Published in final edited form as:

Linguist Inq. 2012 ; 43(1): 97–119. doi:10.1162/LING_a_00075.

On the role of variables in phonology: Remarks on Hayes and Wilson (2008)

Iris Berent,
Northeastern University

Colin Wilson,
Johns Hopkins University

Gary Marcus, and Doug Bemis
New York University

Abstract

A recent computational model by Hayes and Wilson (2008) seemingly captures a diverse range of phonotactic phenomena without variables, contrasting with the presumptions of many formal theories. Here, we examine the plausibility of this approach by comparing generalizations of identity restrictions by this architecture and human learners. Whereas humans generalize identity restrictions broadly, to both native and non-native phonemes, the original model and several related variants failed to generalize to non-native phonemes. In contrast, a revised model equipped with variables more closely matches human behavior. These findings suggest that, like syntax, phonological grammars are endowed with algebraic relations among variables that support across-the-board generalizations.

What is the bare minimum inventory of computational machinery necessary to account for phonological generalizations? In recent years, there has been a surge of interest in a class of statistical models of phonology (e.g., Adriaans and Kager 2010, Albright 2007, Albright 2009, Cole 2009, Coleman and Pierrehumbert 1997, Frisch et al. 2004, Goldsmith 2010, Harm and Seidenberg 1999, Rumelhart and McClelland 1986, Sibley et al. 2008). Much of the excitement surrounding these models stems from their promise to capture native speakers' knowledge of phonology while eliminating many substantive universal constraints assumed by previous theories (e.g., the universal constraint set assumed by classic Optimality Theory, Prince and Smolensky 1993/2004, McCarthy and Prince 1995). The challenge presented by statistical models, however, goes far beyond the question of phonological substance. It is unbounded infinity—the capacity to represent and learn generalizations that apply to *any* member of a given class, both familiar and novel—potentially, a defining feature of the grammar itself (Chomsky 1972, Chomsky 2005) that many such models tacitly deny.

Categories (e.g., “syllable”, “noun”) and variables over categories are necessary for learning and representing certain unbounded generalizations. However, many statistical models of

Address for correspondence Iris Berent Department of Psychology Northeastern University 125 Nightingale Hall 360 Huntington Ave. Boston, MA 02115-5000 i.berent@neu.edu .

⁷Additional simulations compared our revised Maxent_{Gain} to a parallel revision of the Maxent_{O/E} model. Despite the facts that both models were modified so that they could encode identity restrictions, they differed in their capacity to acquire those constraints. While the revised Maxent_{Gain} model acquired a general ban on initial identical consonants, the revised Maxent_{O/E} version only learned a narrower ban on identical *continuants*, a restriction that correctly generalized to the continuant that correctly generalized to the continuant /θ/, but erroneously failed to generalize to /t/ or /d/. The restriction to continuants is a result of overfitting to the training data, which contains a handful (4/1449) of roots with repeated identical elements, all of which are stops.

phonology eliminate variables altogether. Generalizations, in such models, are captured only by constraints on the co-occurrence of features and segments, and these constraints are learned from the lexicon. To the extent such models can account for phonological competence, their success would indicate that phonological learners are potentially limited with respect to the scope of phonological generalizations they ultimately attain. The narrow scope of phonological generalizations would also set the phonological computational component apart from the rest of the grammar.

In this commentary, we wish to draw attention to this tacit assumption and evaluate it by analyzing the performance of the statistical induction model proposed by Hayes and Wilson (2008; henceforth ‘HW08’)—arguably the most successful of its kind. The HW08 model stands apart from its predecessors in that it can demonstrably capture significant generalizations ranging from specific aspects of phonotactics—onset phonotactics, vowel harmony and stress patterns—to the phonology of an entire language. Although the HW08 model has been amply tested against natural linguistic corpora and experimental findings, missing from this evaluation is a systematic assessment of the scope of generalizations attained by the model. Here, we outline such a test and use it to evaluate the model against human behavior. We demonstrate that, despite its numerous achievements, the original HW08 architecture is limited with respect to its capacity to generalize, and that as a result, it is ultimately unable to fully capture human phonological productivity. While such limitations can be remedied by revising the model, the need for such revisions challenges the view of the combinatorial phonological system as narrow in scope.

1. The role of variables in explaining linguistic productivity

The vast majority of contemporary linguistic theory is predicated upon the assumption that many linguistic generalizations extend across the board to unbounded sets of instances. People demonstrably extend their syntactic knowledge to word sequences they have never heard before (Chomsky 1957), for instance. Similarly, word formation applies to utterly unfamiliar bases (Berent et al. 1999, Kim et al. 1994, Marcus et al. 1995, Pinker 1991, Prasada and Pinker 1993), and has been used as a canonical example of the need for operations over variables in human language (Pinker 1991, Marcus, 2001). Whether a grammar is characterized in terms of rules, parameters, principles, or constraints, most linguistic theories attribute unbounded productivity explicitly or implicitly to variables (Chomsky 1980, Marcus 2001, Pinker and Prince 1988, Smolensky and Legendre 2006). In such theories, grammatical statements concern categories (“noun”) rather than any bounded set of tokens that might instantiate those categories (e.g., *dog*, *cat*, as instances of a “noun”). Because variables can stand for unbounded sets of elements, they allow learners to extend generalizations across the board, to any potential members of the class. A learner who acquired the English plural formation rule (Noun_{stem} +S) will generalize relevant grammatical restrictions to any token, irrespective of whether it is familiar (e.g., **rats-eater*) or not, and regardless of whether it is phonologically licit (e.g., **ploons-eater*) or ill-formed (e.g., **ptoons-eater*, Berent and Pinker 2007).

Variables are critical for numerous grammatical functions. Here, it might be useful to distinguish between two types of functions: A simple first-order concatenation of categories (e.g., NP→Det+N; syllable→onset +rhyme); and a second-order *relation* among them, such as Recursion (e.g., S→S+NP) and Identity (XX, where X can stand for a given element). While first-order concatenative functions array the ordering of specific categories (e.g., the category of Nouns, onsets or the labial features), relations such as identity hold for *any* such category, familiar (the class of labials) or novel (the class of segments defined by features that are unattested in one’s language). Such relations therefore cannot be encoded by concatenating categories. Rather, they call for a variable that ranges over a category and that

can appear in more than one place in a single grammatical rule or constraint (thus binding the two positions together by the relation of identity). So unlike first-order concatenation of categories, relational functions *require* the encoding of variables. One can certainly describe sequences of identical elements (e.g., *dog-dog*, *labial-labial*) as the co-occurrence of specific tokens, but the formal relation of identity can only be expressed through the use of variables.

2. Phonology in the absence of variables

Although variables have played a central role in many formal linguistic accounts, the HW08 model can perform surprisingly well without them. Like many statistical models of phonotactics, the HW08 model infers the well-formedness of novel forms from the co-occurrence of their constituents in speakers' linguistic experience. Such inferences are captured by a set of weighted markedness constraints that maximize the probability of observed forms, which Hayes and Wilson take as a proxy of well-formedness. Crucially, constraints can only take two forms—either as bans on the sequencing of feature-matrices, or as logical implications among certain feature values. Variables and relations among variables are not represented in the HW08 model.

Provided only with the set of onset clusters attested in English, the model spontaneously acquired a set of 23 constraints on feature co-occurrence, which allowed it to productively distinguish attested English onsets that were absent in the training set from unattested ones, and even discern the well-formedness of onsets that are all unattested in English: the model's performance on unattested onsets approximated human ratings quite well ($r=.946$) and exceeded both a handcrafted grammar based on Clements and Keyser (1983) and the alternative statistical computational model of Coleman and Pierrehumbert (1997). Although the model was unable to spontaneously learn nonlocal dependencies, when provided with a local representation of vowel harmony and metrical structure (using the vowel projection and metrical grid), the model successfully acquired the vowel-harmony pattern of Shona and extracted the nonlocal stress pattern of Eastern Cheremis as well as 33 other languages. Furthermore, the model's performance is not confined to isolated pieces of the phonological grammar. Given the Australian aboriginal language of Wargamay, the model extracted various aspects of segmental phonotactics, including restrictions on syllable shapes, initial and final consonants, intervocalic consonant clusters and consonant-vowel combinations and learned the metrical pattern.

Not only can the HW08 model capture regularities concerning feature co-occurrence, but in some instances it could conceivably approximate constraints that would in traditional approaches be represented by means of explicitly-represented variables, such as identity restrictions. A restriction on identical places of articulation, for example, could be restated using multiple separate restrictions on feature co-occurrence, against root-adjacent labials (e.g., *labial-labial), dorsals (e.g., *dorsal-dorsal), coronals (e.g., *coronal-coronal), etc. (Coetzee and Pater 2008). A *labial-labial constraint would not be formally distinct from any other first-order restriction on the co-occurrence of non-identical features (e.g., *labial-dorsal), and the fact that the two arguments of a *labial-labial constraint are, indeed, identical would not be directly represented. Nonetheless, constraints on feature co-occurrence might conceivably capture some of the restrictions on the distribution of attested forms.

In a similar way, a variable-free approach might even support some generalizations to novel items. Consider, for example, a hypothetical Semitic-like language consisting of elements such as *smm*, *lff*, and *gpp*, in which identical consonants would be allowed root finally, but not root initially. Although learners that lack variables could not represent identity per se,

HW08-style learners might nonetheless favor the unattested *lbb* over *bbl* in virtue of its labial-labial sequence, relative to *lff* and *gpp*. To the extent that novel items could piggyback onto constraints learned from lexical items, generalization can ensue.

The model's apparent success thus could be viewed as a challenge to the notion that variables are ultimately represented by the phonological mind, and also as an implicit proposal for one way in which the phonological component might be distinct from the rest of the grammar, perhaps even requiring a major rethinking of phonological theory and the relation of phonology to the language system as a whole.

3. The case for variables in phonology

Despite the model's success, there are reasons to question whether the co-occurrence of feature sets is an adequate account of phonological knowledge. Many phonological primitives—entities such as “onset” “syllable”, “foot” and “base”—potentially apply to unbounded classes of elements. Certain phonological constraints on such classes could entail variables.

The case for variables is strongest when it comes to phonological constraints that express second-order relations. Although segmental phonotactics famously lacks the second-order relation of recursion, phonological recursion has been documented at higher prosodic levels (Ladd 1986), and second-order restrictions on identity are common in segmental phonology (Suzuki 1998, Rose and Walker 2004). Identity avoidance is at the core of countless phonological processes, and its effects have been documented at numerous levels, including features (Ito and Mester 1986, McCarthy 1994, Yip 1988), tones (Leben 1973), segments (McCarthy 1981), and prosodic categories (Yip 1988). Similarly, phonological reduplication is widely attested (Moravcsik 1978). Although reduplication can take morphological functions (Inkelas and Zoll 2005), many cases of reduplication occur for purely phonological reasons (Alderete et al. 1999, Inkelas 2008). In a system that lacks variables, such restrictions on identity and reduplication are represented in, at best, a piecemeal and fragmentary way.

A second argument for variables is presented by the scope of phonological generalizations. Numerous studies have shown that people generalize their phonotactic knowledge to novel items. Constraints on identity, for example, have been shown to generalize in both natural (Berent et al. 2001b, Berent and Shimron 2003, Berent et al. 2004, Coetzee 2008, Frisch and Zawaydeh 2001) and artificial languages (Marcus et al. 1999, Nevins 2010, Toro et al. 2008). But as discussed above, many generalizations can be accounted for either with powerful expressions over variables or from expressions consisting only of feature matrices.

How can these two options be empirically differentiated? The hallmark of computational mechanisms that operate on variables is that they allow learners to extend generalizations across the board. A restriction on identical consonants (*XX, where X=a consonant), for instance, would apply to *any* consonant, either familiar or novel.

Marcus (Marcus 2001, Marcus 1998) operationalized across-the-board generalizations in terms of the composition of novel test items and the items attested in learners' linguistic experience (the training set). Some test items overlap with the feature composition of training items, and consequently, generalization to such items (i.e., generalization *within-the-training space*) is ambiguous between mechanisms that relates variables and weaker systems that concatenate bounded feature sets. In contrast, a genuine generalization across-the-board extends to novel items even when these items cannot be exhaustively captured within the representational space of training items—generalization dubbed by Marcus as falling *outside-the-training space*.

Although systems that explicitly represent variables have a mechanism for generalizing outside of training space, we conjectured that a feature-driven architecture that lacks variables, such as HW08, might be expected to generalize in a more piecemeal fashion, depending on the exact relationship between novel elements and specific training examples.

In what follows, we ask two questions: first, is our conjecture correct? Is the HW08 architecture in fact limited in its abilities to generalize outside the training space? And second, how does the performance of the model relate to the available empirical evidence from humans?

3.1.1. Hebrew word-formation: An empirical test case

Our test for the role of variables in phonology comes from the restriction on identical consonants in the Hebrew base morpheme. Like many other Semitic languages, Hebrew allows identical consonants to occur at the right edge of the base (e.g., *simem*), but it bans them at the left edge (e.g., *sisem*; Greenberg 1950). The precise domain of this restriction—whether it constrains a consonantal root or a prosodic base consisting of consonants and vowels (hereafter, stems)—has been the subject of much active research (e.g., Bat-El 1994; Ussishkin 1999; Gafos 2003; for experimental evidence, see Berent et al. 2007). McCarthy's (1981) root based account, for example, attributes the asymmetry to the application of the Obligatory Contour Principle within the consonantal root. In contrast, Bat-El's stem-based proposal (2006) captures the same facts by means of agreement-by-correspondence type constraints that are limited to identical consonants at the right edge. Regardless of the precise nature of this morphological domain, however, a restriction on the location of identical consonants must include mechanisms that specifically target identical elements. Our question here is whether human speakers do, in fact, encode such a constraint, and whether this constraint will be learnable by the HW08 architecture. To assure that our test of the capacity of the HW08 architecture to restrict the location of identical consonants is not contaminated by the orthogonal problem of capturing long-distance restriction, here, we describe this base as a consonantal root—a decision motivated strictly on methodological grounds specific to our present inquiry.

Past research by Berent and colleagues (2002) demonstrated that speakers of Hebrew generalize their knowledge of the constraint on the Hebrew base morpheme to novel bases that include identical consonantal segments that are not native to Hebrew—the consonantal segments /t□/, /d□/, and /θ/ (albeit not to the non-consonantal segment /w/)¹. Hebrew speakers rated /kθθ/-type roots as more acceptable than /θθk/-type controls, and they took longer to determine that such forms do not exist in Hebrew. The generalization to /θ/ is particularly striking: not only is this segment non-native to Hebrew, but its value for the Tongue Tip Constriction Area feature (TTCA, Gafos 1999) does not occur in any native Hebrew segment. A priori it would have been doubtful that restrictions on segment identity can be captured at the feature level—*smm*-type roots are far more frequent and preferable to the homorganic *smb* even though their final consonants share every feature (Berent and Shimron 2003, Berent et al. 2004); the observed generalization to a non-native feature value counters even this remote possibility. Because this generalization exceeds the space of Hebrew segments and features, Berent et al. interpreted these data as implying a constraint on identity that was stated as an operation over variables. Can the HW08 architecture challenge this interpretation and provide an alternative account that does not explicitly appeal to variables?

¹ With the introduction of *chat*, Hebrew has recently borrowed forms such as t it t et (to chat), but it is unlikely that such forms were available to participants in Berent et al.'s (2002) experiments. Moreover, to our knowledge, no borrowing into Hebrew has any reduplicants of the / θ/ phoneme.

3.1.2. Testing the HW08 architecture

To examine this possibility, we tested the Hayes and Wilsons (2008) architecture on the Hebrew OCP data, including both native and non-native phonemes. Each such test was conducted ten times, and the results reported here reflect the mean across the ten runs. In the first stage, we trained the model on a database of tri-consonantal Hebrew roots. Next, we tested the model for its ability to generalize the restriction on segment identity to novel roots—the same set of materials previously tested with human participants (Berent et al. 2001a; Berent et al. 2002). In one condition, novel test roots included phonemes that are all attested in Hebrew. Because such generalizations fall within the model's training space, we expected the original model to successfully generalize to such items. In the second critical condition, novel test items included identical segments with non-native phonemes and features (/w/, /t□/, /d□/, and /θ/), and as such, they fall outside the model's training space. Our conjecture was that because the HW08 architecture lacks relations among variables, it would (unlike human subjects) fail to exhibit such generalizations. If it were to succeed, the model would provide a sharp challenge to the view that variables are necessary for the adequate representation of phonology.

3.1.3. Model Code

The model used in our simulations was identical to the one in HW08 with two exceptions. First, the selection of constraints was determined by the criterion of maximal *gain* (as proposed by Della Pietra et al. 1997); second, the complicated search heuristics proposed in HW08 (pp. 393-395) were replaced by exhaustive search. While this version maintains all of the core properties of the original HW08 architecture, subsequent simulations (reported below) showed that our simulation outperformed the published HW08 version on the Hebrew pattern. As such, this simulation presents a stronger test of the principled limitations of the HW08 architecture. To distinguish this variant from the (very close) version published by HW (Hereafter, the Maxent_{HW}), we refer to the model used in our present simulations as Maxent_{Gain}. (We also evaluated Maxent_{HW} and a third variant, Maxent_{O/E}, on the same training and testing data; results from these simulations, which lead to the same conclusion as those reported below, appear in section 3.2.)

3.1.4. Training corpus

The Maxent_{Gain} model was trained on a set of 1449 productive triliteral Hebrew roots listed in the Even Shoshan Hebrew dictionary (i.e., roots appearing in Hebrew verbs). These roots were represented using a standard set of phonological features of modern Hebrew suggested to us by Outi Bat-El (personal communication, November, 2009)². One additional feature, tongue tip constriction area (Gafos 1999), was added in order to allow for the coding of the unattested phoneme /θ/ (examined in subsequent sections). The entire set of features for native Hebrew phonemes and non-native ones is provided in Table 1.

3.1.5. Generalizations to attested phonemes

We first examined the performance of the model on novel roots consisting entirely of native Hebrew phonemes. The model's harmony scores are presented in Figure 1, alongside their acceptability ratings by Hebrew speakers (data from Berent et al. 2001a, Experiment 3). An inspection of the means suggests that the model's predictions for the attested phonemes closely matched human's performance. Like humans, the model deemed *ssm*-type roots with

²Two differences between this feature set and the one used in Hayes and Wilson's original simulation are noteworthy. First, /j/ is coded as coronal, as Hebrew /l/ alternate with /j/ in Child language (Ben-David 2001). Second, we did not specify /l/ as a lateral, as /l/ and /□/ contrast by their distinct places of articulation. Further simulations showed that the conclusions reported here are robust with respect to various changes in the feature matrix.

initial identical consonants ill-formed. These conclusions are borne out by an ANOVA of the effect of root type (*ssm*-, *smm*-, *psm*-type) on harmony scores. The significant main effect of root type ($F(2, 46)=6.28$; $MSE=2.37$, $p<.004$) in the ANOVA confirmed that the model was sensitive to root structure. Planned comparisons further showed *ssm*-type roots were worse-formed compared to either *smm*- ($t(46)=2.59$, $p<.02$) or *psm*-type ($t(46)=3.39$, $p<.002$) controls, which did not differ from each other ($t(46)<1$, n.s.)³. Thus, within the training space, the model accurately captured the fact that roots with initial identical consonants are ill-formed.

What allowed the model to learn this pattern? Since the original HW08 architecture does not allow constraints containing variables over segments, a fortiori it cannot learn any constraint on segment identity *per se*. Similarly, the model cannot approximate the restriction on segment identity by explicitly banning identical feature matrices, as the model only represents constraints on the co-occurrence of feature matrices, not their relationship (i.e., identity). To the extent that the model captures human performance, we conjectured that the ban on *ssm*-type roots must therefore emerge from restrictions on the co-occurrence of specific feature combinations root-initially. To illustrate this fact, we examined the constraints acquired by one of the ten runs of the model. An inspection of this grammar (see Table 2 for the 30 highest ranked constraints) reveals numerous piecemeal/individual constraints against identical features at the left edge of the root, including bans on identical labials, and identical dorsals. Accordingly, the test root *bbg*, for example, is penalized relative to *bgg* because it violates the highly-ranked constraint $\#[+labial][-\text{continuant}, +labial]$ (see (1)). Note, however, that this constraint does not broadly ban initial identical labials generally. Rather, it narrowly targets only labials that are followed by noncontinuant labials, so sequences such as *mmg* are not penalized by this constraint. Similar piecemeal bans govern the dislike of identical velars in *kkt* (relative to *ktt*, for example), due to the violation of the constraint on non-continuant dorsals root-initially ($\#[-\text{continuant}, +dorsal][-\text{continuant}, +dorsal]$). Once again, however, this narrow ban spares root initial dorsals that are continuant, such as *rrl* which, which in turn, are penalized by another (low-ranked) constraint, ($\#[+sonorant, +dorsal][+sonorant, +dorsal]$). Clearly, root-initial identity is not generally disallowed.

(1) Violations incurred by novel *ssm*- vs. *smm*- type roots

Constraint	Weight	Input				
		bbg	bgg	mmg	kkt	rrl
$\#[+labial][-\text{continuant}, +labial]$	1.89	*				
$\#[-\text{continuant}, +dorsal][-\text{continuant}, +dorsal]$	1.73				*	

The crucial question then is whether such constraints will allow the model to generalize identity restrictions to test items with phonemes that are non-native to Hebrew—the phonemes /w/, /t□/, /d□/, and /θ/.

3.1.6. Generalizations to non-native phonemes

Figure 2 depicts the model's harmony scores for roots with non-native phonemes alongside their acceptability ratings by Hebrew speakers (data from Berent et al. 2002; Experiment 2).

³Similar analyses conducted on human data likewise yielded a significant main effect of root type ($F(2, 46)=36.86$; $MSE=.081$, $p<.001$). Planned comparisons confirmed that *ssm*-type roots were rated significantly lower than either *smm*- ($t(46)=6.46$, $p<.001$) or *psm*-type ($t(46)=8.13$, $p<.0001$) roots, which, in turn, did not reliably differ ($t(46)=1.67$, $p<.11$).

Here, in contrast with the findings obtained with attested phonemes, the model's predictions for non-native phonemes diverge from human behavior. While Hebrew speakers favor roots with final identity over roots with initial identity, the original Maxent_{Gain} model did not reliably differentiate between them. An ANOVA indicated that the main effect of root type was significant ($F(2, 46)=97.89$, $MSE=2.73$, $p<.0001$). However, the dislike of *ssm*- relative to *smm*-type roots was not reliable ($t(46)=1.72$, $p<.10$).

A closer inspection suggests various discrepancies with human behavior (see Figure 3). Humans reliably generalized the *ssm-smm* asymmetry only to consonantal phonemes, but not to nonconsonantal /w/⁴, as Hebrew strictly constrains the co-occurrence of glides (e.g., it disallows identical glides even for the native glide /j/)⁵. Moreover, speakers extended these generalizations to each of the three consonantal phonemes, regardless of whether they fell within the Hebrew feature space (in the case of /t□/, /d□/) or outside it (for /θ/). The Maxent_{Gain} model, by contrast, generalized in a piece-meal fashion. An ANOVA of the harmony scores yielded a reliable root type x phoneme interaction ($F(6, 40)=4.26$, $MSE=1.91$, $p<.003$). Planned comparisons, however, revealed two striking differences relative to human behavior. First, when provided with the nonconsonantal /w/, the model incorrectly favored *wwg*-type roots to their *gww*-type counterparts ($t(40)=2.84$, $p<.008$). Second, though the model successfully generalized within the training space, it failed to extend the generalization beyond it, to the phoneme /θ/ ($t(40)<1$, n.s.).

While the original Maxent_{Gain} model failed to constrain the position of novel identical consonants in the root, it over-penalized any root with non-native identical consonants. Specifically, *smm*-type roots received reliably lower harmony scores compared to *psm* type controls, and the magnitude of the *smm-psm* contrast with non-native phonemes ($\Delta=5.33$) was over sixfold the magnitude of the *ssm-smm* effect ($\Delta=5.33$) and about 15-fold the size of the *smm-psm* difference with native phonemes ($\Delta=.36$). Although humans also penalized roots with two non-native phonemes, the magnitude of the human *smm-psm* effect ($\Delta=.13$) was weaker than the *ssm-smm* asymmetry ($\Delta=.26$) and comparable to the size of the *smm-psm* effect with native phonemes ($\Delta=.14$). The model's dispreference of *smm*-type roots with non-native phonemes must therefore reflect their novelty. And indeed, each such root incurred two violations (one per each non-native segment) of some of the top-ranked constraints in the grammar. Specifically, *xyy*-type roots with identical /θ/ phonemes incurred two violations (one per phoneme) of *[+continuant, -strident]—the most heavily weighted constraints in the grammar; *xyy*-roots with identical /t□/ and /d□/ each incurred two violations of *[-continuant, -anterior] (the second ranked constraint) and those with identical /w/ violated *[-consonantal, +labial] (the eighth ranked constraints). While the ultimate harmony score of a root cannot be ascribed to any single constraint, it nonetheless appears that overall novelty contributed to the harmony of non-native phonemes far more than their root-position. Moreover, the model was unable to generalize the restriction on identical consonants in a consistent fashion to non-native phonemes. Rather than restricting roots with initial identity, the harmony of novel roots was determined by their overlap with the feature-sequences that are attested in the Hebrew lexicon.

3.2. A comparison to other variants of the HW 08 architecture

Why did the Maxent_{Gain} model fail to generalize to novel phonemes? We attribute this limitation to the principled inability of the HW08 architecture to learn relations expressed with variables. Recall, however, that the code used in our simulations differed in some

⁴The 2 (*ssm-smm*) root type x 4 phoneme interaction was also significant in the Human data ($F(3, 20)=3.66$, $MSE=.055$, $p<.03$), and the *ssm-smm* contrast was significant for /t□/ ($t(20)=3.11$, $p<.01$), /d□/ ($t(20)=3.00$, $p<.01$) and /θ/ ($t(20)=2.45$, $p<.03$), but not to the non-consonantal/w/, ($t(20)<1$).

⁵We thank an LI reviewer for pointing out this fact to us.

respects from the published HW08 version ($\text{Maxent}_{\text{HW}}$), so one might worry that these inadequacies result from these specific modifications, rather than limitations that are central to the HW08 architecture.

To rule out this possibility, we submitted the same data to the original code of the published HW08 model (Maxent_{HW}), again performing 10 runs of the model. For further comparison, we also examined 10 runs of another variant of the HW08 model in which constraints were selected using the Observed/Expected criterion (as in HW08, p. 392) but no search heuristics were used (Maxent_{O/E}). Both of these models, however, performed far worse than our Maxent_{Gain} simulations (see Figure 4). The published Maxent_{HW} model was insensitive to the structure of either novel roots with native phonemes (for the *ssm-smm* contrast, $t(46)=1.69$, $p>.09$, n.s.) or non-native ones (for the *ssm-smm* contrast, $t(46)<1$, n.s.).

Similar problems also emerged in the Maxent_{O/E} version. While this model was sensitive to root structure with native phonemes, and correctly rendered *ssm*-type roots less acceptable than *smm*-type controls ($t(46)=3.97$, $p<.003$), when provided with non-native phonemes, the model failed to differentiate between these two root types ($t<1$, n.s.). Thus, the failure to generalize the root-structure constraint to novel phonemes is a systematic problem that is not limited to any particular run, constraint-selection criterion, or search heuristics.

3.3. A revised model

The results presented so far suggest that the original HW08 model and its successors (the $\text{Maxent}_{\text{O/E}}$ and $\text{Maxent}_{\text{Gain}}$) fail to adequately generalize the constraint on Hebrew roots—these models extend the restriction only to roots with native phonemes, but not to non-native ones. We attribute this limitation to the elimination of variables. If our conclusions are correct, and if human phonological grammars do relate among variables, then once the model is revised to support restrictions on segment identity, then its performance should converge with human behavior.

3.3.1. Description of the revised model—To evaluate the role of variables, we next examined the performance of a minimally revised version of the best-performing model, the MaxentGain. This revision was prepared by Colin Wilson to allow constraints that contain variables over segments. Recall that the constraint schemas proposed by Hayes and Wilson are of the form *X, *XY, *XYZ, ..., where X, Y, Z are feature matrices or boundary symbols. For example, the constraint *#[labial] is violated by every instance of a labial segment that occurs at the beginning of the domain demarcated by the boundary symbol ‘#’. The revision created additional schemas by allowing variables bound by feature matrices to appear in the constraints: *X_{i-1}, *X_{i-1}Y, *XY_{i-1}, etc. The constraint *X_{i-1} is violated once for every instance of a segment in the set denoted by X that is immediately followed by an identical segment. The constraint *X_{i-1}Y (or *YX_{i-1}) is evaluated similarly, except that only repeated instances of X in the context __Y (or Y__) incur violations.

To bring out the difference between feature matrices and variables bound by them, consider the constraints $*\#[+labial][+labial]$ and $*\#[+labial]_i-{}_i$. The first constraint is violated by any sequence of two labial segments at the beginning of the domain, regardless of whether the segments are identical (e.g., *bbd*, *bmd*). The second constraint is violated only by identical labial sequences (e.g., *bbd* but not *bmd*). Since the inventory of feature matrices includes $[+segment]$ — denoting the set of all segments — it is also possible to state a constraint $*\#[+segment]_i-{}_i$, which is violated by sequences of identical segments at the beginning of the domain, and a constraint $*[+segment]_i-{}_i\#$, which is violated by identical sequences at the end of the domain.

At a more technical level, the revision works as follows. Given a finite set Σ of segments and a finite set Σ of features that assign values to the segments, the set of feature matrices is determined in the way proposed by HW08. Note that the features and segments could be specified innately or learned by an inductive mechanism not specified here (see the references in Hayes and Wilson 2008, p. 390, note 6); similarly, non-native segments of the kind studied in the Berent et al. (2002) experiments could be present in Σ during the learning of the native language or added to Σ as the result of exposure to non-native input.⁶ For any feature matrix X , the constraint $*X_{i-i}$ as defined above is a function from strings over Σ to violation counts. The constraint can be equivalently represented as a disjunction of the form $*((x_0 x_0)l(x_1 x_1)l...l(x_n x_n))$, where $\{x_0, x_1, \dots, x_n\}$ is the subset of Σ denoted by X . This is the form that is most useful for compiling the constraint to a weighted finite-state machine and combining it with other constraints into a machine that represents the entire phonotactic grammar. Parallel remarks apply to the other schema containing variables.

This revision of the $\text{Maxent}_{\text{Gain}}$ model qualifies as minimal because everything other than the constraint schema — the way in which harmonies and probabilities are assigned to phonological representations, the objective function that determines the weights of the constraints, the method of constraint selection, and the implementation of a grammar as a weighted finite state machine — remains unchanged.

3.3.2. Results—The set of top 30 constraints acquired by the revised model is presented in Table 3. Unlike the original the $\text{Maxent}_{\text{Gain}}$ model implementation, the revised model acquired a restriction that specifically bans identical consonants root-initially, namely a constraint of the type $*\#[+\text{segment}]_i-i$, where $[\text{+segment}]$ is the class containing all segments.

An inspection of the harmony means suggests that the performance of the revised model has improved considerably (see Figure 5). Unlike the original simulation, the revised model correctly banned identical phonemes root-initially for either native and non-native phonemes. These conclusions are supported by one-way ANOVAs of the harmony scores for novel roots, performed separately for roots with native and non-native phonemes. Root structure reliably modulated the predicted harmony scores for both native ($F(2, 46)=24.11$, $\text{MSE}=2.29$, $p<.0001$) and non-native ($F(2, 46)=150.23$, $\text{MSE}=2.41$, $p<.0001$) phonemes. Planned comparison of roots with native phonemes confirmed that *ssm*-type roots were still correctly treated as worse-formed than *smm*- ($t(46)=5.69$, $p<.0001$) and *psm*-controls ($t(46)=6.29$, $p<.0001$). Crucially, however, the ban on *ssm*-type roots now generalized reliably to roots with non-native phonemes: such *ssm*-type roots were reliably worse-formed compared to both *smm*- ($t(46)=5.36$, $p<.0001$) and *psm*-type ($t(46)=16.94$, $p<.0001$) controls.

To further assure that the restriction on root structure generalizes to any novel consonant, we next probed for *ssm-smm* asymmetry for each of the four non-native phonemes. Recall that Hebrew speakers generalize the restriction on identical consonants to any novel consonantal segment, for $/t_\square/$, $/d_\square/$, and even $/\theta/$ — a segment whose TTCA value is unattested in Hebrew, but not to the non-consonantal $/w/$. The performance of the revised model captured this pattern quite well (see Figure 6). A 2 root type (*ssm-smm*) x 4 phoneme ANOVA yielded a reliable interaction ($F(3, 20)=4.90$, $\text{MSE}=1.76$, $p<.02$). Planned comparison confirmed that, like humans, the model generalized the *ssm-smm* asymmetry for each of the three consonantal segments—for $/t_\square/$ ($t(20)=4.45$, $p<.0003$), $/d_\square/$ ($t(20)=5.39$, $p<.00004$), and $/\theta/$ ($t(20)=2.97$, $p<.008$), but not to the non-consonantal $/w/$ ($t(20)<1$, n.s.). In accord

⁶Any analysis of the experimental results must, like ours, take into account the finding that the non-native segments were perceived and rated as such (rather than being completely assimilated to native phonemes).

with human behavior, the revised model banned the reduplication of the nonconsonantal /w/, but allowed all other identical non-native consonantal phonemes to occur root-finally. Thus, the revised model was able to generalize the restriction on root-structure to novel phonemes and its performance closely matched human behavior.

4. Conclusions

Numerous statistical models have captured diverse phonotactic phenomena using mechanisms that track the co-occurrence of features in linguistic experience. The success of such models might prima facie appear to suggest that the computational machinery of phonological grammars is fundamentally limited, devoid of the capacity to relate variables and exhibit unbounded generalizations. Our investigation addressed this possibility by systematically examining the scope of the generalizations in one influential architecture—the architecture of the HW08 model.

Although the original HW08 architecture is remarkably effective at learning various phonotactic phenomena, we found that its failure to learn constraints on the relations among variables ultimately hinders its ability to capture the performance of human learners. This limitation of the original model is not immediately apparent when tested on items consisting of attested phonemes as such generalizations can be approximated from the co-occurrence of features in the lexicon. And indeed, two of the three variants of the HW08 architecture evaluated in our investigation (the Maxent_{Gain} and Maxent_{O/E}, but not the original Maxent_{HW}) generalized to native phonemes adequately. But once these models are challenged with novel phonemes that fall beyond the training space, their failure to generalize becomes evident. Unlike these instantiations of the HW08 proposal, however, humans can generalize the identity function across the board. The discrepancy between human generalizations and those exhibited by the original HW08 architecture, on the one hand, and the strong convergence with the performance of a revised model that implements variables, on the other, strongly supports the view that human phonological grammars include relations among variables.

These conclusions obviously do not undermine the principled adequacy of inductive statistical learning models. In fact, our results show that such models can correctly capture the Hebrew facts once they are equipped with the capacity to encode algebraic restrictions on identity. Such algebraic mechanisms alone, however, might not be sufficient for a full account of phonotactics. First, merely implementing the capacity to encode identity restrictions does not guarantee that such restrictions are effectively acquired. Follow-up simulations with several variants of the HW08 architecture suggest that the ability of a model to correctly restrict segment identity also depends on various additional aspects of its architecture, including the choice of the constraint selection algorithm⁷. Moreover, a large literature suggests the possibility of substantive constraints on human phonotactics. So while the capacity to encode variables and operate over variables might be necessary to capture human phonological generalizations, it might not be sufficient.

Although our results demonstrate that phonological learners extend generalizations to novel members of a phonological class, even when these instances fall beyond the training space of the relevant constraint, these findings do not speak to the actual scope of those classes—whether phonological categories (e.g., “any segment”, “any feature”) are inherently limited to a specific (perhaps even small) number of instances, or unbounded, akin to syntactic categories such as “noun” or “verb”. While some classes such as “any feature” or “any segment” could be a priori confined, this possibility remains controversial (e.g., Lindblom 1998, Port and Leary 2005). Moreover, many other phonological categories are not restricted in such fashion, and they are routinely called by grammatical constraints,

including identity restrictions. One such constraint, Contiguity, requires reduplicants to be a contiguous substring of the base (McCarthy and Prince 1993). To use an example from Hebrew, *vered* (rose) reduplicates as *vrad.rad*, not *vrad.vad*, as the reduplicant *vad* does not form a contiguous substring of the base (see also Bat-El 1996). But unlike segments and features, “base” and “reduplicant” are open-ended categories that cannot plausibly be restricted to a finite set, and experimental results suggest that speakers productively extend the constraints on these classes to novel instances (Berent & Vaknin, unpublished data), suggesting that the relevant classes are not bounded in scope.

While unbounded phonological classes await further evidence, the existing findings clearly demonstrate that some of the core machinery necessary for unbounded productivity—algebraic relations between variables—forms part of phonological grammars. Phonology may well be different from syntax in many important ways (Bromberger and Halle 1989), but at least in this respect, phonology seems much like syntax. Models of phonology that lack the means to represent abstract relationships between variables are, whatever their other virtues, unlikely to be sufficient.

Acknowledgments

We thank Outi Bat-El, Bruce Hayes, Andrew Nevins and John Frampton for discussions of this research. This research was supported by NIDCD grant DC003277.

References

- Alderete, John; Beckman, Jill; Benua, Laura; Gnanadesikan, Amalia; McCarthy, John; Urbanczyk, Suzanne. Reduplication with fixed segmentism. *Linguistic Inquiry*. 1999; 30:327–364.
- Bat-El, Outi. Stem modification and cluster transfer in modern Hebrew. *Natural Language and Linguistic Theory*. 1994; 12
- Bat-El, Outi. Selecting the best of the worst: The grammar of Hebrew blends. *Phonology*. 1996; 13:283–328.
- Bat-El, Outi. Consonant identity and consonant copy: the segmental and prosodic structure of Hebrew reduplication. *Linguistic Inquiry*. 2006; 37:179–210.
- Ben-David, Avivit. Ph.D. dissertation. Tel-Aviv University; 2001. *Language Acquisition and Phonological Theory: Universal and Variable Processes Across Children and Across Languages*.
- Berent, Iris; Pinker, Steven; Shimron, Joseph. Default nominal inflection in Hebrew: Evidence for mental variables. *Cognition*. 1999; 72:1–44. [PubMed: 10520563]
- Berent, Iris; Everett, Daniel L.; Shimron, Joseph. Do phonological representations specify variables? Evidence from the obligatory contour principle. *Cognitive Psychology*. 2001a; 42:1–60. [PubMed: 11161416]
- Berent, Iris; Shimron, Joseph; Vaknin, Vered. Phonological constraints on reading: Evidence from the Obligatory Contour Principle. *Journal of Memory and Language*. 2001b; 44:644–665.
- Berent, Iris; Marcus, Gary F.; Shimron, Joseph; Gafos, Adamantios. I. The scope of linguistic generalizations: evidence from Hebrew word formation. *Cognition*. 2002; 83:113–139. [PubMed: 11869721]
- Berent, Iris; Shimron, Joseph. Co-occurrence restrictions on identical consonants in the Hebrew lexicon : Are they due to similarity? *Journal of Linguistics*. 2003; 39:31–55.
- Berent, Iris; Vaknin, Vered; Shimron, Joseph. Does a theory of language need a grammar? Evidence from Hebrew root structure. *Brain and Language*. 2004; 90:170–182. [PubMed: 15172535]
- Berent, Iris; Pinker, Steven. The dislike of regular plurals in compounds: Phonological familiarity or morphological constraint? *The Mental Lexicon*. 2007; 2:129–181.
- Berent, Iris; Vaknin, Vered; Marcus, F. Gary. Roots, stems, and the universality of lexical representations: evidence from Hebrew. *Cognition*. 2007; 104:254–286. [PubMed: 16890213]
- Bromberger, Sylvain; Halle, Morris. Why phonology is different. *Linguistic Inquiry*. 1989; 20:51–70.

- Chomsky, Noam. Syntactic structures. Mouton; Gravenhage: 1957.
- Chomsky, Noam. Language and mind. Harcourt Brace Jovanovich; New York: 1972.
- Chomsky, Noam. Rules and representations. Behavioral and Brain Sciences. 1980; 3:1–61.
- Chomsky, Noam. Three factors in language design. Linguistic Inquiry. 2005; 36:1–22.
- Clements, George N.; Keyser, Samuel Jay. CV phonology. MIT Press; Cambridge, MA: 1983. Monograph Series (No. 9)
- Coetzee, Andries. Grammaticality and ungrammaticality in phonology. Language. 2008; 84:218–257.
- Coetzee, Andries, W.; Pater, Joe. Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. Natural Language and Linguistic Theory. 2008; 26:289–337.
- Coleman, John; Pierrehumbert, Janet. Stochastic phonological grammars and acceptability. In: Coleman, John, editor. Third meeting of the ACL special interest group in computational phonology: Proceedings of the workshop; East Stroudsburg, PA: Association for computational linguistics; 1997.
- Frisch, Stefan. A.; Zawaydeh, Bushra A. The psychological reality of OCP-place in Arabic. Language. 2001; 77:91–106.
- Gafos, Adamantios, I. The articulatory basis of locality in phonology. Garland publishers; New York: 1999.
- Gafos, Adamantios, I. Greenberg's asymmetry in Arabic: A consequence of stems in paradigms. Language. 2003; 79:317–355.
- Greenberg, Joseph, H. The patterning of morphemes in Semitic. Word. 1950; 6:162–181.
- Hayes, Bruce; Wilson, Colin. A maximum entropy model of phonotactics and phonotactic learning. Linguistic Inquiry. 2008; 39:379–440.
- Inkelas, Sharon; Zoll, Cheryl. Reduplication: Doubling in Morphology: Cambridge studies in linguistics. Cambridge University Press; Cambridge, UK: 2005.
- Inkelas, Sharon. The dual theory of reduplication. Linguistics. 2008; 46:351–401.
- Ito, Junko; Mester, Armin. The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. Linguistic Inquiry. 1986; 17:49–73.
- Kim, John J.; Marcus, Gary F.; Pinker, Steven; Hollander, Michelle; Coppola, Marie. Sensitivity of children's inflection to grammatical structure. Journal of Child Language. 1994; 21:173–209. [PubMed: 8006091]
- Ladd, Robert. Intonational phrasing: The case for recursive prosodic structure. Phonology Yearbook. 1986; 3:311–340.
- Leben, William. Suprasegmental phonology. MIT press; Cambridge, MA: 1973.
- Lindblom, Björn. Systemic constraints and adaptive changes in the formation of sound structure. In: Hurford, James R.; Studdert-Kennedy, Michael; Knight, Chris, editors. Approaches to the Evolution of Language - Social and Cognitive Bases. Cambridge University Press; Cambridge, UK: 1998. p. 242–263.
- Marcus GF, Vijayan S, Bandi Rao S, Vishton PM. Rule learning by seven-month-old infants [Jan 1]. Science. 1999; 283:77–80. [PubMed: 9872745]
- Marcus, Gary. The algebraic mind: Integrating connectionism and cognitive science. MIT press; Cambridge: 2001.
- Marcus, Gary F.; Brinkmann, Ursula; Clahsen, Harald; Wiese, Richard; Pinker, Steven. German inflection: the exception that proves the rule. Cognitive psychology. 1995; 29:189–256. [PubMed: 8556846]
- Marcus, Gary, F. Rethinking eliminative connectionism. Cognitive Psychology. 1998; 37:243–282. [PubMed: 9892549]
- McCarthy, John J. A prosodic theory of nonconcatenative morphology. Linguistic Inquiry. 1981; 12:373–418.
- McCarthy, John, J.; Prince, Alan. Prosodic Morphology I: Constraint Interaction and Satisfaction. Report no. RuCCS-TR-3. Rutgers University Center for Cognitive Science; New Brunswick, NJ: 1993.

- McCarthy, John, J. The phonetics and phonology of Semitic pharyngeals. In: Keating, Patricia, editor. *Papers in laboratory phonology III*. Cambridge University Press, Cambridge; Cambridge: 1994. p. 191-283.
- McCarthy, John J.; Prince, Alan. Prosodic morphology. In: Goldsmith, John A., editor. *Phonological theory*. Basil Blackwell; Oxford: 1995. p. 318-366.
- Moravcsik, Edith. Greenberg, Joseph H. Reduplicative constructions. *Universals of human language: Word structure*. 1978:297-334.
- Nevins, Andrew. Two case studies in phonological universals: A view from artificial grammars. *Biolinguistics*. 2010; 4:218-233.
- Pinker, Steven; Prince, Allan. On language and connectionism: Analysis of parallel distributed processing model of language acquisition. *Cognition*. 1988; 28:73-193. [PubMed: 2450717]
- Pinker, Steven. Rules of language. *Science*. 1991; 253:530-535. [PubMed: 1857983]
- Port, Robert F.; Leary, Adam P. Against Formal Phonology. *Language*. 2005; 81:927-964.
- Prasada, Sandeep; Pinker, Steven. Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes*. 1993; 8:1-55.
- Prince, Alan; Smolensky, Paul. *Optimality theory: Constraint interaction in generative grammar*. Blackwell Pub.; Malden, MA: 1993/2004.
- Rose, Sharon; Walker, Rachel. A typology of consonant agreement as correspondence. *Language*. 2004; 80:475-531.
- Smolensky, Paul; Legendre, Geraldine. Principles of integrated connectionist/symbolic cognitive architecture. In: Smolensky, Paul; Legendre, Geraldine, editors. *The harmonic mind: From neural computation to Optimality-theoretic grammar*. MIT Press; Cambridge, MA: 2006. p. 63-97.
- Suzuki, Keiichiro. A typological investigation of dissimilation. University of Arizona; 1998.
- Toro, Juan M.; Nespor, Marina; Mehler, Jacques; Bonatti, Luca L. Finding words and rules in a speech stream: functional differences between vowels and consonants. *Psychological Science*. 2008; 19:137-144. [PubMed: 18271861]
- Ussishkin, Adam. The inadequacy of the consonantal root: Modern Hebrew denominal verbs and Output-Output correspondence. *Phonology*. 1999; 16:441-442.
- Yip, Moira. The Obligatory Contour Principle and phonological rules: A loss of identity. *Linguistic Inquiry*. 1988; 19:65-100.

Figure 1.

Harmony scores generated by the Maxent_{Gain} model for roots with native Hebrew phonemes and the corresponding acceptability ratings by Hebrew speakers. Error bars reflect confidence intervals constructed for the difference between means.

Figure 2.

Harmony scores generated by the Maxent_{Gain} model for roots with non-native phonemes and the corresponding acceptability ratings by Hebrew speakers. Error bars reflect confidence intervals constructed for the difference between the means.

Figure 3.

Harmony scores generated by the original Maxent_{Gain} model for roots with non-native phonemes and the corresponding acceptability ratings by Hebrew speakers. Error bars reflect confidence intervals constructed for the difference between the means.

Figure 4.

Harmony scores generated by three variants of the HW08 architecture for roots with native and non-native phonemes and the corresponding acceptability ratings by Hebrew speakers. Error bars reflect confidence intervals constructed for the difference between the means.

Figure 5.

The harmony scores generated by the revised Maxent_{Gain} model for native and non-native phonemes. Error bars reflect confidence intervals constructed for the difference between the means.

Figure 6.
The harmony scores generated by the revised MaxentGain model for nonnative phoneme and their acceptability by Hebrew speakers.

Table 1

Feature chart for attested Hebrew phonemes and non-native phonemes

	IPA	Root Boundary	consonantal	sonorant	continuant	nasal	voice	strident	labial	coronal	anterior	dorsal	spread	TTCA.narrow	TTCA.mid	TTCA.wide
Native	#	+	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	p	-	+	-	-	0	-	0	+	0	0	0	0	0	0	0
	b	-	+	-	-	0	+	0	+	0	0	0	0	0	0	0
	t	-	+	-	-	0	-	-	0	+	0	0	0	0	0	0
	d	-	+	-	-	0	+	-	0	+	0	0	0	0	0	0
	k	-	+	-	-	0	-	0	0	0	0	+	0	0	0	0
	g	-	+	-	-	0	+	0	0	0	0	+	0	0	0	0
	m	-	+	+	0	+	0	0	+	0	0	0	0	0	0	0
	n	-	+	+	0	+	0	0	0	+	0	0	0	0	0	0
	v	-	+	-	+	0	+	0	+	0	0	0	0	0	0	0
	s	-	+	-	+	0	-	+	0	+	+	0	0	+	0	0
	z	-	+	-	+	0	+	+	0	+	+	0	0	+	0	0
	ʃ	-	+	-	+	0	-	+	0	+	-	0	0	0	+	0
	ʒ	-	+	-	+	0	-	0	0	0	0	+	0	0	0	0
	□	-	+	-	-	0	-	+	0	+	0	0	0	+	0	0
	j	-	-	+	0	0	0	0	0	+	0	0	0	0	0	0
	h	-	+	-	+	0	-	0	0	0	0	0	+	0	0	0
	l	-	+	+	0	0	0	0	0	+	0	0	0	0	0	0
	□	-	+	+	0	0	0	0	0	0	0	+	0	0	0	0
	ʔ	-	+	-	-	0	-	0	0	0	0	0	-	0	0	0
Non-native	θ	-	+	-	+	0	-	-	0	+	+	0	0	0	0	+
	tʃ	-	+	-	-	0	-	+	0	+	-	0	0	0	0	0
	dʒ	-	+	-	-	0	+	+	0	+	-	0	0	0	0	0
	w	-	-	+	0	0	0	0	+	0	0	0	0	0	0	0

Table 2

The 30 highest-ranked constraints acquired by the Maxent_{Gain} model without identity constraints. (Note: # indicates root boundary).

Constraint	Weight
*[+continuant,−strident]	3.274
*[−continuant,−anterior]	2.898
*#[+continuant,+labial]	2.48
*[+consonantal,+sonorant,+coronal][+sonorant,+dorsal]	2.086
*#[−continuant,+dorsal][−continuant,+dorsal]	1.995
*#[+consonantal,+sonorant,+coronal][+sonorant]	1.955
*[+labial][][−sonorant,+labial]	1.934
*[−consonantal,+labial]	1.916
*#[+labial][−continuant,+labial]	1.899
*[+continuant,+labial]#	1.861
*[][+continuant,+labial]	1.861
*[+voice,+dorsal][−continuant,−voice,+coronal]	1.835
*[−spread][−spread]	1.814
*[−continuant,+labial][+sonorant,+labial]	1.796
*[−spread][][−spread]	1.775
*[−spread][−sonorant][+continuant,+dorsal]	1.771
*[+TTCA.narrow][+consonantal][−voice,+strident]	1.753
*[−consonantal,+labial][]	1.75
*[−consonantal]#	1.731
*[][][−consonantal]	1.731
*[−continuant,+dorsal][][−continuant,+dorsal]	1.729
*[+TTCA.narrow][−anterior]	1.717
*[][+sonorant,+labial][−sonorant,+labial]	1.688
*[+voice,+coronal][−continuant,−voice,+coronal]	1.672
*[+continuant,+dorsal][−sonorant,+dorsal][+coronal]	1.595
*[+sonorant,+dorsal][+consonantal,+sonorant,+coronal]	1.567
*[][−continuant,−anterior]	1.565
*[][−spread][+continuant,+dorsal]	1.554
*[−continuant,−anterior][]	1.551
*[−voice,+labial][−continuant,+voice,+labial]	1.523

Table 3

The 30 highest-ranked constraints acquired by the Maxent_{Gain} model with identity constraints. (Note: # indicates root boundary).

Constraint	weight
*[+continuant,−strident]	3.192
*[−continuant,−anterior]	2.927
*#[+continuant,+labial]	2.358
*[+consonantal,+sonorant,+coronal][+sonorant,+dorsal]	2.093
*[−consonantal,+labial]	2.048
*[+labial][][−sonorant,+labial]	1.989
*[+voice,+coronal][−continuant,−voice,+coronal]	1.964
*#[+consonantal,+sonorant,+coronal][+sonorant]	1.956
*[−continuant,+labial][+sonorant,+labial]	1.923
*[+continuant,+labial]#	1.873
*[][+continuant,+labial]	1.873
*[−consonantal,+labial][]	1.858
*[+voice,+dorsal][−continuant,−voice,+coronal]	1.843
*[+TTCA.narrow][−anterior]	1.84
*[+sonorant,+labial][−continuant,+labial]	1.836
*#[+segment] _i a _i	1.828
*[−spread][][−spread]	1.786
*[+TTCA.narrow][+consonantal][+strident]	1.768
*[−consonantal]#	1.766
*[][−consonantal]	1.766
*[−continuant,+dorsal][][−continuant,+dorsal]	1.741
*[−sonorant,+dorsal][−continuant,+dorsal][+coronal]	1.725
*[+sonorant,+dorsal][+consonantal,+sonorant,+coronal]	1.721
*[][−anterior][+TTCA.narrow]	1.671
*[−spread][−spread]	1.643
*[+voice,+dorsal][][−continuant,−voice,+coronal]	1.59
*[−continuant,−anterior][]	1.59
*[−voice,+labial][−continuant,+voice,+labial]	1.577
*[−continuant,+voice,+labial][−voice,+labial]	1.56
*[][−continuant,−anterior]	1.556