Super Additive Similarity in Dioula Tone Harmony

Stephanie S Shih

1. Introduction[*]

Phonological ganging is a type of cumulativity in constraint interaction where 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 and Rosenbach 2006). In strict-ranking Optimality Theory (OT), phonological ganging is achieved via (local) constraint conjunction (e.g., Smolensky 1993, 2006; Baković 2000; Ito and Mester 2003; cf. Crowhurst and Hewitt 1997). The tableau in (1) illustrates the ganging case in which candidate (a) loses because it violates both C_2 and C_3, thus incurring a violation of the highly-ranked conjunction, $C_2 \& C_3$.

(1)

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

(2)

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

In contrast, recent developments in weighted constraint Harmonic Grammar (HG; e.g., Legendre et al. 1990; Smolensky and Legendre 2006) have argued that phonological ganging can instead be achieved via cumulative constraint addition, without recourse to constraint conjunction (e.g., Farris-Trimble 2008; Potts et al. 2010; Pater 2015). The HG tableau in (2) demonstrates the same ganging effect as shown in (1), but instead captured using the summed weighted violations of simplex constraints, C_2 and C_3, which are additively greater than a single violation of C_1.

This paper argues that the mechanisms of constraint conjunction and additive constraint cumulativity are not mutually exclusive. In fact, to fully capture phonological ganging effects in natural language, cumulativity from additive constraint weights and cumulativity from constraint conjunction co-exist in the weighted constraint world, in *super-cumulativity*.[1] Super-cumulativity is implemented here as *weighted constraint conjunction*, in which each conjoined constraint (e.g., $C_2 \& C_3$ in the following tableau) also receives a weight, above and beyond the cumulative additive effects of singular constraints C_2 and C_3:

(3)

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

This paper presents evidence from a tone harmony phenomenon in Dioula d'Odienné (Mande, Côte d'Ivoire) (henceforth, Dioula) that illustrates the need for super-cumulativity. Information-theoretic model selection and comparison methodology is used to assess the contribution of weighted constraint conjunction to the grammar: such an approach maximizes predictive accuracy while penalising the loss of restrictiveness that comes with the addition of conjoined constraints in CON (see also Wilson and Obdeyn 2009 for a similar approach). The main message here is that there is potentially a

---

[1] also termed 'super-additivity': see e.g., Albright 2009; Green and Davis 2014. I use 'super-cumulativity' here due to the implementation of weighted constraint conjunction, which is not purely additive (see §4.1).

significant loss of information and explanatory power if grammars are *a priori* restricted from constraint conjunction. Instead, thorough, quantitative assessments of the viability of conjunction and cumulative effects must be tested against noisy natural language data.

2. Data

In Dioula, nouns fall into two classes of tonal behavior.[2] In TYPE 1 lexical items (4), a H tone marking definiteness appears on the final vowel of the root.[3] In TYPE 2 lexical items (5), the definite H tone triggers regressive H tone harmony on the final and penultimate syllables.[4]

		indefinite	*definite*			
(4)	a.	fòdà	fòd**á**	'season'	TYPE 1	L.L $\rightarrow$ L.H
	b.	brìsà	brìs**á**	'bush'		
	c.	sὲbὲ	sὲb**έ**	'paper'		
	d.	hàmì	hàm**í**	'concern'		
(5)	a.	kùnà	k**ú**n**á**	'leprosy'	TYPE 2	L.L $\rightarrow$ H.H
	b.	tùrù	t**ú**r**ú**	'oil'		
	c.	bὲgì	b**é**g**í**	'white cotton cloth'		
	d.	bìlì	b**í**l**í**	'flagstone terrace'		
	e.	mèlì	m**é**l**í**	'worm'		
	f.	sànằ	s**á**n**ắ**	'tree'		

The distinction between TYPE 1 and TYPE 2 nouns is predictable on at least three factors, initially observed in Braconnier 1982 and quantitatively shown in Shih 2013: [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, the more similar the segments in the word-final VCV# sequence are, the more likely tone agreement will occur.

Nouns with more sonorous final intervocalic consonants are more likely to be TYPE 2 nouns, which exhibit tone harmony: e.g., *mèlì → mélí* 'worm' (5e). This sonority effect scales quantitatively with increasing sonority (Shih 2013), as illustrated in (6):[5,6] more sonorous consonants are more likely to allow tone assimilation. Nasal similarity also facilitates tone agreement: 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 (e.g., *sànằ → sánắ* 'tree', (5f). Finally, long distance featural identity between vowels also increases the likelihood of tone agreement. A greater proportion of nouns with identical vowels are TYPE 2 (e.g., *bìlì → bílí* 'flagstone terrace', (5d), in comparison with nouns with non-identical vowels.

These similarity preconditions on tone agreement in TYPE 2 nouns "stack" in a cumulative similarity interaction. The more similarity that is exhibited amongst 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 tone agreement between on the final two syllables (e.g., *tùrù → túrú* 'oil', (5b)). This ganging effect is illustrated in (7): the leftmost bar, which has the largest proportion of TYPE 2 items, represents nouns that agree in vowel identity, have a sonorous intervocalic final consonant, and agree in either nasality or orality; the rightmost bar, in contrast, has the largest proportion of TYPE 1 items, and represents nouns with no similarity across the VCV# sequence.

---

[2] All Dioula data is taken from the Braconnier and Diaby 1982 lexicon; n = 1194 nouns.

[3] Braconnier also reports that this pattern can be triggered by a H tone in the following word but provides no indication of the systematicity of that particular environment in triggering tone changes. Thus, this paper focuses on the case triggered by the definite H tone morpheme because evidence is more readily available in the lexicon. This tonally-marked definite versus indefinite alternation is characteristic of several Mande languages.

[4] In underlyingly high-toned roots, the tonal alternation is more complex: tone changes involve contour formation or LH spread. The basic pattern of tone change boundedness on the final two syllables versus the final syllable remains the same and can be modeled, with the addition of ANTI-HOMOPHONY, under the same ABC analysis presented here (for summaries, see Braconnier 1982; Shih 2013).

[5] Note: [g] ~ [ɣ] behaves phonotactically like a sonorant in Dioula (Braconnier 1983).

[6] Width of bars in the following figures (6)–(7) are scaled for proportion of data.

(6)   Final intervocalic consonants by noun TYPE
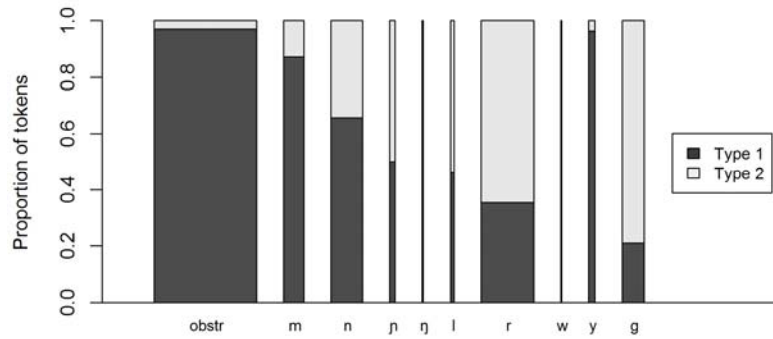
(7)   Cumulative similarity and agreement by noun TYPE
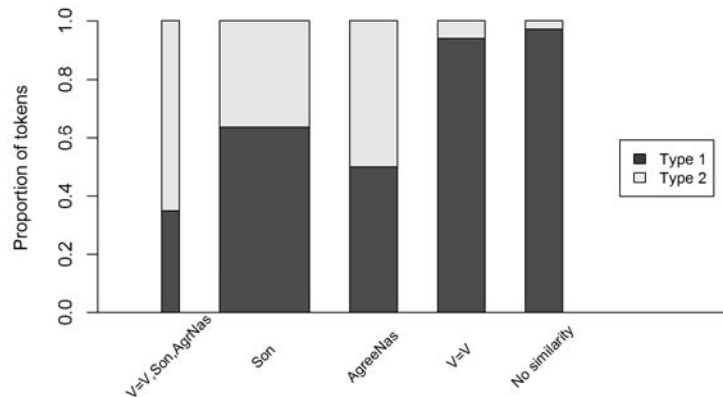
3. ABC and Dioula tone harmony

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 and Walker 2004), it has since been extended to other segmental and tonal interactions (e.g., Sasa 2009; Rhodes 2012; Bennett 2013; Shih 2013; Inkelas and Shih 2014; Lionnet 2014; Shih and Inkelas 2015). ABC builds on the core insight that segments that are similar and proximal are more likely to interact (e.g., Kaun 1995; Zuraw 2002; Frisch et al. 2004). As such, ABC is particularly well-suited for modeling parasitic patterns such as Dioula tone agreement, in which phonological similarity and proximity beget even more similarity.

In ABC, surface correspondence relationships are determined by the phonological similarity and proximity of segments, encoded in CORR constraints. The set of relevant CORR constraints for Dioula tone harmony are given in (8)–(10)[7].

(8)   CORR-X::X {V}   *Segments with highest amount of sonority (i.e., vowels) correspond.*
       CORR-X::X {V,R}   *Vowels, liquids correspond.*
       CORR-X::X {V,R,N}   *Sonorants (vowels, liquids, nasals) correspond.*
       CORR-X::X   *All segments correspond.*
(9)   CORR-VV [F]   *Vowels identical in feature set [F] correspond.*
(10)   CORR-X::X [±nas]   *Segments identical in nasality specification correspond.*[8]

---

[7] "::" denotes immediately adjacent segments. The absence of "::" specification in a CORR constraint denotes segments at any distance, e.g., CORR-VV [F] (9). See Hansson 2001 for proximity scaling in ABC; notation follows e.g., Inkelas and Shih 2014.

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

Unstable surface correspondences occur when corresponding segments are similar enough to interact but are too uncomfortably similar to stably coexist at a certain distance, as mandated by CORR-LIMITER[9] constraints (e.g., IDENT-XX, XX-EDGE) (see Wayment 2009; Inkelas and Shih 2014 for discussion on instability; see Bennett 2013, et seq. for LIMITER constraints). The relevant LIMITER constraint for Dioula tone harmony is given in (11).

(11)   IDENT-XX [tone]   *Corresponding segments must agree in tone specification.*

Attraction (i.e., harmony) and repellence (i.e., dissimilation) are both repairs for unstable similarity- and proximity-driven surface correspondences. The relevant CORR and LIMITER constraints must trump input-output faithfulness (12) for tone harmony as in Dioula.

(12)   IDENT-IO V[tone]   *Maintain input identity of vowel tone specification in the output.*

Hand-weighted HG tableaux are shown here as examples of the ABC system for Dioula tone harmony. In (13), a sonorant final intervocalic consonant (e.g., [l]) facilitates regressive tone agreement between the final vowel and the penultimate vowel by satisfying highly-weighted CORR-X::X {V,R} and mandated tonal identity between corresponding segments (i.e., IDENT-XX [tone]).

(13)

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

In (14), a final intervocalic consonant that is not sonorant (e.g., [s]) does not facilitate tone harmony because it is not sonorant enough to incur violations of CORR-X::X {V,R} alongside the flanking vocalic segments.

(14)

		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$_x$s$_x$á$_x$		-1 (VC)			-4
	c.	brí$_x$s$_x$á$_x$		-2 (VC, CV)	-1		-11
	d.	brí$_x$s$_x$á$_x$			-1		-3
	e.	brì$_x$sá$_x$		-1		-2	-6
	f.	brí$_x$sá$_x$			-1	-2	-5

A ganging effect is illustrated in (15) via cumulative additivity. 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].

---

(15)	*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 (15) is implemented as a weighted constraint conjunction in addition to regular cumulative additivity (16), then we see 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 (15) is 10, the $\Delta\mathcal{H}$ in a super-cumulative approach (16) is 14, denoting a greater likelihood of tone harmony that is predicted, under comparative grammaticality.

(16)	*weight*	2	4	3	4	3	1	$\mathcal{H}$
	tùrú	Cr-X::X {V,R} & Cr-VV	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 details

Maximum Entropy Harmonic Grammar (henceforth MaxEnt; Goldwater and Johnson 2003; Wilson 2006; Jäger 2007; Hayes and Wilson 2008; a.o.) is used here to capture the quantitative Dioula tone pattern. MaxEnt uses weighted constraints, as with non-stochastic Harmonic Grammar, but instead of winner and loser candidates, a MaxEnt grammar produces a probability distribution over the candidate constraint based on the weights, and is thus able to model both variable and (near-)categorical distributions. From candidates' harmony scores ($\mathcal{H}$), which are based on violations and constraint weights (17), the probability of a given candidate can be calculated over all other possible candidates (18).

(17) $\mathcal{H}(y|x) = \sum_{k=1}^{N} w_k C_i(y,x) = \boldsymbol{w_k} \cdot \boldsymbol{C_i}$,
  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 ($C_{i1} \dots C_{iN}$).

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

MaxEnt grammars are a general class of statistical models, also known as log-linear or multinomial logistic regression models, with some minor differences. When there are only two output candidates in a MaxEnt grammar, then MaxEnt is equivalent to binary logistic regression. MaxEnt grammars were fitted using the MaxEnt Grammar Tool (Hayes et al. 2009). For reasons of space, differences between MaxEnt and regression are not discussed here (see e.g., Shih 2015).

4.1 Weighted constraint conjunction as interaction

In regression modeling, interactions occur when the effect of two constraints on a third are not merely additive. Such interactions are quite commonplace in linguistic systems and analyses, particularly in sociolinguistics and psycholinguistic studies (e.g., an interaction between topicality and prototypicality overpowers a strong animacy-based preference in English genitive construction choice; Jäger and Rosenbach 2006). Interactions are implemented in regression modeling as the product of two constraints, C_1*C_2, which can receive its own weight w.

In Optimality-theoretic terms, an interaction is, in essence, a weighted constraint conjunction. When 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., $C_1\&C_2$). When possible violations are positively unbounded, $\in[0, +\infty)$, then C_1*C_2 and $C_1\&C_2$ are no longer equivalent. To remain consistent with statistical regression analyses, weighted constraint conjunctions are implemented in this paper as the product of two constraints; it is left for future work to fully explore the consequences of a multiplicative implementation of weighted constraint conjunction. Further not discussed in this paper is the issue of domain locality in constraint conjunction and how this might (or might not) be restricted in an interaction-based implementation (see e.g., Lubowicz 2005 for discussion).

Though not commonly recognized, there are existing examples of interactions (i.e., weighted constraint conjunctions) in the phonological literature that utilizes MaxEnt and Harmonic Grammar approaches. Hayes et al. (2012) demonstrate the need for interactions between positional and rhythmic constraints in metrical verse. Pater and Moreton (2012) use weighted interactions in HG for feature cooccurrences. Green and Davis (2014) implement weighted conjunction to restrict complex syllable margin phonotactics.

4.2 Model selection and comparison

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

Model comparison has heretofore been underutilized in Harmonic Grammar and Optimality-theoretic approaches (some notable exceptions being Wilson and Obdeyn 2009; Hayes et al. 2012) largely because the usual assumption is that learners have a constraint set already provided by CON, and the question is how to fit constraint weights. Thus, under this mode of operations, the assessment and rejection of the viability of a constraint's existence is largely left to arguments on conceptual (e.g., Occam's Razor) or phonological grounds (e.g., learnability, naturalness). When there is discussion of inductive constraint learning, however, the induction of conjoined forms of markedness constraints is usually allowed (e.g., Hayes and Wilson 2008; Moreton 2010).

When there are competing conceptions of constraint sets provided by CON, then a quantitative way to assess the competing grammars (e.g., whether or not CON contributes conjoined constraints) is necessary. Here, Akaike Information Criterion (AIC) model comparison is used to compare significant improvements between candidate 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 and Leibler 1951; Burnham and 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.[10] For the purposes of this paper, second-order AIC_c, as shown in (19), is used because it penalises for an increasing number of constraints against a sample size.[11]

---

[10] 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 and Anderson 2004:275ff for detailed discussion of AIC versus Bayesian Information Criterion (BIC).

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

where $\mathcal{L}(\hat{\beta}|D)$ = 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.

As a rule of thumb in comparing candidate models, a difference of ≥10 in AIC$_c$ between two candidate models considered large. Translated into an evidence ratio (E; shown in (20), such a 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).

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

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 the sample size increase, while remaining computationally faster than k-fold cross-validation (Stone 1977; et seq.). It is crucial to note that AIC$_c$ is not a statistical test of significance or a stand-alone goodness-of-fit: AIC$_c$ metrics must be taken as *comparison* statistics between more than one candidate model.

Two candidate MaxEnt models were compared: one without constraint conjunction (–Conj) and one with weighted constraint conjunction (+Conj). The models were simplified to having binary outcomes between a candidate with tone agreement via satisfying CORR versus a candidate with no tone agreement via violating CORR conditions.

5. Results

Results from AIC$_c$ model comparisons are given in (21). The comparisons reveal substantial support for the hypothesis that a grammar with weighted constraint conjunction better approximates truth in the indefinite~definite Dioula tone harmony alternation, even after penalising for increased model complexity in the AIC$_c$ calculation.

(21)

– Conj	1101
+ Conj	1083
ΔAIC$_c$	**18**
E	**8103.08**

Constraint weighting results for both MaxEnt models are provided in (22). Testing conjoined constraints reveals that the gang effect of similarity in Dioula is most active for segments that are already highly similar: for example, between liquids and vowels, as evidenced by the weighted conjunction of CORR-VV [F]*CORR-X::X {V,R}, versus the conjunction of CORR-VV [F]*CORR-X::X {V,R,N}, which receives no weight.

---

[11] When $n/K > 40$, AIC and AIC$_c$ begin to converge. Because AIC$_c$ regularizes for sample size, it is the more conservative measure for model comparison in general (Burnham and Anderson 2004:269–270).

[12] Wilson & Obdeyn's maximum a posterior (MAP) approach also explicitly penalizes 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.

(22)

–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

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-cumulativity. 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}). Thus, a model with conjunction assigns correctly harmonic *túrú* 71.04% probability and disharmonic *tùrú* 28.96%. In comparison, a model without conjunction only assigns harmonic *túrú* 64.18% and disharmonic *tùrú* 35.82% probability, because it lacks the additional violations of weighted constraint conjunctions that would otherwise decrease the harmony score and predicted probability of *tùrú*.

6. Discussion and conclusion

This paper has demonstrated a case of super-cumulativity in lexical data, where tone harmony is parasitic on the beyond-additive cumulative similarity of host segments. Weighted constraint conjunction, implemented as an interaction term in Maximum Entropy Harmonic Grammar, captures super-cumulativity. Constraint conjunction is shown to improve the explanatory power of the grammar, even when controlling for the added complexity of conjunction.

Opponents of 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). But, such an argument of theoretical parsimony can cut both ways: restrictiveness can be maintained in the basic theoretical assumptions by allowing for an unrestricted constraint space and letting the grammar do the choosing of relevant constraints, which is arguably the grammar's job. 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. At the very least, it is necessary to entertain the possibility that super-cumulative effects are lurking in natural language data, and to quantitatively test their viability.

Weighted constraint conjunction furthermore provides the 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 function of constraint conjunction is not the same as mere additivity.

It is possible that embedding conjunction into Harmonic Grammar will finally point toward 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 and better accuracy, as measured by hypothesis (i.e., model) comparison. Thus, additive cumulativity in HG can potentially guide learning of weighted constraint conjunction for super-cumulative 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.

BAKOVIĆ, ERIC. 2000. Harmony, dominance, and control. New Brunswick, NJ: 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.

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. *Sociological Methods & Research* 33.261–304.

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.

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. 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. 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. 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 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.

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.

MORETON, ELLIOTT. 2010. Constraint induction and simplicity bias in phonotactic learning. Paper presented at the Workshop on Grammar Induction, Cornell University. http://www.unc.edu/~moreton/Papers/MoretonCornellWkshp2010.pdf.

PATER, JOE. 2015. 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.

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

SHIH, STEPHANIE S. 2015. Super additive phonological similarity as constraint conjunction. Paper presented at the 38th Generative Linguisticcs in the Old World (GLOW), University of Paris Diderot.

SHIH, STEPHANIE S; and SHARON INKELAS. 2015. Autosegmental Aims in Surface Optimizing Phonology. University of California. Merced and University of California, Berkeley, Ms. lingbuzz/002520.

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.

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. Ms.

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