# Strong and Weak Low Tones in Peñoles Mixtec

Thom van Hugte
*Universität Leipzig*

## 1 Introduction

Mixtec languages have some of the most complex tonal systems in the world and pose a challenge to frameworks for tonal analysis. In this chapter, I argue that analysing floating L tone association in Peñoles Mixtec can be done successfully with Gradient Symbolic Representations (Smolensky & Goldrick 2016; Zimmermann 2021), where any phonological object can have a range of strength that is expressed numerically, rather than being categorically present or absent. The analysis presented here shall demonstrate two key advantages of GSR. First, it enriches phonological representations so that complex phenomena such as the Peñoles Mixtec floating L can be accounted for. This additional representational power is offset by keeping the rest of the phonology (e.g. constraints) simple. Second, GSR can also model exceptional morphemes with the same mechanism of gradient representations. By placing the burden of exceptionality on representations, this approach predicts that the same exceptional morphemes tend to be exceptional in multiple contexts. This is exemplified by the Peñoles Mixtec plural enclitic /=kwe/.

The structure of the chapter is as follows. Section 2 provides background information on Peñoles Mixtec and then introduces the data. Section 3 makes the theoretical assumptions explicit and explains the approach, motivating the need for GSR. Section 4 contains a constraint-based analysis of the data. Finally, Section 5 summarises the analysis, discusses the implications, and concludes the chapter.

## 2 Peñoles Mixtec data

**2.1** *Background* Peñoles Mixtec (ISO 639-3: mil) is an Oto-Manguean language spoken in the state of Oaxaca in Mexico. According to Ethnologue, it is in vigorous use and has roughly 5500 speakers (Eberhard et al. 2024; 2000 census). Most linguistic research on Peñoles Mixtec has been done by John and Margaret Daly, spanning a period from the 50's to the 00's. The data for the current analysis all comes from Daly & Hyman (2007), which contains the most developed and complete description of the tonal system. In glosses, it is abbreviated to D&H (2007). The paper lists several phonological rules in the autosegmental format, some involving tonal features. However, not every aspect of the data is formalised or explained so there are still many open questions, including the data discussed here.

Tone Bearing Units (TBUs) are syllables, and each can be H(igh), L(ow), or unspecified Ø. The evidence for underspecification is diverse. The specified tones H and L can form rising LH-contours, whereas Ø cannot form any contours. It is also invisible to OCP-L, which forces deletion of multiple L tones adjacent on the tonal tier, regardless of the number of Ø-syllables in between. Third, Ø-enclitics are transparent to floating L association, as will be discussed in more detail below. Finally, the phonetic realisation of Ø is low-falling unless followed by H, in which case its pitch will match H. Neither the low-falling realisation nor the H anticipation are indicative of a phonological specification as M or another full-fledged tone. Contrary to Daly & Hyman (2007), I will not assume tonal features.

The most important prosodic units in Mixtec languages are the bisyllabic roots called "couplets" in the literature. Most nouns, verbs, and adjectives have this structure. These bisyllabic roots can optionally be preceded and followed by multiple clitics. An example is given in (1).

Proceedings of *AMP 2023/2024*

(1) [bá.nì.ʔí.dé.dɨ]
*bá*$^{L}$= *níʔí*$^{L}$ =*dé* =*dɨ*
NEG.POT= find =3SG.M =animal
‘he will not find the animal’ (D&H, 2007, p.190)

The example also shows the basic VSO word order. The order is the same regardless of whether the subject and object are enclitics on a root or bisyllabic roots themselves. The root is /níʔí$^{\text{L}}$/, which, using the notation from Daly & Hyman (2007), is a H.H$^{\text{L}}$ root. Importantly, this root has only one underlying H; the period indicates a syllable boundary that separates the H-toned TBUs. The superscript $^{\text{L}}$ is a floating tone. The reason for its deletion on the surface in (1) is not the type of floating L deletion tackled in the analysis in the next section but will be addressed briefly. There are 10 types of bisyllabic roots, as shown in Table 1 below.

| Common | Rare |
|---|---|
| H.H$^{\text{L}}$ | |
| H.H | LH.H |
| H.Ø | LH.Ø |
| Ø.H | L.H |
| Ø.Ø | L.Ø |
| Ø.Ø$^{\text{L}}$ | |

**Table 1:** All root types

As can be seen, the rare root types can be understood as derived from the common types, but with an initial L tone. Daly & Hyman (2007) analyse all these initial L tones as floating, but there is no sufficient ground for this, so I assume they are simply associated underlyingly.

### 2.2 *Floating L*

There are two types of roots with floating L: H.H$^{\text{L}}$ and Ø.Ø$^{\text{L}}$. However, floating L after a H-toned root will disappear in some contexts where L after a fully underspecified root will not. Compare the phrases in (2) and (3), both from D&H (2007, p.188).

(2) [níʔí kɨtɨ]
H.H$^{\text{L}}$ Ø.Ø
*níʔí*$^{L}$ *kɨtɨ*
POT.find animal
‘the animal will find’

(3) [koko kɨ̀tɨ]
Ø.Ø$^{\text{L}}$ Ø.Ø
*koko*$^{L}$ *kɨtɨ*
POT.swallow animal
‘the animal will swallow’

Both phrases have the same target for L, and yet L in (2) is unable to surface, a process called floating L tone deletion (FLD). In other contexts, floating L after a H-toned root does surface faithfully. The paradigms in (4) show all combinations of roots, with the case of FLD underlined.

(4) a. H.H$^{\text{L}}$ + H.H$^{\text{(L)}}$ → H.H LH.H$^{\text{(L)}}$
b. H.H$^{\text{L}}$ + H.Ø → H.H LH.Ø
c. H.H$^{\text{L}}$ + Ø.H → H.H L.H
d. H.H$^{\text{L}}$ + Ø.Ø$^{\text{(L)}}$ → $\underline{\text{H.H Ø.Ø}^{\text{(L)}}}$
e. Ø.Ø$^{\text{L}}$ + H.H$^{\text{(L)}}$ → Ø.Ø LH.H$^{\text{(L)}}$
f. Ø.Ø$^{\text{L}}$ + H.Ø → Ø.Ø LH.Ø
g. Ø.Ø$^{\text{L}}$ + Ø.H → Ø.Ø L.H
h. Ø.Ø$^{\text{L}}$ + Ø.Ø$^{\text{(L)}}$ → Ø.Ø L.Ø

Which root precedes L matters for whether FLD happens or not (d.≠h.), but the full paradigms show that the tonal specification of the following root also matters (c.≠d.). Both roots in c. and d. start with Ø, so FLD is sensitive to the full root. Importantly, H.H+L.Ø with a rare L-initial root is an acceptable surface form. In h., the second root’s floating L is deleted due to OCP-L.

There is another context in which FLD happens. Although the standard word order is VSO, noun phrases may be fronted for emphasis and appear in initial position, i.e. SVO/OVS word order. In this case, FLD will also happen if Ø.H is the target. This is shown in (5).

(5) [dútú kiní]
H.H$^{L}$ Ø.H
*dútú*$^{L}$ *kiní*
priest POT.see
'the priest will see' (D&H, 2007, p.189)

From this we can conclude that, in addition to the preceding root and the following root, the syntactic structure also influences FLD, because (4c)≠(5).

**2.3** *Plural enclitic =kwe* The plural enclitic /=kwe/ is exceptional in two regards. First, it is an exceptional undergoer of L tone association in enclitics. Ordinarily, toneless enclitics are no viable targets for floating L tones. Compare (6) with /=kwe/ to (7) with ordinary toneless enclitics, both from D&H (2007, p.178).

(6) [ndu.ku.kwè.ʃi.dɨ]
*nduku*$^{L}$ *=kwe =ʃi =dɨ*
POT.look.for =PL =3SG.F =animal
'they will look for the animal'

(7) ndu.ku.ni.ʃi$^{L}$
*nduku*$^{L}$ *=ni =ʃi*
POT.look.for =only =3SG.F
'she will just look for'

In (6), the floating L of the root simply associates to /=kwe/, so that it surfaces with a low tone. In (7), the floating L is unable to associate to either of the toneless enclitics and will "float past" them, associating to the next word if it is a possible target. However, floating L *does* regularly associate to H-toned enclitics.

(8) [ndu.ku.ni.ʃi.dě]
*nduku*$^{L}$ *=ni =ʃi =dé*
POT.look.for =only =3SG.F =3SG.M
'she will just look for him' (D&H, 2007, p.177)

Here, we see that the floating L of the same root as in the previous examples is able to associate to the high-toned enclitic /=dé/, making it surface with a rising contour. We thus observe that toneless /=kwe/ patterns with H-toned enclitics.

The second unusual characteristic of /=kwe/ is that it is an exceptional non-undergoer of High Tone Spreading (HTS). For the purposes of this chapter, I shall simplify the description of HTS since it is an extremely complex process. However, my analysis of /=kwe/ should be orthogonal to the analytical strategies necessary for a complete analysis of HTS. I will therefore characterise HTS as follows. In HTS, H-final proclitics (H= or H$^{L}$=) and roots (H.H$^{(L)}$ or Ø.H) will spread their H to any following proclitic, root, or enclitic within the same word without any H (i.e. no HTS between H= and Ø.H). A context in which HTS appears to happens across words is when a nominal root is immediately followed by an adjectival root without intervening enclitics.[1] Finally, if HTS takes place, the final H of the span must be followed by a floating L. All the typical characteristics of HTS are exemplified in (9) below.

(9) ní.dá.kú.nú.ʃí.dɨ́$^{L}$
*ní= da*$^{L}$*= kunu =ʃi =dɨ*
CMP= CAUS= run =3SG.F =animal
'she chased the animal away' (D&H, 2007, p.195)

The completive proclitic /ní=/ has a H that will spread onto the next proclitic, to the underlyingly toneless root, and to both toneless enclitics, creating a span of six H tones. This span must be followed by a floating L that will associate to any following word; it is not subject to the rules of FLD discussed in Section 2.2 above.

---

[1] If the process were analysed as phrasal, there would be many more contexts for which it would have to be explained why HTS does *not* happen, most notably between verbs and nouns. Instead, N+Adj. is for some reason treated as a N+encl. sequence.

## 3 Theoretical approach

**3.1** *Assumptions* I assume an architecture of grammar in which words are constructed cyclically with the root as the centre. Every time a clitic is added to the root, the word will pass through the word-level phonology again and undergo optimisation. Phrasal phonology happens after complete words are combined into a full sentence. Cyclic word formation and the distinction between word-level and phrase-level phonology thus provide a restricted amount of ordering. Because the analysis makes reference to words and phrases as the domains of application for phonological processes, I assume the Prosodic Hierarchy (Nespor & Vogel, 1986). However, the analysis does not hinge on any particular tenets of prosodic phonology, so it should be possible to make the same analysis work with direct reference between syntax and phonology. As for the morphological information visible to phonology, it is restricted to Morphological Colour theory (Oostendorp, 2006). That is, the phonology "knows" which phonological objects belongs to the same morpheme, as if each morpheme has its own distinct colour.

I also assume Gradient Symbolic Representations (Smolensky & Goldrick 2016; Zimmermann 2021). This means that every phonological object is not simply present or absent, but can be weakly present. This is expressed as a number in the range of 0–1, where 0 is fully absent, 1 is fully active, and intermediate numbers are weakly present. Weakly present objects cannot be shipped off to the phonetics and must either be deleted or made fully active; in OT terms, there cannot be gradiently active material in the final output. The distinction between weak and strong versions of the same type of phonological object (e.g. a tone) is implemented by means of Gradient Harmonic Grammar (Hsu, 2022), which combines the gradience of GSR with the weighted constraints of Harmonic Grammar (Legendre et al., 1990). For example, if a fully active tone $T_1$ is deleted, it will fully violate the constraint MAX-T that penalises the deletion of tone. If the weight of this constraint is 15, deleting $T_1$ will incur the full penalty of –15. In contrast, deletion of a weakly active $T_{0.4}$ will only incur a penalty of –6 because its lower activity makes it less costly to delete (0.4 x 15 = 6). Conversely, making $T_{0.4}$ fully active would violate DEP-T by 0.6, whereas $T_1$ is already at maximum activity and will not violate DEP-T.

GSR is a very powerful mechanism that greatly increases the representational possibilities, but it comes with its own unique advantages. To start, it means other components of the grammar can be simpler. Because gradience enables unique representations for specific morphemes, morphophonology can essentially be reduced to regular phonology, obviating the need for approaches such as Constraint Indexation (Pater, 2010). The interfaces between phonology and the rest of the grammar can therefore remain limited in scope, which is in line with the hypothesis of modularity (Appelbaum, 2017). Multiple constraint rankings, as in Stratal OT (Bermúdez-Otero, 2018) or Cophonologies (Sande et al., 2020), are also often made redundant by gradience. Constraints themselves can stay simple and interact with strong or weak objects in a consistent way, as the analysis of FLD will evince. GSR also makes specific, falsifiable typological predictions. One of them is that exceptional morphemes should often show multiple types of exceptional behaviour, given that their special representations are evaluated regularly by all constraints. For example, a morpheme that is assumed to have an underlying weak vowel in order to explain some exceptional deletion in some context is *always* weak when it enters the phonology, so that it should be more prone to deletion elsewhere, too. This typological prediction of multiple exceptionalities is borne out by /=kwe/.

**3.2** *Approach* I shall now explain the reasoning behind the constraint-based analysis in the next section, making the need for gradience clear. I start with FLD. There must clearly be some pressure to delete L, but does this apply specifically to $H.H^L+Ø.Ø$ sequences? This is unlikely given that H.H+L.Ø with an underlying associated L surfaces faithfully, and because FLD also happens before Ø.H roots in SVO/OVS word order. It is therefore more likely that the pressure to delete floating L applies more broadly. But out of all the floating L tones of the language ($H^L$=, $Ø^L$=, $Ø.Ø^L$, post-HTS $^L$), FLD only happens to $H.H^L$. When considering the paradigm of FLD in (4), an important observation to make is that in (4a–c) the presence of L avoids adjacent H tones on the tonal tier. It therefore makes sense to assume that floating L after a H-toned root must be deleted unless its presence avoids an OCP-H violation. But this deletion still seems unmotivated and is difficult to achieve formally. Floating low tones are able to associate across words, and are thus active at the phrasal level. If all L tones are representationally identical, then the reason why only L in $H.H^L$ is sometimes deleted must be exclusively due to the phonological context. In order to delete that specific L and none of the others, the responsible rule or constraint(s) must make reference to the fact that the L that undergoes FLD: 1) follows a root with 2) *underlying* high tones, and 3) is floating. Meeting all these criteria in one step at the phrasal level would require extremely complex rules/constraints and FLD would likely still remain unmotivated.

This can be avoided by assuming L in $H.H^L$ is weaker than other L tones. The motivation for this would be a ban on word-final L. On the surface, roots cannot end in L and toneless enclitics cannot take floating L. Given the assumption of cyclic word-formation, these are precisely the contexts that are word-final at some point of the derivation. The only context in which L can appear morpheme-finally is in proclitics, which are never word-final because they are added from the root outwards. $H.H^L$ has a final L during the first cycle of the word-level phonology, which is repaired through weakening it. There is additional evidence for the weakness of this L. In a process that Daly & Hyman (2007) call H-tone delinking, A $H.H^{(L)}$ couplet preceded by a floating low tone in certain contexts will surface as L.H instead of $LH.H^{(L)}$. This can be seen in (1). Although it is unclear why the floating L should be deleted because of H-delinking, this is consistent with its hypothesised status as weak. In conclusion, if L in $H.H^L$ is weakened, the pressure to delete it at the phrasal level is accounted for. How to derive this weakening is shown in the next section. The reason why it does not apply to $Ø.Ø^L$ is because its L is not unambiguously word-final with respect to the final TBU.

The enclitic /=kwe/ exceptionally undergoes low tone association and blocks HTS. Recall that H-toned enclitics can regularly take floating L, creating a LH-contour, and that HTS does not apply to morphemes that already have H. In other words, /=kwe/ consistently patterns with H-toned enclitics. Such clustering of exceptional behaviour is precisely what is predicted by GSR, and GSR offers a way to account for the opposition between its phonological behaviour and phonetic realisation. Let us assume that /=kwe/ actually has a weak underlying $H_{0.1}$ associated to it. A weak phonological object with low activity is still of the same type of object as a fully active one, so that a constraint that categorically evaluates the presence of H will not need to distinguish weak $H_{0.1}$ from fully active $H_1$. However, in order for the weak $H_{0.1}$ to be phonetically realised, it needs to be made fully active. This can easily be made impossible if the penalty for full activation is steep, ultimately favouring deletion instead. But the deletion of $H_{0.1}$ must of course follow the association of floating L and HTS; a counterfeeding order. The restricted ordering that I assume is sufficient to achieve this. Association of L and HTS happen at the word level, whereas outputs without any gradience are required at the phrasal level.

## 4 GHG analysis

Gradient Harmonic Grammar (GHG) combines GSR and Harmonic Grammar. Compared to regular Harmonic Grammar, GHG allows for gradient violations of constraints as already explained above. This means constraints need to be formulated as categorical or sensitive to gradience. Examples of familiar constraints used later are given below.

- MAX-L: assign a violation x where x is deleted activity of L tones
- DEP-A: assign a violation for epenthetic association lines

The highest harmony score is what determines the optimal output candidate. I only use penalising constraints, so that the most optimal candidate will be the one with the least negative score.

**4.1** *Deriving weak $L_{0.5}$* During the first cycle of word-level phonology of a bisyllabic root, the root is the only morpheme in the optimisation window. In that case, $H.H^L$ violates the following constraint:

- *FINAL-$L_1$: assign a violation for each fully active L that is word-final, where word-final means either being the final associated tone or following the final TBU

When word-finality is defined in this way, the constraint targets associated L tones generally, and only floating L tones in $H.H^L$ roots. Unspecified $Ø.Ø^L$ roots and $=Ø^L$ enclitics are left out, because their L is not a priori final, and instead has a more abstract position on the tonal tier. But after H.H, the L's position is undeniably final relative to the final TBU. The constraint also makes reference to the full activity of L. The obvious repair needed to avoid violating the constraint is to make L defective, i.e. to make its activity <1. Naturally, then, the least costly change in GHG is to make it have an activity of 0.99999 —infinitely close to 1. A human learner, however, could not and would not apply such a learning strategy. Phonology tends to maximise its contrasts, as is evident from the vowel inventories, consonantal places of articulation, and tonal inventories of the languages of the world. If this concept is applied to a learner acquiring gradience, then, given that 0 and 1 are already present in the system, the learner should select an activity of 0.5, since this is the furthest removed from the minimum and maximum.

- DifferentiateGradience: assign a violation x where x is the difference between a derived level of activity in the output and the value that would differentiate it the most from the other values

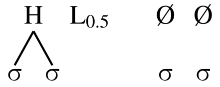
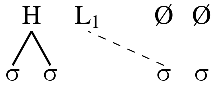
If the null hypothesis of a learner that uses GSR is that all phonological objects are fully active, DiffGrad should be a good alternate hypothesis. When applied to H.H$^{L}$ at the word level, the result is as in Tableau (10).

(10) Word-level phonology: $L_1$ is weakened

| níʔí$^{L1}$ | | *Final-$L_1$<br>100 | DiffGrad<br>50 | Max-L<br>10 | $\mathcal{H}$ |
|---|---|---|---|---|---|
| a. | níʔí$^{L1}$ | –1 | | | –100 |
| ☞b. | níʔí$^{L0.5}$ | | –0 | –0.5 | –5 |
| c. | níʔí$^{L0.999}$ | | –0.499 | –0.001 | –25 |
| d. | níʔí | | | –1 | –10 |

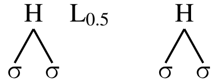
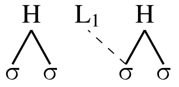
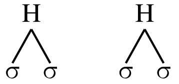
**4.2** *Modelling Floating L Deletion* When underlying H.H$^{L}$ enters the phrase-level phonology, it is already weakened to H.H$^{L0.5}$. At this point, no complicated constraints are necessary. The deletion of L before toneless roots is straightforward.

(11) Phrase-level phonology: a sequence of two roots

|  | | Max-L<br>10 | Dep-L<br>10 | Dep-A<br>1 | $\mathcal{H}$ |
|---|---|---|---|---|---|
| a. |  | | –0.5 | –1 | –6 |
| ☞b. |  | –0.5 | | | –5 |

If Max-L and Dep-L have the same weight, then, all else being equal, deletion is cheaper than association due to the influence of Dep-A. But all else is not equal if an OCP-H violation can be avoided. Tableau (12) shows the contrasting situation where the second root has H, too. Together, the violation of OCP-H and Max-L outweigh Dep-L and Dep-A.

(12)

|  | | Max-L<br>10 | Dep-L<br>10 | OCP-H<br>8 | Dep-A<br>1 | $\mathcal{H}$ |
|---|---|---|---|---|---|---|
| ☞a. |  | | –0.5 | | –1 | –6 |
| b. |  | –0.5 | | –1 | | –13 |

For floating L tones that are not weakened, deletion is never the right option. OCP-H is either irrelevant or outweighed by a fully violable Max-L.

(13)

| Ø Ø $L_1$ Ø Ø<br>σ σ σ σ | Max-L<br>10 | Dep-L<br>10 | OCP-H<br>8 | Dep-A<br>1 | $\mathcal{H}$ |
|---|---|---|---|---|---|
| ☞ a. Ø Ø $L_1$ Ø Ø<br>σ σ σ σ | | | | –1 | –1 |
| b. Ø Ø Ø Ø<br>σ σ σ σ | –1 | | | | –10 |

Finally, the situation in (5) with SVO/OVS word order can be accounted for if one assumes OCP-H applies differently between a fronted NP and a verb. In this context, syntactic movement displaces the NP, so a different syntactic boundary can be assumed between NP+Verb as opposed to canonical Verb+NP. I assume that there is a strict OCP-H active *within* phonological phrases and a more lenient OCP-H(ϕ) active *across* phonological phrases. The strict OCP-H is sensitive to the *next tone* on the tonal tier, whereas the lenient OCP-H(ϕ) is only sensitive to the *next TBU*. If there is a phonological phrase boundary between fronted nouns and verbs, the distinction between (4c) and (5) is captured, i.e. $H.H^{L0.5}{}_{[V]}$+Ø.$H_{[NP]}$ vs. $H.H^{L0.5}{}_{[NP]}$+Ø.$H_{[V]}$. This is shown in Tableau (14) vs. (15).

(14) Canonical Verb + NP order

| H $L_{0.5}$ Ø H<br>σ σ σ σ | Max-L<br>10 | Dep-L<br>10 | OCP-H(ϕ)<br>8 | OCP-H<br>8 | Dep-A<br>1 | $\mathcal{H}$ |
|---|---|---|---|---|---|---|
| ☞a. H $L_1$ Ø H<br>σ σ σ σ | | –0.5 | | | –1 | –6 |
| b. H Ø H<br>σ σ σ σ | –0.5 | | | –1 | | –13 |

(15) Fronted NP + Verb, ϕ-boundaries indicated

| H $L_{0.5}]_\phi$ $_\phi[$Ø H<br>σ σ σ σ | Max-L<br>10 | Dep-L<br>10 | OCP-H(ϕ)<br>8 | OCP-H<br>8 | Dep-A<br>1 | $\mathcal{H}$ |
|---|---|---|---|---|---|---|
| a. H $L_1]_\phi$ $_\phi[$Ø H<br>σ σ σ σ | | –0.5 | | | –1 | –6 |
| ☞b. $H]_\phi$ $_\phi[$Ø H<br>σ σ σ σ | –0.5 | | | | | –5 |

Tableau (14) has the same outcome as (12): FLD cannot take place in order to avoid a violation of OCP-H within the phonological phrase. In contrast, the winning candidate in Tableau (15) *does* involve FLD because the H tones are separated by a ϕ-boundary, so the strict OCP-H is irrelevant. The lenient OCP-H(ϕ) is not at risk of being violated either, because the TBUs to which the H tones are associated are not adjacent. However, in case of the tonal sequence of Tableau (12) across a ϕ-boundary, lenient OCP-H(ϕ) is violable, and hence the floating L must associate.

**4.3** *Modelling /=kwé/* If the correct underlying form is /=kwé/ rather than toneless /=kwe/, the association of floating L is due to *FINAL-L$_1$ not being at risk of violation.

(16) Word-level phonology: weak H$_{0.1}$ prevents violation of *FINAL-L$_1$

| L$_1$ H$_{0.1}$<br>nduku =kwe | *FINAL-L$_1$<br>100 | MAX-L<br>10 | DEP-A<br>1 | $\mathcal{H}$ |
|---|---|---|---|---|
| a. H$_{0.1}$<br>nduku =kwe | | –1 | | –10 |
| ☞b. L$_1$ H$_{0.1}$<br>nduku =kwe | | | –1 | –1 |

When a floating L is followed by regular toneless enclitics, L remains floating. It survives the word-level phonology and may associate to the next word, as was shown in example (7). Floating is preferable to deletion, so *FLOAT must have a lower weight than MAX-L and is gradiently violable. Worst of all is the candidate in which L$_1$ associates. Because it follows a tonally underspecified root, it does not violate *FINAL-L$_1$ when it remains floating. But when L$_1$ associates to a word-final TBU, there will be preceding TBUs within the same word that it is not associated to.

(17) Word-level phonology: L stays floating

| L$_1$<br>nduku =ni | *FINAL-L$_1$<br>100 | MAX-L<br>10 | *FLOAT<br>8 | DEP-A<br>1 | $\mathcal{H}$ |
|---|---|---|---|---|---|
| a. nduku =ni | | –1 | | | –10 |
| b. L$_1$<br>nduku =ni | –1 | | | –1 | –101 |
| ☞c. L$_1$<br>nduku =ni | | | –1 | | –8 |

In case the root is H.H$^{L0.5}$ instead, the constraints against floating and deletion would both be violated by half, so that floating is still the better option. Association would then not only violate *FINAL-L$_1$, but also DEP-L by half. In case multiple suffixes are added after a root and the word-level phonology has additional cycles, L never associates to enclitics from previous cycles. This could be accounted for in multiple ways, for example by restricting association to older cycles or by forcing association to happen to the rightmost target. A constraint of the latter type is needed anyway in some form, because L never associates leftwards in the language. However, I leave this issue open, because it is not directly related to the central claims of this chapter.

The non-application of HTS to /=kwé/ can likewise be explained by an underlying H that never surfaces. I shall assume that HTS is done to satisfy an alignment constraint.

- ALIGN-R(H,]$\omega$: align H on the right edge of a morpheme to the right edge of the word

This constraint captures the fact that the triggers of HTS are morphemes that end in H and that its undergoers are morphemes without any H to the right. For TBUs that already have H, spreading another H there would either mean deletion of the underlying H or multiple associations of H to the same TBU. Both can easily be ruled out with high-weighted MAX-H and an inviolable constraint penalising two tones of the same type associating to the same TBU. HTS happening to the first TBU in Ø.H-roots will not help prevent a violation of ALIGN-R(H,]$\omega$, so it cannot undergo HTS either. Therefore, all the correct triggers and undergoers follow from this alignment constraint. Note that the right edge of the word changes with each cycle where an enclitic is added. HTS thus has the chance to happen iteratively.

In Tableau (18), candidate a. does not undergo HTS because spreading the first H requires deletion of the second one, which is too costly despite its weakness, as shown by the losing candidate. A candidate where /=kwé/ spreads its H is hopeless, because $H_{0.1}$ would ultimately still need to be made fully active.

(18) Word-level phonology, third cycle: $H_{0.1}$ blocks HTS

| H $H_{0.1}$ L<br>kaʃi =kwe =ndo | Max-H | Align-R(H,]$\omega$) | Dep-A | |
|---|---|---|---|---|
| | 1000 | 5 | 1 | $\mathcal{H}$ |
| ☞a. H $H_{0.1}$ L<br>kaʃi =kwe =ndo | | –1 | | –5 |
| b. H L<br>kaʃi =kwe =ndo | –0.1 | | –2 | –102 |

## 5 Discussion

In this chapter, I argued for a GSR-approach to Floating L tone Deletion and the exceptional enclitic /=kwé/ in Peñoles Mixtec. The difficulty with FLD was that deletion happens only in very specific contexts at the phrasal level, and deletion also seemed unmotivated. The solution was to assume that the L tones subject to deletion are weak, and only retained in contexts where their presence prevents an OCP-H violation. Weak $L_{0.5}$ was derived after H-toned roots due to an independently motivated constraint banning word-final L and an innovative constraint that motivates learners to assume maximally contrasting degrees of activity. In this way, the phonology at the phrasal level could remain entirely simple. A straightforward interaction between OCP-H constraints and Dep/Max constraints accounted for all data. The two exceptional phonological behaviours of the enclitic /=kwé/ were given a unified explanation by assuming it has an underlying H that is too weak to ever surface. Although $H_{0.1}$ needs to be deleted at the phrasal level, it consistently interacts with the constraints at the word level. As a H-toned enclitic rather than a toneless one, /=kwé/ can undergo L tone association and resists High Tone Spreading. The analyses therefore demonstrate two advantages of GSR: 1) increased representational options so as to distinguish weak from strong floating L tones and 2) giving a consistent explanation for morpheme-specific exceptionalities by assuming special representations.

As far as I am aware, no analysis of the FLD data exists in the literature. Daly & Hyman (2007) describe the pattern but do not give a formal explanation, so it has not been shown that these data can be adequately analysed without GSR. I consider the observation of the importance of OCP-H to be a theory-neutral contribution of this chapter, but other than that, an analysis of FLD without gradience would surely look very different. The only exception would perhaps be an analysis that relies on tonal features. It is possible to distinguish fully specified floating L from a floating tonal feature [–Upper] (or equivalent thereof). Similarly to the weakening process I assume, one could assume that floating L after a high-toned couplet is forced to become partially underspecified, which makes it easier to delete at the phrasal level. However, using tonal features has disadvantages that GSR does not have. First, there would need to be at least two tonal features, otherwise deleting a feature would delete the tone entirely (or possibly leave only a tonal root node). However, nothing in Peñoles Mixtec suggests that tonal features are necessary or desirable. Note that Hyman's (2011) criticism on tonal features appeared after his work on Peñoles Mixtec that did use them. As such, a learner would have to acquire a subtonal structure consisting of multiple features based solely on extremely abstract evidence. Alternatively, subtonal structure is innate, but assuming that features are innate is a controversial position for its own reasons. A second disadvantage of features is that, even if a two-feature subtonal structure is assumed, there is no way for the learner to determine *which* tonal feature this should be deleted at the word-level; it could be a register feature or tone level feature. Third, deriving the feature deletion at the word level and complete deletion or re-specification at the phrase-level requires more constraints that are also more complex. Fourth, the analysis would rely on derived partial underspecification, which is more abstract than GSR. Fifth, tonal features cannot play a role in segmental analyses. GSR, on the other hand, can do so. If either has to be assumed in order to explain the data, GSR is the more versatile choice. Using both is redundant, so if one avoids using GSR for Peñoles Mixtec, one should commit to never using it.

The analysis also has some implications for the theory of GSR itself. Perhaps the biggest debate among those who use GSR is whether there can be gradience in the output (Hsu, 2022). Arguments against output-gradience are that it massively increases the number of possible candidates, requires new mechanisms to force deletion in some cases (e.g. with output-gradience a weak $H_{0.1}$ could stay as is), and that differently active phonological objects should have a different phonetic implementation. In this chapter, I have taken an intermediate stance. By assuming there can be output-gradience at the word level but not in phrase-level final inputs, arguments against output-gradience do not apply or are at least weakened. The advantage it offers is allowing gradience to be derived, which is very useful for processes that are purely phonological such as FLD. The version of GSR used here therefore strikes a balance between being restricted and increasing analytical options. There is no need to assume gradient activity levels above 1 either, so that straightforward full activity at 1 is retained. The novel DIFFERENTIATEGRADIENCE-constraint further curtailed the number of possibilities by forcing derived degrees of gradience to be different from other degrees. This intuition behind this constraint is that phonology tends to maximise contrasts, reducing the searching space for learners.

In conclusion, this brief chapter has demonstrated that GSR is highly useful in analyses of complex tonal phenomena in Peñoles Mixtec, including morpheme-specific exceptions. The driving force behind the distinction between strong and weak L tones was independently motivated and made deletion of L at the phrasal level simple and consistent. Possible alternative analyses without GSR would require very complicated and unmotivated constraints, so they would fare less well. An implication for the theory itself is that output-gradience does not need to be categorically possible or impossible, but can be restricted to non-final levels/strata. This enables the derivation of contrasting gradience, which is useful for purely phonological processes.

## References

Appelbaum, Irene (2017). Modularity. Bechtel, William and Graham, George (ed.), *A Companion to Cognitive Science*, Blackwell, 625–635.

Bermúdez-Otero, Ricardo (2018). Stratal Phonology. Hannahs, S. J. & A. R. K. Bosch (eds.), *The Routledge Handbook of Phonological Theory*, Routledge, 100–134.

Daly, John P. & Larry M. Hyman (2007). On the Representation of Tone in Peñoles Mixtec. *International Journal of American Linguistics* 73:2, p. 165–207.

Eberhard, David M., Gary F. Simons & Charles D. Fennig (2024). *Ethnologue: Languages of the World*. SIL International, Twenty-seventh edn. Online version: http://www.ethnologue.com.

Hsu, Brian (2022). Gradient Symbolic Representations in Harmonic Grammar. *Language & Linguistics Compass.* e12473.

Hyman, Larry M. (2011). Do tones have features? Goldsmith, John A., Elizabeth Hume & Leo Wetzels (eds.), *Tones and Features*, De Gruyter Mouton, 50–80.

Legendre, Géraldine, Miyata Yoshihiro & Paul Smolensky (1990). Harmonic Grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. Anderson, John R. (ed.), *Proceedings of the twelfth annual conference of the Cognitive Science Society*, 388–395.

Nespor, Marina & Irene Vogel (1986). *Prosodic Phonology*. Foris Publications.

Oostendorp, Marc van (2006). A theory of morphosyntactic colours. Ms, Meertens Institute, Amsterdam. Available online at http://egg.auf.net/06/docs/Hdt.

Pater, Joe (2010). Morpheme-specific phonology: Constraint indexation and inconsistency resolution. Parker, Steve (ed.), *Phonological argumentation: Essays on evidence and motivation*, Equinox, p. 123–154.

Sande, Hannah, Peter Jenks & Sharon Inkelas (2020). Cophonologies by Ph(r)ase. *Nat Lang Linguist Theory* 38, 1211–1261.

Smolensky, Paul & Matthew Goldrick (2016). Gradient symbolic representations in grammar: The case of French liaison. Ms, Johns Hopkins University and Northwestern University. *ROA* 128.

Zimmermann, Eva (2021). Faded copies: Reduplication as distribution of activity. *Glossa: a journal of general linguistics* 6:1. 58.