Complexity and Naturalness Biases in Phonotactics: Hayes and White (2013) Revisited

Brandon Prickett
University of Massachusetts Amherst

1 Introduction

Hayes and White (2013) found that English speakers rate words that violate natural phonotactic constraints[1] as worse than words that violate unnatural ones. Their "natural" constraints both enforced typologically common restrictions and were phonetically grounded, while their unnatural constraints met neither criterion. They used this experimental finding as evidence for a learning bias in favor of natural constraints.

The strength of this conclusion was weakened by the presence of a confound: the unnatural constraints were also more structurally complex.[2] In laboratory studies of artificial language learning, structural complexity seems to have a more consistent effect on ease of learning than phonetic grounding or typological attestedness (see Moreton and Pater 2012 for a review). In this paper, I present results from a version of Hayes and White's experiment that includes both complexity and naturalness as variables.

I consider the constraints of interest to belong to one of four categories: Natural Simple, Natural Complex, Unnatural Simple, and Unnatural Complex. If complexity were solely responsible for Hayes and White's (2013) results, then it should be the only factor affecting a constraint's learnability and the simple constraints should be more easily learned. However, my experiment's results suggest that naturalness and complexity both affect phonological acquisition: supporting the conclusions of Hayes and White (2013), but differing from the findings of most artificial language learning studies.

2 Design

I categorized constraints that were created by the UCLA Phonotactic Learner (Hayes and Wilson 2008) based on a collection of English words[3] as either natural or unnatural (following the methodology of Hayes and White 2013). I labeled constraints as natural if they represented typologically common restrictions and were phonetically motivated (such as a constraint against vowel hiatus). By consulting the existing literature and performing searches of P-Base (Mielke 2008), I confirmed that phonological patterns similar to the natural constraints were attested in languages other than English. The natural constraints that were chosen for this study, as well as their naturalness justifications, appear in Table 1.

In addition to this, I distinguished constraints as either complex or simple based on how many phonetic features were present in their description. This definition of complexity is similar to Chomsky and Halle's Evaluation Procedure (1968) and similar complexity metrics have been shown to have an effect on non-linguistic learning (Bulgarella and Archer 1962). For example, I labeled *[+syllabic][+syllabic] (violated by sequences with two vowels, e.g. [ae]) as simple since it contains only two features, while *[+consonantal,-anterior][+high,-syllabic] (violated by palatal+glide sequences, e.g. [ʃw]) was labeled as

[*] Thanks to UNC's P-Side group, UMass's Sound Workshop, and the AMP 2017 attendees for helpful discussion on this project. I'd also like to thank Jennifer L. Smith, Jeff Mielke, and Brian Dillon for their feedback and guidance. And a big thanks to both Elliott Moreton and Joe Pater for advising this project at different stages. All remaining errors are my own. Research funded by the Steve and Courtney Kapp Fund for Academic Excellence.

[1] The word "constraint" is used in this paper to refer to a learned phonotactic regularity in a language.

[2] While Hayes and White (2013) did perform an ad hoc test checking for the effects of complexity, they describe their analysis as unreliable since its results suggested that participants had a bias against simple patterns.

[3] An edited form of CELEX (Baayen et al. 1995) was used. For more information on this corpus, see Hayes and White (2013:51).

© 2018 Brandon Prickett
Proceedings of *AMP 2017*

complex since it makes use of four. This resulted in each simple group having fewer features in its constraints than its complex counterpart.[4] Unlike Hayes and White (2013), I only included bigram constraints in my experimental design, leaving exploring this variable as an avenue for future research. Table 2 shows the constraints that were used, by their naturalness and complexity.

Table 1. Constraint naturalness justifications

Constraint (Phonetic Justification)	**Typological Justification**
*[+syllabic][+syllabic] (Hiatus avoidance)	"Many languages do not tolerate vowel hiatus." (Casali 1997)
*[-cons.][+cons.] in the onset (Sonority ordering)	See Hayes and White (2013) for a detailed justification.
*[-cont.][-cont.] in the onset (Avoids sonority plateau)	See Hayes and White (2013) for a detailed justification.
*[-son.][+son.] in the coda (Sonority ordering)	See Hayes and White (2013) for a detailed justification.
*[+cons.,-anterior][-son][5] (Sonority ordering in some contexts)	Similar pattern in Sanskrit (Whitney 1889:72-75; see Hayes and White 2013 for discussion).
*[+diph.[6],+round][+labial] (Dissimilation)	Labial sounds are prohibited before [+round] vowels in Kilivila/Kiriwina (Mielke 2008).
*[-low,+tense][+son.,+dors.] (Assimilation)	[-low,+tense] vowels trigger palatalization of dorsals in Ejagham (Mielke 2008).
*[+nas.,+cor][+dors] in the coda (Assimilation)	Nasal clusters are homorganic in Muna (Mielke 2008).

Table 2. Constraints by category

	Constraint	**Description**
Natural Simple	*[+syllabic][+syllabic]	No hiatus
	*[-cons.][+cons.] in the onset[7]	Sonority ordering in onsets
	*[-cont.][-cont.] in the onset	No sonority plateau in onsets
	*[-son.][+son] in the coda	Sonority ordering in codas
Natural Complex	*[+cons.,-anterior][-son.]	No pre-obstruent palatals
	*[+diph., +round][+labial]	No round diphthongs before labial consonants
	*[-low, +tense][+son., +dors.]	No [-low] vowels before low consonants
	*[+nas., +cor.][+dors.] in the coda	No heterorganic nasal clusters in codas
Unnatural Simple	*[+diph.][+cont.,-anterior]	No diphthongs before palatal fricatives
	*[+round,-back][-anterior]	No [oɪ] before palatals or [ɹ]
	*[-son.][-low] in the coda[8]	No obstruents in a coda before glides
Unnatural Complex	*[+round, +high][-cons.,-son.]	No [u, ʊ] before [h]
	*[+diph., +round][+voice., +dors.]	No round diphthongs before [g]
	*[+cons.,-ant.][+high,-syllabic]	No palatals before glides
	*[+diph., +round.][-strid.] in the onset	No round diphthongs before non-stridents, except when the non-strident is in coda position

[4] This kind of feature-counting complexity correlates with Minimum Description Length (Rasin and Katzir 2016). See Heinz et al. (2009) on the role of other complexity metrics in phonology.

[5] While there is little phonetic justification for this constraint, it was primarily considered natural because of typological reasons and because it provided a complex, natural constraint (which are fairly uncommon in the UCLA learner's output). Changing the constraint to "unnatural" doesn't significantly change the effect of naturalness discussed in §4.

[6] Following Hayes and White (2013), I use the feature [±diphthong] to distinguish the diphthongs [oɪ], [aɪ], and [aʊ] from all other vowels. Segments that are phonetically diphthongs, such as [oʊ], were not marked as such.

[7] I counted syllabic information like this as half a feature when performing complexity categorization.

[8] This constraint was categorized as unnatural, although its restriction on onsets that are more sonorous than their preceding codas could be phonetically-motivated. It was considered unnatural because it arbitrarily bans glides, rather than all sonorants. This was dealt with by using stimuli that specifically differed in this respect, e.g. [peb.jɪn] (experimental) vs. [peb.lɪn] (control). In addition to this precaution, an analysis was run in which this constraint was considered natural, and the effect of naturalness discussed in §4 did not significantly change.

3 Methods

3.1 *Stimuli* As in Hayes and White (2013), the stimuli for this experiment consisted of nonce words. Half of the words violated constraints that belonged to one of the four groups shown in Table 2: Natural Simple, Natural Complex, Unnatural Simple, and Unnatural Complex. Each of these violating stimuli had a control group partner. The control group consisted of nonce words that were as similar to their partner as possible without violating any English constraints found by the UCLA Phonotactic Learner (Hayes and Wilson 2008). In many cases, partners only differed in one of the phonetic feature values for one of their segments. Instead of having four categories of stimuli as in Hayes and White (2013), I made use of eight stimulus categories (Natural Simple Control, Natural Simple Experimental, Natural Complex Control, Natural Complex Experimental, Unnatural Simple Control, Unnatural Simple Experimental, Unnatural Complex Control, and Unnatural Complex Experimental)—with the four additional groups being a result of the added variable of interest (i.e. complexity). Table 3 gives a complete list of the stimuli used in the experiment.

The experiment software presented the stimuli through an orthographic representation and an audio recording. Broad IPA transcriptions of the audio recordings are also included in Table 3, beneath their orthographic representations. A native English speaker from the southeastern United States recorded the audio stimuli with a Logitech, Model Number A-00008 headset microphone, using the microphone's default settings.

Table 3. Experiment stimuli

Constraint	Violater 1	Control 1	Violater 2	Control 2
*[+syllabic][+syllabic]	*keeane*	*klane*	*biate*	*brate*
	[kien]	[klen]	[baɪet]	[bɹet]
*[-cons.][+cons.] in onset	*hloop*	*ploop*	*hmit*	*smit*
	[hlup]	[plup]	[hmɪt]	[smɪt]
*[-cont.][-cont.] in onset	*cping*	*sping*	*ctice*	*stice*
	[kpɪŋ]	[spɪŋ]	[ktaɪs]	[staɪs]
*[-son.][+son] in coda	*canifl*	*canift*	*kipl*	*kilp*
	[kanɪfl]	[kanɪft]	[kɪpl]	[kɪlp]
*[+cons.,-anterior][-son.]	*ishty*	*ishmy*	*metchter*	*metchner*
	[ɪʃti]	[ɪʃmi]	[mɛtʃtɹ]	[mɛtʃnɹ]
*[+diph., +round][+labial]	*frowp*	*frope*	*soib*	*soid*
	[fɹaʊp]	[frop]	[soɪb]	[soɪd]
*[-low, +tense][+son., +dors.]	*perng*	*pern*	*plieng*	*pline*
	[pɹŋ]	[pɹn]	[plaɪŋ]	[plaɪn]
*[+nas., +cor.][+dors.] in coda	*quinnk*	*quisk*	*zannk*	*zask*
	[kwɪnk]	[kwɪsk]	[zænk]	[zæsk]
*[+diph.][+cont.,-anterior]	*bowsh*	*boosh*	*coish*	*kish*
	[baʊʃ]	[buʃ]	[koɪʃ]	[kɪʃ]
*[+round,-back][-anterior]	*coitch*	*keetch*	*goige*	*goik*
	[koɪtʃ]	[kitʃ]	[goɪdʒ]	[goɪk]
*[-son.][-low] in coda	*pabyin*	*payblin*	*bidyon*	*binyon*
	[pebjɪn]	[peblɪn]	[bɪdjən]	[bɪnjən]
*[+round, +high][-cons.,-son.]	*pauhin*	*pausin*	*sloohite*	*sleehite*
	[paʊhɪn]	[paʊsɪn]	[sluhaɪt]	[slihaɪt]
*[+diph., +round][+vce., +dors.]	*moig*	*moid*	*towg*	*towk*
	[moɪg]	[moɪd]	[taʊg]	[taʊk]
*[+cons.,-ant.][+high,-syllabic]	*shween*	*sween*	*chwid*	*chrid*
	[ʃwin]	[swin]	[tʃwɪd]	[tʃrɪd]
*[+diph., +round.][-strid.] in onset	*boithin*	*biethin*	*owthat*	*owtat*
	[boɪθIn]	[baɪθIn]	[aʊθæt]	[aʊtæt]

3.2 *Participants* I used Amazon's Mechanical Turk website (https://www.mturk.com; Sprouse 2011) to recruit 77 subjects and limited participants to those that lived in the United States. Subjects were all self-reported native speakers of American English.

3.3 *Procedure* The procedure primarily differed from Hayes and White (2013) in the medium that it was presented in. Rather than participating in person, subjects were directed from Mechanical Turk to a website that hosted the experiment. They were instructed to wear headphones, and then, much like Hayes and White (2013), they were presented with the following instructions:

> Languages have rules that determine how well words sound in that language. For example, in English, 'bzarshk' would sound very odd but the word 'kip' would sound fine, even though neither of them are actual words. For the following words, rate numerically how acceptable they would be in English. Let 'poik' be a value of 100. So anything better than this (for example, 'kip') would receive a higher score and anything worse (such as 'bzarshk') would receive a lower score. Be sure to listen to the audio clips and read the words so that you understand how the word is supposed to be pronounced.

Below these instructions was a button labeled "Start Experiment". Each page after the instructions presented participants with one of the stimuli and they used numerical magnitude estimation to gauge its grammaticality (see Bard et al. 1996 for more on this technique in linguistic experiments). Once all of the stimuli had been presented, subjects reported basic demographic information such as gender and age, although none of these demographic factors had an apparent effect on the results.

4 Results

Figure 1 illustrates the results of the experiment. Each of the stimulus categories' mean log ratings are given, with each constraint-violating category next to its control counterpart. If constraints have an effect on speaker judgments (i.e. if speakers learned the constraints when acquiring English), there should be a significant difference between the ratings of the control and violating stimuli.

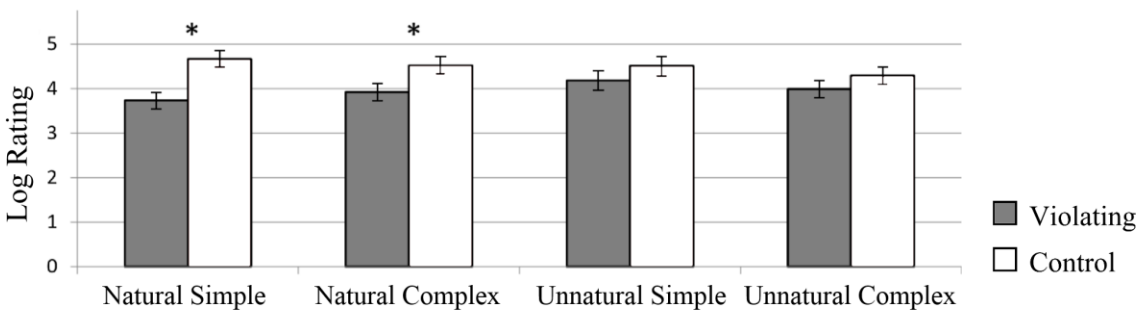

Figure 1. Mean log ratings by stimulus category. Error bars show 95% confidence intervals.

I subtracted the average rating for each violating stimulus from the average rating for its non-violating partner, ran a repeated measures ANOVA (with an error term for subjects) on these data, and found that naturalness, complexity, and their interaction were all significant. These results are given in Table 4:[9]

Table 4. ANOVA Results

	Mean Sq	F value	P-Value	
Natural	78418	176.76	< .001	***
Simple	18936	42.68	<.001	***
Natural:Simple	3647	8.22	0.00453	**

The significant main effects suggest that both complexity and naturalness play a role in constraint learnability. The interaction could mean that both factors have a larger-than-expected effect when acting together, or could be the result of differences in average constraint complexity across the natural and unnatural categories.

[9] I also ran a logistic regression on the data, following Hayes and White's (2013) analysis. These results were similar to those in the ANOVA, although more difficult to interpret due to the experimental design.

5 Discussion

5.1 *Methodological Implications* An important piece of insight that this study provides is the difficulty in constraint choice presented by the Hayes and White (2013) methodology. The fact that it "is not easy to establish firmly the naturalness of constraints" was pointed out by Hayes and White (2013:66) and this difficulty in experimental design only increases when other variables of interest are at play. Since the UCLA Phonotactic Learner (Hayes and Wilson 2008) produces a finite set of constraints, one must always make some compromises in choosing which constraints to use. In my experiment, this primarily affected the difference in complexity between the simple and complex constraint categories, which was very small and inconsistent across different levels of naturalness. While an ideal experimental design would have involved a strict feature cutoff between simple and complex, this was impossible given the constraints provided by the learner. Future work employing this methodology should make an effort to minimize the number of variables being studied so that few, if any, compromises like this must be made.

5.2 *The Effects of Naturalness and Complexity* The results of my experiment not only show an effect of complexity (as predicted by most artificial language learning studies) but also an effect of naturalness on phonotactic constraint learning. When presented with novel words that violated either natural or simple constraints, subjects rated the words as significantly worse sounding than words that violated unnatural or complex constraints. These lower ratings are interpreted, as in Hayes and White (2013), as indicating how well these constraints were internalized by speakers. This is a surprising result, considering the lack of evidence in artificial language learning studies for a naturalness bias in phonology (Moreton and Pater 2012).

5.3 *Future Work* Future research should investigate the discrepancy between bias effects in studies on natural language (such as the current study) and studies that use artificial languages. Factors unique to the Hayes and White (2013) methodology could be exaggerating the strength of a naturalness bias, such as the simplicity bias built into the UCLA learner (Hayes and Wilson 2008) or the fact that orthographic representations are used in stimulus presentation. Another issue could be that gaps in the UCLA learner's data disproportionately bias the learner toward inducing unnatural constraints that humans have evidence against. Examples of English words that violate the unnatural constraints used in this study are relatively easy to think of, and two examples from popular, fictional media are given in (1):

(1) *Words that violate unnatural constraints*
Smaug [smaʊg] (character in *The Hobbit*) violates *[+diphthong, +round][+voice, +dorsal]
Schwifty [ʃwɪfti] (adjective in *Rick and Morty*) violates *[+cons.,-anterior][+high,-syllabic]

Natural language experiments that use a different procedure for finding patterns in the English lexicon could help gauge how methodologically dependent this apparent naturalness bias is.

Alternatively, the discrepancy between these two methodological approaches could stem from artificial language learning hiding the effects of a naturalness bias. Artificial language learning that seeks to more closely replicate the process of natural language acquisition (such as Peperkamp and Martin 2016) has shown more evidence for naturalness bias effects. The results of this study suggest that both naturalness and complexity affect learning, and only through the use of multiple methodological paradigms will we completely understand what contributions both of these biases make to phonological acquisition.

References

Baayen, R. Harald, Piepenbrock, R., and Gulikers, L. 1995. The CELEX Lexical Database (Release 2) [CD-ROM]. Philadelphia, Penn.: Linguistic Data Consortium, University of Pennsylvania [Distributor].

Bard, Ellen Gurman, Dan Robertson, and Antonella Sorace (1996). Magnitude estimation of linguistic acceptability. *Language 72*:32-68.

Bulgarella, Rosaria, and E. James Archer (1962). Concept identification of auditory stimuli as a function of amount of relevant and irrelevant information. *Journal of Experimental Psychology 63(3)*, 254-257.

Casali, Roderic (1997). Vowel elision in hiatus contexts: Which vowel goes? *Language 73(3)*, 493-533.

Halle, Morris, and Noam Chomsky (1968). *The Sound Pattern of English*. New York: Harper & Row.

Hayes, Bruce, and James White (2013). Phonological Naturalness and Phonotactic Learning. *Linguistic Inquiry 44(1)*, 45-75.

Hayes, Bruce, and Colin Wilson (2008). A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *Linguistic Inquiry 39(3)*, 379-440.

Heinz, Jeffrey, Gregory Kobele, and Jason Riggle (2009). Evaluating the Complexity of Optimality Theory. *Linguistic Inquiry, 40(2)*, 277-288.

Mielke, Jeff (2008). *The Emergence of Distinctive Features*. Oxford: Oxford University Press.

Moreton, Elliott, and Joe Pater (2012). Structure and substance in artificial-phonology learning. Part I, Structure. Part II, Substance. *Language and Linguistics Compass 6(11)*, 686-701 and 702-718.

Peperkamp, Sharon and Alexander Martin (2016). Sleep-dependent consolidation in the learning of natural vs. unnatural phonological rules. Presented at *The 15th Conference on Laboratory Phonology*.

Rasin, Ezer and Roni Katzir. (2016). On evaluation metrics in Optimality Theory. *Linguistic Inquiry 47(2)*, 235-282.

Sprouse, Jon (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior research methods 43(1)*, 155-167.

Whitney, William D. (1889). *Sanskrit grammar*. Cambridge, MA: Harvard University Press.