The sonority hierarchy in Choapan Zapotec
Erin Donnelly
University of California, Berkeley
edonnelly@berkeley.edu

1 Introduction

This paper will describe typologically unusual onset clusters in Choapan Zapotec, an Oto-Manguean language spoken in the Sierra Norte of Oaxaca, Mexico. Onset clusters in Choapan Zapotec (CHO) violate sonority sequencing principles (SSP) in theoretically problematic ways; prohibitions in CHO against almost all types of coda consonants combine with these facts to create even more problems for current linguistic theory.

§2 will describe onset clusters in Choapan, and §3 describes coda facts in CHO. §4 discusses some phonological theories that relate to these facts, especially with regards to the sonority hierarchy. I show that while sonority sequencing principles do hold in CHO, they are utilized by the grammar in ways that require us to rethink these principles, and various theories that relate to them\(^1\).

2 Initial consonant clusters in Choapan Zapotec

Word-initial consonant clusters in CHO can violate or adhere to sonority sequencing principles. Cross-linguistically, these principles are based on acoustic properties of segments, for example, energy in a speech signal or wideness of the vocal tract.

Segments that are most sonorous (i.e., vowels) are universally preferred as syllable nuclei. The more segments that separate an onset consonant from the nucleus, the less sonorous it is expected to be. Wright 2004 proposes the following sonority hierarchy:

\[
\text{Vowels} > \text{Glides} > \text{Liquids} > \text{Nasals} > \text{Fricatives} > \text{Stops}
\]

Based on this hierarchy, predictions can be made about what should and should not be allowed in C2 position. Namely, something lower on the scale (a stop, for example) should not be in C2 position if something higher on the scale (a liquid, say) appears in C1 position. It is not cross-linguistically unusual for small violations, like fricative-stop sequences, to occur. However, large violations, like glide-stop or liquid-stop onset clusters, are more unusual and thus interesting typologically.

2.1 Tautomorphic initial consonant clusters

Both SSP-adhering and -violating consonant clusters in CHO can be tautomorphic\(^2\).

\(^1\) I would like to thank my consultants, Antonio Arreola Valentín and AEY, as well as Jonah Katz for his insightful discussion of the data and theory presented here. I am also grateful to members of Fieldwork Forum at University of California, Berkeley for their helpful suggestions. Of course, all mistakes are my own. I am supported by a Graduate Research Fellowship from the National Science Foundation.

\(^2\) CMP stands for ‘completive (aspect)’, 3s stands for ‘third person singular’, 3s.Subj stands for ‘third person singular subject’, 3s.Obj stands for ‘third person singular object’, 1s stands for ‘first person singular’, Pot stands for
The following examples show word-initial consonant clusters that **obey** sonority sequencing principles:

1.  
   \[ \text{bre? zí} \]
   \[ \text{bre? zí} \text{\ ant\ sour} \text{\ ‘sour ant’} \]

2.  
   \[ \text{lwé gosi} \]
   \[ \text{lwé\ gosi} \text{\ together\ time} \text{\ ‘at the same time’} \]

3.  
   \[ \text{pfé} \]
   \[ \text{pfé} \text{\ tomorrow} \text{\ ‘tomorrow’} \]

4.  
   \[ \text{bdáw re?ne} \]
   \[ \text{bdáw\ re?ne} \text{\ child\ tender} \text{\ ‘newborn baby’} \]

The following examples show word-initial consonant clusters that **do not obey** sonority sequencing principles:

5.  
   \[ \text{fko?odzo taka?} \]
   \[ \text{fko?odzo\ tak-a?} \text{\ back\ hand-1s} \text{\ ‘the back of my hand’} \]

6.  
   \[ \text{ngollà kjà?} \]
   \[ \text{ngollà\ kj\ -à?} \text{\ woman\ POSS\ -1s} \text{\ ‘my wife’} \]

7.  
   \[ \text{lbà?a} \]
   \[ \text{lbà?\ -a} \text{\ throat\ -1s} \text{\ ‘my throat’} \]

8.  
   \[ \text{wbidza} \]
   \[ \text{wbidza} \text{\ ‘sun’} \]

‘potential (aspect)’, 1S.OBJ stands for ‘first person singular object’, FUT stands for ‘future (aspect)’, CONT stands for ‘continuative/habitual (aspect)’, 1INCL stands for ‘first person plural, inclusive’, 1EXCL stands for ‘first person plural, exclusive’, and 3ANML stands for ‘third person (animal)’.
2.2 Heteromorphemic initial consonant clusters

Heteromorphemic consonant clusters appear almost exclusively on verbs\(^3\). Like tautomorphemic consonant clusters, these can be SSP-adhering or -violating. There are many more instances of SSP-violating initial consonant clusters across morpheme boundaries than within a single morpheme. The following examples show word-initial heteromorphemic consonant clusters that obey SSP:

(9) \(\text{pkopib'bin?}\)
    \(\begin{array}{c}
    p_{-} & \text{kopi} & -b' & \text{bin?} \\
    \text{CMPL-} & \text{kiss} & -3\text{S.SUBJ} & -3\text{S.OBJ}
    \end{array}\)
    ‘s/he kissed her/him’

(10) \(\text{rwe?bi di?idza?}\)
    \(\begin{array}{c}
    r_{-} & \text{we?} & -b & \text{idza}\? \\
    \text{CONT-} & \text{talk} & -3\text{S.SUBJ} & \text{word}
    \end{array}\)
    ‘to have a conversation’

(11) \(\text{brebí yàgà}\)
    \(\begin{array}{c}
    b_{-} & \text{re} & -b & \text{yàgà} \\
    \text{CMPL-} & \text{climb} & -3\text{S.SUBJ} & \text{tree}
    \end{array}\)
    ‘she climbed a tree’

(12) \(\text{bgozigé?}\)
    \(\begin{array}{c}
    b_{-} & \text{gozi} & -g\? \\
    \text{CMPL-} & \text{get.drunk} & -3\text{S.RESP}
    \end{array}\)
    ‘he got drunk’

There are also heteromorphemic word-initial consonant clusters that do not obey SSP:

(13) \(\text{gokabi? wnébi?}\)
    \(\begin{array}{c}
    go- & \text{kabi?} & w- & \text{né-bi?} \\
    \text{CMPL-be.able-3s} & \text{CMPL-speak-3s}
    \end{array}\)
    ‘he was able to talk’

(14) \(\text{wbifibi?}\)
    \(\begin{array}{c}
    w_{-} & \text{bifi} & -b\? \\
    \text{CMPL-} & \text{fall} & -3\text{S.SUBJ}
    \end{array}\)
    ‘he/she fell’

(15) \(\text{wrja?}\)
    \(\begin{array}{c}
    w_{-} & \text{rj} & -a\? \\
    \text{CMPL-} & \text{climb} & -1\text{s}
    \end{array}\)
    ‘I climbed’

(16) \(\text{rdí ilàbí?}\)
    \(\begin{array}{c}
    r_{-} & \text{díilà} & -bí\? \\
    \text{CONT-} & \text{slip}
    \end{array}\)
    ‘I slipped’

\(^3\)Synchronically, at least, there is no pre-verbal nominal morphology.
2.3 Non-syllabicity of highly sonorous C1s

One possible explanation for liquid- or glide-initial consonant clusters, which have large sonority
reversals, is that these highly sonorous segments might be syllabic. However, there is evidence
indicating that this is not the case.

As in the example waveform and spectrogram below (for example (14) above), /w/ in C1 is much
more comparable to the following consonant, a /b/, than it is to a vowel.

CHO has three primary tones: mid, high, and low, and a tone is assigned to every syllable nucleus.
However, tone is not assigned to sonorous C1s in onset clusters. Consultants can easily whistle the
tones associated with any word or phrase, but never assign tone to C1s in sonority-reversing onset
clusters. For example, (14) and (16) above are both whistled as three tones, not four.

Highly sonorous segments in Choapan correspond to from non-vocalic segments in other closely
related languages. In example (8) above, ‘sun’ is transcribed as wbidza. In Yatzachi el Alto Zapotec,
the cognate form is gwbiŋ, and Yatzachi el Bajo has the form bgwiŋ (Butler 2000). Reconstructions
of proto-Zapotec indicate that /w/, /gw/ or /b/ in these cognate forms originally came from a
velar stop+round vowel sequence. However, it seems that this syllabicity was already lost in proto-
Northern Zapotec, resulting in a non-syllabic onset cluster with unusual sonority reversals.

### 2.4 Summary of word-initial consonant cluster facts

So far, I have given examples wherein word-initial consonant clusters in CHO behave in the following ways:

- **Violate SSPs:**
  - within a single morpheme
  - across a morpheme boundary

- **Obey SSPs:**
  - within a single morpheme
  - across a morpheme boundary

Tables 1 and 2 gives a list of all possible word-initial consonant cluster combinations in CHO, organized by manner of articulation. The left-most column shows C1, and the rest of the columns indicate possible C2s. A checkmark indicates that C1, the consonant type at the very left of a given row, can have the C2, the consonant at the stop of the column, following it in an onset cluster.

### Table 1: Tautomorphemic word-initial consonant clusters

<table>
<thead>
<tr>
<th></th>
<th>Stop</th>
<th>Affricate</th>
<th>Fricative</th>
<th>Nasal</th>
<th>Liquid</th>
<th>Glide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Affricate</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Fricative</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Nasal</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Liquid</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Glide</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

### Table 2: Heteromorphemic word-initial consonant clusters

<table>
<thead>
<tr>
<th></th>
<th>Stop</th>
<th>Affricate</th>
<th>Fricative</th>
<th>Nasal</th>
<th>Liquid</th>
<th>Glide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Affricate</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Fricative</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Nasal</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Liquid</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Glide</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
There are several important generalizations about the phonotactics of the language that are shown in the tables:

- All stop+C combinations are possible within morphemes and across morpheme boundaries, with the exception of tautomorphemic stop+nasal sequences. This seems to be an accidental gap.
- Though affricates are common C2s, the only C2 allowed after an affricate in C1 is a glide.
- Fricatives can’t be C1 before affricates or fricatives, and are C2s after stops and liquids, and after glides across a morpheme boundary.
- Nasals are permitted as C1 (but not before fricatives or liquids) both within and across morpheme boundaries.
- Within a single morpheme, liquids cannot be C1 before affricates or other liquids.
- Liquids and glides can appear as a C1 before any C2 in heteromorphemic environments.

As a general principle, an SSP-violating consonant cluster in CHO is much more likely to be heteromorphemic than to be tautomorphemic.

3 Codas in Choapan Zapotec

Though there are restrictions to what can exist in word-initial onset clusters in CHO, codas are relatively much stricter: only nasals and glides can be in a coda position. Though liquids are more sonorous than nasals, and therefore ought to be allowed as codas in terms of relative sonority, they are not licensed in codas in native CHO words. Some examples of nasal and glide codas in CHO:

(17) lejbi?
   lej- bi?
   teeth-3s
   ‘his teeth’

(18) jò lòw tʃitə
   jò  lòw  tʃitə
   stone face egg
   ‘egg yolk’ (lit. eye of the egg)

(19) kwinró
   kwin-ró
   body-INCL
   ‘our bodies; ourselves (reflexive)’

Glide + nasal sequences can also be found in codas, both word-medially and word-finally. However, they seem to always exist across morpheme boundaries:

(20) lawn
   law- n
   face-INAN
‘its face’

(21) dá gáwndò?

dá  g- áw -ndò?
REL  POT- eat -1EXCL
‘what we will eat’

It is important to note that while all (native) onset nasals in CHO are always alveolar, coda nasals are allophonically neutralized to velar nasals or nasalized vowels.

3.1 Stem-final /g/

Though non-glides, -nasals never appear overtly in coda position, there are instances of obstruents in codas underlyingly. When they would appear in a coda position, these obstruents are deleted, and evidence for them in the representation comes from phonological environments in which they are syllabified as an onset.

A subset of verbs, body part, and kinship terms show this Ø ⇠ /g/ alternation, which is the reflex of a historical stem-final *k (Kaufman 2007). Suffixal person markers that are vowel-initial, like the first person singular, -a?, give evidence to show that deletion is employed in order to avoid having an obstruent in a coda position:

(22) gitfoga?

gitfog -a?
head -1s
‘my head’

(23) gitfoba?

gitfog -ba?
head -3ANML
‘its head (of an animal)’

(24) rtfíloğàn

r- tfílog -ò -n
CONT- scatter -1s -INAN
‘I am scattering them [the seeds]’

(25) rtfílòbín

r- tfílog -bì -n
CONT- scatter-3s-INAN
‘He is scattering them [the seeds]’

The prohibition in CHO against obstruent codas is high enough that phonological information from the stem is deleted due to its position relative to vowel nuclei. If a stem-final /g/ syllabifies as an onset (i.e., when it comes immediately before a vowel), then it is not deleted. However, if it would be a coda in the output, then it must be deleted.

Stem-final /g/ is prevalent in CHO, but there are no other examples of obstruent codas, even underlyingly. This gap, combined with the facts of obstruent coda deletion, indicates that obstruent
codas are very marked within the phonological framework of CHO. The phonotactics actively conspire against having obstruent codas, and the phonology creates repairs to delete those that do exist.

4 What does this mean for phonological theory?

The above data from CHO are interesting phonological typology and theory, because they behave in ways that do not fit well into existing conceptualizations of sonority sequencing principles. Clements 1990 describes the sonority sequencing principle thus:

- Between any member of a syllable and the syllable peak, only sounds of higher sonority rank are permitted. Under this principle, give the sonority scale..., syllables of the type tra, dva, sma, mra are permitted, while syllables like rta, vda, msa, mla are excluded. Cross-linguistic comparison supports the view that clusters conforming to the Sonority Sequencing Principle are ... often the only cluster types permitted in a given language. Clusters violating this principle ... usually occur only in addition to clusters conforming to it.

This idea is not new or controversial; Clements explains that linguists have discussed versions of this since 1865. Every linguistic theory attempts to grapple with facts about sonority, because they so obviously affect the ways syllables are shaped.

4.1 Margin approaches

4.1.1 Optimality theory

Sonority sequencing principles existed before OT, and so from its very beginning, optimality theory constraints were created to deal with these facts. Prince and Smolensky proposed a constraint, called the Margin Hierarchy, to fit SSPs into OT: “it’s less harmonic to parse α as a margin than to parse i as a margin, less harmonic to parse i as a margin than r, and so on down the sonority ordering (2002:141)”.

That is, the less sonorous a segment is, the more a grammar will constrain against it existing in a margin. This could be problematic for various reasons, and in 2003, Baertsch and Davis proposed a revision of this hierarchy, called the split margin approach.

4.1.2 Split margin approach

In Baertsch and Davis’s split margin approach, singleton onsets are the opposite of codas and C2s in clusters, at least from the point of view of the SSP; singleton onsets are preferred to be low-sonority, and C2 and codas ought to be high-sonority. Violations will be assigned to low-sonority segments in C2 or coda position.

In this framework, languages are predicted to constrain against high-sonority segments in a C1 position, especially if the C2 is lower sonority. Furthermore, languages are predicted to license the same consonants in C2 that they license in the coda. The first prediction exists in virtually all
theories of sonority, not just in a margin approach. However, the second, which crucially links the licensing of C2 to the coda, is unique to the split margin approach.

4.2 Problems for general ideas about sonority

CHO phonotactics provide several challenges for general theories about sonority. The biggest is that higher sonority segments are allowed in C1 position when a lower sonority segment is in C2 position. The very sequences that are predicted to not occur, like rta, vda, or wbi, are robustly attested in CHO. Furthermore, many such sequences are not heteromorphemic.

As Clements 1990 predicted, CHO also allows consonant clusters that do not violate the SSP. Though SSP-violating sequences still need to be accounted for, the fact that non-violating sequences exist simplifies the explanation at least somewhat.

4.3 Problems for a split margin approach

CHO creates even more problems for the split margin approach. Within this framework, segments licensed in C2 are also licensed in the coda. However, obstruents in CHO are licensed in C2 and not in coda position. Even more problematically, liquids, which are more sonorous than nasals (which are allowed in coda position), are not licensed in coda position, but are in C2.

Baertsch and Davis 2003 makes many predictions, which hold true for many languages, but which are quite problematic when looking at CHO data. Some of the data can be analyzed to fit better with a split margin approach, but other data show that this approach is perhaps too strong.

4.4 Accounting for CHO phonotactics

The piece of data that can be accounted for most straightforwardly in a theoretical framework is the case of nasals licensed in coda position.

4.4.1 Nasal glides

Ferre 1988 describes in detail nasal ‘absorption’: “... the exceptional susceptibility to “absorption” processes and to processes which neutralize a consonant’s point of articulation which certain nasal consonants have, depending on their position in the word”. For Ferre, absorption occurs when nasals are so weakly articulated that they are no longer obstruent-like at all, and are actually nasal glides. When [N], a placeless velar nasal, is not completely deleted (only to be realized as nasalization on a vowel), it is often taken for a weakly articulated velar nasal, either because it’s actually velarized superficially, or linguists systematically hear and transcribe it that way.

Nasal absorption works well with the CHO data. For one, in a coda (i.e., non-prominent) position, nasal /n/ is realized as a velar nasal or as a word-final nasalized vowel. This alternation is exactly one that Ferre details, and thus fits well into the nasal absorption framework.

By considering coda nasals in CHO to be glides instead of stops, a stronger generalization about codas can be made: only glides are allowed in codas, and nothing else. In terms of sequencing principles, this is an improvement, since liquids are not licensed in coda position. If only glides
are licensed, then there is no need to explain why more-sonorous liquids cannot be in a coda and nasals can.

4.4.2 Nasal glides and the split margin, SSP

Though considering nasals in coda position to be glides makes sense from the standpoint of general sequencing principles, it does not eliminate all of the problems that exist within the split margin hierarchy. That is, though the licensing of nasals (and not liquids) in a coda no longer needs to be explained, the licensing of C2 is still a problem. The split margin hierarchy predicts that consonants licensed in C2 are licensed in the coda; however, only glides (plain and nasal) are licensed in the coda in CHO.

Simply put, the split margin approach cannot account for CHO phonotactics, even if coda nasals are analyzed as glides. The relationship between C2 and coda that seems to exist in other languages does not exist in CHO.

However, sonority sequencing principles are important in CHO, and exist both in the onset and coda. For codas, only glides, the most highly sonorous segments are licensed. In onset position, most, but not all, sonority reversals in consonant clusters are allowed. The types of C1 consonants that are most cross-linguistically prohibited against, liquids and glides, are present in CHO, even before obstruents, which is a large sonority reversal.

These large reversals are most frequently licensed across morpheme boundaries, but not only across morpheme boundaries. This is good and bad for the SSPs: to account for CHO data, constraints relating to the morphology could be successfully utilized. However, since not all large sonority reversals in onset clusters are heteromorphemic, SSPs do not always hold true in CHO; other constraints that ignore SSPs (perhaps faithfulness constraints) would need to be used as well.

5 Conclusions and further research

In CHO, while all consonants are licensed in C2, not every C1+C2 combination is possible, even across morpheme boundaries. Some gaps could be accidental, but it doesn’t seem to be a coincidence that affricates are essentially never C1s, for example. Seemingly systematic gaps should be analyzed for phonotactic conspiracies in the future.

It was pointed out to me that many word-initial clusters in CHO seem to have back → front articulations, where C1 is more back, and C2 is more front. It is possible that instead of following traditional principles of sonority in licensing clusters, CHO licenses them based on place of articulation. This is an area that needs to be investigated much more in the future.

Likewise, this paper has shown that CHO is a prime example of how the split margin approach doesn’t work, and that CHO also creates problems for SSPs in general. However, it has failed to show ways to account for the data in current phonological theory. In the future, I hope to show not only how CHO doesn’t fit with phonological theories, but how theory can be used to account for its phonotactics.
References


