

# Similarity and Correspondence in Chol Mayan<sup>\*</sup>

Jessica Coon and Gillian Gallagher

Massachusetts Institute of Technology

## 1. Introduction

In this paper we examine the implications of new data from Chol (Mayan) for the analysis of consonant cooccurrence restrictions and the computation of similarity between pairs of consonants. As in many languages, certain consonants in Chol are prohibited from cooccurring within a lexical root. Our goal is to account for why some pairs of consonants may cooccur in a CVC root and others may not. We show that differential feature weighting in the similarity calculation is necessary to account for the cooccurrence restrictions found in Chol CVC roots.

The relevant data are shown in (1) and (2). In (1a) we observe that pairs of non-identical ejectives and pairs of non-identical *plain* (non-ejective) stridents may not cooccur in a root. Identical pairs of ejectives and stridents are grammatical, as in (1b). An ejective and a *non*-ejective strident *may* cooccur (if they agree in anteriority) as in (2).

- |     |                       |                                     |                |
|-----|-----------------------|-------------------------------------|----------------|
| (1) | a. <u>unattested</u>  | non-identical ejectives             | *k'ap'         |
|     |                       | non-identical plain stridents       | *tsus          |
|     | b. <u>grammatical</u> | identical ejectives                 | ✓k'ak'         |
|     |                       | identical stridents                 | ✓tsuts         |
| (2) | <u>grammatical</u>    | stridents that differ in ejectivity | ✓ts'as, ✓ʃutʃ' |

We analyze the patterns in (1) and (2) as similarity avoidance, following a growing literature that relates cooccurrence restrictions to similarity (Pierrehumbert

---

<sup>\*</sup> Many thanks to Adam Albright and Donca Steriade for reading and commenting on previous versions of this work. Thanks also to Edward Flemming, Peter Graff, Irene Heim, John McCarthy, Joe Pater, David Pesetsky, Norvin Richards, the members of the Spring 2007 generals workshop at MIT, as well as audiences at the UMass-MIT Meeting and MIT Phonology Circle. We are very grateful to Chol consultants Matilde and Dora Angélica Vázquez Vázquez and Virginia Martínez Vázquez.

Authors' names are listed in alphabetical order.

1993; Kessler & Treiman 1997; MacEachern 1999; Frisch et al 2004; Rose & Walker 2004). As the forms in (1b) illustrate, Chol is characterized by what MacEachern (1999) calls the TOTAL IDENTITY EFFECT: completely identical consonants escape the cooccurrence restrictions. The preference for identical pairs of segments over similar pairs is accounted for in the AGREEMENT BY CORRESPONDENCE (ABC) framework of Rose & Walker (2004) (also see Hansson 2001). In ABC, similar segments are required to stand in correspondence and corresponding segments are required to agree in some or all features. The main challenge posed by the data in (1) and (2) is to explain why stridents differing in ejectivity, as in (2), escape the correspondence requirement (and are thus allowed to cooccur), while pairs of non-ejective stridents do not.

We propose that ejectivity licenses the cooccurrence of non-identical stridents like those in (2) because it renders these two segments adequately *dissimilar*. A successful analysis of Chol requires a method of modeling the interaction of features in a similarity computation. Pairs of plain stridents and pairs of ejectives must be scored as highly similar and thus ungrammatical unless they are identical. Grammatical combinations, such as an ejective and non-ejective strident, must be judged dissimilar enough to escape the identity requirement.

We show that while other similarity metrics have been successful in accounting for some phenomena (Frisch et al 2004; Rose & Walker 2004), they cannot easily account for the importance of ejective (dis)agreement in Chol. Instead, we propose a similarity metric in which certain features receive more weight than others. The rest of this paper is organized as follows: §2 presents the Chol data. The ABC framework and analysis is presented in §3. §4 develops the similarity metric and §5 concludes.

## 2. Background and Data<sup>1</sup>

Chol is a Mayan language spoken in Chiapas, Mexico by around 150,000 people. Chol's consonant inventory contains twenty consonants, shown in (3). There are five ejectives and six stridents, all shown in bold. Chol has six vowels [a, e, i, o, u, ɨ], each with a lengthened aspirated counterpart [a<sup>h</sup>, e<sup>h</sup>, i<sup>h</sup>, o<sup>h</sup>, u<sup>h</sup>, ɨ<sup>h</sup>]. Vowel quality appears to play no role in the cooccurrence restrictions.

### (3) Chol consonant inventory:

	labial	coronal	velar	glottal
implosive	b			
plosive	p	t <sup>i</sup>	k	ʔ
ejective	<b>p'</b>	<b>ts' tf' t<sup>i</sup>'</b>	<b>k'</b>	
fricative		s ʃ		h
affricate		<b>ts tf</b>		
nasal	m	ɲ		
approximant	w	l j		

<sup>1</sup> The Chol data presented here draw heavily on the 1978 Aulie and Aulie dictionary, combined with fieldnotes collected by the first author in Chiapas, Mexico.



The roots with combinations of stridents in (9) are all ungrammatical. In (9a), a plain affricate and fricative cooccur.<sup>2</sup> In (9b), an ejective and a plain strident disagree for [ $\alpha$  anterior]. We do not discuss the role of anteriority harmony in this paper. Elsewhere, we analyze anteriority harmony as articulatory spreading (Coon & Gallagher 2007).

- |     |    |       |    |        |
|-----|----|-------|----|--------|
| (9) | a. | *tsus | b. | *ts'af |
|     |    | *tʃuʃ |    | *tʃ'us |
|     |    | *sats |    | *ʃats' |

The cooccurrence restrictions in (5) and (7) pose three central questions: First, why are pairs of non-identical ejectives and plain stridents disallowed, but not other pairs, e.g. coronals or nasals? Second, why does ejectivity license the otherwise ungrammatical cooccurrence of two non-identical stridents? Finally, why are identical consonants permitted to cooccur while minimally different consonants are not?

To address the first question, we propose that ejectivity and stridency are special because they are more salient than other features. Sharing one of these features thus makes two segments more similar than sharing other features. Similar (but non-identical) consonants are prohibited from cooccurring (see §3). Second, we argue that ejectivity is *more* salient than stridency: disagreement in ejectivity outweighs agreement in stridency, rendering a pair of consonants like [ts'] and [s] dissimilar enough to cooccur. The importance of ejectivity can be seen by the fact that while differing in ejectivity licenses the cooccurrence of two non-identical stridents (✓[ts'as]), differing in stridency does *not* license the cooccurrence of two non-identical ejectives (\*[ts'ak']). We show below that differential feature weighting must be incorporated into the similarity computation. Finally, the preference for identical consonants over similar pairs is accounted for in the Agreement by Correspondence (ABC) framework of Rose & Walker (2004).

### 3. Agreement by Correspondence (ABC)

Rose & Walker (2004) develop the Agreement by Correspondence (ABC) framework to account for languages in which certain similar segments are forced to agree for some or all features. Consonant agreement across unaffected intervening segments is referred to as Long Distance Consonant Agreement (LDCA) and is distinct from local assimilation between adjacent segments.

Rose & Walker propose that LDCA has functional origins in speech planning and production. Speakers form connections (hence, “correspondence”) between similar segments. Segments that are very similar but not identical are more difficult to produce than identical segments. They argue that LDCA is a grammaticalized avoidance of difficult consonant combinations. Similarity, then, plays a crucial role in the system.

---

<sup>2</sup> One counterexample was found to the generalization in (7): [tʃaʃ] ‘mosquito’. There were no counterexamples found to the ejectivity restriction in (5).

### 3.1 Outline of the Framework

In an ABC approach, cooccurrence restrictions with the identity effect are analyzed along with alternation-inducing consonant harmonies as effects of correspondence between the interacting consonants. Similar pairs of consonants are placed in correspondence; identity constraints then force corresponding consonants to agree. Correspondence relations between output consonants are established by  $\text{CORR-C} \Leftrightarrow \text{C}$  constraints, as shown in (10).

(10)  $\text{CORR-C} \Leftrightarrow \text{C}$ :

Let  $S$  be an output string of segments and let  $C_i, C_j$  be segments that share a specified set of features  $F$ . If  $C_i, C_j \in S$ , then  $C_i$  is in a relation with  $C_j$ : that is,  $C_i$  and  $C_j$  are correspondents of one another.

The constraint in (10) is relativized to specific features. Rose & Walker (2004) propose that feature-specific  $\text{CORR-C} \Leftrightarrow \text{C}$  constraints stand in a fixed hierarchy where constraints referring to more similar pairs of consonants outrank those referring to less similar pairs. This reflects the fact, also noted by MacEachern (1999), that more similar pairs of consonants are more likely to exhibit cooccurrence restrictions than less similar pairs. A mini hierarchy showing the interaction of place and ejectives in stops (taken from Rose & Walker 2004, 50b) is given in (11). Here, the highest ranked  $\text{CORR-C} \Leftrightarrow \text{C}$  constraint requires that the most similar pair of consonants, homorganic ejectives, correspond. The similarity of segments required to be in correspondence decreases down the hierarchy: homorganic stops which disagree in ejectivity,  $\text{CORR-T}' \Leftrightarrow \text{T}$ , followed by heterorganic stops which disagree in ejectivity,  $\text{CORR-T}' \Leftrightarrow \text{K}$ .

(11)  $\text{CORR-T}' \Leftrightarrow \text{T}' \gg \text{CORR-T}' \Leftrightarrow \text{T} \gg \text{CORR-T}' \Leftrightarrow \text{K}$

The constraints in (11) enforce correspondence between pairs of output segments. If two segments stand in correspondence, they are required to agree in certain features by constraints from the  $\text{CC-IDENT}[F]$  family, given in (12).

(12)  $\text{CC-IDENT}[F]$ :

Let  $C_i$  be a segment in the output and  $C_j$  be any correspondent of  $C_i$  in the output. If  $C_i$  is  $[\alpha F]$  then  $C_j$  is  $[\alpha F]$ .

For harmony to occur in a language, some  $\text{CORR-C} \Leftrightarrow \text{C}$  and  $\text{CC-IDENT}$  constraint must outrank  $\text{IO-IDENT}$ , as schematized in (13).

(13) Necessary ranking for harmony:

$\text{CORR-C} \Leftrightarrow \text{C}, \text{CC-IDENT}[F] \gg \text{IO-IDENT}[F]$

Languages differ in how similar two consonants may be and still cooccur without harmonizing for some feature. The differences between languages in this respect results from the ranking of input-output faithfulness constraints in the  $\text{CORR-C} \Leftrightarrow \text{C}$  hierarchy. The

ranking of IO-IDENT[F] along hierarchies like the one in (11) will determine which pairs of consonants are subject to cooccurrence restrictions in a given language. For example, by placing IO-IDENT between the last two members of the hierarchy in (11), we would require correspondence between all homorganic stops, but not between heterorganic ones.

Both CORR-C $\leftrightarrow$ C and CC-IDENT[F] are necessary since the features that determine correspondence are distinct from those that must agree. In Aymara, for example, homorganic stops are required to agree in ejectivity and aspiration; heterorganic stops are not subject to this requirement (MacEachern 1999; Rose & Walker 2004). Place features thus determine which consonants stand in correspondence (CORR-C $\leftrightarrow$ C); corresponding consonants must agree for ejectivity and aspiration (CC-IDENT[F]).

The tableau in (14) illustrates an ABC account of the ejective cooccurrence restriction in Chol. CORR-C' $\leftrightarrow$ C' demands correspondence between any two ejectives. The general constraint CC-IDENT is used to stand for all CC-IDENT[F] constraints needed to account for identity in Chol.<sup>3</sup> Matching subscript letters indicate correspondence. Candidates (14b) and (14d) are eliminated by high-ranked CORR-C' $\leftrightarrow$ C' because the two ejectives in the output do not correspond. In (14c), the ejectives correspond but do not agree, violating high-ranked CC-IDENT. The winning form in (14a) correctly places ejectives in correspondence and satisfies CC-IDENT by making the ejectives identical.<sup>4</sup>

(14)

	/k'ap'/	CORR-C' $\leftrightarrow$ C'	CC-IDENT	IO-IDENT	CORR-C' $\leftrightarrow$ C
a.	→ k' <sub>x</sub> ak' <sub>x</sub>			*	
b.	k' <sub>x</sub> ak' <sub>y</sub>	*!		*	
c.	k' <sub>x</sub> ap' <sub>x</sub>		*!		
d.	k' <sub>x</sub> ap' <sub>y</sub>	*!			

### 3.2 The Problem for Chol

While the correspondence-based approach proposed in Rose & Walker (2004) can account for the ejective restrictions in Chol, it is unclear how to explain the more complex strident facts. Recall that in Chol two plain (non-ejective) stridents must be identical to cooccur, but an *ejective* and a plain strident may cooccur: \*[tsas], ✓[ts'as]. In order to capture the ungrammaticality of two non-identical plain stridents, stridents must correspond. This can be done if CORR-S $\leftrightarrow$ S is high ranked, where 'S' stands for all six Chol stridents. As shown in (15), the ungrammatical form \*[tsas] is correctly ruled out.

<sup>3</sup> We discuss the role of CC-Ident[F] in the analysis of Chol in Coon & Gallagher (2007).

<sup>4</sup> [k'ak'] is not the only possible output for an ungrammatical input /k'ap'/. Rather than assimilation, dissimilation could occur, as in /k'ap'/ → [k'ap]. If one of the consonants loses ejectivity, the consonants are no longer required to be in correspondence, since CORR-C' $\leftrightarrow$ C is low-ranked. Other possibilities include [p'ap'] and [kap']. There are no alternations in Chol which tell us the actual output of an ungrammatical input like /k'ap'/.

*Similarity and Correspondence in Chol Mayan*

(15)

/tsas/	CORR-S $\Leftrightarrow$ S	CC-IDENT	IO-IDENT
a. $\rightarrow$ ts <sub>x</sub> ats <sub>x</sub>			*
b. ts <sub>x</sub> ats <sub>y</sub>	* !		*
c. ts <sub>x</sub> as <sub>x</sub>		* !	
d. ts <sub>x</sub> as <sub>y</sub>	* !		

While this ranking can account for the behavior of plain stridents, it incorrectly demands identity between an ejective and a plain strident, as shown in (16), where we see that the attested form  $\checkmark$ [ts'as] is incorrectly ruled out. If [ts'] and [s] are *not* placed in correspondence, as in (16c) the form is ruled out by CORR-S $\Leftrightarrow$ S, which requires all stridents to correspond. If, on the other hand, [ts'] and [s] *are* placed in correspondence, CORR-S $\Leftrightarrow$ S is satisfied but the stridents must then be identical to avoid a fatal violation of CC-IDENT.

(16)

/ts'as/	CORR-S $\Leftrightarrow$ S	CC-IDENT	IO-IDENT
a. $\rightarrow$ ts' <sub>x</sub> ats' <sub>x</sub>			**
b. ts' <sub>x</sub> ats' <sub>y</sub>	* !		**
c. $\checkmark$ ts' <sub>x</sub> as <sub>y</sub>	* !		
d. ts' <sub>x</sub> as <sub>x</sub>		* !	

As discussed above, the role of similarity in Rose & Walker's theory is in the fixed hierarchy of CORR-C $\Leftrightarrow$ C constraints: constraints referring to more similar pairs of segments outrank those referring to less similar pairs, as in (11) above. One solution to the problem in (16) is to propose a more fine-grained hierarchy for pairs of stridents and ejectives. In (17), C stands for non-strident consonants and S for strident consonants. To account for the Chol facts we would place IO-IDENT between the last two members of the hierarchy, requiring correspondence between all pairs of ejectives (strident or not), as well as between plain stridents (S $\Leftrightarrow$ S), but *not* between an ejective and non-ejective strident (S' $\Leftrightarrow$ S).

- (17) Possible ranking of Corr-C $\Leftrightarrow$ C constraints in Chol:  
 Corr-C' $\Leftrightarrow$ C' >> Corr-C' $\Leftrightarrow$ S' >> Corr-S' $\Leftrightarrow$ S' >> Corr-S $\Leftrightarrow$ S >> Corr-S' $\Leftrightarrow$ S

While the hierarchy in (17) can handle the data, there is no formal mechanism in Rose & Walker's theory for creating such a hierarchy based on a set of interacting features. There is no explanation for why [ejective] is more important to similarity than [strident]. There is nothing, for example, to rule out the ranking in (18), where [strident] is more important to similarity than [ejective] (predicting the possibility of a language in which non-identical ejectives are permitted only if they disagree in stridency (\*[k'-p'] but  $\checkmark$ [k'-ts']).

- (18) Impossible ranking for Chol, allowed in Rose & Walker's analysis:  
 Corr-S' $\Leftrightarrow$ S' >> Corr-S $\Leftrightarrow$ S >> Corr-S' $\Leftrightarrow$ S >> Corr-C' $\Leftrightarrow$ C' >> Corr-C' $\Leftrightarrow$ S'

Rose & Walker (2004) suggest that their hierarchies are based partly on the similarity computation of Frisch et al. (2004), which is based on the number of shared natural classes (discussed in §4). Consequently, pairs of segments that differ in the same number of features have similar similarity scores. The Frisch et al. metric does *not* capture the fact that [ejective] is more important than [strident], or that [ejective] and [strident] are more important than [place], though these asymmetries are present in Chol. In the next sections we formalize a notion of similarity that weights features differently, and formulate a constraint which makes direct reference to this computation. While a fixed constraint ranking like (17) captures the Chol facts, our analysis goes a step further by deriving this ranking and making the computation of similarity explicit.

### 3.3 The Solution

We propose a single reformulated CORR-C $\Leftrightarrow$ C constraint, shown in (19). This modified constraint makes direct reference to a numerical similarity score between two consonants.

(19) CORR-C $\Leftrightarrow$ C ( $\Sigma \{C_i-C_j\} > n$ ):

Let S be an output string of segments and  $\Sigma \{C_i-C_j\}$  be the similarity of  $C_i$  and  $C_j$ . If consonants  $C_i, C_j \in S$ , and  $\Sigma \{C_i-C_j\} > n$ , then  $C_i$  and  $C_j$  are in correspondence.

The constraint in (19) demands correspondence between any two consonants whose similarity score,  $\Sigma$ , is greater than the similarity threshold,  $n$ . Both the value of  $n$  and certain aspects of the similarity computation may vary from language to language. The constraint in (19) can account for the cooccurrence restrictions in Chol if the similarity computation gives us the distribution of similarity scores in (20). Pairs of consonants which may not cooccur unless they are identical must receive a similarity score greater than  $n$ , forcing them into correspondence. Pairs which are free to cooccur must have similarity scores less than  $n$ .

(20)  $\Sigma (k'-p', ts-s) > n$  (may not cooccur unless identical)  
 $\Sigma (k-p, ts'-s) < n$  (free to cooccur)

If our similarity metric gives us the distribution in (20), we can account for the identity effect between pairs of ejectives and pairs of plain stridents, shown in (21) and (22). The unattested forms in (21b) and (21c) are eliminated since the similarity values for pairs of ejectives is above the threshold, yet they are not in correspondence. The ejectives in (21d) are correctly placed in correspondence, yet this form is ruled out since the ejectives are not identical, in violation of high-ranked CC-IDENT. The winning output form contains identical ejectives.

(21)

$/k'ap'/$	CORR-C $\Leftrightarrow$ C > n	CC-IDENT	IO-IDENT
a. $\rightarrow k'_xak'_x \quad \Sigma > n$			*
b. $k'_xak'_y \quad \Sigma > n$	*!		*
c. $k'_xap'_y \quad \Sigma > n$	*!		
d. $k'_xap'_x \quad \Sigma > n$		*!	

## Similarity and Correspondence in Chol Mayan

A similar state of affairs is seen with the plain stridents in (22).

(22)

	/tsas/	CORR-C $\Leftrightarrow$ C > n	CC-IDENT	IO-IDENT
a.	$\rightarrow$ ts <sub>x</sub> ats <sub>x</sub> $\Sigma > n$			*
b.	ts <sub>x</sub> ats <sub>y</sub> $\Sigma > n$	* !		*
c.	ts <sub>x</sub> as <sub>y</sub> $\Sigma > n$	* !		
d.	ts <sub>x</sub> as <sub>x</sub> $\Sigma > n$		* !	

Crucially, we can now also account for the grammatical cooccurrence of an ejective and a plain strident, as in the attested form  $\checkmark$ [ts'as]. If our similarity metric places this pair of consonants *below* the similarity threshold, the consonants will not be required to correspond. The winning candidate in (23a) is faithful to the input and incurs no violations of high-ranked CORR-C $\Leftrightarrow$ C > n or CC-IDENT. In the section that follows, we show that differential feature weighting is needed to achieve these similarity scores.

(23)

	/ts'as/	CORR-C $\Leftrightarrow$ C > n	CC-IDENT	IO-IDENT
a.	$\rightarrow$ ts' <sub>x</sub> as <sub>y</sub> $\Sigma < n$			
b.	ts' <sub>x</sub> ats' <sub>y</sub> $\Sigma > n$	* !		**
c.	ts' <sub>x</sub> ats' <sub>x</sub> $\Sigma > n$			** !
d.	ts' <sub>x</sub> as <sub>x</sub> $\Sigma < n$		* !	

### 4. The Similarity Computation

In this section we propose a method of computing similarity between consonants that allows us to set a similarity threshold,  $n$ , to which the constraint CORR-C $\Leftrightarrow$ C > n will refer. Recall that this similarity metric must place all ungrammatical pairs of consonants *above* the threshold (thereby requiring them to stand in correspondence and be fully identical by high-ranked CC-IDENT), and all grammatical ones *below* the threshold (allowing them to cooccur freely), as in (20) above. Our similarity metric differs from that proposed in Frisch et al. (2004) in giving certain features more weight in the computation than others. We show that this differential weighting is necessary in order to account for the Chol data discussed in the preceding sections.

#### 4.1 Frisch et al. (2004)

The similarity metric proposed in Frisch et al. (2004) provides an account of gradience in consonant cooccurrence in Arabic roots. In Arabic, homorganic consonants that agree for more subsidiary features (stricture, sonorancy, voicing) are less grammatical than those that agree for fewer subsidiary features. The goal of Frisch et al. (2004) is to formulate a similarity metric in which the similarity of a pair of consonants correlates with its grammaticality. The similarity computation is based on shared natural classes. Since natural classes are based on features, consonants which share a large number of features are computed as highly similar. The Frisch et al. computation—which Rose & Walker (2004) use to establish the fixed ranking of CORR-C $\Leftrightarrow$ C constraints—is given in (24).

(24) Frisch et al. (2004) similarity metric:

$$\text{Similarity} = \frac{\text{shared natural classes}}{\text{shared natural classes} + \text{unshared natural classes}}$$

We used the similarity metric in (24) to calculate similarity values for all pairs of Chol consonants, in order to determine whether it would allow us to predict which pairs of consonants are subject to cooccurrence restrictions.<sup>5</sup> In the table in (25) we give a sample of similarity scores for several pairs of grammatical and ungrammatical Chol roots. In (25) we observe that the line between grammatical and ungrammatical pairs does not correspond with any possible numerical similarity threshold, since the values above the line are not consistently higher than those below the line. Put differently, there is no possible value for *n* which places all grammatical pairs below *n* and all ungrammatical pairs above *n*.

(25) Similarity values achieved by Frisch et al. (2004) metric:

root form	similarity value	
* k'-p'	.56	
* ts-s	.38	
* ts'-t <sup>j</sup> '	.56	no possible value for <i>n</i> ↓
✓k-p	.62	
✓s-ts'	.32	
✓t <sup>j</sup> -ts	.63	

The problem with Frisch et al.'s metric is that all features are equally weighted. Consequently, two segments that differ in a single feature are given very similar scores. In Chol, however, we have seen that grammaticality is sensitive to *which* features are shared or unshared. Specifically, the features [ejective] and [strident] play an important role in determining the grammaticality of Chol roots.

The problem is made more explicit in (26). Here we observe that the natural class model does not give us the correct correlation between similarity and grammaticality. We expect grammatical pairs to have a *lower* similarity score than ungrammatical pairs. In fact, the grammatical pairs in (26) have *higher* similarity scores than the ungrammatical pairs—the opposite of the desired result.

(26)	consonants	differing features	similarity
a.	* k'-p'	place	.56
	✓k-p	place	.62
b.	* ts-s	continuant	.38
	✓t <sup>j</sup> -ts	strident	.63

<sup>5</sup> Similarity values were calculated with the 'Segmental Similarity Calculator' script developed by Adam Albright (<http://web.mit.edu/albright/www/>). The full results can be found in Coon & Gallagher (2007).

## 4.2 A Weighted Similarity Metric

Above we saw that Frisch et al.'s (2004) metric did not give us the correct results. Pairs that differ in the same number of features—regardless of *which* features—receive similar similarity scores. Instead, we require a similarity computation in which certain features receive more weight than others, such as the one proposed in Tversky (1977). Tversky provides experimental evidence that more salient features contribute more to perceptions of similarity. Frisch et al. formalized this idea by basing their computation on shared natural classes, as opposed to shared features (Pierrehumbert 1993). In their system, contrastive features contribute more than redundant features (e.g. [voice] in sonorants). While appealing to natural classes instead of individual features eliminates redundant features from the computation, this is not enough to account for the Chol data.

Specifically, agreeing in certain features (like ejectives) must make two segments more similar than agreeing in other features (like continuancy). We also know that *disagreeing* in ejectivity licenses the cooccurrence of two (otherwise very similar) stridents, as seen by the contrast between \*[tsas] and ✓[ts'as]. To compute similarity in Chol, then, we need the features [strident] and [ejective] to have more weight than other features. Furthermore, we need [ejective] to be worth more than [strident], to account for the fact that *not* agreeing in ejectivity makes two consonants more dissimilar than agreeing in stridency makes them similar. The feature weights we use are given in (27).<sup>6</sup>

- (27) a. [ejective] > [strident] > others  
 b. [ejective] 3  
     [strident] 2  
     others 1 ([continuant], [voice], [sonorant], [labial], [coronal], [dorsal],  
               [glottal], [anterior], [nasal], [lateral])

The numerical values in (27) allow the effect of [ejective] and [strident] to be decisive, even when a pair of segments share or differ in a large number of other features. For example, the consonant pairs \*[ts-s] and ✓[t'-t] each differ in exactly one feature—[continuant] and [strident] respectively—but their similarity scores are different (5 vs. 2) because disagreeing in continuancy is less important to similarity than disagreeing in stridency. The similarity of two consonants is equal to the sum of the weights of shared features, minus the sum of the weights of unshared features, as formalized in (28).

- (28) Similarity  
 Let  $x$  and  $y$  be segments consisting of the feature sets  $x = \{x_1, x_2, \dots\}$  and  $y = \{y_1, y_2, \dots\}$ ;  
 Let  $shared = x \cap y$  and  $unshared = x \setminus y \cup y \setminus x$  (where  $A \setminus B$  denotes the relative complement of  $A$  in  $B$ ,  $\{b \in B \mid b \notin A\}$ );  
 Let  $w(f)$  be the weight of feature  $f$ , as given in (26b);  
 Then,  $similarity(x, y) = \sum w(shared) - \sum w(unshared)$ .

---

<sup>6</sup> Exact numerical values are unimportant. Rather, we are concerned with *relative* feature weights.

We propose that [ejective] and [strident] are privative features and that unvalued features are ignored by the similarity computation. This accounts for the fact that while sharing ejectivity renders two consonants very similar (e.g. \*[kʰ]-[pʰ]), sharing *non*-ejectivity does not (e.g. ✓[k]-[p]). Since [ejective] is privative, the pair [k]-[p] does not agree for [-ejective]. This is shown by the contrast in (29). The similarity value of the consonants [kʰ] and [pʰ] is 4, as illustrated in (29a), while the similarity value of [k]-[p] is 1, as in (29b). Despite the fact that each pair of consonants differs in exactly two features, [labial] and [dorsal], the weighted similarity metric correctly assigns them different scores due to the high weight and privativity of ejectivity.

- (29) a. similarity (kʰ,pʰ) = (3+1+1+1) – (1+1) = 4
- |         |                    |           |           |          |    |
|---------|--------------------|-----------|-----------|----------|----|
| shared: | [-continuant]      | +1        | unshared: | [labial] | -1 |
|         | [-voice]           | +1        |           | [dorsal] | -1 |
|         | [-sonorant]        | +1        |           |          |    |
|         | <b>[+ejective]</b> | <b>+3</b> |           |          |    |
- b. similarity (k,p) = (1+1+1) – (1+1) = 1
- |         |                     |            |           |          |    |
|---------|---------------------|------------|-----------|----------|----|
| shared: | [-continuant]       | +1         | unshared: | [labial] | -1 |
|         | [-voice]            | +1         |           | [dorsal] | -1 |
|         | [-sonorant]         | +1         |           |          |    |
|         | <b>[0 ejective]</b> | <b>n/a</b> |           |          |    |

A sample of the results of the similarity computation is given in (30).<sup>7</sup> The values in (30) show that by using the weighted similarity metric, we can now set a similarity threshold that determines whether a pair of consonants in Chol may or may not cooccur. All pairs which must be in correspondence (and are thus required to be identical) have similarity values greater than 3, while all pairs which are free to cooccur have values lower than 3. If the similarity threshold is set at 3, pairs of ejectives and pairs of plain stridents will necessarily correspond (and thus agree) via our modified CORR-C ⇔ C > n constraint, but an ejective strident and a plain strident will not.

(30) Weighted similarity metric results:

root form	weighted similarity
* kʰ-pʰ	4
* ts-s	5
* tsʰ-tʃʰ	8
* tsʰ-tʰ	4
similarity threshold $n=3$ ↓	
✓k-p	1
✓tʰ-tsʰ	-2
✓tʰ-ts	1
✓m-n	2
✓l-j	2
✓tsʰ-s	2

<sup>7</sup> The full results may be found in Coon & Gallagher (2007).

## *Similarity and Correspondence in Chol Mayan*

The table in (31) compares a sample of results from our weighted metric with those from Frisch et al.'s for pairs of minimally different consonants. In the weighted metric ungrammatical pairs are assigned values consistently higher than those of grammatical pairs; this is not the case for the Frisch et al. unweighted metric.

(30)

	consonants	differing features	weighted similarity	Frisch et al. similarity
a.	* kʔ-pʔ	place	4	.56
	✓k-p	place	1	.62
b.	* ts-s	continuant	5	.38
	✓tʰ-ts	strident	1	.63

### 5. Conclusion

The interaction between stridency and ejectivity in Chol consonant cooccurrence restrictions highlighted a shortcoming of Rose & Walker's (2004) theory of Agreement by Correspondence. Specifically, we showed that when a restriction involves multiple interacting features, as is the case for stridents in Chol, it is unclear how to determine what the fixed hierarchy of CORR-C $\Leftrightarrow$ C constraints should be. While Rose & Walker write that their fixed hierarchy is based on the similarity computation in Frisch et al. (2004), it was shown that a similarity metric of this type, in which all features contribute the same weight to the computation, does not achieve the correct results for Chol.

We argued instead that certain features must contribute more to the similarity computation than others (Tversky 1977). A weighted similarity computation, in which the features [ejective] and [strident] are weighted more than other features, enabled us to set a similarity threshold to which our modified CORR-C $\Leftrightarrow$ C > n constraint can refer, and correctly accounts for the difference in grammaticality between pairs of segments that differ in the same number of features.

While we suggest that the high weights assigned to ejectivity and stridency reflects a high degree of saliency for these features, a number of questions remain open to future work on feature weighting in similarity computations. Among these is the question of how and why feature weightings differ from language to language. In Chol, for example, [sonorant] has a low weighting. In Arabic, however, Frisch et al. (2004) note that sonorancy plays an important role in the similarity of coronals. Further work also remains to be done to understand why different languages have cooccurrence restrictions on different classes of segments. For instance, while Chol, like many other languages (see MacEachern 1999 and Hansson 2001), has a restriction on ejectivity (a laryngeal feature), other languages like Arabic primarily have restrictions on homorganic segments. Whether these differences can be predicted from other facts about these languages remains an open question.

## References

- Aulie, Wilbur and Evelin Aulie. 1978. *Diccionario Ch'ol-Español, Español-Ch'ol*. México: Summer Institute of Linguistics.
- Ayres, Glenn. 1991. *La gramática Ixil*. Guatemala: Centro de Investigaciones Regionales de Mesoamérica.
- Coon, Jessica and Gillian Gallagher. 2007. Similarity and correspondence in Chol roots. Ms, MIT. URL: <http://web.mit.edu/~jcoon/www/papers.html>
- Cowan, Marion. 1969. Tzotzil grammar. In *Summer Institute of Linguistics Publications*. Norman: University of Oklahoma.
- Dayley, Jon. 1985. *Tzutujil grammar*. Berkely: University of California Press.
- Frisch, Stefan, Janet Pierrehumbert, and Michael Broe. 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22:179-228.
- Hansson, Gunnar Olafur. 2001. Theoretical and typological issues in consonant harmony. Doctoral Dissertation, University of California, Berkeley.
- Kaufman, Terrence. 1971. Tzeltal phonology and morphology. In *University of California publications in linguistics*, 61. Berkeley: University of California Press.
- Keller, Kathryn C. 1959. The phonemes of Chontal (Mayan). *International Journal of American Linguistics* 25:44-53.
- Kessler, Brett and Rebecca Treiman. 1997. Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language* 37:295-311.
- MacEachern, Margaret. 1999. *Laryngeal co-occurrence restrictions*. New York: Garland.
- Pierrehumbert, Janet. 1993. Dissimilarity in Arabic verbal roots. in A. Schafer (ed.), *NELS 23: Proceedings of the North East Linguistic Society*, GLSA, Amherst, pp. 367-381.
- Rose, Sharon and Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80: 475—531.
- Straight, H. Stephen. 1976. *The acquisition of Maya phonology: Variation in Yucatec child language*. New York: Garland.
- Tversky, Amos. 1977. Features of similarity. *Psychological Review* 84:327-352.
- Weathers, Nadine. 1947. Tsotsil phonemes with special reference to allophones of b. *International Journal of American Linguistics* 13:108-111.

Massachusetts Institute of Technology  
Department of Linguistics and Philosophy  
77 Massachusetts Ave. 32-D808  
Cambridge, MA 02139

jcoon@mit.edu  
gilliang@mit.edu