Constraint conjunction models super-additivity in weighted probabilistic grammar

Stephanie S Shih
University of California, Merced
shih@ucmerced.edu

[Accepted to *Phonology* (to appear 2017, pending minor revisions)
Please note this version is the submitted draft that does not yet include revisions.]

This paper examines a key difference between constraint conjunction and constraint weight additivity, arguing that the two do not share equivalent empirical coverage. In particular, constraint conjunction in weighted probabilistic grammar allows for *super-additive* constraint interaction, where the effect of violating two constraints goes beyond the additive combination of the two constraints' weights alone (e.g., Albright 2009; Green & Davis 2014). A case study from parasitic tone harmony in Dioula d'Odienné demonstrates super-additive local and long-distance segmental feature similarities that increase the likelihood of tone harmony. The super-additivity in Dioula d'Odienné is formally captured via weighted constraint conjunction in Maximum Entropy Harmonic Grammar (Goldwater & Johnson 2003). Counter to previous approaches that supplant constraint conjunction with weight additivity in Harmonic Grammar (e.g., Potts et al. 2010), information-theoretic model comparison herein reveals that weighted constraint conjunction improves the grammar's explanatory power when modeling quantitative natural language patterns.

Keywords: constraint conjunction, superadditivity, Maximum Entropy, Harmonic Grammar, Dioula d'Odienné

1 INTRODUCTION[1]

Phonological ganging is a type of cumulativity in constraint interaction wherein one strong constraint C_1 can be overtaken by two weaker constraints C_2 and C_3 together but not by C_2 or C_3 independently (e.g., Jäger & Rosenbach 2006). In traditional, strict-ranking Optimality Theory (OT; Prince & Smolensky 2004), phonological ganging is achieved via the (local) conjunction of two constraints—e.g., $C_2\&C_3$—as defined in (1) (Smolensky 1993, 2006; Ito & Mester 1998, 2003; Baković 2000; a.o.; cf. Crowhurst & Hewitt 1997). Conjunctions of constraints usually come with locality domain restrictions (e.g., Lubowicz 2005), and are assumed to be ranked above their simplex counterparts: $C_2\&C_3 \gg C_2$, C_3 (e.g., Baković 2000:27–28ff).

(1) $C_2\&_DC_3$ Assign a violation if both C_2 and C_3 are violated (in local domain D).

The pair of tableaux below illustrates ganging using constraint conjunction. In (2), the winning candidate (a) violates only one of the lower-ranked constraints, C_2 or C_3, and the losing candidate

---

[1] Acknowledgements to Kie Zuraw, Brian Smith, Laura McPherson, Sharon Inkelas, Gunnar Hansson, Robert Daland, Joe Pater, Rachel Walker, Anne-Michelle Tessier, Christopher Green, Eric Baković, Lauren Hall-Lew, Colin Wilson, and audiences at WCCFL 33, GLOW 38, UC Merced, and UC Berkeley for discussion on various portions of this work.

violates the more highly-ranked constraint, C_1. But, if both C_2 and C_3 are violated, as in candidate (3a) in the second tableau, then the highly-ranked conjunction of $C_2\&C_3$ is also violated. In this way, the combined ganging effect of violating both C_2 and C_3 overtakes the C_1.

(2)

	$C_2\&C_3$	C_1	C_2	C_3
☞ a. Winner			1	
b. Loser		1		

(3)

	$C_2\&C_3$	C_1	C_2	C_3
a. Loser	1		1	1
☞ b. Winner		1		

In contrast to strict-ranking OT, weighted constraint approaches in Harmonic Grammar (HG; e.g., Legendre et al. 1990; Smolensky & Legendre 2006) can achieve ganging via cumulative constraint addition, leading many to posit that constraint conjunction may not be a necessary mechanism in HG (e.g., Farris-Trimble 2008; Potts et al. 2010; Pater 2016). The HG tableaux in (4)–(5) illustrate the same ganging effect as shown in (2)–(3). In tableau (4), the winning candidate (a) violates only one of the more lowly weighted constraints, C_2 or C_3, incurring a lower harmony ($\mathcal{H}$) score than the losing candidate (b), which violates a more highly weighted constraint C_1. Ganging occurs, as in (5), when both C_2 and C_3 are violated: the sum of the weighted violations of the simplex constraints C_2 and C_3 is additively greater than a single violation of C_1. So a candidate (5a) that violates both lower-weighted constraints loses to a candidate (b) that only violates the higher-weighted constraint.

(4)

	C_1	C_2	C_3	$\mathcal{H}$
	3	2	2	
☞ a. Winner		1		2
b. Loser	1			3

(5)

	C_1	C_2	C_3	$\mathcal{H}$
	3	2	2	
a. Loser		1	1	4
☞ b. Winner	1			3

This paper argues that additive constraint cumulativity in Harmonic Grammar does not entirely supplant constraint conjunction, and that constraint conjunction is need in HG to capture quantitative phonological patterns. There are at least two major points of difference between constraint conjunction and additive cumulativity that have been previously noted. The first is structural, in that additive cumulativity in HG has a different way of modeling locality conditions on the domains of constraint interaction than constraint conjunction does (Pater 2016; on locality conditions in conjunction, see e.g., Baković 2000; Lubowicz 2005). The second point of difference—and the focus of this paper—is that conjunction in HG has the ability to capture *super-additivity* effects in quantitative phonological patterns (suggested by Albright 2009; also implemented by Green & Davis 2014). Super-additivity occurs when the cumulative effect of two constraints goes above and beyond the summation of the two simplex constraints' weights. Super-additivity can be captured in HG as *weighted constraint conjunction*, in which each conjoined constraint, $C_2\&C_3$, receives an independent weight, above and beyond the additive effects of singular constraints C_2 and C_3. This is illustrated in (6), where the conjunction $C_2\&C_3$ is also assigned a weight independent of the weights of C_2 and C_3. The losing candidate, which violates the conjunction $C_2\&C_3$ and simplex constraints (C_2, C_3), receives a harmony score of 6. Compared to the HG tableau in (5) that does not have weighted constraint conjunction, the winning candidate in (6b) wins by a larger harmony score margin over its competitor ($\Delta\mathcal{H} = 3$) than the winning candidate in (5b) does ($\Delta\mathcal{H} = 1$).

(6)

	$C_2\&C_3$	C_1	C_2	C_3	$\mathcal{H}$
	2	3	2	2	
a. Loser	1		1	1	6
☞ b. Winner		1			3

Being able to capture super-additive effects using weighted constraint conjunction in HG advantageously provides more explanatory power in modeling patterns found in variable natural language data. This paper presents evidence from a case in Dioula d'Odienné tone (henceforth, Dioula; Mande, Côte d'Ivoire) that illustrates such a need for super-additivity. An information-theoretic model selection and comparison methodology is introduced to assess the contribution of weighted constraint conjunction to the grammar, measuring the trade-off between the loss of restrictiveness that comes with adding conjoined constraints to Con and the gain of predictive accuracy over the observed data (see also Wilson & Obdeyn 2009 for a similar approach). The main message here is that we suffer a potentially significant loss of information and explanatory power if grammars are *a priori* restricted from having constraint conjunction. Instead, the necessity and viability of conjunction must be quantitatively assessed against noisy natural language data.

The paper is organized as follows. Section 2 introduces the implementation of weighted constraint conjunction used herein, with toy grammars to demonstrate the differences between Harmonic Grammar approaches with and without conjunction. Section 3 presents data from a tone alternation phenomenon in Dioula d'Odienné nouns that illustrates the need for super-additivity. In §4, grammars for Dioula with and without constraint conjunction are tested and compared using information-theoretic model selection. Section 5 discusses extensions and consequences of weighted constraint conjunction, as it is implemented here: how relaxing the non-negativity requirement in constraint weights enhances the conjunction effect (§5.1), and how the weight of evidence test can also identify cases in which super-additivity is not necessary (§5.2). Section 6 concludes.

2 MODELING CUMULATIVE GANGING

To capture quantitative phonological patterns, Maximum Entropy Harmonic Grammar is used here (henceforth MaxEnt; Goldwater & Johnson 2003; Wilson 2006; Jäger 2007; Hayes & Wilson 2008; a.o.). MaxEnt uses weighted constraints, as with non-stochastic HG; however, instead of winner and loser candidates in OT and non-stochastic HG, a MaxEnt grammar produces a probability distribution over the candidate set based on constraint weights. As such, MaxEnt grammars are able to model both variable and (near-)categorical distributions. From candidates' harmony scores ($\mathcal{H}$), which are based on multiplying constraint violations with respective constraint weights (7), the probability of a given candidate can be calculated over all other possible candidates (8).

(7) $\mathcal{H}(y|x) = \sum_{k=1}^{N} w_k C_i(y,x) = \boldsymbol{w_k} \cdot \boldsymbol{C_i}$,

where y = candidate for input x,
w_k= weight of constraint C_i,
$C_i(y,x)$ = number of violations of C_i that (y,x) incurs, and
N = vector of constraints $(C_{i1} \ldots C_{iN})$.

(8) $P(y|x) = \exp\big(-\mathcal{H}(y|x)\big)/Z(x)$,

where $Z(x) = \sum_{y \in \mathcal{Y}(x)} \exp(-\mathcal{H}(y|x))$, and
$\mathcal{Y}(x)$ = set of candidate output forms given input x.

The MaxEnt results reported herein were fitted using the MaxEnt Grammar Tool (Hayes et al. 2009). MaxEnt grammars are a part of a general class of statistical models, known also as log-linear or multinomial logistic regression models. Some differences between MaxEnt Harmonic Grammar that are imposed by Optimality-theoretic assumptions are discussed in §5.1.

Weighted constraint conjunction for super-additive ganging effects is akin to a standard feature in regression models: interaction terms. In regression modeling, interactions occur when the effect of two constraints on a third are not merely additive—that is, when the combination of two constraints has an effect that is over and above the mere addition of those two constraints on their own. Such interactions are commonplace in linguistic systems and analyses, and have been demonstrated in numerous phonetic, sociolinguistic, psycholinguistic, and corpus studies that computationally model quantitative linguistic patterns. For example, an interaction between topicality and prototypicality can overpower a strong animacy-based preference in English genitive construction choice (Jäger & Rosenbach 2006).

Interactions are implemented in regression models as the product of two constraints, $C_1 {*} C_2$, which can receive its own weight w, independent of the weights assigned to C_1 and C_2 separately, as shown in (9):

(9) $w_1C_1 + w_2C_2 + w_3(C_1 \times C_2) + \cdots + w_kC_{iN} = \boldsymbol{w_k} \cdot \boldsymbol{C_i} = \mathcal{H}(x,i)$

When the possible violations of constraints are limited to $\in \{0, 1\}$, then the multiplicative interaction term—i.e., $C_1 {*} C_2$—is equivalent to previously implemented forms of constraint conjunction (i.e., $C_1 \& C_2$), locality conditions aside. But, when the possible violations are positively unbounded, $\in [0, +\infty)$, then $C_1 {*} C_2$ and $C_1 \& C_2$ are no longer equivalent. In the interaction implementation, $C_1 {*} C_2$, violations are multiplied: if C_1 receives 2 violations and C_2 receives one violation, then $C_1 {*} C_2$ has 2 violations. In contrast, standard assumptions for constraint conjunction, $C_1 \& C_2$, only assign one violation the conjunction, regardless of the number of violations for C_1 and C_2 independently: even if C_1 receives 2 violations and C_2 receives one violation, $C_1 \& C_2$ still only incurs a single violation. To remain consistent with statistical regression analyses, weighted constraint conjunctions are implemented in this paper as the product of two constraints. Exploring the full consequences of this multiplicative implementation of weighted constraint conjunction, however, is left for future work.

The utility of weighted constraint conjunctions to model interactions has already been demonstrated—though sparingly—in the phonological literature couched in MaxEnt and Harmonic Grammar approaches (see Albright 2009 for a first proposal in capturing super-additivity). For example, Hayes et al. (2012) demonstrate the need for interactions between positional and rhythmic constraints in metrical verse. Pater & Moreton (2012) use weighted interactions in Harmonic Grammar for feature co-occurrences. Green & Davis (2014) implement weighted conjunction to restrict complex syllable margin phonotactics.

2.1 Toy grammar illustration

While additive cumulativity of the weights of simplex constraints in HG captures standard ganging phenomena well, super-additive phenomena that exhibit independent effects that are not merely additive reveals the short-coming of additive cumulativity. The toy language in (10) illustrates this case where HG without conjunction makes inaccurate predictions, but where HG with conjunction makes accurate predictions, as shown later in this section. Weights for grammars in this section were obtained using the MaxEnt Grammar Tool (Hayes et al. 2009), unless otherwise specified.

(10)

			freq	*weight*	1.39	1.39	1.39	ℋ	*Pred %*
					C1	C2	C3		
a.	/Input 1/	☞	60	Winner		1		1.39	50
			40	Loser	1			1.39	50
b.	/Input 2/	☞	60	Winner			1	1.39	50
			40	Loser	1			1.39	50
c.	/Input 3/		10	Loser		1	1	2.78	20
		☞	90	Winner	1			1.39	80
d.	/Input 4/	☞	70	Winner				0	80
			30	Loser	1			1.39	20

In this toy language, candidates that violate only C_2 or C_3 are observed to be more frequent than candidates that violate C_1 (10a, b), occurring with a 60 to 40 ratio. A candidate that violates both C_2 and C_3 is much less frequent than a candidate that violates only C_1 (10c), occurring at a 10 to 90 ratio. Finally, a candidate that does not violate any constraint is better than one that violates C_1 (10d), occurring at a 70 to 30 ratio.

Given this input pattern, the constraints are weighted evenly because the grammar is attempting to reconcile to competing forces. First, the singular weights of C_2 and C_3 must be low enough such that candidates that violate C_2 and C_3 individually can win in (10a) and (10b). But at the same time, the sum of C_2 and C_3 must be high enough to overtake C_1 with a large enough margin to achieve the extreme 10-to-90 gang effect in (10c). The weight of C_1 is kept in check by the pair in (10d).[2] Under this schema, regular additive ganging cannot capture the extreme effect shown in (10c), where the candidate violating both C_2 and C_3 loses by a large margin. The predictions of the grammar, given in the rightmost column in (10), are wrong because the grammar does not have enough degrees of freedom to capture the independent slope variances of the gang effect versus the simplex effect.

---

[2] Hat tip to Kie Zuraw, *p.c.*, for this component.

Decreasing the weights of C_2 and C_3 to try to accurately map the surface frequencies for Inputs 1 and 2 in the toy language results in a greater inaccuracy in predicting the extreme gang effect. This adjustment is shown in tableau (11).[3]

(11)

			freq	*weight*	1.39	1	1	ℋ	*Pred %*
					C1	C2	C3		
a.	/Input 1/	☞	60	Winner		1		1	60
			40	Loser	1			1.39	40
b.	/Input 2/	☞	60	Winner			1	1	60
			40	Loser	1			1.39	40
c.	/Input 3/		10	Loser		1	1	2	35
		☞	90	Winner	1			1.39	65
d.	/Input 4/	☞	70	Winner				0	80
			30	Loser	1			1.39	20

The decrease in C_2 and C_3 weights now allows for an accurate prediction in (11a, b), where the winning candidates only violate C_2 or C_3 independently. However, the additive effect of C_2 and C_3 is also now insufficiently large to capture the gang effect in (11c). The grammar relying on standard additive cumulativity mispredicts a ratio of 35-to-65 between the loser and winner, whereas the observed surface pattern is 10-to-90. The grammar does achieve a gang effect, but crucially, the grammar cannot achieve *enough* of a gang effect.

Weighted constraint conjunction allows the grammar to specify a separate, independent weight above and beyond the additive effect of C_2 and C_3. A grammar with weighted constraint conjunction is given in (12).

(12)

			freq	*weight*	2.16	0.84	0.44	0.44	ℋ	*Pred %*
					C2&C3	C1	C2	C3		
a.	/Input 1/	☞	60	Winner			1		0.44	60
			40	Loser		1			0.84	40
b.	/Input 2/	☞	60	Winner				1	0.44	60
			40	Loser		1			0.84	40
c.	/Input 3/		10	Loser	1		1	1	3.04	10
		☞	90	Winner		1			0.84	90
d.	/Input 4/	☞	70	Winner					0	70
			30	Loser		1			0.84	30

The conjoined constraint $C_2\&C_3$ shoulders the burden here of the super-additive gang effect: it adds a weight to the harmony score of the losing candidate in (12c) that goes above and beyond the sum of C_2 and C_3 weights. This allows C_2 and C_3 to be properly weighted below C_1 so that the correct output candidates for (12a, b) can be achieved. Furthermore, because C_1 is now correctly weighted above C_2 and C_3, the correct surface output of (12d) can also be modeled, where a candidate that does not violate any constraint wins over a candidate that violates C_1 with greater likelihood than a candidate that violates C_2 or C_3, cf. (12a, b).

---

[3] The weights of C_2 and C_3 were decreased by hand.

Thus, cases that will necessitate weighted constraint conjunction will be instances of extreme, super-additive gang effects in quantitative data. The following section turns to one such case, from Dioula d'Odienné, which shows cumulative effects that necessitate weighted constraint conjunction.

3 DIOULA D'ODIENNÉ DEFINITENESS TONE ALTERNATION

The Dioula data in the following study comes from the Braconnier & Diaby 1982 lexicon. There are 1,194 nouns.

In Dioula, nouns fall into two classes of tonal behavior for the indefinite-definite alternation.[4] In Type 1 lexical items, there is a tone change in the final vowel of the root from indefinite to definite nouns: e.g., L.L → L.H. In Type 2 lexical items, the tone change from indefinite to definite form involves both the final and penultimate syllables: e.g., L.L → H.H. I follow Braconnier 1982 in assuming that the indefinite is the underlying form from which the definite is derived. The pattern is illustrated with nouns that are underlyingly L toned in the indefinite form, in (13) and (14):

				Indefinite	*Definite*	
(13)	Type 1	L.L → L.H	a.	fòdà	fòd**á**	'season'
			b.	brìsà	brìs**á**	'bush'
			c.	sɛ̀bɛ̀	sɛ̀b**ɛ́**	'paper'
			d.	hàmì	hàm**í**	'concern'
(14)	Type 2	L.L → H.H	a.	kùnà	kúná	'leprosy'
			b.	tùrù	**túrú**	'oil'
			c.	bègì	b**égí**	'white cotton cloth'
			d.	bìlì	**bílí**	'flagstone terrace'
			e.	mèlì	m**élí**	'worm'
			f.	sànã̀	s**ánã́**	'tree'

In the underlyingly L toned nouns (i.e., nouns that end with two /L.L/ syllables), Type 1 items feature a tone change in only the final syllable, from /L.L/ → [L.H]. The domain of tone change in Type 2 nouns includes both the penultimate and final syllables, /L.L/ → [H.H].

In nouns with underlying H tone (15)–(16), the tonal alternation is more complex, but the tone changes occur in the same domains as underlyingly L toned nouns. Underlying H tone nouns are ones that have a H tone in the penultimate or in both the penultimate and final syllables. There are no /L.H/ indefinite forms in Dioula.

				Indefinite	*Definite*	
(15)	Type 1	H.H → H.LH	a.	bɛ́sɛ́	bɛ́s**ɛ̌**	'machete'
			b.	dáfé	dáf**ě**	'horse'
			c.	jámú	jám**ǔ**	'clan name'

---

[4] Braconnier (1982) reports that this tone behavior can also be triggered by a H tone in the following word but provides no indication of the systematicity of that particular environment in conditioning the tone changes across the word boundary. Thus, this paper focuses on the data available for definiteness in the lexicon. This tonally-marked definite versus indefinite alternation is characteristic of several Mende languages.

(16)	Type 2	H.H → L.H	a.	múrú	m**ù**r**ú**	‘knife’
			b.	jégí	j**è**g**í**	‘hope’
			c.	télú	t**è**l**ú**	‘tree’[5]
			d.	nyúmán	ny**ù**m**á**n	‘bond’

Underlyingly H toned, Type 1 items feature a LH contour tone on the final syllable in the definite form: /H.H/ → [H.LH]. Type 2, H-toned nouns feature a change across the final two syllables: /H.H/ → [L.H].

Type 1 items are ones in which only the final syllable is the domain of tone alternation in definite forms, and Type 2 items are ones in which the final two syllables are the domain of tone alternation. Following Braconnier 1982, I assume here that the definite is marked by a H tone morpheme that must occur on at least the final syllable. The pattern is summarised in (17).

(17)		Type 1	Type 2
Underlyingly L-toned	/L.L/	L.**H**	**H.H**
Underlyingly H-toned	/H.H/	H.**LH**	**L.H**

3.1 Predicting Type 1 versus Type 2 items

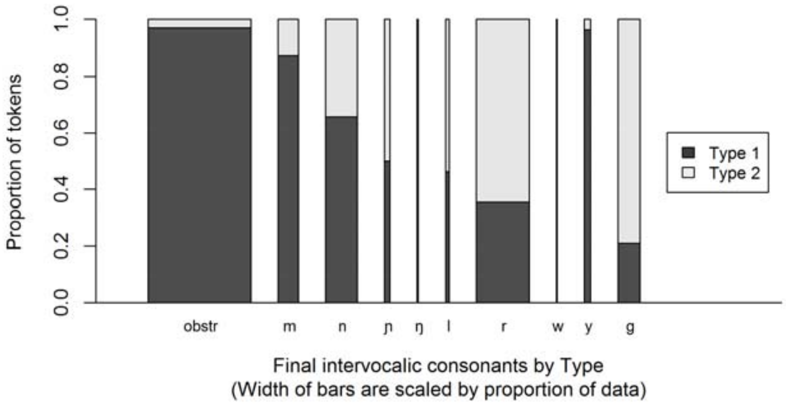
Whether an item is a Type 1 versus Type 2 noun is, to an extent, predictable. In his original description of the Dioula tone system, Braconnier 1982 points to three factors that likely contribute to distinguishing between Type 1 and Type 2: [1] the sonority of the final intervocalic consonant (C_f); [2] the place identity of the final two vowels; and [3] the nasality or orality of the final vowel and final intervocalic consonant. Crucially, here I show quantitatively that the more similar the segments in the word-final VCV# sequence are along these independent planes of similarity, the more likely the noun is Type 2. That is, the more similarity in the final two syllables, the more likely they will be in the same tonal domain (i.e., Type 2).

Nouns with more sonorous final intervocalic consonants are more likely to be Type 2 nouns, with the domain of tone change being the final two syllables: e.g., *mèlì* → *mélí* ‘worm’, (14e). This sonority effect scales quantitatively with increasing sonority (Shih 2013), as illustrated in (18).

---

[5] Braconnier and Diaby (1982: 114) list the definite of this word as variable, either *tèlú* or *télǔ*.

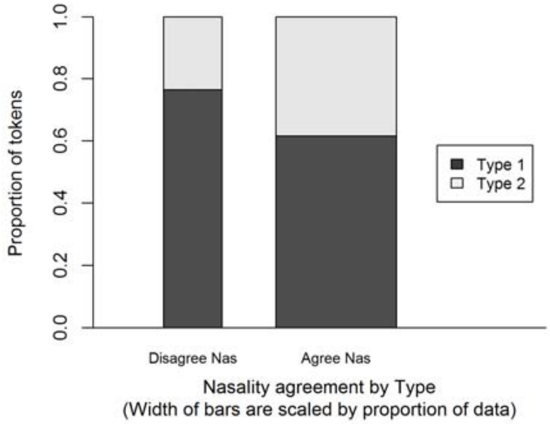
(18)



Thus, as the sonority of the final intervocalic consonant becomes more similarity to the sonority of its surrounding vowels, the tonal domain for the definite alternation is more likely to include the entire VCV# sequence (i.e., final two syllables). This is especially true comparing nasals [m, n, ɲ, ŋ] with liquids [l, r] and [g], which behaves phonotactically like sonorant [ɣ] in Dioula (Braconnier 1983). Braconnier 1983:50 notes that [y] behaves phonotactically like a fricative in the language, which aligns to its behavior here in distinguishing Type 1 versus Type 2 nouns: nouns with [y] in the final intervocalic consonant position are more likely Type 1. There are too few tokens of [w] (*n*=3; Type 1=2 token, Type 2=1 token) to make any definite conclusion of its behavior. Obstruent consonants are the least similar in sonority to the flanking vowels, and so items with obstruent final intervocalic consonants are the least likely to be Type 2: there are only 4 exceptions, out of 294 total obstruent C_f items.

Nasal similarity also contributes to distinguishing Type 1 versus Type 2 items, shown in (19).

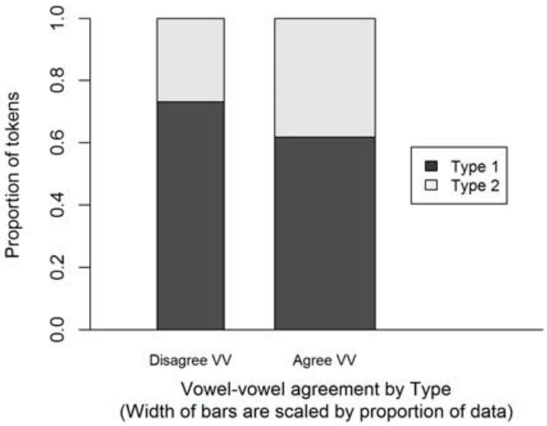
(19)



Nouns in which the final vowel and intervocalic consonant are both nasal or both oral are more likely to be Type 2 than nouns in which the final vowel and intervocalic consonant disagree in nasality or orality: *sànã̀* → *sánã́* ‘tree’, (14f).

The third predictor of Type 1 versus Type 2 items noted by Braconnier 1982 is long distance featural identity between vowels, in (20):

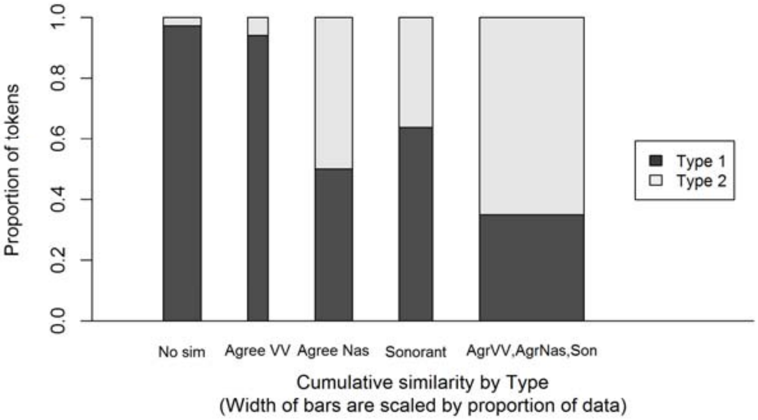
(20)



If the final two vowels are featurally identical (e.g., they agree in height, rounding, and backness), then the likelihood of being Type 2 increases: e.g., *bìlì* → *bílí* ‘flagstone terrace’, (14d).[6]

These similarity preconditions on definite tone behavior “gang up” in a cumulative similarity interaction. The more similarity that is exhibited amongst the members of the word-final VCV# sequence in terms of sonority, nasality, and vowel identity, the more likely the noun will be Type 2, with the tonal domain extending over the final two syllables: e.g., *tùrù* → *túrú* ‘oil’ (14b). This ganging effect is illustrated in (21): the rightmost bar, which has the largest proportion of Type 2 items, represents nouns that agree in vowel identity, have a sonorant intervocalic final consonant, and agree in either nasality or orality. The leftmost bar, in contrast, has the largest proportion of Type 1 items, and represents nouns that share no similarity across the VCV# sequence.[7] The bars in between represent items that only share similarity along a single dimension: from left to right, in VV identity but not in nasality or sonority, in nasality but not in vowel identity or sonority, or in sonority but not in vowel identity or nasality.

(21)



The ganging up effect demonstrated by the Dioula nouns amounts to a super-additive phenomenon, wherein the more similar the vowels and consonant of the final two syllables are, the more likely

---

[6] Dioula vowels: [i, e, ɛ, a, ɔ, o, u] and nasalized counterparts.

[7] The dip in Type 2 representation in the graph for sonorants arises from the effect that nasal sonorants are less likely to be Type 2 than other sonorants. Lumped together, this lowers the proportion of Type 2 items for all sonorants.

they belong in the same tonal domain. The ganging effect is made up of three (somewhat) independent planes of similarity. They are independent because each contributes individual effects on Type 1 versus Type 2 likelihood: for example, as is evident in (21), agreement in nasality/orality is much more likely to predict Type 2 behavior than agreement in vowel featural identity alone. We will see this borne out in the constraint weighting results in §4.

3.2 An OT approach to Dioula tone harmony

The Dioula tone pattern is formalised using Agreement by Correspondence theory (ABC), adapted from the analysis presented in Shih 2013. While ABC was originally developed for long distance consonant harmony phenomena (Walker 2000; Hansson 2001, 2010; Rose & Walker 2004), it has since been extended to other segmental and tonal interactions, including vowel harmony (e.g., Sasa 2009; Walker 2009, 2014; Rhodes 2012), dissimilation (e.g., Bennett 2013, 2014, 2015), local harmonies (e.g., Wayment 2009; Inkelas & Shih 2014; Lionnet 2014), and tone (e.g., Shih & Inkelas 2014; Inkelas & Shih, forthcoming). ABC builds on the core insight that segments that are similar and proximal are more likely to interact (e.g., Kaun 1995; Burzio 2002; Zuraw 2002; Frisch et al. 2004). As such, ABC is particularly well-suited for modeling parasitic patterns such as the similarity-based tone domains in the Dioula definiteness alternation, in which phonological similarity and proximity beget phonological interactions via correspondence domains. Here, I focus the analysis on the similarity bases that determine the tonal domains for the definite alternation. For further details of the Dioula analysis that do not immediately pertain to the similarity issue, see Shih 2013.

In ABC, surface correspondence relationships are determined by the phonological similarity and proximity of segments, encoded in CORRespondence constraints. In essence, segments with similar features project a correspondence set, which is the domain of phonological interaction such as agreement or dissimilation. Under an ABC approach for Dioula, the tone domains that distinguish Type 1 versus Type 2 items are correspondence domains over which tone interactions take place. The set of relevant CORR constraints for Dioula nouns are given in (22)–(23). The constraints here follow pairwise correspondence and identity assessment (Hansson 2007; Rhodes 2012).[8]

(22) a. CORR-X::X {V} Assign a violation for each immediately adjacent pair of segments in the output that are within the {Vowel} range of the sonority scale and do not share a correspondence relationship.

b. CORR-X::X {V,R} Assign a violation for each immediately adjacent pair of segments in the output that are within the {Vowel, Liquid} range of the sonority scale and do not share a correspondence relationship.

c. CORR-X::X {V,R,N} Assign a violation for each immediately adjacent pair of segments in the output that are within the {Vowel, Liquid, Nasal} range of the sonority scale and do not share a correspondence relationship.

---

[8] "::" denotes immediately adjacent segments. The absence of "::" specification in a CORR constraint denotes segments at any distance, e.g., CORR-VV [F]. See Hansson 2001 for proximity scaling in ABC; notation follows e.g., Inkelas & Shih 2014 for local correspondence relationships.

d. CORR-X::X — Assign a violation for each immediately adjacent pair of segments in the output that do not share a correspondence relationship.

(23) CORR-[X::X]$_\sigma$ [±Nas] — Assign a violation for each immediately adjacent pair of segments in the output that are within a syllable, both [αNas], and do not share a correspondence relationship.[9]

(24) CORR-VV [F] — Assign a violation for each adjacent pair of vowels in the output that are identical in feature set [F] and do not share a correspondence relationship.

Correspondence constraints are needed for each independent dimension of similarity involved in determining the tone domain for Type 1 versus Type 2 items. The constraints in (22) mandate that segments that are sufficiently similar in sonority with adjacent segments be in a correspondence relationship; given the scalar nature of sonority, a set of scaled correspondence constraints is used here. The constraint in (23), CORR-[X::X]$_\sigma$ [±Nas], mandates that segments within a syllable that share nasality or orality must be in a correspondence relationship. The constraint in (24), CORR-VV [F], mandates that vowels that are featurally identical be in a correspondence relationship. For simplicity in tracking the individual similarity-based correspondence relationships, I assume here that each correspondence constraint (or set of correspondence constraints, in the case of the scalar sonority constraints) projects its own correspondence set, as demonstrated in (25) (see also Walker 2014 for a similar approach; cf. Bennett 2013 et seq. for single correspondence sets in a single word). Correspondence relationships are indicated by subscripts:

(25)	/t	u	r	u/		
	t	u_x	r_x	u_x	*sonority*	e.g., CORR-X::X {V,R}
	t	u_y	r	u_y	*vowel identity*	CORR-VV [F]
	t	u	r_z	u_z	*nasality*	CORR-[X::X]$_\sigma$ [±Nas]

While correspondence constraints set up similarity-conditioned correspondence relationships, phonological interactions predicated on these correspondence sets are triggered by CORR-LIMITER constraints (Bennett 2013; et seq.). Limiter constraints create unstable surface correspondence relationships, wherein corresponding segments are similar enough to interact but are too uncomfortably similar to stably coexist at a certain distance (see Wayment 2009; Inkelas & Shih 2014 for discussion on instability). Unstable correspondences give rise to repairs by harmony (i.e., attraction of similar segments to be more similar) or dissimilation (e.g., repellence of similar segments to become sufficiently dissimilar).

The Dioula definiteness alternation can be characterised by mandating that corresponding segments agree in tone, using the Limiter constraint in (26).

---

[9] Nasality is formulated here as a bivalent feature so that segments agreeing in *orality* also mandate similarity-induced correspondence.

(26) Ident-XX [tone] Assign a violation for each immediately adjacent pair of segments in a correspondence relationship that do not agree in tone specification.

Under this analysis, Type 2 behavior in underlyingly L toned items means that the H tone of the definite morpheme spreads onto both penultimate and final syllables: e.g., /L.L/ → [$H_x.H_x$]. If the relevant Corr and Limiter constraints trump input-output faithfulness to tone (27), then similar segments will correspond and assimilate in tone, meaning that if the segments are similar enough, the noun will be Type 2.

(27) Ident-IO V[tone] Assign a violation for each vowel that has a different tonal specification in the output than in the input.

Shih 2013 argues that underlyingly H toned items can also follow the same correspondence-based analysis, even though the surface tonal alternations are different. The contour tone formation in Type 1 items and lack of H tone agreement in Type 2 items for underlyingly H toned nouns stems from Antihomophony and OCP avoidances of adjacent, heteromorphemic H tones. For simplicity here, I set aside the surface tonal differences between underlyingly H toned and L toned nouns, and focus on the similarity-based correspondence relationships because these determine the relevant tonal domains in Type 1 versus Type 2 items, which cut across any differences in underlying tones.

Hand-weighted HG tableaux are shown below as illustrations of the ABC system for Dioula tone harmony. In (28), a sonorant final intervocalic consonant (e.g., [l]) facilitates tone agreement between the final vowel and the penultimate vowel by satisfying highly-weighted Corr-X::X {V,R} and the mandated tonal identity between corresponding segments (i.e., Ident-XX [tone]). The relevant Corr constraints that give the necessary precondition for correspondence and tone agreement must outweigh input-output faithfulness in order for a corresponding and tonally-agreeing candidate, (28d), to win.

(28)

	weight	4	4	3	1	ℋ
	mèlí	Corr-X::X {V,R}	Id-XX [tone]	Id-IO V[tone]	Corr-X::X	
	a. mèlí	2			2	10
	b. mè$_x$ĺ$_x$í$_x$		1 (VC)			4
	c. mé$_x$l$_x$í$_x$		2 (VC, CV)	1		11
☞	d. mé$_x$ĺ$_x$í$_x$			1		3
	e. mè$_x$lí$_x$	2	1		2	14
	f. mé$_x$lí$_x$	2		1	2	13

In (29), a final intervocalic consonant that is not sonorant (e.g., [s]) does not facilitate tone harmony because it is insufficiently sonorous to incur violations of Corr-X::X {V,R} when alongside the flanking vocalic segments. The winner when there is an obstruent, then, is a disharmonic candidate with no correspondence (29a).

(29)

weight	4	4	3	1	$\mathcal{H}$
brìsá	Corr-X::X {V,R}	Id-XX [tone]	Id-IO V[tone]	Corr-X::X	
☞ a. brìsá				2	2
b. brì$_x$ś$_x$á$_x$		1 (VC)			4
c. brí$_x$s$_x$á$_x$		2 (VC, CV)	1		11
d. brí$_x$ś$_x$á$_x$			1		3
e. brì$_x$sá$_x$		1		2	6
f. brí$_x$sá$_x$			1	2	5

The tableau in (30) illustrates a ganging effect via cumulative addition of constraint weights. In this case, sonorant consonants and featurally-identical vowels facilitate tone spread due to the additive effect of two highly weighted Corr constraints, Corr-X::X {V,R} and Corr-VV [F].

(30)

weight	4	3	4	3	1	$\mathcal{H}$
tùrú	Corr-X::X {V,R}	Corr-VV [F]	Id-XX [tone]	Id-IO V[tone]	Corr-X::X	
a. tùrú	2	1			2	13
b. tù$_{xy}$ŕ$_x$ú$_{xy}$			2 (VC, VV)			8
c. tú$_{xy}$r$_x$ú$_{xy}$			2 (VC, CV)	1		11
☞ d. tú$_{xy}$ŕ$_x$ú$_{xy}$				1		3
e. tù$_y$rú$_y$	2		1		2	14
f. tú$_y$rú$_y$	2			1	2	13

If the ganging effect shown in (30) is implemented as a weighted constraint conjunction in addition to regular cumulative additivity (31), then there is an increase in the magnitude of harmony score differences ($\Delta\mathcal{H}$) between the winning candidate and the loser candidates. Where the $\Delta\mathcal{H}$ between the harmonic winner and completely non-harmonic, non-corresponding candidate in a merely additive approach (30) is 10, the $\Delta\mathcal{H}$ between in a super-additive approach (31) is 14, denoting a greater predicted likelihood of tone harmony, under comparative grammaticality.

(31)

weight	2	4	3	4	3	1	ℋ
tùrú	CR-X::X {V,R} * CR-VV [F]	CR-X::X {V,R}	CR-VV [F]	ID-XX [tone]	ID-IO V[tone]	CR-X::X	
a. tùrú	2	2	1			2	17
b. tù$_{xy}$ŕ$_{x}$ú$_{xy}$				2 (VC,VV)			8
c. tú$_{xy}$r$_{x}$ú$_{xy}$				2 (VC,CV)	1		11
☞ d. tú$_{xy}$ŕ$_{x}$ú$_{xy}$					1		3
e. tù$_{y}$rú$_{y}$		2		1		2	14
f. tú$_{y}$rú$_{y}$		2			1	2	13

4 MODELING DIOULA SUPER-ADDITIVITY

The question at hand is whether weighted constraint conjunctions are necessary to capturing the probabilistic similarity- and proximity-determined Dioula definite alternation tonal domains (Type 1 = final syllable; Type 2 = final two syllables). Two candidate MaxEnt models are compared here: one without constraint conjunction (–Conj) and one with weighted constraint conjunction (+Conj) of pairwise constraints. The models were simplified by limiting output candidates to a binary choice between a Type 2 candidate that has a two-syllable correspondence domain (i.e., satisfies correspondence and all correspondence limitations) and a Type 1 candidate that has a one-syllable correspondence domain (i.e., does not satisfy correspondence). For example, a Type 2, underlyingly L toned noun input (/L.L/) has two output candidates that are considered: one that maximally satisfies all correspondence constraints and agrees in tone ([H.H]) and one that fails to satisfy correspondence constraints and is thus disharmonic ([L.H]). Because IDENT-XX [tone] will be vacuously satisfied whenever CORR constraints are satisfied under this assumption, IDENT-XX [tone] is not shown in the results below.

Constraint weighting results for both MaxEnt models are provided in (32). Model statistics are given in (33).

(32)

–Conj		**+Conj**	
Constraint	*Weight*	*Constraint*	*Weight*
IDENT-IO V[tone]	3.985	IDENT-IO V[tone]	3.588
CORR-X::X	0.0	CORR-X::X	0.0
CORR-X::X {V,R,N}	1.341	CORR-X::X {V,R,N}	1.266
CORR-X::X {V,R}	0.589	CORR-X::X {V,R}	0.118
CORR-X::X {V}	2.343	CORR-X::X {V}	2.516
CORR-VV [F]	0.187	CORR-VV [F]	0.0
CORR-X::X [±nas]	0.521	CORR-X::X [±nas]	0.0
		CORR-VV [F] * CORR-X::X {V,R,N}	0.0
		CORR-VV [F] * CORR-X::X {V,R}	0.263
		CORR-VV [F] * CORR-X::X [±nas]	0.259
		CORR-X::X [±nas] * CORR-X::X {V,R}	0.466

(33)[10]

	C	D_{xy}		*Likelihood (-2LogLik)*
– Conj	0.8576	0.7152		1085
+ Conj	0.8654	0.7309		1059
			Δ-2LogLik	26 (***, *df*=4, *p* <0.0001)

Testing conjoined constraints reveals that the gang effect of similarity in Dioula is most active for segments that are already highly similar. For example, the conjunction of Corr-VV [F] * Corr-X::X {V,R} receives a weight of 0.263 while the conjunction of Corr-VV [F] * Corr-X::X {V,R,N} receives no weight: this difference indicates that the effect of similarity for Type 2 items is further heightened when the VCV# segments are vowels and liquids ({V,R}), rather than vowels and nasals ({V,R,N}).

The additive effect of conjoined constraints involving liquids is further ganging. Because weighted violations of constraint conjunctions are part of the additive harmony scores, this effect amounts to super-additivity. For example, a non-agreeing vowel-liquid-vowel sequence such as **tùrú* incurs the additive violations of not only the simplex constraints (i.e., Corr-X::X, Corr-X::X {V,R}, Corr-X::X {V,R,N}, Corr-VV [F], Corr-X::X [±nas]) but also the conjoined constraints (i.e., Corr-VV [F]*Corr-X::X {V,R,N}, Corr-VV [F]*Corr-X::X {V,R}, Corr-VV [F]*Corr-X::X [±nas], Corr-X::X [±nas] * Corr-X::X {V,R}). A model with conjunction therefore assigns harmonic *túrú* 71% probability and disharmonic **tùrú* 29% probability.

(34) +Conj (zero-weighted constraints not shown)

weight	0.263	0.466	0.259	3.588	2.516	0.118	1.266	$\mathcal{H}$	*pred %*
/tùrú/	Corr-X::X {V,R} & Corr-VV [F]	Corr-X::X {V,R} & Corr-X::X [±N]	Corr-VV [F] & Corr-X::X [±N]	Id-IO V[tone]	Corr-X::X {V}	Corr-X::X {V,R}	Corr-X::X {V,R,N}		
☞ a. tú$_{xy}$ŕ$_{x}$ú$_{xy}$				1				3.588	71%
b. tùrú	2	2	1			2	2	4.485	29%

In comparison, a model without conjunction only assigns harmonic *túrú* 64% and disharmonic **tùrú* 36% probability, because it lacks the additional violations of weighted constraint conjunctions that would otherwise decrease the harmony score and predicted probability of **tùrú*.

---

[10] *C* is the concordance statistic, which indicates the probability that predicting the outcome is better than chance. *C* values range between 0.5 and 1.0, with good models usually considered ones where $C > 0.8$ (Baayen 2008). D_{xy} is the Somers' Dxy rank correlation, which measures the association between predicted and observed values. D_{xy} measures run between 0 and 1.

(35) –Conj

weight	3.985	2.343	0.589	1.341	0.521	0.187	$\mathcal{H}$	*pred %*
/tùrú/	Id-IO V[tone]	Corr-X::X {V}	Corr-X::X {V,R}	Corr-X::X {V,R,N}	Corr-X::X [±N]	Corr-VV [F]		
☞ a. tú$_{xy}$ŕ$_{x}$ú$_{xy}$	1						3.985	64%
b. tùrú			2	2	1	1	4.568	36%

With the conjoined constraints, the model (e.g., (34)) can capture the super-additive effect that goes above and beyond the addition of the independent constraint weights.

4.1 Comparing weighted conjunction and additive cumulativity

The necessity of weighted constraint conjunction can be tested using model comparison of grammars with and without conjoined constraints. If a gang effect of similarity-driven Dioula tone alternation domain is merely additive, then a grammar with conjoined constraints should not contribute any additional explanatory power. If, however, a gang effect is super-additive, then a grammar with conjoined constraints will demonstrate improved explanatory power over a grammar without conjunction.

Model comparison tests for the necessity of constraints has been heretofore used sparingly in Harmonic Grammar and Optimality-theoretic approaches, largely due to the usual assumption that Con provides a constraint set and the grammar's primary concern is constraint weighting or ranking. Under this mode of operations, the assessment and rejection of the viability of a constraint's existence is largely left to arguments on conceptual or phonological grounds: e.g., Occam's Razor, naturalness. There are notable exceptions in the MaxEnt literature, however, that appeal to model comparison for constraint assessment: see e.g., Wilson & Obdeyn 2009 using a maximum a posterior approach, and Hayes et al. 2012 using model log likelihood comparison.

One quantitative approach to assessing competing grammars—i.e., one which allows conjunction and one which does not—is Akaike Information Criterion (AIC)-based model comparison, which allows for the comparison of significant improvements in capturing information between competing grammars. AIC model comparison is an approach founded on the idea that all models (i.e., grammars) are mere approximations of full reality, an ideal for which the true parameters (*β*) remain unknown (Kullback & Leibler 1951; Burnham & Anderson 2002, 2004; a.o.). The aim in AIC model comparison is to reduce the amount of information loss in a candidate grammar: the less information that a candidate grammar loses, the more weight of evidence there is in favor of that particular grammar.

Information criteria measures come in various forms.[11] I use second-order AIC$_C$, as shown in (36) here, because it penalises for an increasing number of constraints against a sample size, with adjustments for sample size *n* and the number of constraints *K* in the grammar (Burnham & Anderson 2002, et seq.):

(36) $$AIC_c = -2log\left(\mathcal{L}\left(\hat{\beta}\middle|D\right)\right) + 2K + \frac{2K(K+1)}{n-K-1},$$

where $\mathcal{L}\left(\hat{\beta}\middle|D\right)$ = maximum likelihood of observed data D given fitted parameters $\hat{\beta}$,
K = number of estimable parameters (i.e., constraints) in the model, and
n = sample size.

When *n/K* > 40, AIC$_C$ begins to converge with AIC. Because AIC$_C$ regularises for sample size, it is the more conservative measure for model comparison in general (Burnham & Anderson 2004:269–270).

As a rule of thumb in comparing candidate models, a difference of any amount greater than 10 in AIC$_c$ between two candidate models is considered large. Translated into an evidence ratio *E*, as calculated in (37), a 10-point difference between two candidate models is equivalent to about a 150 to 1 odds that the second best model has essentially no evidential support of being as good as the best candidate model (e.g., Anderson 2008:89–90).

(37) $$E_{i,j} = \frac{1}{e^{\left(-(1/2)\Delta_j\right)}},$$

for models i and j,
where $\Delta_j = AIC_{c_j} - AIC_{c_i}$.

AIC$_C$ is based on the standard likelihood ratio test, which maximizes descriptive accuracy given the observed data (see e.g., Hayes et al. 2012 for use in phonology). But unlike a likelihood ratio, AIC$_C$ penalises for a loss of restrictiveness in the grammar that potentially comes with the addition of constraint conjunction.[12,13] A further advantage of AIC$_C$ comparison is that its results for assessing restrictiveness and generalisability have also been shown to asymptotically converge with *k*-fold cross-validation as sample size increases, while remaining computationally faster than *k*-fold cross-validation (Stone 1977; et seq.).

It is crucial to note that the AIC$_C$ number on its own is not a statistical test of significance or of stand-alone goodness-of-fit: AIC$_C$ metrics must be taken as *comparison* statistics between more than one candidate model and between at least two AIC$_C$ measures.

---

[11] Comparisons between available information criteria measures is beyond the scope of this paper; the reader is referred to the rich existing literature on this topic: see e.g., Burnham & Anderson 2004:275ff for detailed discussion of AIC versus the other common IC measure, Bayesian Information Criterion (BIC).

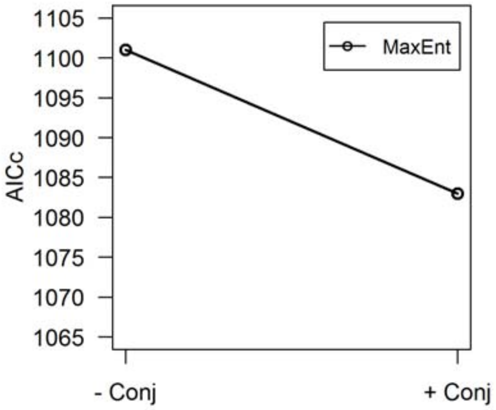
[12] Wilson & Obdeyn 2009 use a maximum a posterior (MAP) approach, which in addition explicitly penalises for extreme values of estimated parameters using an assumed prior.

[13] Also unlike likelihood ratios, AIC comparison can compare the weight of evidence for non-nested models over the same data, making it a more versatile tool in assessing models (for use in phonology considering non-nested models, see e.g., Shih 2014).

4.2 AIC$_C$ Results

Results from AIC$_C$ model comparison are given in (38).

(38)



A model with constraint conjunctions has an AIC$_C$ value of 1101, while a model without conjunction has an AIC$_C$ value of 1083. This 18-point difference between AIC$_C$ scores indicates a rough 8000 to 1 odds (E=8103.08) that there is substantially more support that a grammar with weighted constraint conjunctions better approximates truth in predicting Type 1 versus Type 2 items in Dioula, even after penalising for increased model complexity ($\Delta K = 4$).

5 DISCUSSION

5.1 Conjunction effect in non-bounded Maximum Entropy

One significant point of difference between MaxEnt grammars and their multinomial logistic regression counterpart is that MaxEnt HG restricts constraint weights to non-negative numbers, under the assumption that violations of constraints should only penalise rather than reward (though see e.g., Goldrick & Daland 2009 for a possible application of negative weights in MaxEnt). Regression models as general statistical models do not have this *a priori* restriction on constraint weighting.

Removing the non-negativity restriction allows for the rewarding of the avoidance of correspondence when segments are insufficiently similar (cf. Kimper 2011 for rewards in vowel harmony). This case is simply the flip side to penalising for the lack of correspondence when segments are sufficiently similar, and follows the observation by, e.g., Bennett 2013 that segments can escape correspondence domains in ABC by becoming even more dissimilar (though Bennett does not utilize negative constraint weighting). In this way, negative constraint weights actually help to highlight the similarity bases of ganging effects for tone harmony.

To demonstrate the utility of non-bounded MaxEnt constraint weights and weighted constraint conjunction, binary logistic regression models for Dioula was fitted using `bayesglm()` from the `arm R` package (Gelman et al. 2013).[14] As with the MaxEnt models above, two models were

---

[14] When there are only two output choices considered in a MaxEnt grammar, a non-bounded version of MaxEnt is a binary logistic regression.

fitted for comparison: one with conjunction and one without. For the regression models, the intercept is the equivalent of input-output faithfulness: the CORR constraints must outweigh the intercept and be additively greater than 0 to condition correspondence for Type 2 nouns. The resulting constraint weights, associated errors, and z scores are given in (39) and (40), with model statistics in (41).

(39) –Conj (logistic regression)

Constraint	*Weight*	*Std. Error*	*z value*	*Pr(>|z|)*
(Intercept)	-4.0756	0.3214	-12.68	<0.0001
CORR-X::X {V, R, N}	1.3144	0.1590	8.27	<0.0001
CORR-X::X {V, R}	0.5766	0.0948	6.08	<0.0001
CORR-X::X {V}	2.4274	0.3707	6.55	<0.0001
CORR-VV [F]	0.4502	0.1606	2.80	0.0051
CORR-X::X [±Nas]	0.5097	0.1714	2.97	0.0029

(40) +Conj (logistic regression)

Constraint	*Weight*	*Std. Error*	*z value*	*Pr(>|z|)*
(Intercept)	-3.8181	0.5506	-6.93	<0.0001
CORR-X::X {V, R, N}	1.5524	0.2824	5.50	<0.0001
CORR-X::X {V, R}	0.0501	0.1795	0.28	0.78019
CORR-X::X {V}	2.7530	0.4209	6.54	<0.0001
CORR-VV [F]	0.7478	0.6161	1.21	0.22481
CORR-X::X [±Nas]	-0.5937	0.3347	-1.77	0.07608
CORR-VV [F] * CORR-X::X {V, R, N}	-0.5789	0.3228	-1.79	0.07291
CORR-VV [F] * CORR-X::X {V, R}	0.3258	0.1958	1.66	0.09613
CORR-VV [F] * CORR-X::X [±Nas]	0.5942	0.3531	1.68	0.09243
CORR-X::X [±Nas] * CORR-X::X {V, R}	0.6443	0.1816	3.55	0.00039

(41)

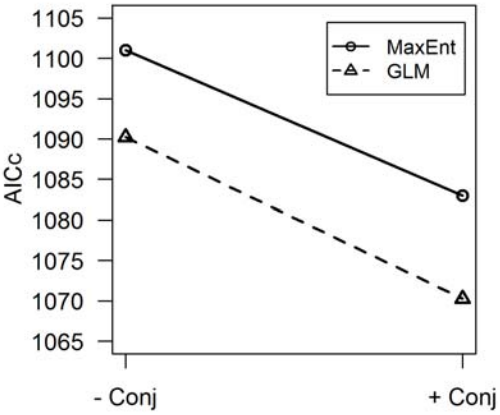
	C	*D_{xy}*	*R^2*		*Likelihood (-2LogLik)*
– Conj	0.8399	0.6797	0.4939		1078.155
+ Conj	0.8479	0.6959	0.5199		1050.074
				Δ-2LogLik	28 (***, *df*=4, *p* <0.0001)

The negative estimated parameter values in the logistic regression model with conjunction demonstrates rewards for avoiding correspondence relationships when segments are insufficiently similar. For example, the negative estimated weight for CORR-X::X [±Nas] (β = –0.5937) rewards the lack of correspondence between nasals, obstruents, and their surrounding vowels, as compared to penalising the lack of correspondence between liquids and their surrounding vowels with the conjunction CORR-X::X [±Nas] * CORR-X::X {V, R} (β = 0.6443). The interaction of CORR-X::X [±Nas] * CORR-X::X {V, R} allows the effect of nasality similarity for nasals and obstruents to be modeled independently of the effect for liquids and vowels.

Similarly, the negative estimated weight for the conjunction of CORR-VV [F] * CORR-X::X {V, R, N} (β =–0.5937) indicates that nasals are less likely to have a similarity ganging effect with the surrounding vowels than liquids, which have a positive weighting of CORR-VV [F] * CORR-X::X {V, R} ($\beta = 0.3258$). This negative weight is a case of constraint *disjunction* (Crowhurst & Hewitt 1997), where correspondence is avoided if not enough similarity in the VCV# sequence exists to precondition a shared tonal domain.

The AIC$_C$ comparison results for the two logistic regression models are provided in (42). The results are similar to the MaxEnt AIC$_C$ comparison, showing that weighted constraint conjunctions allow for better representations of the quantitative pattern in Dioula, even after adjusting for the number of added parameters in the grammar.

(42)



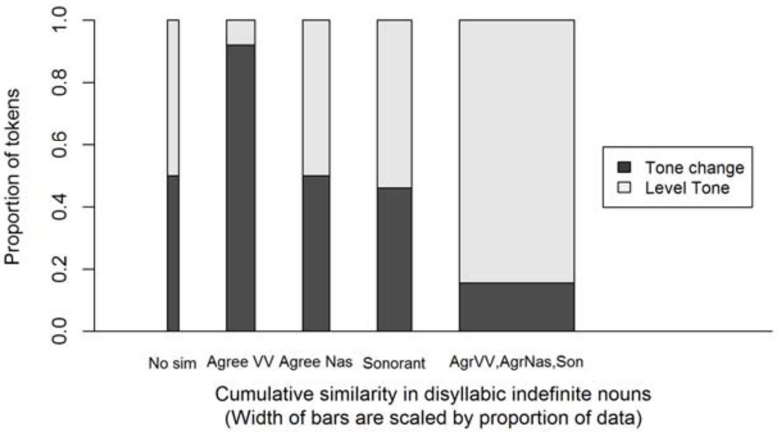
A logistic regression model with conjunction has an AIC$_C$ value of 1090.225, and a model without conjunctions has an AIC$_C$ value of 1070.26. This 20-point difference translates into 22000 to 1 odds (E=22026.47) that a grammar with weighted constraint conjunctions better captures Type 1 and Type 2 distinctions, given the available weight of evidence.

5.2 Weight of evidence for conjunction

It is not the case that AIC$_C$ will in every instance support a more complex model. As a demonstration that AIC$_C$ will support a simpler grammar without constraint conjunction when there is not enough weight of evidence to justify the added parameters of conjoined constraints, I turn here to tone domains within Dioula noun stems as an illustrative case.

Indefinite forms in Dioula exhibit a similar morpheme structure constraint on tone domains to the definite alternation tone domains of Type 2 nouns. Level tone patterns across syllables—i.e., H.H and L.L—tend to correlate with syllables in which the segments are similar in sonority, vowel identity, and nasality/orality. Syllables in which these segments are not sufficiently similar are more likely have a tone change at the syllable boundary: e.g., H.L. The effect is illustrated in (43), using all disyllabic indefinite noun forms in the Dioula data (n=667).

(43)



In (43), the rightmost bar shows nouns that agree in vowel identity, have a sonorant intervocalic final consonant, and agree in either nasality or orality. These disyllabic nouns are the ones that also demonstrate the greatest proportion of level tone surface patterns in the indefinite form, either H.H or L.L, suggesting that segmental similarity correspondence relationships determine tone agreement domains across syllable boundaries. The individual effects on their own in indefinites, however, do not appear to precondition a two-syllable tone domain more than having no similarity in the disyllabic sequence. As (43) shows, the similarity-conditioning of tone domain is largely concentrated in the gang effect of the rightmost bar.

Using the same constraints and methodology introduced above, it is possible to examine whether constraint conjunction is needed here to achieve a super-additive effect, or whether the ganging effect observed in (43) is merely additive. Two models are tested for disyllabic indefinite items. Output candidates for the models were level versus non-level tone patterns: that is, a candidate that corresponds and agrees in tone ($H_x.H_x$) versus a candidate that does not satisfy correspondence ($H_x.L_y$). Because these models are predicting lexical information, no input-output faithfulness (i.e., Ident-IO V[tone]) was included. Resulting constraint weights are given in (44), with model statistics in (45).

(44)

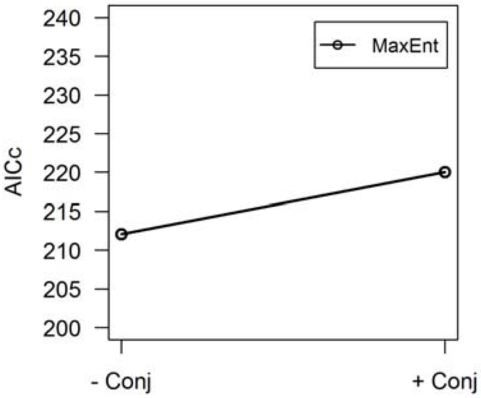
–Conj		**+Conj**	
Constraint	*Weight*	*Constraint*	*Weight*
Corr-X::X	0.481	Corr-X::X	0.524
Corr-X::X {V,R,N}	1.054	Corr-X::X {V,R,N}	1.038
Corr-X::X {V,R}	0.346	Corr-X::X {V,R}	0.225
Corr-X::X {V}	9.536	Corr-X::X {V}	9.597
Corr-VV [F]	0.948	Corr-VV [F]	0.836
Corr-X::X [±nas]	0.857	Corr-X::X [±nas]	0.699
		Corr-VV [F] * Corr-X::X {V,R,N}	0
		Corr-VV [F] * Corr-X::X {V,R}	0.041
		Corr-VV [F] * Corr-X::X [±nas]	0.232
		Corr-X::X [±nas] * Corr-X::X {V,R}	0.204

(45)

	C	D_{xy}		*Likelihood (-2LogLik)*
– Conj	0.983	0.967		195.95
+ Conj	0.984	0.968		195.77
			Δ-2LogLik	0.1837 (*N.S.*,*df*=4, *p* <0.999)

The high concordance statistics indicate that the models are highly accurate. A noticeable difference from the definiteness alternation results is that the weights (i.e., effect sizes) for the conjunctions are lower for this static lexical pattern. This suggests that the ganging effect here is not so extreme as to be super-additive. The AIC$_C$ comparison results, given in (46), clearly corroborate this conclusion:

(46)



The AIC$_C$ comparison demonstrates that a model with conjunction (AIC$_C$ = 220.004) actually receives less support than a model without conjunction (AIC$_C$ = 212.061). The difference in AIC$_C$ is not large, suggesting that the effect is starting to border on super-additive, but for disyllabic indefinite forms in the lexicon, there is insufficient weight of evidence to justify the addition of more complexity in the grammar through constraint conjunction. The support is moderate (*E* = 53.056; i.e., 50 to 1 odds) that conjunctions are not needed to describe the lexical pattern: here, we have a standard case of straightforward additive ganging, that can be sufficiently captured via additive cumulativity in HG without recourse to constraint conjunction.

6 CONCLUSION

This paper has demonstrated a case of super-additivity in quantitative data, where tonal domains are parasitic on the beyond-additive cumulative similarity of host segments. Weighted constraint conjunctions, implemented as interaction terms in Maximum Entropy Harmonic Grammar, capture super-additivity. Constraint conjunction is shown approve the explanatory power of the grammar, even when controlling for the added complexity of conjunction.

Those who have suggested that HG additivity can supplant constraint conjunction often argue that CON should *a priori* not provide constraint conjunctions so as to maintain restrictiveness (i.e., reduce complexity) in the constraint space (e.g., Potts et al. 2010; Jesney 2014). Such an argument based on theoretical parsimony, however, can cut both ways: restrictiveness can also be maintained in the basic theoretical assumptions by allowing for an unrestricted constraint space and by letting the grammar do the choosing of relevant constraints, which is arguably the grammar's task in OT. The information-theoretic model comparison method presented here gets at the best of both worlds, permitting only constraint conjunctions that are shown to improve the model, even after penalising for increased model complexity from added the conjunction parameters. At the very least, it is necessary to entertain the possibility that super-additive effects are lurking in natural language data, and to quantitatively test their viability.

Weighted constraint conjunction furthermore provides a theory of conjunction a fair chance to be evaluated in a probabilistic phonological approach. Previous comparisons of conjunction and Harmonic Grammar have been confounded by comparing only strict ranking Optimality Theory with conjunction, versus Harmonic Grammar without conjunction (e.g., Potts et al. 2010). As seen here, once conjunction is allowed on equal footing in a weighted grammar, it is evident that the function of constraint conjunction is not the same as mere additivity, and that the differences run beyond noted ones of co-reference (e.g., Pater 2016).

It is possible that embedding conjunction into Harmonic Grammar will finally point towards a solution for the long-cited problem that constraint conjunction lacks an associated learnability model. If a learner sees enough weight of evidence that there are cumulative effects from additive constraint interactions, then a separate and independent conjunction can be posited, with the result of reducing the extreme values of simplex constraints in favor of a grammar with justifiably more complex parameters that lead to better accuracy, as measured by hypothesis (i.e., model) comparison. Thus, additive cumulativity in Harmonic Grammar can potentially guide the learning of weighted constraint conjunction for super-additive effects.

REFERENCES

ALBRIGHT, ADAM. 2009. Cumulative violations and complexity threshold: Evidence from Lakhota. Paper presented at the 17th Manchester Phonology Meeting, Manchester, UK.

ANDERSON, DAVID R. 2008. *Model Based Inference in the Life Sciences: A Primer on Evidence*. New York, NY: Springer.

BAAYEN, R. HARALD. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge, UK: Cambridge University Press.

BAKOVIĆ, ERIC. 2000. Harmony, dominance, and control. Rutgers University ph.d. dissertation.

BENNETT, WILLIAM G. 2013. Dissimilation, consonant harmony, and surface correspondence. New Brunswick, NJ: Rutgers, The State University of New Jersey ph.d. dissertation.

BENNETT, WILLIAM G. 2014. Assimilation, dissimilation, and surface correspondence in Sundanese. *Natural Language & Linguistic Theory*.

BENNETT, WM. G. 2015. *The phonology of consonants: Harmony, dissimilation, and correspondence*. Cambridge Studies in Linguistics 147. Cambridge, UK: Cambridge University Press.

BRACONNIER, CASSIAN. 1982. *Le système tonal du dioula d'Odienné*. Abidjan, Ivory Coast: Université d'Abidjan.

BRACONNIER, CASSIAN. 1983. *Phonologie du dioula d'Odienné*. Agence de cooperation culturelle et technique.

BRACONNIER, CASSIAN.; and S DIABY. 1982. *Dioula d'Odienné: Material Lexical*. 96. Institut de Linguistique Appliquee: Agence de cooperation culturelle et technique.

BURNHAM, KENNETH P.; and DAVID R. ANDERSON. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd edition. New York, NY: Springer-Verlag.

BURNHAM, KENNETH P.; and DAVID R. ANDERSON. 2004. Multimodel Inference: Understanding AIC and BIC in Model Selection. \iSociological Methods & Research 33.261–304._

BURZIO, LUIGI. 2002. Many Morphologies, ed. by P. Boucher, 142–177. Somerville, MA: Cascadilla Press.

CROWHURST, MEGAN.; and MARK HEWITT. 1997. *Boolean operations and constraint interaction in Optimality Theory*. Chapel Hill, NC, and Waltham, Mass.

FARRIS-TRIMBLE, ASHLEY W. 2008. Cumulative faithfulness effects in phonology. Indiana University ph.d. dissertation.

FRISCH, STEFAN A.; JANET B. PIERREHUMBERT.; and BROE, MICHAEL B. 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22.179–228.

GELMAN, ANDREW.; YU-SUNG SU.; MASANAO YAJIMA.; JENNIFER HILL.; MARIA GRAZIA PITTAU.; JOUNI KERMAN.; and TIAN ZHENG. 2013. *Package "arm."* http://CRAN.R-project.org/package=arm.

GOLDRICK, MATTHEW.; and ROBERT DALAND. 2009. Linking speech errors and phonological grammars: insights from Harmonic Grammar networks. *Phonology* 26.147–185.

GOLDWATER, SHARON.; and MARK JOHNSON. 2003. Learning OT Constraint Rankings Using a Maximum Entropy Model. *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, ed. by Jennifer Spenader, Anders Eriksson, and Östen Dahl, 111–120. Stockholm: Stockholm University.

GREEN, CHRISTOPHER.; and STUART DAVIS. 2014. Superadditivity and Limitations on Syllable Complexity in Bambara Words. *Perspectives on Phonological Theory and Acquisition: Papers in Honor of Daniel Dinnsen*, ed. by Ashley W. Farris-Trimble and Jessica Barlow. Philadelphia/Amsterdam: John Benjamins Co.

HANSSON, GUNNAR ÓLAFUR. 2001. Theoretical and Typological Issues in Consonant Harmony. University of California, Berkeley ph.d. dissertation.

HANSSON, GUNNAR ÓLAFUR. 2007. Blocking Effects in Agreement by Correspondence. *Linguistic Inquiry* 38.395–409.

HANSSON, GUNNAR ÓLAFUR. 2010. *Consonant Harmony: Long-Distance Interaction in Phonology*. UC Publications in Linguistics. Berkeley, CA: University of California Press.

HAYES, BRUCE.; and COLIN WILSON. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39.379–440.

HAYES, BRUCE.; COLIN WILSON.; and BEN GEORGE. 2009. *Maxent Grammar Tool.* http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/.

HAYES, BRUCE.; COLIN WILSON.; and ANNE SHISKO. 2012. Maxent grammars for the metrics of Shakespeare and Milton. *Language* 88.691–731.

INKELAS, SHARON.; and STEPHANIE S SHIH. forthcoming. Tone melodies in the age of surface correspondence. *Proceedings of the 51st Annual Meeting of the Chicago Linguistic Society*. University of Chicago.

INKELAS, SHARON.; and STEPHANIE S SHIH. 2014. Unstable surface correspondence as the source of local conspiracies. *NELS 44: Proceedings of the 44th Meeting of the North East Linguistic Society*, ed. by Jyoti Iyer and Leland Kusmer, 1:191–204. University of Massachusetts, Amherst: GLSA Publications.

ITO, JUNKO.; and ARMIN MESTER. 1998. *Markedness and word structure: OCP effects in Japanese*.

ITO, JUNKO.; and ARMIN MESTER. 2003. *Japanese Morphophonemics: Markedness and Word Structure*. Cambridge, Massachusetts: MIT Press Linguistic Inquiry Monograph Series 41.

JÄGER, GERHARD. 2007. Maximum Entropy Models and Stochastic Optimality Theory. *Architectures, Rules, and Preferences: A Festschrift for Joan Bresnan*, ed. by Jane Grimshaw, Joan Maling, Chris Manning, Jane Simpson, and Annie Zaenen. Stanford, CA: CSLI.

JÄGER, GERHARD.; and ANETTE ROSENBACH. 2006. The winner takes it all—almost. *Linguistics* 44.937–971.

JESNEY, KAREN. 2014. Complex process interaction: weighted constraints vs. ranked constraints and rules. Paper. Paper presented at the Interaction of Grammatical Building Blocks (IGRA): Workshop on Building Blocks, Universität Leipzig. http://www-bcf.usc.edu/~jesney/Jesney2014IGRA_handout.pdf.

KAUN, ABIGAIL R. 1995. The Typology of Rounding Harmony: An Optimality-theoretic Approach. Los Angeles, CA: University of California, Los Angeles ph.d. dissertation.

KIMPER, WENDELL A. 2011. Competing triggers: Transparency and opacity in vowel harmony. Amherst, MA: University of Massachusetts, Amherst ph.d. dissertation.

KULLBACK, SOLOMAN.; and RICHARD A. LEIBLER. 1951. On Information and Sufficiency. *Annals of Mathematical Studies* 2.79–86.

LEGENDRE, GERALDINE.; YOSHIRO MIYATA.; and PAUL SMOLENSKY. 1990. Harmonic Grammar – A formal multi-level connectionist theory of linguistic well-formedness: An Application. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, 884–891. Mahwah, NJ: Lawrence Erlbaum Associates.

LIONNET, FLORIAN. 2014. Doubly Triggered Harmony in Laal as Subphonemic Agreement by Correspondence. *Supplemental Proceedings of the 2013 Meeting on Phonology*, ed. by John Kingston, Claire Moore-Cantwell, Joe Pater, and Robert Staubs. Washington, DC: Linguistic Society of America. http://journals.linguisticsociety.org/proceedings/index.php/amphonology/issue/view/1.

LUBOWICZ, ANNA. 2005. Locality of Conjunction. *Proceedings of WCCFL 24*. Somerville, MA: Cascadilla. http://www.lingref.com/cpp/wccfl/24/paper1230.pdf.

PATER, JOE. 2016. Universal Grammar with Weighted Constraints. *Harmonic Grammar and Harmonic Serialism*, ed. by John J McCarthy and Joe Pater. London: Equinox Press.

PATER, JOE.; and ELLIOT MORETON. 2012. Structurally biased phonology: Complexity in learning and typology. *The English and Foreign Languages Journal* 3.1–44.

POTTS, CHRISTOPHER.; JOE PATER.; KAREN JESNEY.; RAJESH BHATT.; and MICHAEL BECKER. 2010. Harmonic Grammar with linear programming: from linear systems to linguistic typology. *Phonology* 27.77–117._

PRINCE, ALAN.; and PAUL SMOLENSKY. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden, Mass., and Oxford, UK: Blackwell.

RHODES, RUSSELL. 2012. *Vowel Harmony as Agreement by Correspondence*. UC Berkeley Phonology Lab Annual Report. University of California, Berkeley. http://linguistics.berkeley.edu/phonlab/annual_report/annual_report_2012.html.

ROSE, SHARON.; and RACHEL WALKER. 2004. A Typology of Consonant Agreement by Correspondence. *Language* 80.475–531.

SASA, TOMOMASA. 2009. Treatment of vowel harmony in Optimality Theory. Iowa City, IA: University of Iowa ph.d. dissertation.

SHIH, STEPHANIE S. 2013/Ms. Consonant-tone interaction as Agreement by Correspondence. Stanford University and University of California, Berkeley. http://stanford.edu/~stephsus/shih-ctoneABC-draftms_1-18-13.pdf.

SHIH, STEPHANIE S. 2014. Towards Optimal Rhythm. Stanford, CA: Stanford University phd dissertation.

SHIH, STEPHANIE S.; and SHARON INKELAS. 2014. A subsegmental correspondence approach to contour tone (dis)harmony patterns. *Proceedings of Phonology 2013*. University of Massachusetts, Amherst. http://journals.linguisticsociety.org/proceedings/index.php/amphonology/article/view/22.

SMOLENSKY, PAUL. 1993. Harmony, markedness, and phonological activity. New Brunswick, NJ.

SMOLENSKY, PAUL. 2006. Optimality in Phonology II: Harmonic Completeness, Local Constraint Conjunction, and Feature Domain Markedness. *The Harmonic Mind*, 27–160. MIT Press.

SMOLENSKY, PAUL.; and GÉRALDINE LEGENDRE. 2006. *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*. Cambridge, MA: MIT Press.

STONE, M. 1977. An Aymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion. *Journal of the Royal Statistical Society. Series B (Methodological)* 39.44–47.

WALKER, RACHEL. 2000. Long-distance Consonantal Identity Effects. *Proceedings of WCCFL 19*, ed. by Roger Billerey and Brook Danielle Lillehaugen, 532–545. Somerville, MA: Cascadilla Press.

WALKER, RACHEL. 2009. Similarity-sensitive blocking and transparency in Menominee. Paper. Paper presented at the 83rd Annual Meeting of the Linguistic Society of America, San Francisco, California. http://www-bcf.usc.edu/~rwalker/MenomineeLSAHdt.pdf.

WALKER, RACHEL. 2014. Prominence-control and Multiple Triggers in Vowel Harmony: An ABC Analysis. *ABCConference Archive*, UC Berkeley Phonology Lab Annual Report:202–213. University of California, Berkeley. http://linguistics.berkeley.edu/phonlab/annual_report/documents/2014/ABCC/
Walker.pdf.

WAYMENT, ADAM. 2009. Assimilation as attraction: Computing distance, similarity, and locality in phonology. Baltimore, MD: Johns Hopkins University ph.d. dissertation.

WILSON, COLIN. 2006. Learning Phonology with Substantive Bias: An Experimental and Computational Study of Velar Palatalization. *Cognitive Science* 30.945–982.

WILSON, COLIN.; and MARIEKE OBDEYN. 2009. Simplifying subsidiary theory: statistical evidence from Arabic, Muna, Shona, and Wargamay.

ZURAW, KIE. 2002. Aggressive reduplication. *Phonology* 19.395–439.