A method to evaluate systemic constraints in probabilistic grammars*

Benjamin Storme (Leiden University)

To appear in *LI Squibs*

Abstract

Systemic constraints evaluating several input-output mappings jointly have been proposed in phonology and morphology to account for dispersion effects (e.g., contrast enhancement) and paradigmatic effects (e.g., homophony avoidance). However these constraints are not directly compatible with current probabilistic models for constraint-based grammars, as these models typically assume constraints that evaluate input-output mappings *individually*. As a consequence, analyses involving systemic constraints cannot currently be fit to noisy data (e.g., corpus frequencies, experimental data). This squib proposes a simple method to address this issue. The method consists in reconstructing the probability of individual input-output mappings from the joint probability distribution of all input-output mappings through marginalization. This method provides a way to make inferences about systemic properties of input-output mappings from the frequencies of individual input-output mappings, therefore allowing for a quantitative assessment of analyses of dispersion and paradigmatic effects. The approach is illustrated with a concrete case study involving variable homophony avoidance in Spoken Persian.

*I would like to thank two anonymous reviewers for very helpful suggestions. This work also greatly benefited from exchanges with Guillaume Guex, François Bavaud, Giorgio Magri, and participants in the 20th Old World Conference in Phonology in Tours (France). All errors are my own. This work is sponsored by the Leiden University Fund/LUF (www.luf.nl).

1 Introduction

Over the last two decades, the theory of constraint-based grammars proposed by Prince & Smolensky (1993) has witnessed two independent technical innovations from very different research corners. On one hand, probabilistic grammars have been developed to deal with within-speaker variation (e.g., Boersma & Hayes 2001, Goldwater & Johnson 2003, Hayes & Wilson 2008). By allowing a given input to be mapped to more than one output, this probabilistic turn made it possible to infer grammatical parameters from noisy data (e.g., corpus frequencies, experimental data) and therefore was key to establish constraint-based grammars as a viable approach to real-world linguistic data.

Domain	Effects	Examples
phonology	dispersion effects	contrast enhancement
		contrast neutralization
morphophonology/	paradigmatic effects	homophony avoidance
morphology		syncretism

Table 1: Evidence for the role of surface similarity in languages.

On the other hand, the original constraint set with constraints evaluating individual input-output mappings (Prince & Smolensky 1993) has been enriched with a new type of constraints, i.e. systemic constraints. Systemic constraints evaluate properties of *systems* of output forms rather than properties of *individual* input-output mappings. This innovation was motivated by the discovery that surface similarity is relevant in a range of dispersion effects and paradigmatic effects listed in Table 1. For instance, patterns of contrast enhancement whereby perceptually more distinct phonological outputs are preferred over less distinct ones have motivated analyses involving constraints penalizing distinct output forms according to their perceptual distance (i.e. distinctiveness constraints; e.g., Flemming 2004, 2017). Similarly, the observation that phonological processes might be introduced or blocked to avoid homophony within a paradigm (e.g., Crosswhite 1999, Munteanu 2021) has motivated researchers to posit anti-homophony constraints penalizing identical outputs for distinct underlying forms. Intraparadigmatic anti-homophony constraints are also motivated by the typology of morphological syncretism: languages are less likely to map two distinct cells of a morphological paradigm to the same exponent if the resulting homophony would be harder to disambiguate (Storme 2022). However the adoption of this new type of systemic constraints (distinctiveness constraints, anti-homophony constraints) requires an important departure from the classical mode of constraint evaluation: in order to compare output forms, input-output mappings can no longer be evaluated individually but must be evaluated jointly (Magri & Storme 2021).

Unfortunately, joint evaluation is not directly compatible with the probabilistic frameworks that have been developed independently for constraint-based grammars. Indeed, these frameworks assume the original mode of constraint evaluation where input-output mappings are evaluated individually: they specify a probability distribution over individual input-output mappings. This inadequation has resulted in an unfortunate lack of connection between these two domains of the theory of constraint-based grammars. Works that use probabilistic grammars typically assume constraints evaluating individual input-output mappings whereas works that make use of systemic constraints and joint evaluation typically focus on idealized patterns without variation. As a result, constraint-based grammars are currently unable to model dispersion effects and paradigmatic effects using realistic, noisy data. Moreover, analyses using systemic constraints and joint evaluation cannot be evaluated quantitatively against analyses that do not postulate these effects.

This squib proposes a simple method to allow for the evaluation of systemic constraints in probabilistic grammars. This method, known as marginalization, exploits a theorem of probability theory stating that the joint distribution of any given number of random variables encodes the marginal distributions, i.e. the distributions of each of the individual random variable (e.g., Evans & Rosenthal 2009: section 2.7.3). In the context of constraint-based grammars, this means that the probability of any given individual input-output mapping can be reconstructed from the joint probability of all input-output mappings by marginalizing over all other mappings. Marginalization therefore provides a way to make inferences about systemic properties of languages from the frequencies of individual linguistic forms, therefore allowing for a quantitative assessement of analyses of dispersion and paradigmatic effects.

The squib is organized as follows. Section 2 presents the method in more details. Section 3 provides a concrete application of this approach using a probabilistic pattern of vowel-hiatus resolution conditioned by homophony avoidance in Spoken Persian, as described by Ariyaee & Jurgec (2021). Section 4 briefly addresses two other concerns that have been raised about systemic constraints beyond the problem of variation that this squib focuses on: (i) how to deal with the combinatorial explosion implied by joint evaluation and (ii) whether systemic constraints should really be part of the synchronic grammar.

2 Combining joint evaluation and variation

2.1 Joint evaluation in the categorical case: constraint summation

In classical constraint-based grammars, input-output mappings are evaluated independently for distinct input forms. For instance, the mapping from input a to output candidates α_1 and α_2 and the mapping from input b to output candidates β_1 and β_2 are evaluated independently, as shown in Tables 2a and 2b, respectively. By contrast, joint evaluation evaluates the same four input-output mappings all at once, as shown in Table 3. As previewed above, joint evaluation is motivated by the existence of constraints that evaluate the similarity between output forms corresponding to different input forms. For instance, an anti-homophony constraint will require that output forms for inputs a and b be non-identical. When input-output mappings are evaluated jointly, the constraint-violation profile of a set of input-output mappings is obtained by summing the violation profiles of individual input-output mappings inside that set, as shown in Table 3. Constraint summation (also known as the Minkowski sum; Prince 2015) has been shown to be typologically innocuous in the categorical case: for a given constraint ranking (in Optimality Theory) or set of weights (in Harmonic Grammar), the same output candidates are predicted to win whether input-output mappings are evaluated independently as in Table 2 or jointly as in Table 3 (Prince 2015, Magri & Storme 2021).

a	Constraint$_1$	Constraint$_2$
α_1	1	0
α_2	0	1

(a) Evaluation for input *a*

b	Constraint$_1$	Constraint$_2$
β_1	1	0
β_2	0	1

(b) Evaluation for input *b*

Table 2: Classical constraint evaluation: constraints evaluate individual input-output mappings

a,	b	Constraint$_1$	Constraint$_2$
α_1,	β_1	2	0
α_1,	β_2	1	1
α_2,	β_1	1	1
α_2,	β_2	0	2

Table 3: joint constraint evaluation: constraints evaluate sets of input-output mappings

As noted by Prince (2015) and Magri & Storme (2021: Footnote 1), joint evaluation and constraint summation require that the number of inputs be *finite*, as in the example above where there are only two inputs (a and b) or as in the case study in section 3 where, as we will see, homophony avoidance only applies within a restricted morphological paradigm. The assumption of a finite number of inputs will be further discussed in section 4.1.

2.2 Joint evaluation in the probabilistic case: marginalization

Categorical constraint-based grammars derive a single winner candidate for a given input (e.g., α_1 will categorically win over α_2 if Constraint$_2$ is higher-ranked or more weighted than Constraint$_1$). However real-world linguistic data typically show within-speaker variation, with one input form corresponding to more than one output form in the speech of a single speaker (e.g., Bayles, Kaplan & Kaplan 2016). Probabilistic implementations of constraint-based grammars have been proposed to deal with this issue. These grammars do not define a single winner but a probability distribution over output candidates. This probability distribution is defined based on the constraint-violation profile of individual input-output mappings and the rankings or weights of the corresponding constraints.

However these probabilistic approaches do not generalize directly to joint evaluation. Indeed, a probabilistic implementation of a grammar assuming a joint mode of constraint evaluation will define the probability of a set of input-output mappings (e.g., the joint probability that a is mapped to α_1 and b to β_2) but not the probability of individual input-output mappings (e.g., the probability that a is mapped to α_1). Yet frequencies of sets of input-output mappings are not readily available from linguistic data. What is available is the frequency distribution of output forms for a specific input.

To solve this issue, I propose to reconstruct the probability of individual input-output mappings by marginalizing over the probabilities of all other input-output mappings, as detailed in (1). This method builds on a theorem of probability theory according to which the probability that a random variable takes a particular value can be calculated by summing the joint distribution over all values of the *other* random variables (e.g., Evans & Rosenthal 2009: section 2.7.3). This theorem itself follows from Kolmogorov's third axiom of probability theory stating that the probabibility of a set of mutually exclusive random variables is equal to the sum of the probability of each random variable (Kolmogorov 1956). The relation between linguistic forms and random variables is the following: an output set for a single input can be thought of as a single random variable, and the linguistic system consists of a set of random variables, one for each input. In (1), output α_i (one of the possible outputs for input a_i) is fixed throughout the summand and every possible combination of outputs for the *other* inputs (a_1, ..., a_{i-1}, a_{i+1}, ... a_n) are summed over. In words, (1) says that the probability of a specific output α_i is equal to the sum of the probabilities of all combinations of output forms involving output α_i.

(1) Marginalization

Let a_1, ..., a_n be underlying forms, A_1, ..., A_n the corresponding sets of output candidates, and α_i a specific output candidate in A_i. Then,

$$P(\alpha_i) = \sum_{\alpha_1,\ldots,\alpha_{i-1},\alpha_{i+1},\ldots,\alpha_n} P((\alpha_1, \ldots, \alpha_i, \ldots, \alpha_n))$$

Marginalization is an extra component that can be added to the pipeline of probabilistic constraint-based grammars when systemic constraints are used and joint evaluation is therefore required. Concretely, a grammar including systemic constraints will evaluate sets of input-output mappings (as in Table 3) and derive a probability distribution over these sets. Marginalization will then make it possible to go from this probability distribution to the probability distribution over *individual* input-output mappings. Probabilities of individual input-output mappings and other grammatical parameters (e.g., constraint weights) can then be estimated using corpus or experimental frequencies.

For instance, this method makes it possible to derive the probability of individual mappings in Tables 2a and 2b by summing the probabilities of all rows that include those individual mappings in Table 3:

$$\begin{cases} P(\alpha_1) = P((\alpha_1,\beta_1)) + P((\alpha_1,\beta_2)) \\ P(\alpha_2) = P((\alpha_2,\beta_1)) + P((\alpha_2,\beta_2)) \\ P(\beta_1) = P((\alpha_1,\beta_1)) + P((\alpha_2,\beta_1)) \\ P(\beta_2) = P((\alpha_1,\beta_2)) + P((\alpha_2,\beta_2)) \end{cases}$$

For concreteness, suppose that a probabilistic grammar with joint evaluation derives a probability of 0.2 for (α_1, β_1) and a probability of 0.1 for (α_1, β_2) based on the violation profiles of those sets of mappings and constraint weights/rankings. Then, the probability of α_1 predicted by the grammar is equal to $0.2 + 0.1 = 0.3$.

Marginalization should be compatible with any probabilistic framework for constraint-based grammars that provides a probability distribution over input-output mappings, regardless of how this probability distribution is derived. In section 3, I show through an example that it is compatible with MaxEnt (Goldwater & Johnson 2003, Hayes & Wilson 2008). The question whether it is also compatible with other frameworks (e.g., Noisy Harmonic Grammar, Stochastic OT; see Hayes 2017) is left for further research.

3 Application: homophony avoidance in Spoken Persian

This section shows how marginalization can be applied to a concrete case study involving both variation and joint evaluation. The case study is a probabilistic pattern of vowel-hiatus resolution conditioned by homophony avoidance in Spoken Persian (Ariyaee & Jurgec 2021). The pattern is described in section 3.1 and the analysis in section 3.2. The code to run the analysis in `R` (R Core Team 2021) is available in the supplementary materials, along with a prose description of the implementation details.

3.1 Motivations for joint evaluation and variation

In Spoken Persian, vowel hiatuses are generally avoided at stem-suffix boundaries by deletion of the suffix vowel. For instance, a suffixed word /hut͡ʃɑ-emun/ (built with the possessive first person plural suffix /-emun/) would typically be realized as [hut͡ʃɑmun] with deletion of the suffix vowel. However monosegmental suffixes do not follow this generalization: their suffix vowel is more resistant to deletion. For instance, a suffixed word /hut͡ʃɑ-e/ (built with the definite suffix -e) would typically be realized as [hut͡ʃɑe] with a vowel hiatus.

This pattern can be analyzed as motivated by homophony avoidance: the general tendency to delete suffix vowels is overriden in the specific case of monosegmental suffixes by the need to maintain suffixed forms distinct from the unsuffixed base. Indeed vowel deletion results in homophony with the base form (e.g., /hut͡ʃɑ/ [hut͡ʃɑ]) for monosegmental suffixes (e.g., /hut͡ʃɑ-e/ realized as [hut͡ʃɑ]) but not for polysegmental suffixes.

Homophony avoidance can be modeled with a systemic constraint penalizing identical output forms corresponding to distinct underlying forms (e.g., Crosswhite 1999 for an application in morphophonology, Storme 2022 for an application in core morphology). This constraint is a systemic constraint that requires joint evaluation because it evaluates the surface similarity of two outputs corresponding to distinct underlying forms. In the case of Spoken Persian, this constraint more specifically penalizes identical output forms for a suffixed form and the corresponding base. For instance, it penalizes mapping the base /hut͡ʃɑ/ and the corresponding suffixed form /hut͡ʃɑ-e/ to the same output form [hut͡ʃɑ].[1] There is a

[1]The analysis proposed by Ariyaee & Jurgec (2021) does not use an anti-homophony constraint but a constraint that penalizes a null realization for a morpheme (RealizeMorpheme; see Kurisu 2001). Crucially, this constraint does not require joint evaluation. However RealizeMorpheme cannot provide a general account of homophony avoidance across languages because not all patterns of homophony avoidance involve the non-realization of a morpheme (e.g., Trigrad Bulgarian, as described by Crosswhite 1999, and Northern Haitian, as described by Storme & Otilien 2022).

finite number of suffixes in Persian (see the exhaustive list of suffixes provided by Ariyaee & Jurgec 2021 in Table 2). Therefore the assumption of a finite number of inputs required by joint evaluation (see section 2.1) can be satisfied.

Moreover, hiatus resolution in Spoken Persian is probabilistic. Indeed, although vowel deletion is favored in polysegmental suffixes, vowel hiatuses are also attested (e.g., /hut͡ʃɑ-emun/ may be realized as [hut͡ʃɑemun]). Conversely, although vowel hiatuses are favored in monosegmental suffixes, vowel deletion is also attested (e.g., /hut͡ʃɑ-e/ may be realized as [hut͡ʃɑ]). Moreover a third pronunciation with glottal-stop epenthesis is attested for both polysegmental and monosegmental suffixes (e.g., /hut͡ʃɑ-emun/ may be realized as [hut͡ʃɑʔemun] and /hut͡ʃɑ-e/ as [hut͡ʃɑʔe]).

To summarize, the analysis of hiatus resolution in Spoken Persian as conditioned by homophony avoidance requires a model that can deal with both joint evaluation and variation. Spoken Persian is therefore a good test case for the method proposed in this squib.

3.2 Analysis

The analysis closely follows Ariyaee & Jurgec (2021). The only difference lies in the use of an anti-homophony constraint (*Homophony) instead of a constraint that penalizes null exponents for morphemes (RealizeMorpheme; see Footnote 1 for a discussion). The analysis compares the behavior of monosegmental suffixes (e.g., /hut͡ʃɑ-e/) and polysegmental suffixes (e.g., /hut͡ʃɑ-emun/). The same three outputs are considered for each type of suffix as in Ariyaee & Jurgec (2021): vowel hiatus (e.g., [hut͡ʃɑe]), glottal-stop epenthesis (e.g., [hut͡ʃɑʔe]), and vowel deletion (e.g., [hut͡ʃɑ]). Vowel hiatus is penalized by *Hiatus, glottal-stop epenthesis by Dep, and vowel deletion by Max. The unsuffixed base /hut͡ʃɑ/ is also included in the present analysis because *Homophony evaluates the similarity between suffixed forms and the base. In the absence of information on patterns of variability in the

/hut͡ʃɑ/,	/hut͡ʃɑ-e/,	/hut͡ʃɑ-emun/	*Homophony $w = 2.27$	Dep $w = 2.47$	*Hiatus $w = 1.89$	Max $w = 1.00$	P
[hut͡ʃɑ],	[hut͡ʃɑe],	[hut͡ʃɑemun]			2		0.14
[hut͡ʃɑ],	[hut͡ʃɑe],	[hut͡ʃɑʔemun]		1	1		0.08
[hut͡ʃɑ],	[hut͡ʃɑe],	[hut͡ʃɑmun]			1	1	0.34
[hut͡ʃɑ],	[hut͡ʃɑʔe],	[hut͡ʃɑemun]		1	1		0.08
[hut͡ʃɑ],	[hut͡ʃɑʔe],	[hut͡ʃɑʔemun]		2			0.04
[hut͡ʃɑ],	[hut͡ʃɑʔe],	[hut͡ʃɑmun]		1		1	0.19
[hut͡ʃɑ],	[hut͡ʃɑ],	[hut͡ʃɑemun]	1		1	1	0.03
[hut͡ʃɑ],	[hut͡ʃɑ],	[hut͡ʃɑʔemun]	1	1		1	0.02
[hut͡ʃɑ],	[hut͡ʃɑ],	[hut͡ʃɑmun]	1			2	0.09

Table 4: Constraint-based analysis of hiatus resolution in Spoken Persian

pronunciation of the base, a single output candidate is considered for this form: the fully faithful candidate ([hut͡ʃɑ]). The constraint *Homophony requires joint evaluation. Given the above assumptions about the set of inputs and outputs, the set of candidates for this joint evaluation contains nine triplets, represented in Table 4.

The MaxEnt framework (Goldwater & Johnson 2003, Hayes & Wilson 2008) was used to get from constraint violations to probabilities. Constraint weights were inferred from the frequencies of individual input-output mappings reported in Ariyaee & Jurgec (2021: section 4.3) using a Bayesian multinomial logistic regression implemented in `rjags` (Plummer 2016). A multinomial regression was used because inputs have up to three corresponding outputs, and therefore binary random variables and simple logistic regression are not enough. A Bayesian approach was adopted (rather than a frequentist approach) because it provides more intuitive and meaningful inferences (Kruschke & Liddell 2018). However this choice is not key here and a frequentist approach could very well be adopted.[2]

The posterior means for the weights of the first three constraints (*Homophony, Dep, *Hiatus) are reported in Table 4 along with the predicted frequencies for each of the nine

[2] A reviewer was able to reproduce the analysis using the Excel solver.

			Predicted frequency
Base	/hut͡ʃɑ/	[hut͡ʃɑ]	1.00
Monosegmental suffix	/hut͡ʃɑ-e/	[hut͡ʃɑe]	0.55
		[hut͡ʃɑʔe]	0.31
		[hut͡ʃɑ]	0.14
Polysegmental suffix	/hut͡ʃɑ-emun/	[hut͡ʃɑemun]	0.25
		[hut͡ʃɑʔemun]	0.14
		[hut͡ʃɑmun]	0.61

Table 5: Model predictions for individual input-output mappings

paradigms.[3] The frequencies predicted for individual input-output pairs can be reconstructed from these frequencies through marginalization, as shown in Table 5. The frequencies predicted by the analysis for individual input-output mappings perfectly match the frequencies observed by Ariyaee & Jurgec (2021), indicating that the analysis combining variation and joint evaluation is successful at accounting for the Persian data.

This model was also compared to a model that does not include the anti-homophony constraint and therefore does not require joint evaluation, using the Deviation Information Criterion (DIC; Gelman et al. 2013: pp. 172–173). The model with *Homophony was found to provide a better fit to the data than the model without this constraint (DIC $= -496.62$, Sample Standard Error $= 51.34$).

4 Open issues about systemic constraints

This squib fills a gap in the literature by providing a method to fit grammars including systemic constraints to noisy data. However the methodological challenge to combine joint evaluation and variation is probably not the only reason or even the main reason why analyses involving systemic constraints have not been more widely adopted in linguistics.

[3]The weight of Max was not estimated by the model but set to 1 to help with model convergence.

There are two additional and more substantial concerns about systemic constraints that come up in the literature. Although these additional concerns are not directly relevant to the main focus of this squib, it is important to address them briefly as they are likely to also play a role in the linguist's decision whether to use systemic constraints in their analyses of variable linguistic phenomena.

4.1 How to deal with the combinatorial explosion implied by systemic constraints?

Systemic constraints require joint evaluation and joint evaluation is complex as it potentially involves optimizing over entire languages (Ní Chiosáin & Padgett 2010, Flemming 2017). In particular, in its most unconstrained form, joint evaluation requires evaluation of a number of candidates that is equal to the number of combinations of output forms for all input forms. Suppose there are n inputs and k outputs for each input. The number of candidates to evaluate in one tableau is then equal to k^n. This number grows exponentially with the number of inputs in the language. In the toy case study discussed in section 3, this number ($3^2 = 9$) was rather small because the focus was on a very small number of inputs ($n = 2$) and outputs ($k = 3$). (The unsuffixed base is ignored here as there was only one possible output.) However, in more realistic applications, this number would be much higher. For instance, even in the unlikely scenario of a language with only 10,000 inputs and 3 outputs for each input, the number of candidates to evaluate would already be unrealistically high ($3^{10,000}$).[4] The problem becomes of course even more acute if the number of inputs is infinite (Ní Chiosáin & Padgett 2010).

Different strategies have been proposed to reduce the size of the candidate set to a

[4]Note that a similar problem involving a combinatorial explosion of candidates was observed by Mascarenhas (2021) in the computation of the set of scalar alternatives in semantics/pragmatics.

more manageable size, but this issue is still open (Flemming 2017: section 3.1). One approach consists of imposing restrictions on the type of output-output similarity that systemic constraints can evaluate. This has been proposed explicitly for phonological patterns of homophony avoidance. For instance, Kaplan & Muratani (2015) observe that these patterns always imply comparisons between words in the same morphological paradigm, i.e. words that share the same stem (e.g., Spoken Persian in section 3.2 or Trigrad Bulgarian in footnote 1). This observation is compatible with a restricted constraint set with only *intraparadigmatic* anti-homophony constraints.

This restriction on the constraint set can lead to a significant decrease in the size of the candidate set. To illustrate, consider two underlying stems a and b (with candidate sets A and B) and two underlying suffixes c and d (with candidate sets C and D). The restriction that homophony avoidance only applies to outputs sharing a stem means that outputs based on stem a and stem b do not need to be evaluated in the same tableau but can be evaluated independently in two tableaux. This reduces the number of total candidates to evaluate for inputs /*a*-*c*/, /*a*-*d*/, /*b*-*c*/, and /*b*-*d*/ from the number in (2a) to the more manageable number in (2b) ($|A|$ denotes the cardinality of A). If candidate sets for each individual input include two outputs, this corresponds to $2^8 = 256$ and $2^4 + 2^4 = 32$ sets of outputs to evaluate, respectively (i.e. 256 lines in the tableau vs. 32 lines). In the more constrained approach, the number of candidates therefore grows exponentially with the number of inputs *in a morphological paradigm*. Future work should establish formally whether this kind of pruning provides a general and viable solution to the problem of the scalability of systemic constraints.[5]

(2) a. Size of the candidate set with unconstrained homophony avoidance:

[5] A reviewer also suggests that the idea expressed by Prince (2015) that every typology is representable as a finite collection of finite candidate sets might also be exploited to reduce the combinatorial issue implied by joint evaluation.

$$\underbrace{|A| \times |C|}_{\text{nb of outputs for /}a\text{-}c\text{/}} \times \underbrace{|A| \times |D|}_{\text{nb of outputs for /}a\text{-}d\text{/}} \times \underbrace{|B| \times |C|}_{\text{nb of outputs for /}b\text{-}c\text{/}} \times \underbrace{|B| \times |D|}_{\text{nb of outputs for /}b\text{-}d\text{/}}$$

b. Size of the candidate set with intraparadigmatic homophony avoidance:

$$\underbrace{|A| \times |C|}_{\text{nb of outputs for /}a\text{-}c\text{/}} \times \underbrace{|A| \times |D|}_{\text{nb of outputs for /}a\text{-}d\text{/}} + \underbrace{|B| \times |C|}_{\text{nb of outputs for /}b\text{-}c\text{/}} \times \underbrace{|B| \times |D|}_{\text{nb of outputs for /}b\text{-}d\text{/}}$$

4.2 Do we need systemic constraints in synchronic grammars?

Alternative analyses of dispersion and paradigmatic effects have been proposed that do not require systemic constraints in *synchronic* grammars (e.g., Boersma & Hamann 2008, Blevins & Wedel 2009, Wedel 2012). These analyses are *diachronic*: dispersion and paradigmatic effects emerge gradually across generations due to the way listeners categorize word forms. For instance, in Blevins & Wedel (2009) and Wedel (2012), listeners discard phonetically ambiguous word tokens when constructing two word categories, in particular when the two words are in strong lexical competition, e.g., if they occur in the same type of morphosyntactic environment (Wedel, Kaplan & Jackson 2013). Listeners discard these ambiguous tokens because they fail to correctly categorize them due to their overlap in phonetic and distributional space. On the production side, this gives rise to hyperarticulation: the removal of ambiguous tokens results in more distinct phonetic realizations for the relevant words than what listeners were initially exposed to, and therefore in patterns of contrast enhancement/homophony avoidance. The analysis is implemented in agent-based models where language is transmitted from one generation to the next.

One appeal of these alternative diachronic analyses is that they do not require the kind of cognitively complex joint evaluation that is implied by analyses using systemic constraints in synchronic grammars (see section 4.1). However it is unclear that they can account for all the patterns that systemic constraints can account for. For instance, there is evidence for an anti-homophony bias in artificial language learning, and the fact that this bias emerges over the course of an experiment seems more compatible with an explicit cognitive bias against homophony (Yin & White 2018). Moreover, it is unclear how diachronic analyses can jointly account for contrast enhancement and contrast neutralization in the typology of the world's languages, as pointed out by Flemming (2017: section 5.7). Some languages that are otherwise very similar might differ in whether they neutralize a phonological/morphological contrast or not. This kind of differences is unexpected if all speakers/listeners share the exact same biases and there is no speaker-specific trade-off mechanism, as in Wedel (2012) for instance. In synchronic analyses using systemic constraints, both contrast enhancement and neutralization can be derived in this case as different speakers are allowed to trade-off conflicting biases (e.g., effort minimization and distinctiveness maximization) differently. I made a similar argument for synchronic systemic constraints based on the typology of morphological syncretism (Storme 2022: section 5.1.3).

The method proposed in this paper to evaluate systemic constraints in probabilistic grammars will hopefully foster new empirical and modeling studies documenting variable paradigmatic effects and dispersion effects, and help to settle whether systemic constraints are really needed in synchronic grammars.

References

Ariyaee, Koorosh & Peter Jurgec. 2021. Variable hiatus in Persian is affected by suffix length. In Ryan Bennett, Richard Bibbs, Mykel Loren Brinkerhoff, Max J. Kaplan, Stephanie Rich, Amanda Rysling, Nicholas Van Handel & Maya Wax Cavallaro (eds.), *Proceedings of the 2020 Annual Meeting on Phonology*. Washington: Linguistic Society of America. https://doi.org/10.3765/amp.v9i0.4919.

Bayles, Andrew, Aaron Kaplan & Abby Kaplan. 2016. Inter-and intra-speaker variation in French schwa. *Glossa: a journal of general linguistics* 1. https://doi.org/10.5334/gjgl.54.

Blevins, Juliette & Andrew Wedel. 2009. Inhibited sound change: An evolutionary approach to lexical competition. *Diachronica* 26. 143–183. https://doi.org/10.1075/dia.26.2.01ble.

Boersma, Paul & Silke Hamann. 2008. The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25. 217–270. https://doi.org/10.1017/S0952675708001474.

Boersma, Paul & Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32. 45–86. https://doi.org/10.1162/002438901554586.

Crosswhite, Katherine M. 1999. Intra-paradigmatic homophony avoidance in two dialects of Slavic. In Matthew Gordon (ed.), *UCLA Working Papers in Linguistics 1: Papers in phonology 2*, 48–67. Los Angeles: UCLA.

Evans, Michael J. & Jeffrey S. Rosenthal. 2009. *Probability and statistics: The science of uncertainty*. Second edition. New York: W. H. Freeman.

Flemming, Edward. 2004. Contrast and perceptual distinctiveness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically based phonology*, 232–276. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486401.008.

Flemming, Edward. 2017. Dispersion theory and phonology. In Mark Aronoff (ed.), *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.110.

Gelman, Andrew, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari & Donald B Rubin. 2013. *Bayesian data analysis*. New York: Chapman & Hall/CRC. https://doi.org/10.1201/b16018.

Goldwater, Sharon & Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In Jennifer Spenader, Anders Eriksson & Östen Dahl (eds.), *Pro-*

ceedings of the Stockholm Workshop on Variation within Optimality Theory, 111–120. Stockholm: Stockholm University, Department of Linguistics.

Hayes, Bruce. 2017. Varieties of Noisy Harmonic Grammar. In Karen Jesney, Charlie O'Hara, Caitlin Smith & Rachel Walker (eds.), *Proceedings of the 2016 Annual Meeting on Phonology*. Washington: Linguistic Society of America. https://doi.org/10.3765/amp.v4i0.3997.

Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39. 379–440. https://doi.org/10.1162/ling.2008.39.3.379.

Kaplan, Abby & Yuka Muratani. 2015. Categorical and gradient homophony avoidance: Evidence from Japanese. *Laboratory Phonology* 6. 167–195. https://doi.org/10.1515/lp-2015-0006.

Kolmogorov, Andreaei Nikolaevich. 1956. *Foundations of the theory of probability*. Second English edition. New York: Chelsea.

Kruschke, John K. & Torrin M. Liddell. 2018. Bayesian data analysis for newcomers. *Psychonomic bulletin & review* 25. 155–177. https://doi.org/10.3758/s13423-017-1272-1.

Kurisu, Kazutaka. 2001. *The phonology of morpheme realization*. UC Santa Cruz dissertation.

Magri, Giorgio & Benjamin Storme. 2021. Constraint summation in phonological theory. *Journal of Language Modelling* 8. 251–294. https://doi.org/10.15398/jlm.v8i2.216.

Mascarenhas, Salvador. 2021. A note on the cardinalities of sets of scalar alternatives. *Journal of Semantics* 38. 473–482. https://doi.org/10.1093/jos/ffab011.

Munteanu, Andrei. 2021. Homophony avoidance in the grammar: Russian nominal allomorphy. *Phonology* 38. 401–435. https://doi.org/10.1017/S0952675721000257.

Ní Chiosáin, Máire & Jaye Padgett. 2010. Contrast, comparison sets, and the perceptual space. In Steve Parker (ed.), *Phonological argumentation: Essays on evidence and motivation*, 103–121. London: Equinox.

Plummer, Martyn. 2016. *rjags: Bayesian Graphical Models using MCMC*. R package version 4-5. https://CRAN.R-project.org/package=rjags.

Prince, Alan. 2015. *One tableau suffices*. Manuscript (Rutgers University). Available from the Rutgers Optimality Archive as ROA1250. http://roa.rutgers.edu/content/article/files/1453_alan_prince_4.pdf.

Prince, Alan & Paul Smolensky. 1993. *Optimality theory: constraint interaction in generative grammar*. New Brunswick, NJ: Rutgers Center for Cognitive Science, Rutgers University.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/.

Storme, Benjamin. 2022. Implicational generalizations in morphological syncretism: the role of communicative biases. *Journal of Linguistics* 58. 381–421. https://doi.org/10.1017/S0022226721000207.

Storme, Benjamin & Jocelyn Otilien. 2022. Phonologically conditioned suppletive allomorphy in Haitian as morphological optimization. In Peter Jurgec, Liisa Duncan, Emily Elfner, Yoonjung Kang, Alexei Kochetov, Brittney K. O'Neill, Avery Ozburn, Keren Rice, Nathan Sanders, Jessamyn Schertz, Nate Shaftoe & Lisa Sullivan (eds.), *Supplemental Proceedings of the 2021 Annual Meeting on Phonology*. Washington: Linguistic Society of America. https://doi.org/10.3765/amp.v9i0.5155.

Wedel, Andrew. 2012. Lexical contrast maintenance and the organization of sublexical contrast systems. *Language and Cognition* 4. 319–355. https://doi.org/10.1515/langcog-2012-0018.

Wedel, Andrew, Abby Kaplan & Scott Jackson. 2013. High functional load inhibits phonological contrast loss: a corpus study. *Cognition* 128. 179–186. https://doi.org/10.1016/j.cognition.2013.03.002.

Yin, Sora Heng & James White. 2018. Neutralization and homophony avoidance in phonological learning. *Cognition* 179. 89–101. https://doi.org/10.1016/j.cognition.2018.05.023.