NATURAL AND UNNATURAL CONSTRAINTS IN HUNGARIAN VOWEL HARMONY

Bruce Hayes
University of California, Los Angeles

Kie Zuraw
University of California, Los Angeles

Péter Siptár
Eötvös Loránd University Budapest and *Hungarian Academy of Sciences*

Zsuzsa Londe
University of California, Los Angeles

Phonological constraints can, in principle, be classified according to whether they are natural (founded in principles of universal grammar (UG)) or unnatural (arbitrary, learned inductively from the language data). Recent work has used this distinction as the basis for arguments about the role of UG in learning. Some languages have phonological patterns that arguably reflect unnatural constraints. With experimental testing, one can assess whether such patterns are actually learned by native speakers. Becker, Ketrez, and Nevins (2007), testing speakers of Turkish, suggest that they do indeed go unlearned. They interpret this result with a strong UG position: humans are unable to learn data patterns not backed by UG principles.

This article pursues the same research line, locating similarly unnatural data patterns in the vowel harmony system of Hungarian, such as the tendency (among certain stem types) for a final bilabial stop to favor front harmony. Our own test leads to the opposite conclusion of Becker and colleagues': Hungarians evidently do learn the unnatural patterns.

To conclude we consider a bias account—that speakers are able to learn unnatural environments, but devalue them relative to natural ones. We outline a method for testing the strength of constraints as learned by speakers against the strength of the corresponding patterns in the lexicon, and show that it offers tentative support for the hypothesis that unnatural constraints are disfavored by language learners.*

Keywords: Hungarian, vowel harmony, naturalness, wug test, variation

1. Introduction: the ug problem. Theoretical study in linguistics has long sought to discover principles of universal grammar (UG)—aspects of language that are grounded in human nature, are (presumably) genetically coded at some level, and are claimed to assist children in language acquisition. In its narrow sense, the term universal grammar includes only characteristics specific to the language faculty; in its broad sense it can include any innate properties of humans (cognitive, phonetic, etc.) that determine aspects of language. Our interest here lies not in distinguishing between these two, but rather in finding evidence concerning innate properties of language in general.

Historically, work in generative grammar has pursued UG with an analytic/typological approach. Grammarians submit individual languages to formal analysis under some particular theoretical framework incorporating principles of UG. When multiple languages are analyzed, the theory is put to the test, since it must provide apparatus sufficient for the analysis of every existing language, yet make falsifiable predictions about phenomena that should not occur. Examples of work in phonology that has

* We would like to thank Adam Albright, Michael Becker, Robert Daland, Matt Goldrick, Sarah Finley, Edward Flemming, Shigeto Kawahara, John McCarthy, Michael Mitchell, Joe Pater, Colin Wilson, Kristine Yu, and members of audiences at the University of Massachusetts–Amherst, the Seoul International Conference on Linguistics, Northwestern University, and the University of California, Los Angeles for helpful advice. This research was supported by a grant from the Committee on Research of the UCLA Academic Senate.

Web-based resources for this article (data, simulations, links to software) may be found at http://www.linguistics.ucla.edu/people/hayes/hungarianvh/.

pursued this goal include parametric approaches to metrical stress theory (e.g. Halle & Vergnaud 1987, Hayes 1995), or studies that compute and assess a factorial typology (e.g. Kaun 1995, Alderete 1998, Kager & Elenbaas 1999, Gordon 2002) in the framework of optimality theory (Prince & Smolensky 2004 [1993]:Ch. 3).

Although the analytic/typological approach has been very fruitful, there is reason to believe that at least in phonology, it is unlikely to provide final answers to questions of UG. The reason is that the languages we have available for analysis have in most cases evolved from earlier diachronic stages, and it is a strong possibility that the channels of change themselves, notably phonetic variation and reinterpretation by new generations of speakers, have a strong influence on language typology. Moreton (2008) particularly emphasizes the role of 'channel bias', in which the possibility of a phonetic precursor will strongly influence the observed crosslinguistic frequency of a particular phonological pattern. An example can be seen in cases where the normal phonetic lowering of pitch after voiced obstruents evolves into a categorical phonological pattern ('tonogenesis'; see e.g. Hombert et al. 1979). It is a logical possibility that phonological typology is largely governed by whether such precursors exist, rather than by innate limits on the language faculty.

The most forceful statements of the phonetic-precursor idea attribute all phonological patterning to channel bias (e.g. Ohala 1981, Blevins 2004). There are also mixed views, in which a role is assigned to UG, perhaps in regulating the way that phonetic variation is restructured into phonological change (Baroni 2001, Myers 2002, Koo & Cole 2006, Wilson 2006, Finley 2008, Kawahara 2008, Moreton 2009, etc.). Under any of these views, however, research based solely on analysis and typology will not suffice to establish UG theories on a solid basis.

2. Searching for UG with experiments. A natural consequence is that many phonologists interested in questions of UG have become involved in experimental work, which could in principle be made to bear on these questions free of the confounding factor of diachronic origin. 'UG experiments' have been conducted in phonology at least since the 1970s, but with a considerable upsurge in recent times. A partial list of references would include Schane et al. 1974, Pertz & Bever 1975, Bailey et al. 1999, Jusczyk et al. 2002, Albright & Hayes 2003, Jusczyk et al. 2003, Pater & Tessier 2003, 2006, Pycha et al. 2003, Saffran & Thiessen 2003, Wilson 2003a, 2006, Pertsova 2004, Gilkerson 2005, Seidl & Buckley 2005, Kawahara 2006, Koo & Cole 2006, Peperkamp et al. 2006, Pycha et al. 2006, Zhang & Lai 2006, Albright 2007, Becker et al. 2007, Berent et al. 2007, Finley & Badecker 2007, Graff 2007, Peperkamp & Dupoux 2007, Thatte 2007, Zhang et al. 2007, Zuraw 2007, Berent et al. 2008, and Finley 2008.

The study reported here is also intended as a UG experiment. As background to our work, we first offer a survey of the kinds of experiments that have been done.

2.1. Poverty-of-the-stimulus experiments. Many UG experiments in phonology rely on a 'poverty of the stimulus' argument (Chomsky 1980, Pullum & Scholtz 2002). The idea is that an experiment will reveal something about UG if it shows that the participants exhibit consistent behavior that could not be traced either to their own language background or to the information made available to them during the experiment.

For instance, it is possible to expose experimental participants to miniature artificial languages, then test them on additional data to see what generalizations they inferred. If participants consistently generalize in some particular directions and not in others (as in Wilson 2006), a plausible inference is that some UG principle is guiding the process of generalization—the data alone were too impoverished to have governed the participants' responses. The same strategy can be followed using the native language of the participants, by constructing nonce forms that test what generalizations they formed in the course of language acquisition (Zuraw 2007).

A less direct form of a poverty-of-the-stimulus experiment (e.g. Wilson 2003a, Holt et al. 2004, Carpenter 2005, Moreton 2008) uses more than one artificial language, comparing how easily experimental participants learn them. The poverty-of-the-stimulus argument here is based on the premise that if one such language is learnable and another is not, then it must be a principle of UG that is making the difference.

2.2. UG AS ABSOLUTE LIMIT VS. UG AS BIAS. Such UG interpretations come in strong and weak versions. A strong-UG interpretation is that any grammatical rule or constraint not countenanced under UG is simply unlearnable; such a claim might be based, for instance, on the universal constraint hypothesis adopted in many versions of optimality theory (Tesar & Smolensky 1998, 2000, Prince & Smolensky 2004 [1993]). Alternatively, theorists have suggested a BIAS interpretation of UG (Ohala 1974, Wilson 2006, Zuraw 2007, Kawahara 2008, Moreton 2008): UG makes it easier to learn certain data patterns, perhaps by directing the learner's attention in particular ways; but people do have the ability to learn phonology inductively, accessing data patterns outside the scope of UG.

For phonological constraints that have some plausible foundation in UG, we employ the term NATURAL, acknowledging that this word has other uses in phonology as well. At present, it is impossible to distinguish natural from unnatural constraints with certainty, but there is often a sufficiently clear picture from typology to provide a reasonable basis for proceeding; see §§6, 7 below.

2.3. SURFEIT-OF-THE-STIMULUS EXPERIMENTS. The flip side of the poverty-of-the-stimulus idea is the 'surfeit of the stimulus' (term from Becker et al. 2007). Here, we suppose that the data to which the language learner is exposed contain patterns that are clear enough to the linguist (who sees the system 'from the outside'), but are not characterizable under the theory of UG. In a strong-UG interpretation, such patterns are predicted to be simply unlearnable; under a bias interpretation, they will be harder to learn. Comparative artificial-language experiments, mentioned above, implicitly involve a surfeit-of-the-stimulus argument: the core argument is that one of the two compared artificial languages is harder to learn because it has no UG backing. Substantial differences in learnability are similarly explained under a bias interpretation of UG.

Somewhat surprisingly, it turns out to be possible to carry out a surfeit-of-the-stimulus experiment using the native language of the experimental participants. The lexicon, whose content is to some degree arbitrary, may by sheerest accident include data patterns that have no basis in UG. Such patterns could be carried down through history in the inherited lexicon, with the speakers of each generation oblivious to their presence. Hayes and Londe (2006:§4.3) point out a data pattern from Hungarian that, according to their experimental findings, goes unnoticed by speakers of this language, and offer a surfeit-of-the-stimulus explanation. A more extensive example from Turkish is discussed below.

To summarize: experiments testing principles of UG can be classified according to whether they use artificial languages or test speakers' productive extensions of their own language; and whether the outcome is interpreted on the basis of a poverty-of-the-stimulus argument, a surfeit-of-the-stimulus argument, or some combination of the two.

The experiment we report here is a real-language experiment (Hungarian) that tests for knowledge of unnatural constraints. We find evidence that Hungarian speakers are in fact able to apprehend such constraints in the data they encounter during acquisition, but this is not the whole story. Detailed examination of the experimental results with computational modeling indicates that the speakers may UNDERVALUE unnatural patterns relative to the strength with which they are manifested in the lexicon, a possible UG bias effect.

In what follows, we introduce the research that forms the direct ancestry of our own efforts, cover the crucial aspects of Hungarian, outline our experiment and its results, and then interpret the experimental results with modeling and give our conclusions.

3. Two previous studies.

3.1. Frequency matching in phonology: Ernestus & Baayen 2003. Our research is based on a frequency-matching ability that is evidently possessed by human speakers. As an example of this ability we review the findings of Ernestus & Baayen 2003.

Ernestus and Baayen's study investigates the process of FINAL DEVOICING in Dutch (1a), which is seen in data pairs like 1b.

(1) a. Dutch final devoicing
$[-\text{sonorant}] \rightarrow [-\text{voice}]\ /\ \_\_\ ]_{\text{word}}$

b. Examples

BEFORE VOWEL		FINAL	
[vɛr'ʋɛid-ən]	'to widen'	[vɛr'ʋɛit]	'widen'
[vɛr'ʋɛit-ən]	'to reproach'	[vɛr'ʋɛit]	'reproach'

Final devoicing is a standard textbook process in phonology, and the usual analysis sets up underlying voiced and voiceless obstruents (/vɛrʋɛid/, /vɛrʋɛit/), devoicing the voiced ones when they occur in final position: /vɛr'ʋɛid/ → [vɛr'ʋɛit]. In Dutch, there are many pairs like [vɛr'ʋɛid-ən] ~ [vɛr'ʋɛit] that are related by final devoicing, as well as many nonalternating pairs like [vɛr'ʋɛit-ən] ~ [vɛr'ʋɛit].

Ernestus and Baayen extended the study of this process by examining the distribution of alternating and nonalternating forms in the Dutch lexicon. They found that the alternation patterns for final devoicing are unevenly distributed according to factors like the place and manner of articulation of the stem-final consonant. For instance, Dutch has many nonalternating pairs like [p#] ~ [pV] (that is, [p] word-finally, [p] before a vowel), but few instances of [p#] ~ [bV]. In the case of the labial fricatives, the statistics go the other way, with few instances of [f#] ~ [fV], many of [f#] ~ [vV]. The idea behind Ernestus and Baayen's study was to see if Dutch speakers are tacitly aware of these statistical regularities. This was tested by checking if speakers use them when they guess the suffixed form from the isolation form for novel stems, that is, in a 'wug' test (Berko 1958). Since the Dutch lexicon has many cases of [p#] ~ [pV] and few of [p#] ~ [bV], one might expect participants who heard [ɪk dɑup] 'I *daup*' to reply with *daupen* rather than *dauben* when queried for the infinitive,[1] but they should provide an infinitive [v] when probed with [ɪk tɑf] 'I *taff*', again reflecting the lexicon.

In fact, the results were more interesting. Rather than always guessing the most frequent option, participants gave mixed responses, and these responses taken in the

[1] We give *daupen* and *dauben* to simplify the presentation; in fact the forms elicited were past tense *daupte* and *daubde*; these are equally informative about the base form of a stem.

aggregate matched the frequencies of the lexicon: for any particular stem-final consonant, the more strongly the lexicon supported alternation, the more responses favored alternation. Compare, for example, the five cases in Figure 1, taken from Ernestus & Baayen 2003.

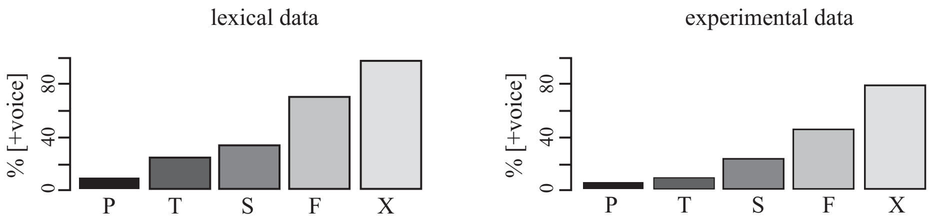


FIGURE 1. Frequency matching in Dutch voicing alternations: Ernestus & Baayen 2003.

Although Ernestus and Baayen do not address this point, their findings are unlikely to reflect a 'dialect split' among the participants. It is more likely that the variation is within rather than across participants; this is what has been found in comparable studies in which participants responded to the same items multiple times (Baroni 1999, Ringen & Heinämäki 1999:314–15). The overall result of the experiment was to show that Dutch speakers do tacitly know the quantitative alternation pattern of the lexicon, and that they demonstrate their knowledge with frequency-matching behavior.

Ernestus and Baayen's study is an especially meticulous demonstration of what other experiments have also revealed. The finding is sufficiently important (both to what follows, and we think for linguistics in general) that we give it a name (2).

(2) LAW OF FREQUENCY MATCHING: Speakers of languages with variable lexical patterns respond stochastically when tested on such patterns. Their responses aggregately match the lexical frequencies.

Some other phonological experiments whose results support this law are reported in Eddington 1996, 1998, 2004, Coleman & Pierrehumbert 1997, Berkley 2000, Zuraw 2000, Bailey & Hahn 2001, Frisch & Zawaydeh 2001, Albright 2002, Albright & Hayes 2003, Pierrehumbert 2006, and Jun & Lee 2007. Moreover, sociolinguistic study demonstrates frequency matching by children during real-life phonological acquisition (Labov 1994:Ch. 20).[2]

Having characterized 2 as a 'Law', we wish immediately to add that it holds only as an approximation, and that indeed it is the DEVIATIONS from the law that are often of greatest interest (Pertsova 2004, Goldrick & Larson 2010). In particular, if UG facilitates certain forms of frequency matching, but inhibits or suppresses others, then the law will not hold in the latter cases. In what follows, we cover studies claiming effects of this kind.[3]

[2] In his 1994 discussion, Labov strikingly situates such behaviors in a broader context: frequency-matching behavior is quite common in the animal kingdom (Gallistel 1990) and is even to be expected from the viewpoint of natural selection (e.g. creatures who sometimes select a less abundant feeding site will have less competition there). For further discussion of frequency matching in phonology, see Idsardi 2006, Silverman 2006, and Nevins & Vaux 2007.

[3] While the law of frequency matching seems fairly valid as a default characterization of wug-testing behavior, its broader validity remains to be explored. For possible cases of 'prefer the most frequent' behavior, see Hudson Kam & Newport 2005. For a puzzling case where native speakers systematically 'fail' a wug test, see Vance 1987 and Griner 2001.

3.2. A SURFEIT-OF-THE-STIMULUS EXPERIMENT: BECKER ET AL. 2007. Turkish resembles Dutch in having a regular process of final devoicing as well as a lexicon that distributes the various patterns of alternation and nonalternation asymmetrically with respect to place of articulation and other factors. Becker, Ketrez, and Nevins (2007) studied these asymmetries in detail by consulting the Turkish Electronic Living Lexicon (Inkelas et al. 2000). Becker and colleagues found, for example, that in Turkish there are many stems that alternate [p#] ~ [bV], but few with [p#] ~ [pV] (the opposite of Dutch). For alveolar stops they found that there are few cases of [t#] ~ [dV], but many of [t#] ~ [tV]. Thus Turkish is in principle a good venue for further study of the law of frequency matching. Becker and colleagues pursued the question from a slightly different angle, however, namely the possibility of 'surfeit of the stimulus' (see §2.3 above). In particular, when they inspected the lexicon, they considered environments that were more complex than just the place and manner of the final consonant, particularly the quality of the preceding vowel. Their search found that the lexicon of Turkish is indeed skewed for this factor. For obstruents in general, voicing alternation is more common after high vowels, and for the particular obstruent pair [tʃ]/[dʒ], alternation is more common after back vowels.

To find out whether Turkish speakers internalize these environments, Becker and colleagues carried out a wug test, which employed twenty-four participants. There were seventy-two wug items. The test found that with respect to place and manner of the stem-final consonant, Turkish speakers did indeed follow the law of frequency matching. However, when Becker and colleagues examined their test results with regard to the other environments, involving the features of the rightmost stem vowel, they got a null result: there is no evidence from the wug test that the speakers made any use of these environments at all.[4]

Interpreting these findings, Becker and colleagues adopt a strong UG stance (§2.1): 'The possibility of consonant voicing being determined or affected by vowel height or vowel backness is excluded, or highly disfavored [by UG] to the point that even significant evidence for such a relationship in the lexicon is not enough [for it to be learned]' (2007:62). As for the UG principle that renders the consonant-vowel interactions inaccessible to learners, they speculate that some kind of '*same-feature constraint* on vowel-consonant interactions' (p. 61) is at work, to the effect that 'the phonetic feature in the consonant that triggers the change is identical to the changed feature on the vowel (or vice-versa)' (p. 60). Thus, processes like nasalization and palatalization (shown schematically in 3a,b) would be possible under this constraint, but voicing governed by vowel height (3c) would not.

(3) Illustrations of the same-feature constraint on consonant-vowel interactions

a. Nasalization—possible

$$V \rightarrow [+\text{nasal}] \;/\; \_\_ \begin{bmatrix} \text{C} \\ +\text{nasal} \end{bmatrix}$$

b. Palatalization—possible

$$C \rightarrow [-\text{back}] \;/\; \_\_ \begin{bmatrix} \text{V} \\ -\text{back} \end{bmatrix}$$

[4] The authors suggest (p. 7) that the same finding is also observable for similar phenomena in Ernestus and Baayen's data.

c. Height-triggered voicing—held impossible

$$C \rightarrow [+\text{voice}] \,/\, \begin{bmatrix} \text{V} \\ +\text{high} \end{bmatrix} \text{—}$$

The authors note that their principle would also explain the experimental results of Moreton (2008), who in a comparative artificial-language learning study demonstrated that a different-feature, vowel-to-consonant dependency is relatively difficult to learn. (For Moreton's own account of the relevant UG principle, see Moreton 2009.)

The strong UG stance adopted by Becker and colleagues makes a clear prediction about the law of frequency matching: it should be obeyed only for phonological patterns that can be expressed in UG. For other patterns, one should find that language learners are at a loss, because (by hypothesis) there can be no learning without UG support.

4. Prospect. While we admire Becker and colleagues' work, we are skeptical of the strong UG stance that they adopt. There is a fairly substantial literature arguing that phonological processes are NOT always natural; see for example Bach & Harms 1972, Hellberg 1978, Anderson 1981, Buckley 2003, Hansson 2007, Odden 2007, and Blevins 2008. Phonological processes can be rendered unnatural by diachronic developments yet retain some productivity: see Ohala 1974 and Pierrehumbert 2006 on English velar softening and Kawahara 2008 on Lyman's Law in Japanese. The phenomenon of 'islands of reliability', studied by Albright (2000, 2002, 2003), frequently reveals productive unnatural environments. For example, Albright and Hayes (2003) observe that every verb of English that ends in a voiceless fricative is regular (i.e. ending in a voiceless fricative is an island of reliability for the regular past); their wug-test evidence indicates that this generalization, despite its lack of a natural basis, is productively internalized by native speakers.

We judge that such cases indicate an ability in people to locate and internalize phonological generalizations on an inductive basis, without having them prespecified in UG. This does not imply, however, that endogenous factors (i.e. UG, broadly construed) could not play a role in phonological learning, and we judge that some of the experimental work cited in §2 above in fact supports this view. Thus, the theory we ultimately advocate is a mixed, bias-based theory, in which learning can be assisted by UG, but is not limited to a strict UG-specified form.[5]

In what follows, we support this 'middle way' empirically with a study of natural and unnatural constraints in Hungarian vowel harmony. Like our predecessors Ernestus and Baayen and Becker and colleagues, we combine a corpus study (examining the generalizations in the Hungarian lexicon) with a wug-test study examining the degree to which the generalizations are internalized by native speakers. We begin with the relevant aspects of Hungarian phonology.

5. Hungarian vowel harmony: the data. The Hungarian vowel harmony system has been extensively studied; references include Esztergár 1971, Vago 1974, 1976, 1980, Ringen 1975, Kontra & Ringen 1986, Hare 1990, Kornai 1991, Ringen & Vago

[5] In various forms, this position has a long pedigree. The approach of NATURAL PHONOLOGY (Stampe 1973 et seq.) drew a distinction between natural 'processes' and learned 'rules'; for similar distinctions see Schane 1972 and Hooper 1976.

1995, 1998, Dienes 1997, Siptár & Törkenczy 2000, Benus 2005, Benus & Gafos 2005, 2007, and Hayes & Londe 2006.

Backness harmony in Hungarian is a highly productive process; the language abounds in suffixes and most of them alternate. Harmony depends on the following classification of vowels.

(4) Hungarian vowels

back	[u, uː, o, oː, ɔ, aː]	abbreviated 'B'
front rounded	[y, yː, ø, øː]	abbreviated 'F'
front unrounded, often called 'neutral'	[i, iː, eː, ɛ]	abbreviated 'N'

To illustrate harmony, we use examples with the dative suffix, whose behavior is representative. The dative has two allomorphs, back [-nɔk] and front [-nɛk], which are distributed according to the principles of harmony. For further discussion and illustration, see Siptár & Törkenczy 2000:157–70, Hayes & Londe 2006:61–69.

To begin, whenever the rightmost vowel of a stem is [+back] ('B'), the stem must take back suffixes. This generalization, illustrated in 5, is exceptionless.

(5) Closest vowel back: back suffixes

B**B**	[ɔblɔk-nɔk]	'window-DAT'
N**B**	[biːroː-nɔk]	'judge-DAT'
F**B**	[glykoːz-nɔk]	'glucose-DAT'

Likewise exceptionless is the generalization that if the closest stem vowel is front rounded ('F'), the stem must take front suffixes.

(6) Closest vowel front rounded: front suffixes

F	[yʃt-nɛk]	'cauldron-DAT'
N**F**	[sɛmøltʃ-nɛk]	'wart-DAT'
B**F**	[ʃoføːr-nɛk]	'chauffeur-DAT'

Moreover, front suffixes are obligatory when a front rounded vowel occurs separated from the end of the word by one or more neutral vowels.

(7) F + N*: front suffixes

FN	[fyːsɛr-nɛk]	'spice-DAT'
FNN	[øːrizɛt-nɛk]	'custody-DAT'

This completes the taxonomy of Hungarian stem types in which harmony is fully predictable. The remaining stems fall into what we call the ZONES OF VARIATION. In a zone of variation, individual stems vary in the kind of harmony they take; this is a fundamentally unpredictable property of these stems. Moreover, the zones of variation also contain a fair number of 'vacillators': stems for which either front or back suffixes are acceptable, and occur in various proportions. There are two zones of variation in Hungarian, one consisting of the stems in which a back vowel is followed by one or more front unrounded vowels (. . .BN, . . .BNN), the other consisting of all-neutral stems (N, NN, . . .).[6]

Some representative data showing the lexical arbitrariness of harmony within the BN(N) zone of variation are given in Table 1, with Google hits used to estimate frequency. All four stems contain [o] followed by [eː], and there is evidently no way to predict their differing harmonic behavior. Many similar cases exist.

[6] More precisely, only the monosyllables and (marginally) disyllables fall into this zone; NNN and longer predictably take front harmony.

WORD ([o] + [eː])		GLOSS	GOOGLE HITS (Sept. 2008)[7]	%
doménnak	[domeːn-nɔk]	'domain (on Web)-DAT'	5	2.1
doménnek	[domeːn-nɛk]		234	97.9
bohémnak	[boheːm-nɔk]	'easy-going-DAT'	433	24.4
bohémnek	[boheːm-nɛk]		1,340	75.6
honvédnak	[honveːd-nɔk]	'Hungarian soldier-DAT'	8,820	74.1
honvédnek	[honveːd-nɛk]		3,084	25.9
poénnak	[poeːn-nɔk]	'punch line-DAT'	56,400	99.9
poénnek	[poeːn-nɛk]		36	0.1

TABLE 1. Examples: lexical arbitrariness of harmony within the zones of variation.

In the other zone of variation, the all-neutral stems, the norm is for the suffix to be front (as one would expect, since neutral vowels are front phonetically). Examples of this kind are given in 8. There is, however, a modest number of all-N stems that take back suffixes, rendering harmony in this zone not entirely predictable; for discussion see Vago 1980:8–10, Siptár & Törkenczy 2000:68. Examples are given in 9.

(8) All-neutral stems: normally front suffixes

N	[kɛrt-nɛk]	'garden-DAT'
N	[tsiːm-nɛk]	'address-DAT'
NN	[rɛpɛs-nɛk]	'splinter-DAT'

(9) Examples of all-neutral stems taking back suffixes

N	[hiːd-nɔk]	'bridge-DAT'
N	[ʃiːp-nɔk]	'whistle-DAT'
NN	[dɛreːk-nɔk]	'waist-DAT'

For detailed study of the zones of variation, it is helpful to have systematic frequency data. Hayes and Londe (2006) addressed this need by machine searching for the Google hit counts of the [-nɔk] and [-nɛk] forms of several thousand Hungarian nominal stems.[8] We have since heavily edited this set, removing more compounds and other misleading forms, and reducing to a single entry those instances in which a particular suffix (such as the always-transparent nominalizing suffix [-eːk], e.g. *maradék(nak)* 'remainder (-DAT)') was contributing a great number of items. The revised list, with 8,915 stems, is posted at http://www.linguistics.ucla.edu/people/hayes/HungarianVH/. In discussion below we refer to this list as our 'lexical database'.

5.1. STATISTICAL PATTERNS WITHIN THE ZONES OF VARIATION. As inspection of our lexical database demonstrates, the zones of variation are statistically structured in a way that renders harmony 'semi-predictable'—much like with the patterns of voicing alternation in Dutch and Turkish, discussed above. The most significant patterns, discussed in the literature (e.g. Siptár & Törkenczy 2000:70–71) and documented by Hayes and Londe, are summarized below.

Within the zone consisting of stems ending in BN or BNN, there are two salient quantitative patterns. There is a COUNT EFFECT, whereby stems ending in BNN (aggre-

[7] There is a nonsignificant trend toward higher [-nɔk] rates for higher-frequency words.

[8] Changes at Google have prevented us from updating the 2004 Hayes/Londe hit counts; the current counts, checked by hand for just a few stems, are much higher.

A referee asks to what degree the Google counts might be contaminated by text from nonnative speakers. This seems unlikely, given the small size of the linguistic minorities that inhabit Hungary (see *Ethnologue*, http://www.ethnologue.com). The phonological effects found in the lexical database were also checked, for a crucial subset of the words, against the judgments of three native speakers, who showed very similar patterns (Hayes & Londe 2006:§3.2).

gately) take front suffixes more than do stems ending in BN. Similarly, there is a HEIGHT EFFECT, whereby stems ending in BN or BNN (aggregately) take more front suffixes if their last vowel is [ɛ] than if it is [eː], and likewise more front suffixes if their last vowel is [eː] than if it is [i] or [iː]. Both the count effect and the height effect are very strong in the lexicon, as the frequencies from our lexical database, graphed in Figures 2 and 3, show.

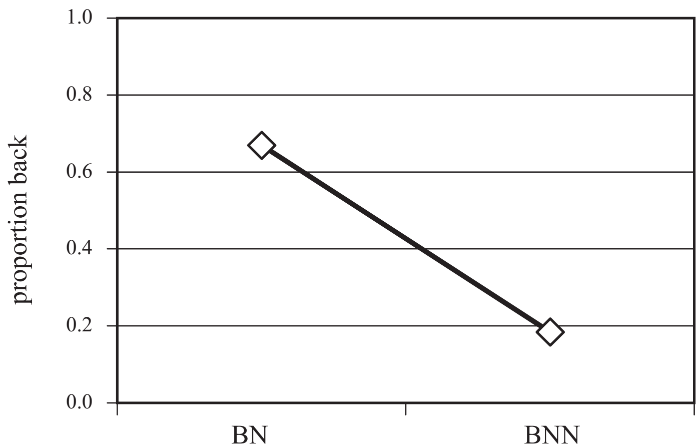


FIGURE 2. The count effect in Hungarian vowel harmony.

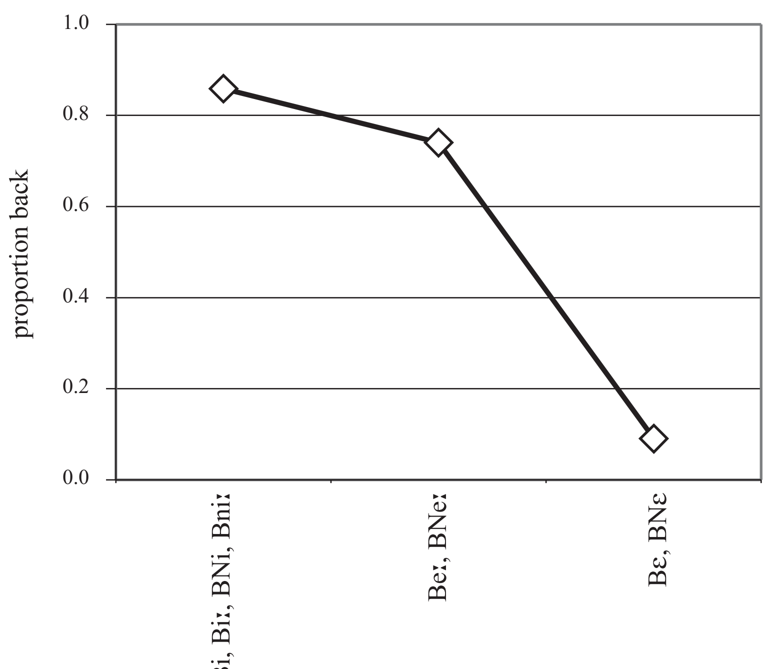


FIGURE 3. The height effect in Hungarian vowel harmony.

Moreover, as Figure 4 indicates, the height and count effects combine in a fairly consistent way.

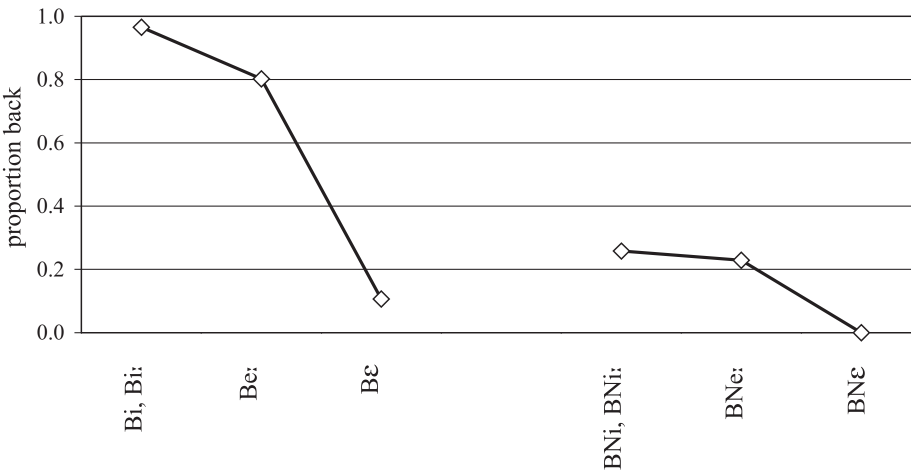


FIGURE 4. The height and count effects together.

There is also statistical structuring present in the all-neutral zone of variation; we defer discussion of this to 11 below.[9]

The height and count effects are demonstrably phonologically productive. This was shown by the wug test carried out by Hayes and Londe, which resulted in a fairly good match between lexical patterns and native-speaker responses. This is summarized in Figure 5.

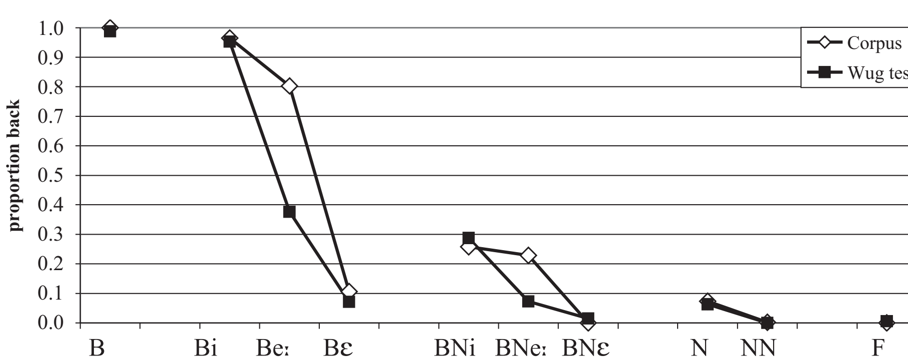


FIGURE 5. Hayes and Londe's wug test: participant responses largely match lexical counts.

To sum up so far: the Hungarian vowel harmony system consists primarily of inviolable phonological generalizations, but incorporates zones of variation in which harmony is unpredictable. Within these zones, the statistical generalizations embodied in the

[9] We also found, in both the lexicon and our wug test (below), that there is a weaker height effect for the first N in BNN words, and that there are lower rates of [-nɔk] after BBN words than after BN (for all three heights of N). These two effects are not discussed further, however, as they make only modest improvements in fitting the data.

height and count effects can be discerned. In Hayes and Londe's wug test, these generalizations were demonstrated to be internalized by native speakers, following the law of frequency matching.

6. ANALYSIS OF THE HUNGARIAN DATA WITH NATURAL CONSTRAINTS. Hungarian vowel harmony can be analyzed using constraints that we argue to be natural. Although in the present state of our knowledge assessing the naturalness of a constraint is necessarily a tentative procedure, we think that useful assessments are possible. There are two ways to assess naturalness. First, a constraint should match typological data, particularly when it is related to the other constraints in the construction of a factorial typology, as noted in §1. Second, a constraint can be asserted to be natural on phonetic grounds, when it can be shown to increase the ease of articulation or the saliency of contrasting forms in perception. The role of phonetics in defining naturalness has long been a theme in phonological theorizing; for recent overviews see Hume & Johnson 2001, Blevins 2004, and Hayes, Kirchner, & Steriade 2004.

We begin by asserting pretheoretical principles of naturalness against which the naturalness of the specific constraints governing vowel harmony can be evaluated.[10]

(i) SINGLE-FEATURE ASSIMILATION. First, we observe the naturalness of vowel harmony itself: vowels assimilate to one another iteratively in a particular feature (here, [back]), and they do so across strings of intervening consonants. Thus defined, vowel harmony can be observed in a great number of languages throughout the world (Vago 1994). Harmony is often thought (see e.g. Beddor et al. 2001) to originate in the phenomenon of vowel-to-vowel coarticulation, documented in Öhman 1966 and subsequent work.

(ii) LOCALITY. When disagreeing triggers for harmony are present in a stem, it is normally the LOCAL (closer) trigger that wins. This pattern is seen plainly in Hungarian, where in borrowed stems F competes with B, as in [ʃoføːr] 'chauffeur' (BF) or [glykoːz] 'glucose' (FB). For these, the dative forms must be [ʃoføːr-nɛk] and [glykoːz-nɔk], with harmony governed by the local vowel. The only exceptions to the locality principle involve the neutral vowels, as in cases like [poeːn-nɔk] from Table 1 above. Here, the neutral vowel may plausibly be regarded as a 'weak' harmony trigger: alone (as in forms like 8) it would govern front harmony, but in BN stems a preceding stronger trigger often will override.

(iii) TRIGGER STRENGTH. Based on an extensive survey of harmony systems, Kaun (1995, 2004) asserts a general principle: if a particular vowel is deficient in the acoustic correlates for the harmonizing feature, it is typologically more likely to be a harmony trigger. Initially, this seems almost paradoxical, but Kaun suggests a functional explanation: the vowels that are most likely to be misidentified are the most likely to take advantage of harmony to spread their harmonizing feature onto neighboring vowels, thus providing a redundant cue for perception. The weak trigger effect is also invoked by Walker (2005) in an account of Romance metaphony, and was observed to some extent in an artificial-grammar learning experiment by Finley (2008:Ch. 10).

[10] A quite different approach to naturalness in Hungarian vowel harmony is proposed by Benus (2005) and Benus and Gafos (2005, 2007). The key idea in their work, supported by phonetic measurement, is that the neutral vowels coarticulate extensively with neighboring vowels in backness, and that higher vowels coarticulate more. In this view, backness can be considered as propagating across ALL the vowels of a word, as opposed to the present conception here involving rival triggers. We have not explored this approach here, but any constraint-based analysis that quantitatively fits the Hungarian data would be likely to suffice for our present purpose. What we need is a baseline analysis involving only natural constraints, against which the unnatural constraints of §7 can be assessed.

In Hungarian, the most powerful triggers for front harmony are the front rounded vowels, which always force harmony when they are the last nonneutral vowel; these are acoustically inferior bearers of the feature [–back] since their rounding lowers F2, where a high F2 is the principal cue for frontness. Likewise, the height effect can be attributed to Kaun's principle: the lower front vowels, which are the strongest triggers of front harmony in BN or BNN stems, are the ones that have the lowest F2.[11] A language also showing a height effect is Old Mongolian (Svantesson et al. 2005:114), where of the front unrounded vowels [i, e], only [i] was neutral.

(iv) TRIGGER COUNT. Lastly, Hayes and Londe suggest that having two trigger vowels in a row is a stronger harmony environment than having just one. This is at least intuitive as a concept, and there are two known cases (in Classical Manchu and Oroqen) in which harmony occurs only when there are two consecutive triggers present (Dresher & Zhang 1996, Zhang 1996, Walker 2001). A Finnish analogue to the count effect of Hungarian is noted by Campbell (1980:252), and could be analyzed as a trigger-count effect.

With these four principles in mind, we turn to the formulation of a set of natural constraints for the analysis of Hungarian vowel harmony. The inventory in Table 2 is closely based on the constraint set used by Hayes and Londe (2006:76–77).[12]

CONSTRAINT	FORMALIZATION	COMMENT
a. AGREE(back, local)	*[+back][−back]	Vowels must agree in backness with a preceding back vowel.
b. AGREE(back, nonlocal)	*[+back] N_0 [−back]	Nonlocal version of (a)
c. AGREE(front rounded, local)	*[−back, +round][+back]	Vowels must agree in backness with a preceding front rounded vowel.
d. AGREE(front rounded, nonlocal)	*[−back, +round] N_0 [+back]	Nonlocal version of (c)
e. AGREE(front, local)	*[−back][+back]	Vowels must agree in backness with a preceding front vowel.
f. AGREE(nonhigh front, local)	*[−back, −high][+back]	Specific version of (e), limited to nonhigh triggers
g. AGREE(low front, local)	*[−back, +low][+back]	Specific version of (f), limited to low triggers[a]
h. AGREE(double front, local)	*[−back][−back][+back]	Two-trigger harmony

TABLE 2. Natural constraints for Hungarian vowel harmony.

[a] We treat [ɛ], the lowest front vowel of Hungarian, as [+low]: in suffixes it alternates with a low vowel, and phonetically it is often rather lower than the IPA symbol suggests.

Here are the formal assumptions behind the constraints. First, it is assumed that constraints may access some kind of 'vowel tier' (Clements 1976) or 'vowel projection' (McCarthy 1979, Hayes & Wilson 2008), which displays all the vowels of a word adjacently, permitting vowel-to-vowel interaction in spite of intervening consonants. Second, we allow constraints to specify the agreeing feature as well as restrictions on the trigger based on the principle of trigger strength given above. Third, we allow

[11] These commonplace acoustic effects are evident in Hungarian; see Magdics 1997:6.

[12] The constraint type AGREE originates with Lombardi (1999). AGREE constraints suffer from problems of overgeneration—that is, generation of typologically unattested patterns. These are discussed, with possible remedies, in Wilson 2003b, McCarthy 2004, and Finley 2008. The overgeneration problem appears to be independent of the issues at hand here, so we stick with AGREE for expository simplicity.

constraints to specify whether they apply locally, or permit certain intervening neutral vowels.

These are quite similar to the constraints used by Hayes and Londe (2006). In the case of Table 2(e–g), we have made a minor change: instead of referring to specific heights (e.g. 'AGREE(mid vowel)'), we adopt the alternative approach of the 'stringency hierarchy' (Prince 1997, De Lacy 2004), where agreement is required when a vowel is of a particular height OR LOWER. These are essentially equivalent as far as the grammatical analysis goes, but are easier to work with in statistical testing (see §8.3). For AGREE(back, nonlocal) and AGREE(front rounded, nonlocal), in principle our system should assess violations whenever the stem contains B or F (respectively); however, difficulties involving the theory of locality (see Hayes & Londe 2006:§5.12 for discussion and a proposal) led us to employ an ad hoc simplification, which simply assigns a violation whenever the closest nonneutral vowel is B or F respectively.

Inspection indicates that all of these constraints reflect the typologically supported, proposed UG principles given above. All of them enforce agreement in the feature [back], reflecting the principle of single-feature assimilation. Five of the eight constraints are specifically limited to local triggers (a, c, e, f, g), respecting locality. Wherever the trigger class is limited by features, it is limited in a way that designates particular vowel classes that in accordance with Kaun's principles should be strong triggers: front rounded vowels (c, d) or front vowels of a particular height (f, g). Lastly, the constraint in (h) reflects the principle of trigger count.

A complete analysis of Hungarian vowel harmony would have to cover several aspects not discussed here—for instance, the need to limit harmony targets to suffixes and to enforce any lexically specified patterns for stems within the zones of variation. For discussion, see Hayes & Londe 2006:§5. Here, we are concerned only with how a ranking of the crucial constraints yields the aggregated data patterns. Three of the constraints are never violated and may be ranked at the top of the grammar: AGREE(back, local), AGREE(front rounded, local), and AGREE(front rounded, nonlocal). The remaining five constraints are illustrated below with schematic tableaux, leaving open for the moment the constraint ranking and the winning candidates.

At an intuitive level, Table 3 shows how the height and count effects can be derived: the lower the last stem vowel, the more violations penalize the [-nɔk] candidate, increasing the likelihood of [-nɛk]. BNN forms with [-nɔk] are likewise penalized by one more constraint than the corresponding BN forms.

This general scheme can be made precise in a framework that generates quantitative predictions. Hayes and Londe used STOCHASTIC OPTIMALITY THEORY (Boersma 1998, Boersma & Hayes 2001), in which the constraints are assigned numerical 'ranking values' and the grammar generates multiple outputs probabilistically. Their analysis achieved a fairly close quantitative match to the wug-test data they gathered (see graph and discussion, Hayes & Londe 2006:89). We implement a somewhat different analysis below in §9.

The upshot of the discussion so far is this: not only do Hungarian speakers tend to match the frequencies of the lexicon when tested on vowel harmony, but it is also possible to achieve a good first-pass approximation to their behavior using a probabilistic, constraint-based framework, the constraints being limited to those that would plausibly be described as natural in phonological theory.

7. HUNGARIAN BEYOND UG: ARE THERE UNNATURAL CONSTRAINTS? We turn next to the question of whether there might also be unnatural environments governing the choice of front and back suffixes in the Hungarian zones of variation. A good place

		AGREE(back, nonlocal)	AGREE(front, local)	AGREE(nonhigh front, local)	AGREE(low front, local)	AGREE(double front, local)
/Bi/	Bi-nɔk		*			
	Bi-nɛk	*				
/Beː/	Beː-nɔk		*	*		
	Beː-nɛk	*				
/Bɛ/	Bɛ-nɔk		*	*	*	
	Bɛ-nɛk	*				
/BNi/	BNi-nɔk		*			*
	BNi-nɛk	*				
/BNeː/	BNeː-nɔk		*	*		*
	BNeː-nɛk	*				
/BNɛ/	BNɛ-nɔk		*	*	*	*
	BNɛ-nɛk	*				

TABLE 3. Schematic tableaux for the five violable constraints.

to search is among the consonants at the end of the stem; at least in published work, consonants have not been considered as possible influences on Hungarian vowel harmony.[13] We therefore performed an intensive search for consonant effects in our Hungarian lexical data, using spreadsheets and small computer programs. Adopting a fairly standard feature set for Hungarian, we checked every natural class of consonants C_i that it defines, assessing whether a constraint preferring front suffixes or back suffixes in the context / C_i ___ has noticeable effects. We also checked as many other reasonable-seeming unnatural environments that we could, simply relying on our intuitions of what might work.

Our search was successful, in the sense that we found four consonant environments that proved to have considerable influence on the outcome of harmony within the zones of variation. As it happens, all four favor front harmony; they are listed in 10.

(10) Four unnatural vowel harmony constraints for Hungarian
- a. Prefer front suffixes when the stem ends in a bilabial noncontinuant ([p, b, m]).
- b. Prefer front suffixes when the stem ends in a sibilant ([s, z, ʃ, ʒ, ts, tʃ, dʒ]).
- c. Prefer front suffixes when the stem ends in a coronal sonorant ([n, ɲ, l, r]).
- d. Prefer front suffixes when the stem ends in a sequence of two consonants.[14]

[13] A web-posted abstract, Trón 2007, claims that there are consonant environments for Hungarian vowel harmony (without identifying them), and Anderson (1980) claims consonant environments for Finnish; both study the lexicon only. The case of Turkish (Clements & Sezer 1982) is quite different: the consonants that influence harmony in this language are palatalized, and hence themselves bear the harmonizing feature.

[14] A referee suggested splitting 10d in two: one version for final geminates and one for true clusters. We found that both significantly favor [-nɛk], and model fits (see below) were not improved by the split.

We judge that one of these environments (10d) is incontestably unnatural;[15] and they may all be. The contexts 10b and 10c might conceivably be said to favor front harmony given the slight affiliation between coronal place of articulation and front vowels (Clements 1991, Hume 1992, Flemming 2003). However, the limitation to sonorants and sibilants (observe the exclusion of coronal [t, d]) makes little sense in terms of typology. Moreover, it is particularly puzzling that the coronality of the stem-final consonant could influence the choice of suffix vowel, given that the suffix we examined begins with [n]: the suffix vowel is always adjacent to a coronal sonorant. The effect in 10a could be some form of labial dissimilation ([ɔ] is rounded), as suggested by a referee, but note that labiodentals ([f, v]) are excluded from the effect. In order for 10a,b,c to all be phonetically based there would have to be both consonant-vowel assimilation and consonant-vowel dissimilation.

For completeness, we note another apparently unnatural constraint, which governs the minor all-neutral zone of variation.

(11) Constraint: monosyllables with [iː]
Prefer back suffixes when the stem is monosyllabic with the vowel [iː].

In the database we worked with, there are 1,955 all-neutral stems, of which only twenty take majority back harmony. Of the latter, eighteen are monosyllabic, and of these, sixteen have the vowel [iː] (the other two have [eː]). Hence, being of the form [C_0iːC_0] nearly qualifies as a necessary condition for back harmony in all-neutral stems. It is hardly a sufficient condition, however, since there are also seventeen monosyllabic stems with [iː] that take the expected majority front harmony. For an exhaustive listing of the relevant stems, leading to similar conclusions, see Vago 1980:8.

Though vowel dissimilation does sometimes occur in phonology (see e.g. Itô 1984), the limitations to a particular vowel and to syllable count seen here have no natural basis of which we are aware. The data pattern expressed in 11 has been claimed to be traceable to a common diachronic path whereby the words that originally had the form *CɯCV (and took back harmony) evolved into CiːC (Kálmán 1972:63–64).

Thus, we have a total of five plausibly unnatural constraints that can be evaluated for their influence on Hungarian vowel harmony.

7.1. Quantitative assessment of the constraints. Before proceeding with a wug test based on these constraints, we sought some preliminary assurance that the skewing of the Hungarian lexicon they represent is large enough that we might expect to find their influence in the testing data.

To begin, when the constraints are taken in the aggregate, they produce a noticeable quantitative effect in the zones of variation. In Table 4, we divide up the stems in various lexical subzones according to whether at least one unnatural environment is present. It can be seen that for certain stem types, the increase in [-nɛk] counts found

[15] John McCarthy (p.c.) offered an intriguing scenario under which 10d would emerge as natural: in the BN/BNN zone of variation, a CC sequence increases the distance between the B trigger and the suffix, thus perhaps attenuating the influence of B. We checked this by examining whether having two or more consonants between the penult and ultima had a similar effect, since it also adds distance between B and the suffix. In contrast to final CC, it had an essentially zero effect. This suggests that raw distance does not explain the final-CC constraint.

in our lexical database is quite substantial. The column labeled *N* indicates the number of items over which the fractions in the preceding columns are calculated.

STEM TYPE	FRACTION [-nɛk]		*N*
	no unnatural consonant environment present	at least one unnatural consonant environment favoring [-nɛk]	
B + high	0.03	0.03	97, 85
Beː	0.08	0.35	43, 32
Bɛ	0.34	0.91	4, 109
BNN	0.74	0.88	27, 36
N	0.84	0.96	68, 191

TABLE 4. Aggregated effects of the unnatural environments in the lexical database.

Statistics for the individual constraints are provided in Table 5. Where a stem shows divided behavior in the Google counts, we apportion the values between the two options; this is why the third column contains noninteger values.

	total stems meeting the description	total taking predicted harmony value	%
USE FRONT / bilabial __	41	36.1	88.0
USE FRONT / sibilant __	136	88.5	65.1
USE FRONT / [+cor, +son] __	182	116.8	64.2
USE FRONT / CC __	168	157.6	93.8
(compare: all in zones of variation)	*692*	*279.1*	*40.3*
USE BACK / [C_0iːC_0] __	33	15.6	47.1
(compare: all monosyllabic N)	*259*	*18.8*	*7.3*

TABLE 5. Lexical statistics for five unnatural constraints (zones of variation only).

We also sought to guard against the possibility that some of the unnatural constraints could be artifacts. Words are often subject to several constraints, both natural and unnatural, and some of the relevant properties may not be independent. For example, if the height of front vowels differs statistically before consonants of various types, we might get a consonant effect that was an epiphenomenon of the height effect.[16] To address this problem and assess the independent contribution of each constraint, we fitted a generalized linear model of [-nɔk] rates in the lexicon. Because [-nɔk] rates in the lexicon are not at all normally distributed, but instead clustered very strongly at 0 and 1, we could not use a simple linear regression. Instead, we collapsed all rates to 0 or 1, and performed a logistic regression.[17] The model has a set of coefficients, one for each constraint, plus an intercept (baseline value). These are shown in Table 6.

[16] Indeed, almost all [Bɛ] stems meet at least one unnatural environment; see Table 4.

[17] Values ≤ 0.1 were treated as zero, values ≥ 0.9 were treated as 1, and the remaining sixty-seven words with values in the middle were discarded. We used the *bayesglm*() function of the *arm* package (Gelman et al. 2009) of the R statistics program (R Development Core Team 2007), with default values (*prior.scale* = 2.5, *prior.df* = 1), following Gelman and colleagues' (2008) recommendation for logistic regression when independent variables strongly predict each other (collinearity) and when some combinations of values for the independent variables fully determine the outcome (separation).

	COEFFICIENT	p
(Intercept)	−2.479	0.0004
Natural constraints		
AGREE(back, local)	6.330	< 0.0001
AGREE(back, nonlocal)	6.629	< 0.0001
AGREE(front rounded, local)	−1.095	0.5751
AGREE(front rounded, nonlocal)	−1.370	0.4834
AGREE(front, local)	*(redundant)*[a]	
AGREE(nonhigh front, local)	−1.384	0.0046
AGREE(low front, local)	−4.553	< 0.0001
AGREE(double front, local)	−4.996	< 0.0001
Unnatural constraints		
USE FRONT / bilabial ___	−2.356	0.0095
USE FRONT / [+cor, +son] ___	−1.290	0.0053
USE FRONT / sibilant ___	−1.210	0.0205
USE FRONT / CC ___	−1.266	0.0570
USE BACK / [C_0iːC_0] ___	3.187	< 0.0001

TABLE 6. Evaluating harmony constraints for strength in the lexicon: a generalized linear model.

[a] AGREE(front, local) is not assigned a value because its violation pattern is entirely predictable (the complement set of AGREE(back, local)). When we later turn to maxent modeling of the data, however, we see that AGREE(front, local) does play a role, accounting for the small number of Bi forms with front suffixes. In the regression model, the intercept takes care of these cases.

In using the model to predict the probability that a given stem will take [-nɔk], we begin with the intercept, then add the coefficient for each constraint that is applicable to that stem. The result is the logit, which can be converted into a probability: *probability* $= 1/(1 + e^{-logit})$. Some constraints' coefficients are positive, which means that they increase the probability that the stem takes [-nɔk], while others' are negative, decreasing this probability. For example, the model's logit value for [diɔdeːm-nɔk] 'diadem-DAT' is computed as follows: −2.479 (intercept) + 6.629 (nonlocal back vowel encourages [-nɔk]) − 1.384 (local nonhigh front vowel discourages [-nɔk]) − 2.356 (final bilabial discourages [-nɔk]) = 0.410. This logit is converted to a probability with the formula just given: $1/(1 + e^{-0.41}) = 0.601$ (cf. the observed rate for [diɔdeːm-nɔk] in the corpus, 0.727).

The model demonstrates that the effects of the unnatural constraints are unlikely to be artifacts: had this been the case, the model would have assigned coefficients to the unnatural constraints at or near zero.[18] Our local conclusion is that Hungarian includes a number of phonological environments that appear to be unnatural, are correlated with harmonic behavior in the lexicon, and whose effects are independent of the natural constraints.

We emphasize that these findings pertain to the Hungarian LEXICON; the issues of whether Hungarian speakers are aware of the unnatural environments and the degree to which the unnatural environments play a role in forming their linguistic intuitions remain to be covered.

[18] Significance values for the coefficients are shown in the right column of Table 6. These p-values can be used to evaluate the independence of the unnatural constraints from the others, but would not be appropriate as a test of whether the lexicon contains unnatural constraints at all, since we are showing only a selection of the strongest ones. AGREE(front rounded, local) and AGREE(front rounded, nonlocal) receive surprisingly high p-values, considering that their phonological status in Hungarian is impeccable. Each of these two factors is highly predictable from the other factors, leading to a large standard error and thus a high p-value.

8. A WUG TEST FOR UNNATURAL CONSTRAINTS. To gain insight into these questions, we conducted a wug experiment. The focus of the experiment was the four unnatural consonant-based environments listed in 10, though we also report our findings on the remaining constraints. Our particular interest in the consonant-based environments lies in the fact that, following the work of Becker and colleagues (2007) summarized above, we would expect them to fall outside the range of learnable constraints.

8.1. CHOICE OF WUG STEMS. All of the wug stems in our experiment were from the zones of variation. To increase the sensitivity of our study, we employed a large number of wug stems, by sampling with replacement from the set of possible stems, as detailed below. With 1,602 items distributed over 1,703 trials (131 subjects × 13 trials each), most items that a given participant saw were unique. Our purpose was to reduce the chance of unwanted factors about particular stems playing a role; whatever factors may have been present would be likely to be washed out among such a large number of varying stems.[19]

We arranged the wug stems to have final consonants or consonant clusters that would test the unnatural consonant environments given in 10. There were equal numbers of monosyllables, disyllables, and trisyllables. Otherwise, we tried to make the wug stems statistically resemble Hungarian stems in every possible respect by generating them with a procedure that matched the frequencies of their initial, medial, and final consonant sequences to the frequencies found in our lexical corpus. Each stem was required to belong to one of the vowel schemata BNN, (B)BN (shown as BN in graphs),[20] and N, and within each vowel type (B or N) the vowels were again chosen at a frequency that matched their frequency in the lexical corpus. We then filtered out real stems as carefully as we could, both with comparison to large electronic Hungarian dictionaries and with visual inspection by the two native-speaker authors. Because so many possible monosyllables are already real words, there ended up being sixty-eight monosyllabic stimuli that were used for two or more participants.

With this procedure, we obtained enough wug stems to give each consultant a balanced set of thirteen stems, which were chosen as indicated in Table 7.[21]

# OF STIMULI	FINAL BILABIAL	FINAL CORONAL SONORANT	FINAL SIBILANT	FINAL CC	EXAMPLE(S)
2	no	no	no	no	kóde, ráltényé
2	no	no	no	**yes**	Prett, Bótáredt
2	no	no	**yes**	no	csandendes, Vórmámlés
1	no	no	**yes**	**yes**	szalhárécc
2	no	**yes**	no	no	nen, Cir
1	no	**yes**	no	**yes**	Vureldern
2	**yes**	no	no	no	hárakem, Ém
1	**yes**	no	no	**yes**	kébb

TABLE 7. Basis for selecting each participant's set of thirteen wug stems.

[19] An alternative method to achieve this goal is to have a great number of wug items, each one given to every participant, and then carry out a by-items analysis on the results. This would not be feasible with our web survey method—our experience in a pilot study was that most participants would not complete more than about fifteen items.

[20] We included BBN in order to expand the list of phonotactically ordinary possibilities for stems ending in BN. For monosyllables, there was no alternative but to give some of the stems to more than one participant.

[21] In Hungarian orthography, í, é, á, ó, ú = [iː, eː, aː, oː, uː], a = [ɔ]; e = [ɛ], c = [ts], cs = [tʃ], s = [ʃ], sz = [s], ny = [ɲ].

8.2. Subjects and procedure. We ran our test on the World Wide Web, using scripts adapted from Zuraw 2007. Our server displayed the experimental materials and recorded the participants' responses.

Participants were recruited using Google AdWords, a commercial web advertising service. This service caused a brief advertisement to be displayed to web users whose internet addresses indicated that they had logged on to the web from a site within Hungary. The advertisement appeared in some cases on the Google search engine, but mostly on affiliated independent web sites. The target words that were used to guide the placement of the advertisement were the Hungarian versions of 'Hungarian language', 'language', 'dictionary', 'thesaurus', 'etymology', 'linguistics', 'vowel', 'consonant', 'dialect', 'letter game', and 'word game'.

Google ran our ad about 2,000,000 times, resulting in 872 visitors to the site. Of these, 131 ultimately met our qualifications as participants: they completed all the items in the test, they gave no noncongruent responses (see below), and on the information page they told us that they were native speakers, eighteen or older, and living in Hungary.[22] Of these 131 subjects, seventy-six were male and fifty-five female, and the median age was thirty-three (range: eighteen to seventy-five). We used Google Analytics software (www.google.com/analytics/) to monitor the location of our participants. They turned out to be from all over Hungary, with about 56% from Budapest.

Following the introductory screen, the experiment displayed a series of screens giving the wug words in Hungarian orthography, which is unambiguous in the grapheme-to-phoneme direction. Each screen included a paragraph using the word twice; paragraphs were drawn from Hayes & Londe 2006 and were intended to promote the sense that the words could be native Hungarian words that had become obsolete with the passage of time, rather than loanwords.[23] Toward the end of the paragraph, a blank appeared, which following the principles of Hungarian grammar would naturally be filled with the dative form ([-nɔk] or [-nɛk]) of the wug stem. Participants first clicked on one of two buttons to indicate a binary choice between [-nɔk] and [-nɛk], then rated each form on a 1–7 scale. A sample page (rendered into English; the original is entirely in Hungarian) is given in Figure 6.

The order of frame sentences was randomized for each subject. Each stimulus-frame combination was automatically adjusted for upper vs. lower case, low vowel lengthening (Hayes & Londe 2006:64), and definite-article allomorphy. For each item, the order of the two candidate suffixes was randomized, though it was held constant between the forced-choice task and the rating task.

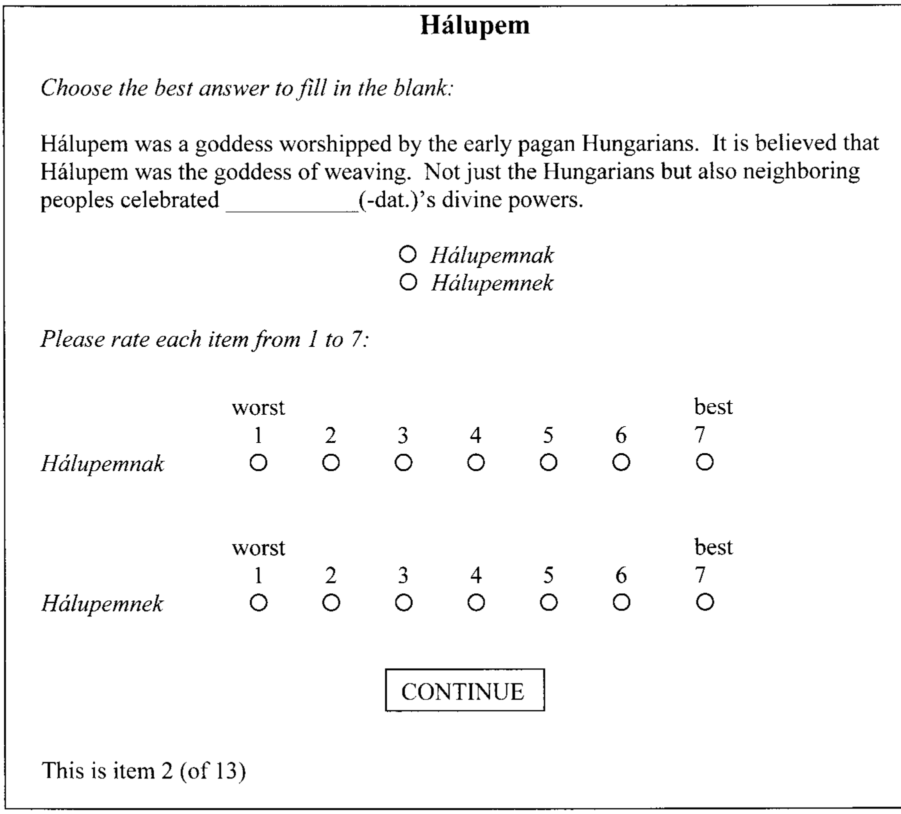
Participants had to complete the choice task and both rating tasks on each page before the program would allow them to continue to the next page. We deemed a response 'congruent' if the suffixed form chosen was rated at least as high as the other option. Participants were allowed by the program to continue even if they gave a noncongruent answer, but no data for such participants was included in the results reported below.[24]

Participants were rewarded for continuing the experiment with a 'fun facts' scheme: in the introductory screen, a question of popular interest about the Hungarian language was posed, and participants were shown the answer only after they had provided re-

[22] For the importance of controlling for country of residence, see Zhang et al. 2007.

[23] We judged that this would best serve in encouraging participants to access their native intuitions, rather than treating the task as an abstract exercise or a form of word play.

[24] Ninety-three subjects completed the survey but were excluded because they gave at least one noncongruent response (usually just one or two).



Hálupem

Choose the best answer to fill in the blank:

Hálupem was a goddess worshipped by the early pagan Hungarians. It is believed that Hálupem was the goddess of weaving. Not just the Hungarians but also neighboring peoples celebrated ___________(-dat.)'s divine powers.

O *Hálupemnak*
O *Hálupemnek*

Please rate each item from 1 to 7:

	worst 1	2	3	4	5	6	best 7
Hálupemnak	O	O	O	O	O	O	O

	worst 1	2	3	4	5	6	best 7
Hálupemnek	O	O	O	O	O	O	O

CONTINUE

This is item 2 (of 13)

FIGURE 6. Example of a page from the web-based wug test.

sponses for two wug forms; they were then shown the next question, with the answer coming two wug forms later, and so on.

8.3. RESULTS. First, we checked to see if our experiment yielded results similar to those found by Hayes and Londe (2006), particularly with respect to the height and count effects (§5). With a caveat, this turned out to be true.

The major difference is that the new wug-test results are less polarized: see in particular the series Bi/Beː/Bɛ, where every value in the present wug test is closer to the middle value .5 than in the Hayes-Londe wug test. There is a plausible explanation for this. The Hayes-Londe test was entirely of the 'volunteering' type, where the participants must make up the responses on their own. In contrast, the present test also included the task of rating both the [-nɔk] and [-nɛk] forms. Thus, it modeled to the participants a number of forms that are of very low probability in Hungarian: for instance, [BNɛ-nɔk] forms. We suggest that participants are more likely to accept an unusual form presented to them than to volunteer it. The modeling effect has been found in earlier work that used both rating and volunteering (Albright & Hayes 2003, Becker et al. 2007, Becker 2009).

Whatever the explanation for the less-polarized responses, the new test, just like the old one, demonstrates the psychological reality of the height and count effects. This is

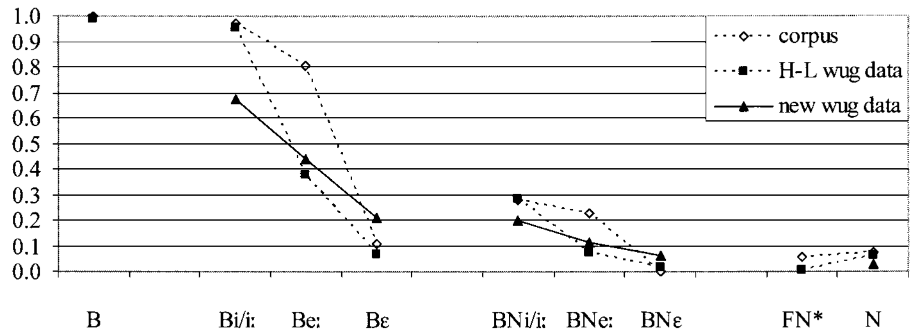


FIGURE 7. Hayes-Londe experiment replicated.

evident in the graph above in Figure 7. Our statistical analysis, which confirms this, is delayed to the following paragraphs where we discuss the results concerning unnatural constraints. The results also reassure us that participants understood the task and applied their knowledge of existing words to the wug words.

We now turn to the purpose for which the experiment was designed, namely determining whether the phonologically unnatural environments are internalized by Hungarian speakers. We present results first for the forced-choice task and then for the rating task.

We analyze the forced-choice results with logistic regression, much as in Table 6 above, except that here we use a mixed-effects model (see Quené & van den Bergh 2004): as before, there are a baseline (intercept) value and a coefficient for each phonological constraint (fixed effects), but this model also assigns each subject his or her own baseline probability of choosing [-nɔk] (random effect). The model was fitted using the *lmer*() function of the *lme4* package in R (Bates 2007). As in the logistic-regression model of Table 6, we obtain the predicted logit of the probability of a [-nɔk] answer for a given stimulus by adding to the intercept the coefficients for the constraints applicable to that form, but here we also add the subject intercept. The subject intercepts average to 0, so on average, without knowing the identity of the subject, the predicted logit will be the same as if there were no subject intercept added. The logit is then transformed into a probability ($probability = 1/(1 + e^{-logit})$). The coefficients found are in Table 8; a positive value means that the property in question encourages a [-nɔk] answer, and a negative value the opposite.

The results can be summarized as follows. All of the natural constraints have strong, reliable effects. Moreover, four of the five unnatural constraints also emerged as significant in predicting the forced-choice responses. The only exception was USE BACK / [C_0iːC_0] __, a constraint that our experiment was not set up to test (not every participant saw a [C_0iːC_0] stimulus). It remains to be seen whether a wug test with sufficient items of this type would reveal effects of this constraint as well.

Our experiment also gathered rating data for both [-nɔk] and [-nɛk] forms. These data require a different statistical approach. We define '[-nɔk] advantage' as the observed rating for the [-nɔk] option minus the rating for the [-nɛk] option. Since each response ranged from 1 to 7, the [-nɔk] advantage ranges from -6 to 6. The model is like that in Table 8 (and was fitted using the same function, except that *p*-values were estimated using *pvals.fnc*() in the *languageR* package (Baayen 2007)), but here the value obtained by adding the relevant coefficients to the intercept is the predicted

	COEFFICIENT	*p*
Intercept	−2.447	< 0.0001
Natural constraints		
AGREE(back, local)	*(omitted)*	
AGREE(back, nonlocal)	4.024	< 0.0001
AGREE(front rounded, local)	*(omitted)*	
AGREE(front rounded, nonlocal)	*(omitted)*	
AGREE(front, local)	*(omitted)*	
AGREE(nonhigh front, local)	−1.229	< 0.0001
AGREE(low front, local)	−1.132	< 0.0001
AGREE(double front, local)	−2.200	< 0.0001
Unnatural constraints		
USE FRONT / bilabial ___	−1.211	< 0.0001
USE FRONT / [+cor, +son] ___	−0.519	0.0074
USE FRONT / sibilant ___	−0.500	0.0230
USE FRONT / CC ___	−0.762	< 0.0001
USE BACK / [C_0iːC_0] ___	0.610	0.3672

TABLE 8. Logistic-regression model of the forced-choice results.

[-nɔk] advantage. The results are reported in Table 9. As before, a positive coefficient means that the constraint favors [-nɔk]; negative means it favors [-nɛk].

	COEFFICIENT	ESTIMATED *p*
Intercept	−2.247	< 0.0001
Natural constraints		
AGREE(back, local)	*(omitted)*	
AGREE(back, nonlocal)	3.727	< 0.0001
AGREE(front rounded, local)	*(omitted)*	
AGREE(front rounded, nonlocal)	*(omitted)*	
AGREE(front, local)	*(omitted)*	
AGREE(nonhigh front, local)	−1.583	< 0.0001
AGREE(low front, local)	−1.286	< 0.0001
AGREE(double front, local)	−2.678	< 0.0001
Unnatural constraints		
USE FRONT / bilabial ___	−1.015	< 0.0001
USE FRONT / [+cor, +son] ___	−0.597	0.0046
USE FRONT / sibilant ___	−0.537	0.0138
USE FRONT / CC ___	−0.785	< 0.0001
USE BACK / [C_0iːC_0] ___	−0.854	0.1612

TABLE 9. Mixed-effects model of the numerical rating results.

Again, all the constraints other than USE BACK / [C_0iːC_0] ___ emerge as having a statistically significant effect.

We conclude that our experiment's participants had learned the four unnatural generalizations about final consonants and applied this implicit knowledge to the novel words presented to them.

8.4. DISCUSSION. Our finding that unnatural constraints affect native-speaker judgment is the opposite of what Becker and colleagues (2007) found. This is despite the fact that our constraints, like theirs, involved a relationship between vowels and consonants not reducible to dependency in a single feature. We cannot offer any firm conclusions on why our study came out differently. One possibility is that the effects we found are subtle enough that they emerge only with a fairly large-scale experiment (we used 131 participants).[25] And of course, the particular phonological patterns examined were different (Turkish: local influence of vowels on following consonants; Hungarian: nonlocal effect of consonants on the following vowel).

The study of Moreton (2008), which likewise addressed non-same-feature V-C dependencies, actually did find a modest effect for vowel environment: this emerged in his experiment 2. Moreton suggested, therefore, that the role of naturalness in governing the responses of his subjects involved not an absolute prohibition on learning vowel-consonant dependencies (i.e. those involving distinct features) but rather a bias, in the sense laid out in §2.2. Under this view, speakers perhaps have a harder time locating the consonant environments, or they locate them but discount them to some degree because of their unnaturalness. The following section, covering modeling, represents our attempt to explore these possibilities more rigorously.

9. Modeling our results. In this final section, we attempt to develop a computational model of how the relevant aspects of Hungarian phonology are learned. The goal is to explain why Hungarian speakers, whose life experience has exposed them to the data patterns described above, would arrive at the intuitions that they displayed in our wug test. Our modeling focuses on two questions.

First, there is the problem of whether such an arbitrary set of constraints as our unnatural set is learnable at all. We assume that they would be learned through induction, and show that one inductive learning model is capable of detecting them.

Second, we are interested in to what degree constraints are overlearned or underlearned: that is, represented with a greater or lesser strength in the mental grammar than would be justified purely by the strength of the corresponding patterns in the Hungarian lexicon. We outline a method for detecting and assessing such deviations, and suggest that unnatural constraints tend to be underlearned.

9.1. Ensuring the learnability of the unnatural constraints. We ran a subset of our lexical database—the forms in the zones of variation—through the 'MinGen' (minimal generalization) inductive morphophonemic learner proposed by Albright and Hayes (2002, 2003). We equipped the MinGen learner with the phonological features mentioned in §7 and assigned it the task of finding the best set of rules for predicting the [-nɔk] and [-nɛk] suffixes from the segmental make-up of the stem. Of the four unnatural environments that our wug test indicates were productively internalized by native speakers (i.e. 10a–d), three showed up in the list of [-nɛk] environments that the algorithm learned, specifically as in 12.

(12) Unnatural harmony rules learned by the MinGen model[26]

	RULE	CORRECT/TOTAL CASES APPLICABLE
a.	$\emptyset \rightarrow$ nɛk / $\begin{bmatrix}+\text{labial}\\-\text{continuant}\end{bmatrix}$ __	38/44
b.	$\emptyset \rightarrow$ nɛk / [+sibilant] __	93/146
c.	$\emptyset \rightarrow$ nɛk / $\begin{bmatrix}+\text{coronal}\\+\text{sonorant}\end{bmatrix}$ __	127/211

[25] Post-hoc Monte Carlo simulations, fitting regression models to randomly sampled subsets of our data, suggest that about 125 subjects would be needed in order to reliably observe significant effects of all four unnatural constraints.

[26] The implemented model's output files describe natural classes in maximally verbose form; for clarity, redundant features have been removed here. Numbers are rounded, reflecting a requirement of MinGen input files.

The consonant-cluster environment of 10d is not one that MinGen readily learns, because it cannot express rules containing a sequence of natural classes. However, when we provided the model with a representation of the training data consisting solely of CV strings (cf. the 'CV' tier in some models of phonology; Clements & Keyser 1982), it readily learned the cluster environment (13).

(13) Postcluster environment learned by the MinGen model given CV information[27]

Ø → nɛk / CC __ 158/169

We note two caveats. First, the generalizations given above are not the only ones that the MinGen model picked up from the data; dozens of others were found. Many of these embody subsets of the environments in 12–13; others involve supersets or overlapping sets. In principle, we could model Hungarian learning with all of these generalizations included; however, given that our wug test examined only the specific generalizations given in 12–13, we limit ourselves to the unnatural environments already discussed here.

Second, the MinGen model could not learn the 'monosyllables with [iː]' environment for harmony (11), even when we provided it with a vowels-only tier. Whether this is an important failing would depend on further evidence, not yet in hand, that this environment is in fact internalized by native speakers.

We conclude that the unnatural constraint environments discussed here are indeed accessible from a fairly simple inductive procedure.

9.2. Modeling the underlearning and overlearning of constraints. To assess whether constraints are overlearned or underlearned, our strategy was to employ a computational model that learns a grammar for Hungarian vowel harmony on the basis of the lexical data. This grammar is then tested on the same wug test forms we gave to our human participants, permitting us to compare the strength of a constraint as it is objectively manifested in the lexicon versus the strength of a constraint as it guides human intuition.

Such modeling requires a number of initial assumptions. We first assume that our lexical database is representative of Hungarian as a whole and can serve as a reasonable stand-in for the learning experience that has been accumulated over time by a fluent adult Hungarian speaker. We use the lexical database as the training data for our learning simulation.

Second, we assume that the learning model has access to phonological constraints. The constraint set we used consisted of thirteen constraints, the eight natural constraints given in Table 2 and the five unnatural constraints in 10 and 11. We suppose that the natural constraints could be part of UG or constructed by learners on the basis of their superior phonetic properties, as in Hayes 1999 and Steriade 2008 [2001]. The unnatural constraints, if we are right that they are unnatural, can only be learned through induction, as in the preceding section.

Lastly, we must assume a framework for grammars and for learning; this is described below.

9.3. Choosing a framework. The grammatical model we choose must be able to make gradient predictions. Plausibly, these would be based on the assignment of numeri-

[27] The full environment learned was / VCC __, the result of there being no triple final clusters in the input data.

cal values to the constraints reflecting their strengths.[28] The gradient predictions would take the form of probabilities assigned to the members of a candidate set (here, trivially, the [-nɔk] and [-nɛk] suffixed versions of each stem). Such probabilities can then be matched against the gradient data seen in our wug test.

There are various theories of this type available. Hayes and Londe (2006) primarily employed stochastic optimality theory (OT), with its affiliated learning algorithm, the gradual learning algorithm (GLA; Boersma 1998, Boersma & Hayes 2001). In recent work, however, Pater (2008a) has shown that the GLA sometimes fails to find the correct grammar.[29] For this reason, we shift here to the alternative framework of maximum entropy (maxent) grammars (e.g. Goldwater & Johnson 2003, Fischer 2005, Wilson 2006, Jäger 2007, Hayes & Wilson 2008). This framework is described in the following section.

9.4. Maxent grammars. As deployed here, maxent grammars share the overall architecture of optimality theory (Prince & Smolensky 2004 [1993]): a generator (GEN) component creates a set of candidates for each input, and the grammar selects from this set. More precisely, since maxent is a probabilistic approach, it assigns a probability (often vanishingly small) to every member of the set.

In a maxent grammar the constraints are not ranked, but instead have weights, which determine their influence in assigning probabilities to candidates: the larger a constraint's weight, the lower the predicted probability of candidates that violate it. Given the weights, candidates, and violation counts, the predicted probability of each candidate x is $P(x) = \exp(-\sum_{i=1}^{N} w_i C_i(x)) / Z$. This formula is explicated in 14.

(14) Probability computation in maxent for a particular candidate x
  a. For each constraint C_i, multiply its weight w_i by the number of times the candidate violates it, $C_i(x)$.
  b. Sum the result of (a) over all N constraints. This sum, $\sum_{i=1}^{N} w_i C_i(x)$, is often called the harmony of the candidate (Legendre et al. 1992, Smolensky & Legendre 2006, Pater et al. 2007).
  c. Take e to the negative power of the harmony. Hayes and Wilson (2008) call this the maxent value of the candidate.
  d. Compute the maxent values of all candidates assigned by GEN to the input and sum them. This value is conventionally called **Z**.[30]
  e. $P(x)$, the predicted probability of candidate x, is its maxent value divided by Z.

[28] Kiparsky (2005) rejects such numerical models on grounds of insufficient restrictiveness, favoring instead the theory of Anttila (1997), in which there are no quantitative parameters, only freely ranked strata. In our efforts to model our data with this system, we found that it somewhat underperforms the maxent simulations given in §9.5 below (best correlation to wug ratings data: $r = .522$), though in fairness it has fewer degrees of freedom. More problematically, inclusion of the unnatural constraints actually decreases accuracy ($r = .507$). It appears that including the unnatural constraints in the same stratum as the violable natural ones gives them too much force. Placing them in a lower stratum, however, is equivalent to omitting them.

[29] Indeed, our own simulation may be such a case; the stringency hierarchy of Table 2 (not used by Hayes and Londe) creates a 'credit problem' (Pater 2008a, Dresher 1999), which leads the GLA under some settings to form a grammar assigning 100% [-nɔk] values to Bi and Beː forms.

[30] Where GEN creates an infinite set of candidates, some form of idealization is needed to perform the computation—for instance, setting a large but finite length limit. We presuppose here a very small set of just two candidates.

Several aspects of maxent grammars should be noted. First, the theory is in a certain sense a superset of classical optimality theory: every nonstochastic OT grammar with a finite bound on the number of violations has a maxent translation that derives the same outputs with a probability as close to 1 as is desired (Johnson 2002). Second, like all versions of HARMONIC GRAMMAR, maxent has the property of GANGING (Pater et al. 2007, Pater 2008b, Hayes & Wilson 2008): when a candidate violates two constraints together, its probability will go down more than if it violated just one. Third, maxent is affiliated with a learning procedure (Berger et al. 1996) that is proven to converge to the optimal weights (i.e. those that maximize the probability of the training data). To our knowledge, it is the only current theory of stochastic grammar that has this property.[31]

9.5. A FIRST-PASS LEARNING SIMULATION. For our maxent simulations, we developed an input file for grammar training based on our lexical database (§5). In this file, each stem included both a [-nɔk] candidate and a [-nɛk] candidate, along with the observed fraction for both of these candidates in the Google survey data. This was intended as our best available approximation of what real Hungarian speakers encounter while learning harmony. We used type frequencies, not token frequencies, since (as is usually the case in modeling intuitions based on the lexicon; Hayes & Wilson 2008:395), the type frequencies result in a better fit to the wug-test data. Our file included the violations for all thirteen natural and unnatural constraints examined here.

Submitting this file to maxent software (we used a program created by Colin Wilson and Benjamin George), we found the constraint weights given in Table 10.[32] We refer to this combination of constraints and weights as GRAMMAR L, for 'lexicon-trained'.

	WEIGHT
Natural constraints	
AGREE(back, local)	3.95
AGREE(back, nonlocal)	5.39
AGREE(front rounded, local)	3.72
AGREE(front rounded, nonlocal)	1.70
AGREE(front, local)	1.69
AGREE(nonhigh front, local)	1.46
AGREE(low front, local)	3.00
AGREE(double front, local)	4.04
Unnatural constraints	
USE FRONT / bilabial __	2.45
USE FRONT / [+cor, +son] __	1.08
USE FRONT / sibilant __	.91
USE FRONT / CC __	1.73
USE BACK / [C_0iːC_0] __	2.42

TABLE 10. Grammar L: constraints and weights.

[31] The HG-GLA algorithm, which learns NOISY HARMONIC GRAMMAR, is proven to converge for cases where there is no free variation by Boersma and Pater (2008). The question of whether it reliably converges for gradient data remains open.

[32] Our program (available from http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/) employs the Conjugate Gradient method (Press et al. 1992) to find the weights. It includes a Gaussian prior term, explained in Goldwater & Johnson 2003. We set this to be weak enough ($\sigma = 100{,}000$) to have almost nil effect on the weights.

As a minimal check on Grammar L, we assessed whether it was capable of reproducing the frequencies for the data on which it was trained. In fact, there is close agreement between its predictions and the original values: $r = .992$. This simply means that the constraint set largely sufficed to capture the generalizations present in the lexical data. When we confined the comparison to words of the form . . .BN(N) and N (as in the wug test), r emerged as .892, indicating that not only the exceptionless generalizations but also the gradient ones are well captured by these constraints.

We find that Grammar L also shows some ability to predict the wug-test data. We obtained slightly higher correlations with our '[-nɔk] advantage' scores ([-nɔk] rating minus [-nɛk] rating), and report these here. The correlation found is not as high, but still substantial: $r = .546$. This matchup indicates that at a basic level, the participants were following the law of frequency matching, projecting the patterns of the lexicon into their own wug-test intuitions.

The degree of matching can be seen in Figure 8. The model's values shown for patterns that were in the wug test are the average for the complete set of wug-test forms falling within each pattern; the model's predicted [-nɔk] rates are lower than those in the lexicon, since the wug items are loaded with [-nɛk]-preferring final consonantal environments. The model-prediction values for remaining items are averages for lexical-corpus items.

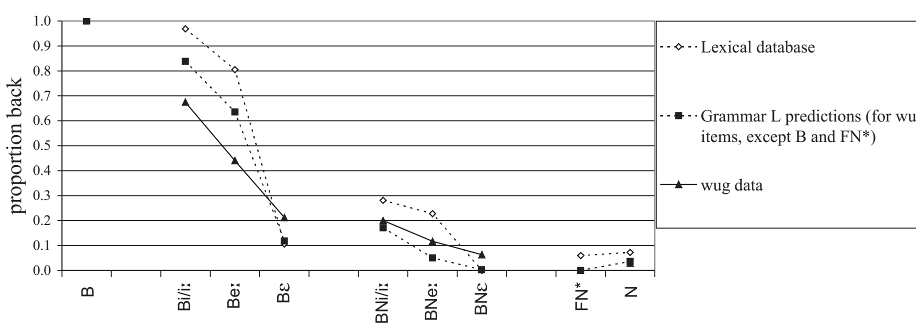


FIGURE 8. Predictions of Grammar L.

Through further modeling, we found that excluding the unnatural constraints reduces the fit to the data. Let us call the grammar with just the eight natural constraints of Table 2, and with weights again learned by the maxent algorithm from the lexical data, GRAMMAR L_{NAT}. The correlations of the predictions of Grammar L_{nat} against the wug-test data are lower, $r = .521$. This is just as we would expect, given that our statistical testing earlier demonstrated that the unnatural constraints play a role in our participants' judgments.

9.6. DETECTING OVERLEARNING AND UNDERLEARNING. In our next simulation, we trained the weights directly on the forced-choice wug-test data. This certainly does not simulate the process of language learning (children seldom access their parents' well-formedness intuitions), but rather is intended as an approximation of the final, learned state of the grammar, insofar as this can be assessed by our wug test. The purpose is to obtain weights we can compare against the weights of Grammar L, in order to detect overlearning and underlearning.

In this simulation, we left out the constraints AGREE(back, local), AGREE(front rounded, nonlocal), and AGREE(front rounded, local). Since the wug data consisted solely of BN, BNN, BBN, and N forms, these constraints are violated by neither the [-nɔk] nor the [-nɛk] candidates, so their weights are irrelevant. The grammar thus obtained, which we call GRAMMAR W (for 'wug'), is as in Table 11.

	WEIGHT
Natural constraints	
AGREE(back, local)	NA
AGREE(back, nonlocal)	3.59
AGREE(front rounded, local)	NA
AGREE(front rounded, nonlocal)	NA
AGREE(front, local)	2.24
AGREE(nonhigh front, local)	.95
AGREE(low front, local)	1.06
AGREE(double front, local)	1.94
Unnatural constraints	
USE FRONT / bilabial ___	1.03
USE FRONT / [+cor, +son] ___	.44
USE FRONT / sibilant ___	.37
USE FRONT / CC ___	.67
USE BACK / [C_0i:C_0] ___	.81

TABLE 11. Grammar W: constraint weights matched to the wug-test results.

We note the following. First, the correlation to the wug-test data obtained with Grammar W was $r = .575$: higher, but not greatly so, than Grammar L (recall $r = .546$). Some improvement is to be expected, since these weights are the mathematical optimum for the data to which they were fitted.

Second, the weights of Grammar W (Table 11) are considerably lower than those of Grammar L (Table 10): leaving out the constraints not shared by both grammars, we find that the average weight in Grammar W was 1.31 and in Grammar L was 2.42, a ratio of .54. We are not certain why the wug-fitted weights should be lower, but our guess is that it has to do with a factor mentioned in §8.3: because the wug test presented unlikely outputs to the participants (rather than seeing if they would be volunteered), such outputs were treated more favorably. In general, lowering the weights depolarizes the output distribution. Thus, the lower weights in Grammar W produce a better fit to the wug-test data.

Keeping this general effect in mind, we now have a means of roughly estimating the degree of underlearning and overlearning for the various constraints, specifically the degree of 'shrinkage' one observes for a given constraint weight in comparing Grammar L with Grammar W. Table 12 expresses this as a ratio, *Grammar W/Grammar L*, for each constraint.

In rough terms, we see that AGREE(back, nonlocal) and AGREE(nonhigh front, local) have rather middling ratios, of about .65. All the unnatural constraints' ratios are rather lower than this, around .4, and AGREE(low front, local) and AGREE(double front, local) also have rather low ratios. Lastly, AGREE(front, local) stands out for its high ratio, 1.32.[33]

Before we try to make sense of these patterns, we first propose a means of testing such differences more rigorously.

[33] The differences apparently do not arise from Grammar W and Grammar L having different constraints, a possible confound. We set up a Grammar $L_{BNN/BN/N}$, with the same constraints as Grammar W but trained using the . . .BNN, . . .BN, and N stems of the lexicon, which were the patterns used in the wug test. Comparison of Grammar W and Grammar $L_{BNN/BN/N}$ yielded ratios similar to Table 12.

	GRAMMAR L WEIGHT	GRAMMAR W WEIGHT	W/L RATIO
Natural constraints			
AGREE(back, local)	3.95	NA	
AGREE(back, nonlocal)	5.39	3.59	.67
AGREE(front rounded, local)	3.72	NA	
AGREE(front rounded, nonlocal)	1.70	NA	
AGREE(front, local)	1.69	2.24	1.32
AGREE(nonhigh front, local)	1.46	.95	.65
AGREE(low front, local)	3.00	1.06	.35
AGREE(double front, local)	4.04	1.94	.48
Unnatural constraints			
USE FRONT / bilabial __	2.45	1.03	.42
USE FRONT / [+cor, +son] __	1.08	.44	.41
USE FRONT / sibilant __	0.91	.37	.40
USE FRONT / CC __	1.73	.67	.39
USE BACK / [C_0iːC_0] __	2.42	.81	.34

TABLE 12. Comparing Grammar L and Grammar W.

9.7. ASSESSING THE CONSTRAINT WEIGHTS WITH A MONTE CARLO SIMULATION. The weights in Grammars L and W are based on different sets of words, and thus are not expected to be the same. To take an extreme example, AGREE(back, local) has a weight of 3.95 in Grammar L, which was fitted to the lexicon. But we do not expect this constraint to have a strong weight in Grammar W, no matter what the experimental data: Grammar W was fitted to a data set in which no candidates ever violated that constraint, so regardless of participant responses, the constraint's weight remains at whatever the default value has been set to (typically 0). More generally, even if participants used exactly Grammar L to generate their experimental responses, Grammar W is expected to be different from Grammar L because each constraint applies to different numbers of words in the two training sets.

Therefore, in order to determine just how different the experimental responses are from the lexicon, we cannot simply compare Grammars L and W. Instead, we must ask how the observed Grammar W weights compare to those that would be expected if experimental participants' responses faithfully reflect the lexicon: are some of the constraints reflected more weakly (or more strongly) than would be expected? Our null hypothesis is that Hungarian speakers are bias-free; that is, their responses in the wug test are based on the law of frequency matching, and deviate from it only insofar as there is noise. One might imagine a population of idealized Hungarian speakers who have learned Grammar L and use it as the basis of their vowel harmony intuitions.

In order to test the null hypothesis, we adopted a Monte Carlo approach (Metropolis & Ulam 1949), approximating the distribution of expected wug-test responses by repeatedly generating simulated sets of responses using Grammar L. Pseudo-Grammar Ws fitted to these simulated responses yield the distribution of the expected weight for each constraint, which can then be compared to its observed, Grammar W weight.

The details are as follows. Grammar L is used to assign a probability to the [-nɔk] and [-nɛk] output for each of the 1,703 wug stems. We conducted 10,000 simulated wug tests, simulating how 10,000 groups of idealized speakers would have responded to the forced-choice task. As noted above in §9.5, the observed wug responses tended to be closer to 50% [-nɔk] than the lexical rates, perhaps as a result of both choices being shown to the participants. To model this, we introduced a parameter of temperature (Ackley et al. 1985:150–52), which appears as the parameter T in equation 15, replacing the formula discussed above in 14.

(15) Maxent with temperature

$$P(x) = \exp(-\sum_{i=1}^{N} - w_i C_i(x) / \mathbf{T}) / Z'$$

where Z′ is the sum of $\exp(-\sum_{i=1}^{N} w_i C_i(x) / \mathbf{T})$ over all candidates.

If temperature is equal to 1, then the probability that the [-nɔk] candidate is chosen on any given simulation run is exactly equal to that candidate's probability under Grammar L. As the temperature climbs, the probability moves toward 0.5. With temperature of 1.5 (the value used in the simulation below), the Grammar-L generated probabilities 0.90, 0.75, 0.50, 0.25, and 0.10 are transformed to 0.81, 0.68, 0.50, 0.32, and 0.19.

At this stage we have a large population of simulated wug tests, each created with temperature 1.5 to better match the behavior of the real participants. The next step is to take each simulated wug test and construct a maxent grammar for it, with the constraints of Grammar W. Since this is done 10,000 times, this creates 10,000 values for each weight, approximating the distribution of weights that would be expected from idealized (unbiased) speakers. For example, Figure 9 gives the probability distribution for the weight of AGREE(back, nonlocal). For reference, the solid line shows the weight of this constraint in Grammar W and the dashed line shows its weight in Grammar L.

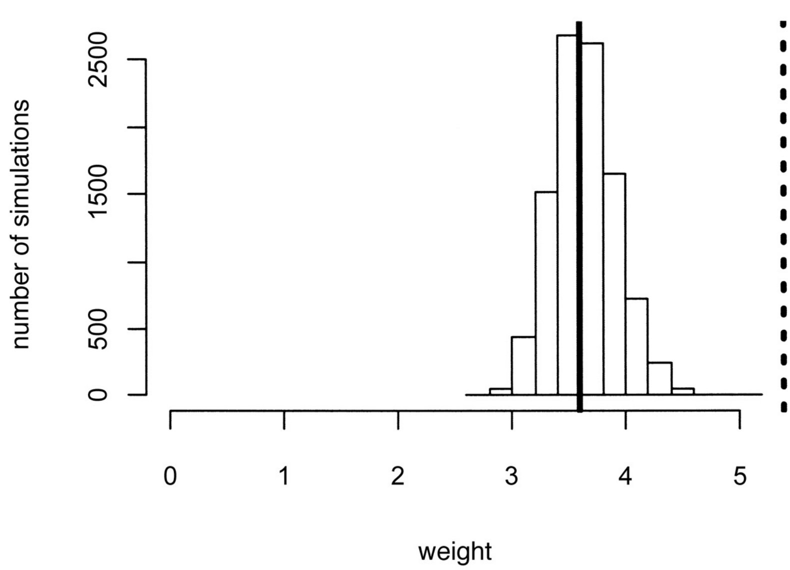


FIGURE 9. Simulated distribution of expected weights for AGREE(back, nonlocal).

We selected the temperature value of 1.5 because it causes the observed weight of this uncontroversial constraint (3.59 in Grammar W) to be at the peak of the expected distribution, as seen in Fig. 9.

The last step is to see where the weights of Grammar W, based on real speakers, fall within these null-hypothesis distributions. If the weight of a constraint C in Grammar

W falls in the left tail of the distribution, we take this as evidence that C is underlearned; that is, it has a lower weight than the null hypothesis predicts. Similarly, if it falls in the right tail, we diagnose it as overlearned. These characterizations are quantified by determining what proportion of simulations yielded a weight lower than the observed weight. For instance, if only 1% of simulations yield a weight to the left of C's Grammar W weight, then we can say that the estimated probability of observing such a low weight—this degree of underlearning—under the null hypothesis of unbiased speakers is 0.01; that is to say, $\hat{p}$ (estimated p) $= 0.01$.

The results for all constraints are shown in Figure 10. Again, the solid line shows the Grammar W weight for each constraint, and the dashed line shows the Grammar L weight.

The numerical results for under- or overlearning are given in the last two columns of Table 13. A small $\hat{p}$ value indicates underlearning, and a large $\hat{p}$ value indicates overlearning. For the unnatural constraints, we test the one-tailed hypothesis that they are underlearned, so we adopt a cut-off value of 0.05 for significance. For the natural constraints, we test the two-tailed hypothesis that they are either under- or overlearned, with cut-offs of 0.025 for underlearning and 0.975 for overlearning. For comparison, we include also the weights and ratios of Table 12.

Our test generally confirms what was said in §9.6. Agree(back, nonlocal) and Agree(nonhigh front, local) are neither overlearned or underlearned. The unnatural constraints emerge as underlearned or tending in this direction. Agree(low front, local) and Agree(double front, local) are underlearned, and Agree(front, local) is overlearned.

What might these data be telling us? Clearly, interpretations at this stage are tentative, but we offer the following conjectures.

First, the underlearning of the unnatural constraints may be reflecting a naturalness bias, of the sort suggested by Wilson (2006) and Moreton (2008).

Second, concerning differences among the natural constraints, we think it is useful to make a distinction between harmony in its raw form—backness harmony, skipping across neutral vowels—and the more subtle modulations of it seen in the height and count effects. Specifically, the details of the height effect are not reflected in the wug data. This can be seen graphically in Fig. 7, where the lexicon shows a larger difference between low and mid N triggers than between mid and high, but the wug responses—including the responses obtained by Hayes and Londe—are more evenly spaced. This discrepancy is robust in that it occurs in both BN and BNN words. Correspondingly, our simulation finds that Agree(low front, local) was underlearned, meaning that in Grammar W the additional [-nɛk]-favoring effect of a low N vowel as compared to a mid N vowel, though substantial, is smaller than it is in the lexicon. The underlearning of Agree(double front, local) likewise reflects the lesser salience of the count effect for speakers, relative to its prevalence in the lexicon. Looking again at Fig. 7, this underlearning too is robust, in that it occurs in both Bi/iː vs. BNi/iː and Beː vs. BNeː (the difference between Bɛ and BNɛ is small to begin with in the lexicon). The underlearning is largely in the direction of BN words taking more [-nɛk] suffixes, rather than BNN words taking fewer; this is reflected in the overlearning seen for Agree(front, local).

What may be emerging is a simplicity bias: that is, a bias for phonological generalizations that embody relatively few segments and features. The basic harmony constraints refer only to backness (of trigger and target), whereas the elaborated constraints refer in addition to height or to the number of triggers present. The data do not permit any more detailed assessment of this hypothesis, which our experiment was not designed

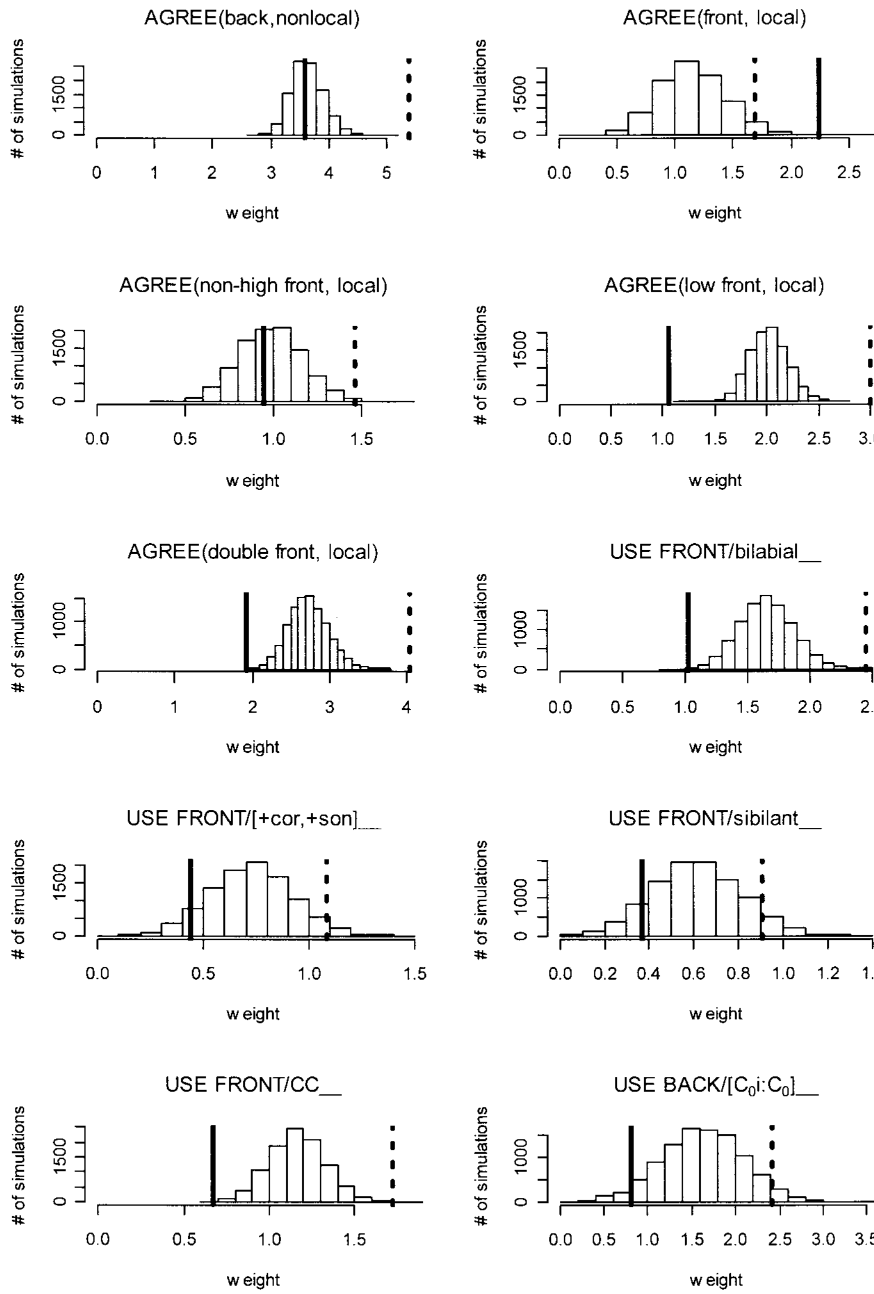


FIGURE 10. Simulated distributions of expected weights for all Grammar W constraints.

	GRAMMAR L	GRAMMAR W	W/L RATIO	MONTE CARLO TEST	$\hat{p}$
Natural constraints					
AGREE(back, local)	3.95	NA			
AGREE(back, nonlocal)	5.39	3.59	0.67	neither overlearned nor underlearned	0.448
AGREE(front rounded, local)	3.72	NA			
AGREE(front rounded, nonlocal)	1.70	NA			
AGREE(front, local)	1.69	2.24	1.32	overlearned	0.999
AGREE(nonhigh front, local)	1.46	.95	0.65	neither overlearned nor underlearned	0.421
AGREE(low front, local)	3.00	1.06	0.35	underlearned	0.0000
AGREE(double front, local)	4.04	1.94	0.48	underlearned	< 0.0001
Unnatural constraints					
USE FRONT / bilabial __	2.45	1.03	0.42	underlearned	< 0.001
USE FRONT / [+cor, +son] __	1.08	.44	0.41	nonsignificantly underlearned	0.069
USE FRONT / sibilant __	.91	.37	0.40	nonsignificantly underlearned	0.103
USE FRONT / CC __	1.73	.67	0.39	underlearned	< 0.001
USE BACK / [C_0iːC_0] __	2.42	.81	0.34	underlearned	0.043

TABLE 13. Monte Carlo test of underlearning and overlearning.

to test. The idea of a simplicity bias in phonology has a long history, going back to the feature-counting evaluation metric proposed in Chomsky & Halle 1968. We believe that further study of the kind carried out here could shed further light on how phonological learning is constrained by simplicity considerations.

It also seems possible that a well-supported theory of simplicity bias might ultimately explain the underlearning of our unnatural constraints as well. Again, careful assessment of this possibility would require much broader data than we have at present.

9.8. ANALOGY? Before concluding, it is appropriate to address one other possibility, which would render our experimental results compatible with the idea that phonology is exclusively based on a universal set of natural constraints. The idea is that what we have posited as 'unnatural constraints' are not grammatical constraints at all, but rather are only approximations for what in reality is analogy. This idea has been repeatedly proposed (Ohala 1974, Daelemans & van den Bosch 2005, Skousen et al. 2002, and, as a straw man, Albright & Hayes 2003): when speakers must make a grammatical choice for a novel stem (such as that of [-nɔk] vs. [-nɛk]), they consult their lexicons, locating stems that are phonetically similar to the stems in question, and the novel stem is then classified so as to behave like the preponderant similar stem(s).

In the present context, the key idea would be that such analogical effects have a somewhat weaker influence on native-speaker judgment than true grammatical constraints. The view that grammatical rules/constraints work in tandem with analogy is a hallmark of the 'dual mechanism' theory of grammar; see Pinker & Prince 1988, Pinker 1999, and for further discussion Albright & Hayes 2003.

We addressed the analogy hypothesis using the well-known TiMBL model of analogy (Daelemans & van den Bosch 2005, http://ilk.uvt.nl/timbl/). We adopted the same phonological coding that the TiMBL 6.1 reference guide employs for its illustrative Dutch diminutive simulation. Exploring various combinations of learning parameters, we found the best TiMBL correlation to the rating data to be .500. This is substantially below what we could obtain with the best temperature-adjusted version of Grammar L, which was $r = .559$.

We evaluated the dual-mechanism hypothesis as follows. We started with a grammar containing only the eight natural constraints of Table 2, optimally adjusted for temperature. This grammar's predicted scores correlated somewhat less with the wug data, $r = .529$. We thought that by blending the scores of this all-natural grammar with the scores of the analogical system we might obtain the best of both; and we tried this by examining all possible weighted averages of their predictions. However, the best of these (81% all-natural constraints, 19% analogy) received a correlation of only .532. Thus, adding in unnatural constraints increases the correlation by .030, whereas folding in analogy increases it by only .003. In sum, our modeling efforts so far offer modest support for an approach in which the effects of the unnatural environments are embodied by highly specific grammatical constraints, not analogy.

10. SUMMARY AND CONCLUSION. Our original interest in this topic arose from the issue of the surfeit of the stimulus: can a language include data patterns to which native speakers are oblivious because the patterns are unlearnable? The question is central to current UG theorizing, since it sharply distinguishes strong versions of the UG hypothesis from theories that also permit a degree of inductive learning. To this end, we located a group of arguably unnatural constraints that hold true statistically of the Hungarian lexicon, then checked the ability of Hungarian speakers to learn these unnatural constraints by conducting a wug test.

Our test got a different result from the earlier findings of Becker and colleagues (2007) on Turkish: the behavior of our participants clearly shows the effect of the unnatural constraints. The ability of speakers to learn unnatural constraints has been asserted before (§4), and our experimental findings further support such claims. Why the two experiments obtained contrasting results must remain a topic for further research.

Our results are also of interest because, although they generally reflected the law of frequency matching, there were notable deviations. We examined these deviations more closely through modeling: we fitted maxent grammars to both the lexical data and the wug-test responses, and, in a Monte Carlo simulation, compared the constraint weights fitted to the observed data against those expected under the null hypothesis that responses are faithfully generated by the lexicon-trained grammar. We found that unnatural constraints were underlearned, giving modest support to the idea (Wilson 2003a, Moreton 2008) that people show a learning bias against unnatural constraints. We also found underlearning for some of the natural constraints, however, namely those responsible for the count effect and part of the height effect. This suggests perhaps a role for a simplicity bias as well. We suggest that study of underlearning and overlearning along the lines we have laid out, for a variety of phonological systems, could help shed further light on these issues.

REFERENCES

ACKLEY, DAVID; GEOFFREY HINTON; and TERRENCE SEJNOWSKI. 1985. A learning algorithm for Boltzmann machines. *Cognitive Science* 9.147–69.

ALBRIGHT, ADAM. 2000. The productivity of infixation in Lakhota. Los Angeles: University of California, Los Angeles, MS.

ALBRIGHT, ADAM. 2002. *The identification of bases in morphological paradigms*. Los Angeles: University of California, Los Angeles dissertation.

ALBRIGHT, ADAM. 2003. A quantitative study of Spanish paradigm gaps. *West Coast Conference on Formal Linguistics* 22.1–20.

ALBRIGHT, ADAM. 2007. Natural classes are not enough: Biased generalization in novel onset clusters. Cambridge, MA: MIT, MS.

ALBRIGHT, ADAM, and BRUCE HAYES. 2002. Modeling English past tense intuitions with minimal generalization. *Proceedings of the 2002 Workshop on Morphological Learning*, ed. by Mike Maxwell, 58–69. Philadelphia: Association for Computational Linguistics.

Albright, Adam, and Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90.119–61.

Alderete, John. 1998. *Morphologically governed accent in optimality theory*. Amherst: University of Massachusetts dissertation. [Published, London: Routledge, 2001.]

Anderson, Lloyd B. 1980. Using asymmetrical and gradient data in the study of vowel harmony. *Issues in vowel harmony*, ed. by Robert M. Vago, 271–340. Amsterdam: John Benjamins.

Anderson, Stephen R. 1981. Why phonology isn't 'natural'. *Linguistic Inquiry* 12.493–539.

Anttila, Arto. 1997. *Variation in Finnish phonology and morphology*. Stanford, CA: Stanford University dissertation.

Baayen, R. Harald. 2007. LanguageR: Data sets and functions with 'Analyzing linguistic data: A practical introduction to statistics'. R package version 0.2. Online: http://cran.r-project.org/web/packages/languageR/index.html.

Bach, Emmon, and Robert T. Harms. 1972. How do languages get crazy rules? *Linguistic change and generative theory*, ed. by Robert P. Stockwell and Ronald K. S. Macaulay, 1–21. Bloomington: Indiana University Press.

Bailey, Todd M., and Ulrike Hahn. 2001. Determinants of wordlikeness: Phonotactics or lexical neighborhoods. *Journal of Memory and Language* 44.568–91.

Bailey, Todd M.; K. Plunkett; and E. Scarpa. 1999. A cross-linguistic study in learning prosodic rhythms: Rules, constraints and similarity. *Language and Speech* 42.1–38.

Baroni, Marco. 1999. The representation of prefixed forms in the Italian lexicon: Evidence from the distribution of intervocalic [s] and [z] in northern Italian. *Yearbook of Morphology* 1999.121–52.

Baroni, Marco. 2001. How do languages get crazy constraints? Phonetically-based phonology and the evolution of the Galeata Romagnolo vowel system. *UCLA Working Papers in Linguistics* 7.152–78.

Bates, Douglas. 2007. lme4: Linear mixed-effects models using S4 classes. R package version 0.99875–6.

Becker, Michael. 2009. *Phonological trends in the lexicon: The role of constraints*. Amherst: University of Massachusetts, Amherst dissertation.

Becker, Michael; Nihan Ketrez; and Andrew Nevins. 2007. The surfeit of the stimulus: Grammatical biases filter lexical statistics in Turkish voicing deneutralization. Amherst: University of Massachusetts, Amherst; New Haven, CT: Yale University; and Cambridge, MA: Harvard University, ms. Online: http://ling.auf.net/lingbuzz/000544.

Beddor, Patrice; Rena Krakow; and Stephanie Lindemann. 2001. Patterns of perceptual compensation and their phonological consequences. In Hume & Johnson, 55–78.

Benus, Stefan. 2005. *Dynamics and transparency in vowel harmony*. New York: New York University dissertation.

Benus, Stefan, and Adamantios Gafos. 2005. Qualitative and quantitative aspects of vowel harmony: a dynamics model. *Proceedings of the 27th annual conference of the Cognitive Science Society*, ed. by Bruno G. Bara, Lawrence Barsalou, and Monica Bucciarelli, 226–31. Mahwah, NJ: Lawrence Erlbaum.

Benus, Stefan, and Adamantios Gafos. 2007. Articulatory characteristics of Hungarian transparent vowels. *Journal of Phonetics* 35.271–300.

Berent, Iris; Donca Steriade; Tracy Lennertz; and Vered Vaknin. 2007. What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* 104.591–630.

Berent, Iris; Tracy Lennertz; Jongho Jun; Miguel A. Moreno; and Paul Smolensky. 2008. Language universals in human brains. *Proceedings of the National Academy of Sciences* 105.5321–25.

Berger, Adam; Stephen Della Pietra; and Vincent Della Pietra. 1996. A maximum entropy approach to natural language processing. *Computational Linguistics* 22.39–71.

Berkley, Deborah. 2000. *Gradient OCP effects*. Evanston, IL: Northwestern University dissertation.

Berko, Jean. 1958. The child's learning of English morphology. *Word* 14.150–77.

Blevins, Juliette. 2004. *Evolutionary phonology*. Cambridge: Cambridge University Press.

Blevins, Juliette. 2008. Consonant epenthesis: Natural and unnatural histories. *Language universals and language change*, ed. by Jeff Good, 79–109. Oxford: Oxford University Press.

Boersma, Paul. 1998. How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences* 21.43–58.

Boersma, Paul, and Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32.45–86.

Boersma, Paul, and Joe Pater. 2008. Convergence properties of a gradual learning algorithm for harmonic grammar. Amsterdam: University of Amsterdam, and Amherst: University of Massachusetts, ms. Rutgers Optimality Archive 970.

Buckley, Eugene. 2003. Children's unnatural phonology. *Berkeley Linguistics Society* 29.523–34.

Campbell, Lyle. 1980. The psychological and sociological reality of Finnish vowel harmony. *Issues in vowel harmony*, ed. by Robert Vago, 245–70. Amsterdam: John Benjamins.

Carpenter, Angela C. 2005. Acquisition of a natural vs. unnatural stress system. *Proceedings of the 29th annual Boston University Conference on Language Development*, ed. by Alejna Brugos, Manuella R. Clark-Cotton, and Seungwan Ha, 134–43. Somerville, MA: Cascadilla.

Chomsky, Noam. 1980. *Rules and representations*. New York: Columbia University Press.

Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.

Clements, George N. 1976. Neutral vowels in Hungarian vowel harmony: An autosegmental interpretation. *North East Linguistic Society* 7.49–64.

Clements, George N. 1991. Place of articulation in consonants and vowels: A unified theory. *Working Papers of the Cornell Phonetics Laboratory* 5.77–123.

Clements, George N., and Samuel Jay Keyser. 1982. *CV phonology: A generative theory of the syllable*. Cambridge, MA: MIT Press.

Clements, George N., and Engin Sezer. 1982. Vowel and consonant disharmony in Turkish. *The structure of phonological representations*, part 2, ed. by Harry van der Hulst and Norval Smith, 213–55. Dordrecht: Foris.

Coleman, John, and Janet Pierrehumbert. 1997. Stochastic phonological grammars and acceptability. *Computational phonology: The third meeting of the ACL Special Interest Group in Computational Phonology*, 49–56. Somerset, NJ: Association for Computational Linguistics.

Daelemans, Walter, and Antal van den Bosch. 2005. *Memory-based language processing*. Cambridge: Cambridge University Press.

De Lacy, Paul. 2004. Markedness conflation in optimality theory. *Phonology* 21.145–99.

Dienes, Péter. 1997. Hungarian neutral vowels. *The Odd Yearbook* 4.151–80.

Dresher, B. Elan. 1999. Charting the learning path: Cues to parameter setting. *Linguistic Inquiry* 30.27–67.

Dresher, B. Elan, and Xi Zhang. 1996. Labial harmony in written Manchu. *Saksaha: A review of Manchu Studies* 1.13–24.

Eddington, David. 1996. Diphthongization in Spanish derivational morphology: An empirical investigation. *Hispanic Linguistics* 8.1–35.

Eddington, David. 1998. Spanish diphthongization as a non-derivational phenomenon. *Rivista di Linguistica* 10.335–54.

Eddington, David. 2004. *Spanish phonology and morphology: Experimental and quantitative perspectives*. Amsterdam: John Benjamins.

Ernestus, Mirjam, and R. Harald Baayen. 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79.5–38.

Esztergár, Marianne. 1971. *A generative phonology of nouns and vowel harmony in Hungarian*. San Diego: University of California, San Diego dissertation.

Finley, Sarah. 2008. *The formal and cognitive restrictions on vowel harmony*. Baltimore: Johns Hopkins University dissertation.

Finley, Sarah, and William Badecker. 2007. Towards a substantively biased theory of learning. *Berkeley Linguistics Society* 33, to appear.

Fischer, Markus. 2005. A Robbins-Monro type learning algorithm for an entropy maximizing version of stochastic optimality theory. Berlin: Humboldt University masters thesis. Rutgers Optimality Archive 767.

Flemming, Edward. 2003. The relationship between coronal place and vowel backness. *Phonology* 20.335–73.

Frisch, Stefan, and Bushra Zawaydeh. 2001. The psychological reality of OCP-Place in Arabic. *Language* 77.91–106.

Gallistel, Charles R. 1990. *The organization of learning*. Cambridge, MA: Bradford Books/MIT Press.

Gelman, Andrew; Alek Jakulin; Maria Grazia Pittau; and Yu-Sung Su. 2008. A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics* 2.1360–83.

Gelman, Andrew; Yu-Sung Su; Masanao Yajima; Jennifer Hill; Maria Grazia Pittau; Jouni Kerman; and Tian Zheng. 2009. arm: Data analysis using regression and multilevel/hierarchical models. R package. Online: http://cran.r-project.org/web/packages/arm/index.html.

Gilkerson, Jill. 2005. Categorical perception of natural and unnatural categories: Evidence for innate category boundaries. *UCLA Working Papers in Linguistics (Papers in psycholinguistics 2)* 13.34–58.

Goldrick, Matthew, and Meredith Larson. 2010. Constraints on the acquisition of variation. *Papers in laboratory phonology 10: Variation, detail and representation*, ed. by C. Fougeron and MariaPaola D'Imperio. Berlin: Mouton de Gruyter, to appear.

Goldwater, Sharon, and Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, ed. by Jennifer Spenader; Anders Eriksson, and Östen Dahl, 111–20. Stockholm: Stockholm University Department of Linguistics.

Gordon, Matthew. 2002. A factorial typology of quantity insensitive stress. *Natural Language and Linguistic Theory* 20.491–552.

Graff, Peter. 2007. The nature of systematic sound change: An experiment. Edinburgh: University of Edinburgh masters thesis.

Griner, Barry. 2001. Productivity of Japanese verb tense inflection: A case study. Los Angeles: University of California, Los Angeles masters thesis.

Halle, Morris, and Jean-Roger Vergnaud. 1987. *An essay on stress*. Cambridge, MA: MIT Press.

Hansson, Gunnar. 2007. On the evolution of consonant harmony: The case of secondary articulation agreement. *Phonology* 24.77–120.

Hare, Mary. 1990. The role of similarity in Hungarian vowel harmony: A connectionist account. *Connection Science* 2.123–50.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Hayes, Bruce. 1999. Phonetically-driven phonology: The role of optimality theory and inductive grounding. *Functionalism and formalism in linguistics*, vol. 1, ed. by Michael Darnell, Edith Moravcsik, Michael Noonan, Frederick Newmeyer, and Kathleen Wheatly, 243–85. Amsterdam: John Benjamins.

Hayes, Bruce, and Zsuzsa Cziráky Londe. 2006. Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* 23.59–104.

Hayes, Bruce; Robert Kirchner; and Donca Steriade (eds.) 2004. *Phonetically based phonology*. Cambridge: Cambridge University Press.

Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39.379–440.

Hellberg, Staffan. 1978. Unnatural phonology. *Journal of Linguistics* 14.157–77.

Holt, Lori; Andrew Lotto; and Randy Diehl. 2004. Auditory discontinuities interact with categorization: Implications for speech perception. *Journal of the Acoustical Society of America* 116.1763–73.

Hombert, Jean-Marie; John Ohala; and William Ewan. 1979. Phonetic explanations for the development of tones. *Language* 55.37–58.

Hooper, Joan. 1976. *An introduction to natural generative phonology*. New York: Academic Press.

Hudson Kam, Carla, and Elissa Newport. 2005. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development* 1.151–95.

Hume, Elizabeth. 1992. *Front vowels, coronal consonants and their interaction in nonlinear phonology*. Ithaca, NY: Cornell University dissertation.

Hume, Elizabeth, and Keith Johnson (eds.) 2001. *The role of speech perception in phonology*. San Diego: Academic Press.

Idsardi, William. 2006. A Bayesian approach to loanword adaptations. Poster presented at the annual meeting of the Linguistic Society of America, Albuquerque, NM.

Inkelas, Sharon; Aylin Kuntay; John Lowe; Orhan Orgun; and Ronald Sprouse. 2000. Turkish Electronic Living Lexicon (TELL). Database. Online: http://linguistics.berkeley.edu/TELL/.

Itô, Junko. 1984. Melodic dissimilation in Ainu. *Linguistic Inquiry* 15.505–13.

Jäger, Gerhard. 2007. Maximum entropy models and stochastic optimality theory. *Architectures, rules, and preferences: Variations on themes by Joan W. Bresnan*, ed. by Annie Zaenen, Jane Simpson, Tracy Holloway King, Jane Grimshaw, Joan Maling, and Chris Manning, 467–79. Stanford, CA: CSLI Publications.

Johnson, Mark. 2002. Optimality-theoretic lexical functional grammar. *The lexical basis of sentence processing: Formal, computational and experimental issues*, ed. by Paula Merlo and Susan Stevenson, 59–74. Amsterdam: John Benjamins.

Jun, Jongho, and Jeehun Lee. 2007. Multiple stem-final variants in Korean native nouns and loanwords. *Journal of the Linguistic Society of Korea* 47.159–87.

Jusczyk, Peter W.; Paul Smolensky; and Theresa Allocco. 2002. How English-learning infants respond to markedness and faithfulness constraints. *Language Acquisition* 10.31–73.

Jusczyk, Peter W.; Paul Smolensky; Karen Arnold; and Elliott Moreton. 2003. Acquisition of nasal place assimilation by 4.5-month-old infants. *Jusczyk lab final report*, ed. by D. Houston, A. Seidl, G. Hollich, E. Johnson, and A. Jusczyk. Online: http://hincapie.psych.purdue.edu/Jusczyk/pdf/Nasal.pdf.

Kager, René, and Nine Elenbaas. 1999. Ternary rhythm and the lapse constraint. *Phonology* 16.273–329.

Kálmán, Béla. 1972. Hungarian historical phonology. *The Hungarian language*, ed. by Loránd Benkő and Samu Imre, 49–83. Budapest: Akadémiai Kiadó.

Kaun, Abigail. 1995. *The typology of rounding harmony: An optimality theoretic approach*. Los Angeles: University of California, Los Angeles dissertation.

Kaun, Abigail. 2004. The phonetic foundations of the rounding harmony typology. In Hayes et al., 87–116.

Kawahara, Shigeto. 2006. Mimetic gemination in Japanese: A challenge for evolutionary phonology. *Theoretical Linguistics* 32.411–24.

Kawahara, Shigeto. 2008. Phonetic naturalness and unnaturalness in Japanese loanword phonology. *Journal of East Asian Linguistics* 17.317–30.

Kiparsky, Paul. 2005. Where stochastic OT fails: A discrete model of metrical variation. *Berkeley Linguistics Society* 31.409–30.

Kontra, Miklós, and Catherine Ringen. 1986. Hungarian vowel harmony: The evidence from loanwords. *Ural-Altaische Jahrbücher* 58.1–14.

Koo, Hahn, and Jennifer Cole. 2006. On learnability and naturalness as constraints on phonological grammar. *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics*, ed. by Antonis Botinis, 174–77. Athens: University of Athens.

Kornai, András. 1991. Hungarian vowel harmony. *Approaches to Hungarian 3*, ed. by István Kenesei, 183–240. Szeged: Attila József University.

Labov, William. 1994. *Principles of linguistic change, vol. 1: Internal factors*. Oxford: Blackwell.

Legendre, Géraldine; Yoshiro Miyata; and Paul Smolensky. 1992. Harmonic grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. (Technical report 92–16.) Boulder: Institute of Cognitive Science, University of Colorado.

Lombardi, Linda. 1999. Positional faithfulness and voicing assimilation in optimality theory. *Natural Language and Linguistic Theory* 17.267–302.

Magdics, Klara. 1997. *Studies in the acoustic characteristics of Hungarian speech sounds.* Richmond, UK: Curzon.

McCarthy, John. 1979. *Formal problems in Semitic phonology and morphology*. Cambridge, MA: MIT dissertation.

McCarthy, John. 2004. Headed spans and autosegmental spreading. Amherst, MA: University of Massachussetts, Amhert, ms.

Metropolis, Nicholas, and Stanislaw Ulam. 1949. The Monte Carlo method. *Journal of the American Statistical Association* 44.335–41.

Moreton, Elliott. 2008. Analytic bias and phonological typology. *Phonology* 25.83–128.

Moreton, Elliott. 2009. Underphonologization and modularity bias. *Phonological argumentation: Essays on evidence and motivation*, ed. by Steve Parker. London: Equinox, to appear.

Myers, Scott. 2002. Gaps in factorial typology: The case of voicing in consonant clusters. Austin: University of Texas, ms.

Nevins, Andrew, and Bert Vaux. 2007. Underlying representations that do not minimize grammatical violations. *Freedom of analysis?*, ed. by Sylvia Blaho, Patrik Bye, and Martin Krämer, 35–62. Berlin: Mouton de Gruyter.

Odden, David. 2007. The unnatural tonology of Zina Kotoko. *Tones and tunes, vol. 1: Typological studies in word and sentence prosody*, ed. by Tomas Riad and Carlos Gussenhoven, 63–89. Berlin: Mouton de Gruyter.

Ohala, John J. 1974. Experimental historical phonology. *Historical linguistics 2: Theory and description in phonology*, ed. by John M. Anderson and Charles Jones, 353–87. Amsterdam: North-Holland.

Ohala, John J. 1981. The listener as a source of sound change. *Chicago Linguistic Society* 17.2.178–203.

Öhman, Sven. 1966. Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America* 39.151–68.

Pater, Joe. 2008a. Gradual learning and convergence. *Linguistic Inquiry* 39.334–45.

Pater, Joe. 2008b. Optimization and linguistic typology. Amherst: University of Massachusetts, ms.

Pater, Joe; Rajesh Bhatt; and Christopher Potts. 2007. Linguistic optimization. Amherst: University of Massachusetts, ms.

Pater, Joe, and Anne-Michelle Tessier. 2003. Phonotactic knowledge and the acquisition of alternations. *Proceedings of the 15th International Congress on Phonetic Sciences*, ed. by María-Josep Solé, Daniel Recasens, and Joaquín Romero, 1177–80. Barcelona: Casual Productions.

Pater, Joe, and Anne-Michelle Tessier. 2006. L1 phonotactic knowledge and the L2 acquisition of alternations. *Inquiries in linguistic development: Studies in honor of Lydia White*, ed. by Roumyana Slabakova, Silvina A. Montrul, and Philippe Prévost, 115–31. Amsterdam: John Benjamins.

Peperkamp, Sharon, and Emmanuel Dupoux. 2007. Learning the mapping from surface to underlying representations in an artificial language. *Laboratory phonology 9*, ed. by Jennifer Cole and José Ignacio Hualde, 315–38. Berlin: Mouton de Gruyter.

Peperkamp, Sharon; Katrin Skoruppa; and Emmanuel Dupoux. 2006. The role of phonetic naturalness in phonological rule acquisition. *Proceedings of the 30th annual Boston University Conference on Language Development*, ed. by David Bamman, Tatiana Magnitskaia, and Colleen Zaller, 464–75. Somerville, MA: Cascadilla.

Pertsova, Katya. 2004. Distribution of genitive plural allomorphs in the Russian lexicon and in the internal grammar of native speakers. Los Angeles: University of California, Los Angeles masters thesis.

Pertz, Doris L., and Thomas G. Bever. 1975. Sensitivity to phonological universals in children and adults. *Language* 51.149–62.

Pierrehumbert, Janet. 2006. The statistical basis of an unnatural alternation. *Laboratory phonology 8: Varieties of phonological competence*, ed. by Louis Goldstein, D. H. Whalen, and Catherine Best, 81–106. Berlin: Mouton de Gruyter.

Pinker, Steven. 1999. *Words and rules: The ingredients of language*. New York: Basic Books.

Pinker, Steven, and Alan S. Prince. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28.73–193.

Press, William H.; Brian P. Flannery; Saul A. Teukolsky; and William T. Vetterling. 1992. *Numerical recipes in C: The art of scientific computing*. Cambridge: Cambridge University Press.

PRINCE, ALAN. 1997. Stringency and anti-Paninian hierarchies. Talk given at Cornell University, Ithaca, NY. Online: http://ling.rutgers.edu/people/faculty/prince.html.

PRINCE, ALAN, and PAUL SMOLENSKY. 2004 [1993]. *Optimality theory: Constraint interaction in generative grammar*. Cambridge, MA: Blackwell. [Technical report CU-CS-696–93, Department of Computer Science, University of Colorado at Boulder, and Technical report TR-2, Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, NJ, 1993.]

PULLUM, GEOFFREY K., and BARBARA C. SCHOLTZ. 2002. Empirical assessment of stimulus poverty arguments. *The Linguistic Review* 19.9–50.

PYCHA, ANNE; PAWEL NOWAK; EURIE SHIN; and RYAN SHOSTED. 2003. Phonological rule-learning and its implications for a theory of vowel harmony. *West Coast Conference on Formal Linguistics* 22.423–35.

PYCHA, ANNE; EURIE SHIN; and RYAN SHOSTED. 2006. Directionality of assimilation in consonant clusters: An experimental approach. *UC Berkeley Phonology Lab Annual Report*, 152–66.

QUENÉ, HUGO, and HUUB VAN DEN BERGH. 2004. On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication* 43.103–21.

R DEVELOPMENT CORE TEAM. 2007. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. http://www.R-project.org.

RINGEN, CATHERINE O. 1975. *Vowel harmony: Theoretical implications*. Bloomington, IN: Indiana University dissertation. [Published, New York: Garland, 1988.]

RINGEN, CATHERINE O., and ORVOKKI HEINÄMÄKI. 1999. Variation in Finnish vowel harmony: An OT account. *Natural Language and Linguistic Theory* 17.303–37.

RINGEN, CATHERINE O., and ROBERT M. VAGO. 1995. A constraint based analysis of Hungarian vowel harmony. *Approaches to Hungarian: Levels and structures*, vol. 5, ed. by István Kenesei, 309–19. Szeged: Attila József University.

RINGEN, CATHERINE O., and ROBERT M. VAGO. 1998. Hungarian vowel harmony in optimality theory. *Phonology* 15.393–416.

SAFFRAN, JENNY R., and ERIK D. THIESSEN. 2003. Pattern induction by infant language learners. *Developmental Psychology* 39.484–94.

SCHANE, SANFORD. 1972. Natural rules in phonology. *Linguistic change and generative theory*, ed. by Robert P. Stockwell and Ronald K. S. Macaulay, 199–229. Bloomington: Indiana University Press.

SCHANE, SANFORD; BERNARD TRANEL; and HARLAN LANE. 1974. On the psychological reality of a natural rule of syllable structure. *Cognition* 3.351–58.

SEIDL, AMANDA, and EUGENE BUCKLEY. 2005. On the learning of arbitrary phonological rules. *Language Learning and Development* 1.289–316.

SILVERMAN, DANIEL. 2006. The diachrony of labiality in Trique, and the functional relevance of gradience and variation. *Laboratory phonology 8: Varieties of phonological competence*, ed. by Louis Goldstein, D. H. Whalen, and Catherine Best, 133–52. Berlin: Mouton de Gruyter.

SIPTÁR, PÉTER, and MIKLÓS TÖRKENCZY. 2000. *The phonology of Hungarian*. Oxford: Oxford University Press.

SKOUSEN, ROYAL; DERYLE LONDALE; and DILWORTH B. PARKINSON. 2002. *Analogical modeling: An exemplar-based approach to language.* Amsterdam: John Benjamins.

SMOLENSKY, PAUL, and GÉRALDINE LEGENDRE. 2006. *The harmonic mind*. Cambridge, MA: MIT Press.

STAMPE, DAVID. 1973. *A dissertation on natural phonology*. Chicago: University of Chicago dissertation. [Published, New York: Garland, 1979.]

STERIADE, DONCA. 2008 [2001]. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. *The nature of the word: Studies in honor of Paul Kiparsky*, ed. by Kristin Hanson and Sharon Inkelas, 151–80. Cambridge, MA: MIT Press.

SVANTESSON, JAN-OLOF; ANNA TSENDINA; ANASTASIA MUKHANOVA KARLSSON; and VIVA FRANZÉN. 2005. *The phonology of Mongolian*. Oxford: Oxford University Press.

TESAR, BRUCE, and PAUL SMOLENSKY. 1998. Learnability in optimality theory. *Linguistic Inquiry* 29.229–68.

TESAR, BRUCE, and PAUL SMOLENSKY. 2000. *Learnability in optimality theory*. Cambridge, MA: MIT Press.

THATTE, VICTORIA. 2007. Phonetic naturalness as a learning bias in phonological acquisition. Los Angeles: University of California, Los Angeles masters thesis.

TRÓN, VIKTOR. 2007. Vowels and consonants in harmony: Naturalness in phonology and the role of consonants in Hungarian front/back harmony. Abstract for the 8th International Conference on the Structure of Hungarian, CUNY.

VAGO, ROBERT M. 1974. *Hungarian generative phonology*. Cambridge, MA: Harvard University dissertation.

VAGO, ROBERT M. 1976. Theoretical implications of Hungarian vowel harmony. *Linguistic Inquiry* 7.243–63.

VAGO, ROBERT M. 1980. *The sound pattern of Hungarian*. Washington, DC: Georgetown University Press.

VAGO, ROBERT M. 1994. Vowel harmony. *The encyclopedia of language and linguistics*, ed. by Ronald E. Asher, 4954–58. Oxford: Pergamon.

VANCE, TIMOTHY J. 1987. *An introduction to Japanese phonology*. Albany: State University of New York Press.

WALKER, RACHEL. 2001. Round licensing, harmony, and bisyllabic triggers in Altaic. *Natural Language and Linguistic Theory* 19.827–78.

WALKER, RACHEL. 2005. Weak triggers in vowel harmony. *Natural Language and Linguistic Theory* 23.917–89.

WILSON, COLIN. 2003a. Experimental investigation of phonological naturalness. *West Coast Conference on Formal Linguistics* 22.533–46.

WILSON, COLIN. 2003b. Analyzing unbounded spreading with constraints: Marks, targets, and derivations. Los Angeles: University of California, Los Angeles, MS.

WILSON, COLIN. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30.945–82.

ZHANG, JIE, and YUWEN LAI. 2006. Testing the role of phonetic naturalness in Mandarin tone sandhi. *Kansas Working Papers in Linguistics* 28.65–126.

ZHANG, JIE; YUWEN LAI; and CRAIG SAILOR. 2007. Effects of phonetics and frequency on the production of Taiwanese tone sandhi. *Chicago Linguistic Society* 43.273–86.

ZHANG, XI. 1996. *Vowel systems of the Manchu-Tungus languages of China*. Toronto: University of Toronto dissertation.

ZURAW, KIE. 2000. *Patterned exceptions in phonology*. Los Angeles: University of California, Los Angeles dissertation.

ZURAW, KIE. 2007. The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from Tagalog. *Language* 83.277–316.

Hayes, Zuraw, and Londe
Department of Linguistics
University of California, Los Angeles
Los Angeles, CA 90095-1543
[bhayes@humnet.ucla.edu]
[kie@ucla.edu]
[zsuzsa@ucla.edu]

[Received 24 September 2008;
accepted 7 May 2009]

Siptár
Eötvös Loránd University &
Research Institute for Linguistics
Hungarian Academy of Sciences
H-1068 Budapest
Benczúr utca 33.
Hungary
[siptar@nytud.hu]