación del estado de Veracruz (SEFIPLAN). 2013. *Cuadernillos Municipales: Coahuitlán, Mexico*.

Smythe Kung, Susan. 2007. *A Descriptive Grammar of Huehuetla Tepehua*. Ph.D. dissertation, University of Texas, Austin.

Tadmor, Uri, Martin Haspelmath, and Bradley Taylor. 2010. Borrowability and the notion of basic vocabulary. In *Quantitative approaches to linguistic diversity: Commemorating the centenary of the birth of Morris Swadesh*, ed. Søren Wichmann and Anthony P. Grant. *Diachronica* 27(2): 226–246.

Watters, James K. 1988. *Topics in Tepehua Grammar*. Doctoral dissertation, University of California, Berkeley.

Watters, James K. 2007. *Diccionario tepehua de Tlachichilco–español*. Ms. in possession of David Beck, University of Alberta.

Wichmann, Søren, André Müller, Annkathrin Wett, Viveka Velupillai, Julia Bischoffberger, Cecil H. Brown, Eric W. Holman, Sebastian Sauppe, Zarina Molochieva, Pamela Brown, Harald Hammarström, Oleg Belyaev, Johann-Mattis List, Dik Bakker, Dmitry Egorov, Matthias Urban, Robert Mailhammer, Agustina Carrizo, Matthew S. Dryer, Evgenia Korovina, David Beck, Helen Geyer, Pattie Epps, Anthony Grant, and Pilar Valenzuela. 2013. The ASJP Database (version 16), available at http://asjp.clld.org.

**APPENDIX A**

## Language Abbreviations and Sources

Abbreviations of names for languages used in this study are listed below in alphabetical order. The primary source for information on each language is given in parentheses. Other sources are given in the text.

A Apapantilla Totonac (Reid and Bishop 1974)
Ch Coahuitlán Totonac (author's fieldwork)
Co Coyutla Totonac (lexical database prepared by H. Aschmann)
Ct Coatepec Totonac (McQuown 1990 and unpublished dictionary by McQuown)
CX Cerro Xinolatépetl (communication between David Beck and Gerry Anderson)
FM Filomeno Mata Totonac (lexical database prepared by T. McFarland)
H Huehuetla Tepehua (Smythe Kung 2007)
M Misantla Totonac (MacKay and Trechsel 2005)
Ol Olintla Totonac (word list prepared by Jorge Tino)
Oz Ozelonacaxtla Totonac (word list prepared by Gabriela Román Lobato)
P Papantla Totonac (Aschman 1973a, Levy 1990)
Pf Pisaflores Tepehua (communication between David Beck and Albert Daveltshin, and James Watters, MacKay and Trechsel 2010)
T Tlachichilco Tepehua (Watters 1988, 2007)
U Upper Necaxa Totonac (Beck 2011)
Z Zapotitlán Totonac (Aschmann 1973b)
Zh Zihuateutla Totonac (Michelle García-Vega, p.c.)