FOOT STRUCTURE AND ANALYTIC BIAS*

RYAN BENNETT
Yale University

1 Introduction

This paper addresses two outstanding questions in prosodic theory. The first concerns the nature of prosodic domains. Stress, tone, phonotactics, and phonetic patterning are often domain-sensitive: their effect on a segmental string may depend on its larger prosodic or morpho-syntactic context. One tradition holds that these contextual domains reflect abstract, nested prosodic constituents: syllables, feet, prosodic words, and so on (Liberman and Prince 1977, Selkirk 1980a,b, Nespor and Vogel 1986, Pierrehumbert and Beckman 1988, to name a few). This view, known as PROSODIC HIERARCHY THEORY (PHT), takes prosodic domains to be layered and hierarchical in character.

An alternative line of thought eschews the rich phonological representations assumed by PHT, and assigns a greater role to simple linear position as a conditioning factor for prosodic phenomena (e.g. Steriade 1999, Seidl 2001, Gordon 2002, Samuels 2009, Scheer 2012, among others). The details of these proposals vary, but they all share the view that prosodic domains are 'flatter' than the nested constituents proposed in PHT, and depend to a greater extent on plain string adjacency.

A second question concerns the universality of prosodic domains. Are prosodic domains fixed across languages, being determined *a priori* by Universal Grammar or some set of broad functional principles (e.g. Itô and Mester 2013 and earlier work)? Or are prosodic domains language-specific, showing idiosyncratic variation from language to language (Hyman 1985, Schiering et al. 2010)? Though a long-standing issue in prosodic theory, this debate has seen renewed interest in recent years (e.g. Jun 2005, Hyman 2011, Padgett 2014). PHT is committed to both the universality of prosodic domains and to their hierarchical character, but the question of universality is conceptually independent of the hierarchical vs. linear organization of prosodic domains themselves.

My contribution to these debates comes in the form of an artificial grammar experiment, designed to test whether speakers learn a novel phonotactic pattern in terms of abstract prosodic

*I am grateful to Jaye Padgett, Junko Itô, Grant McGuire, Armin Mester, René Kager, Boris Harizanov, and Bruce Hayes for advice on this project. Special thanks to Junko Itô, Tomo Yoshida, Shin-ichiro Sano, Mikio Giriko, Kayo Takahashi, Kazumi Onnagawa, Mami Maeno, and the staff of NINJAL for helping with the logistics of the experiment. I also thank audiences at UCSC, the Tokyo Circle of Phonologists, the Stanford Phonetics and Phonology Workshop, the 2012 LSA meeting, Yale University, and FAJL 7 for their insightful comments. This research was partially funded by a Summer Fellowship from the Institute for Humanities Research at the University of California, Santa Cruz.
Email: ryan.bennett@yale.edu, website: http://pantheon.yale.edu/∼rtb27/.

structure or simple linear context. I report on a study, conducted with Japanese speakers, which finds evidence of a learning bias favoring a structure-dependent formulation of a stress-based vowel phonotactic. On the basis of these results, I argue (i) that abstract foot structure exists (so prosody is hierarchical), and (ii) that the metrical foot may be a prosodic universal, in that speakers make use of foot structure even when the empirical evidence for higher-order metrical grouping is weak. Further details can be found in Bennett (2012), though this paper supercedes that work.

2 Background

2.1 Stress and constituency

The metrical foot is an abstract prosodic constituent which intervenes between the syllable and the prosodic word. Within generative phonology, the foot was first proposed as a means of modeling stress placement (e.g. Liberman and Prince 1977, Selkirk 1980b, Hayes 1995). The central idea is that syllables are grouped into higher-order constituents (feet) which serve as the domains in which stress is assigned. Following Prince (1985) and Hayes (1995), I assume that feet are maximally binary, consisting of no more than two syllables: ($\sigma\sigma$). One syllable in each foot is the 'strong' or 'head' syllable; in languages with stress accent, this is the syllable that bears phonetic stress, as in English *Mississippi* [(mì.sə)(sí.pi)].[1]

Competing with the foot-based ('metrical') theory of stress are various proposals which reject foot-level constituency altogether. Non-metrical theories of stress assignment include Chomsky and Halle (1968), Prince (1983), and Gordon (2002). A shared assumption of these works is that stress assignment does not depend on abstract metrical constituents, but is instead computed with direct reference to syllables, word-edges, and morpheme boundaries.

These two approaches to stress assignment make different predictions about the interaction of stress with other linguistic phenomena. If stress is computed over abstract, hierarchical constituent structure, then other grammatical regularities could also be sensitive to those same constituents. Metrical stress theory thus predicts that morphological, phonological, and phonetic patterns can be directly conditioned by abstract feet, without mediation by stress as such (see Hayes 1995, Bennett 2012 for references). Since non-metrical theories of stress do without foot-level constituency, they make no such prediction. This difference forms the basis of the artificial grammar study presented below.

2.2 Artificial grammar experiments

I take both metrical and non-metrical theories of stress to be proposals about linguistic competence, committed to the psychological reality of the structures that they assume (see also Goldrick 2011). That is, I interpret these theories as claims about the cognitive machinery that speakers use to learn linguistic generalizations. Given that metrical and non-metrical theories of stress rely on different representational primitives, they make different, hopefully testable predictions about how learners identify and internalize novel phonological patterns.

The ARTIFICIAL GRAMMAR (AG) paradigm is a common experimental technique for exploring linguistic competence (Esper 1925, Reber 1967, Wilson 2006, Moreton 2008, Cristià et al. 2013,

[1] There is also compelling evidence for foot structure and foot-headedness in several languages which do not have phonetic stress (e.g. Awajún, McCarthy 2008). I return to this point in §4 below.

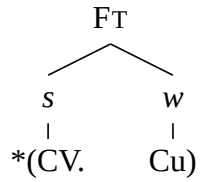
and many others; see Moreton and Pater 2012a,b for an overview of recent findings in phonology). AG studies attempt to uncover the implicit generalizations that learners extract (or fail to extract) from novel linguistic data. An important premise of this paradigm is that cognitive constraints on language learning — so-called LEARNING BIASES — can affect the generalizations that speakers make during an AG experiment. Here, I ask whether Japanese speakers are biased toward learning a prosodically-determined vowel phonotactic in terms of foot structure or simple linear context.

3 The present study

This study investigated the implicit learning of an artificial stress-conditioned vowel phonotactic. Participants in this study were exposed to an array of auditory stimuli, composed of CV syllables where C ∈ [p t k s] and V ∈ [u i] (see §3.1 for more details on stimulus construction). In the initial training phase of the experiment, every item was trisyllabic and carried audible stress on either the first or second syllable, [CV́.CV.CV] or [CV.CV́.CV]. If participants were to impose foot structure on these stimuli (consistent with metrical stress theory) trochaic footing [($\acute{\sigma}\sigma$)σ]∼[σ($\acute{\sigma}\sigma$)] would be the natural choice; iambic footing [($\sigma\acute{\sigma}$)σ)] would require either a degenerate foot [($\acute{\sigma}$)$\sigma\sigma$] or trochaic-iambic variation for initial stress, both of which are marked (Hayes 1995).

The training phase stimuli obeyed a key phonotactic restriction: the post-tonic syllable always contained [i], and never [u]. This meant that participants heard words like [tí.ki.su] and [pu.tú.si] during the training phase, but not words like *[tí.ku.su] or *[pu.tú.su]. This phonotactic was chosen because it is systematically ambiguous between a structure-dependent formulation referring to foot structure (1a) and a purely linear formulation (1b) that refers directly to the position of stress. Given trochaic footing, the linear relationship 'post-tonic' is equivalent to the structural relationship 'in the weak branch of the foot': the two predicates pick out exactly the same syllables in the trisyllabic training items.[2]

(1) a. STRUCTURE-DEPENDENT PHONOTACTIC:
The weak branch of the foot cannot contain [u]:



b. NON-STRUCTURAL, LINEAR PHONOTACTIC:
The post-tonic syllable cannot contain [u]: *CV́.Cu

The ambiguity of this phonotactic allows us to evaluate the predictions made by different theories of stress and prosodic parsing. If stress is assigned on the basis of foot structure, then participants should sometimes (or even mostly) learn a structure-dependent statement of the phonotactic (1a) rather than a purely linear formulation of the same pattern (1b). If stress placement does not make reference to abstract constituent structure, then partipants should uniformly extract the non-structural version of the phonotactic (1b). As with classic 'poverty of the stimulus' arguments, the learning data is compatible with either analysis; a preference for (1a) would therefore diagnose an ANALYTIC BIAS favoring structure-dependence in prosody (see Wilson 2006, Moreton 2008, and §4). The test phase of the experiment (§3.3) was designed to tease apart which version of the phonotactic participants actually learned during the initial training phase of the study.

[2]For examples of foot-based vowel phonotactics in natural language, see Gouskova (2003), Bennett (2012) and references there.

3.1 Stimulus construction

The auditory stimuli used in this experiment were produced with the MBROLA speech synthesis program (`http://tcts.fpms.ac.be/synthesis/mbrola.html`) and a series of Python scripts (`http://www.python.org/`). Stress was cued by vowel length and by a pitch peak on the stressed syllable. Duration and pitch are reliable cues to stress cross-linguistically (Cutler 2005). Stressed vowels were about 215ms long, while unstressed vowels were about 140ms each. These durations are comparable to values reported for [u i] in American English (Hillenbrand et al. 1995). Pitch peaks occurred at roughly 140Hz on stressed syllables; pitch was 90Hz or below elsewhere. Other potential cues to stress (such as consonant length and intensity) were held constant across syllables.

The participants in this experiment were all native Japanese speakers (§3.2). Japanese is a pitch accent language rather than a stress accent language, and the pitch properties of stressed syllables in this experiment were similar to the pitch characteristics of accented moras for male speakers of Tokyo Japanese (Beckman 1986, Sugiyama 2008 and references there). This holds both for pitch height and for the pitch fall following the stressed syllable. Duration is not a cue for pitch accent in Japanese.[3] While participants may have perceived the stressed syllables as bearing pitch accent rather than stress, there is strong evidence that pitch accent in Tokyo Japanese is itself assigned on the basis of trochaic footing (Kubozono 2008). As such, the phonetic characteristics of the stimuli should not adversely affect the interpretation of the results.

Lastly, the distribution of segments was balanced across stimuli to avoid accidental phonotactic regularities that could interfere with the experimental results. The vowels [u i] occurred with equal frequency in all non-post-tonic syllables. The four consonants [p t k s] occurred with roughly equal frequency in all syllables. Previous work suggests that stimulus variation of this sort encourages speakers to abstract over individual items and extract higher-order generalizations (see Cristià et al. 2013 for references). While no stimulus was an actual word of English, some items may have been words in Japanese; this possibility was not checked systematically.

3.2 Participants

This paper reports results from 32 participants, all native Japanese speakers with limited L2 English (English proficiency was assessed with a questionnaire and informal conversation). Participants were recruited from undergraduate courses at International Christian University (ICU) and from the National Institute for Japanese Language and Linguistics (NINJAL), both in Tokyo, Japan. The experimental sessions took place at these two recruitment sites, and were conducted in English.

Japanese speakers were recruited precisely because Japanese does not have phonological or phonetic stress. Given that there are no stress-based segmental phonotactics in Japanese, the native language background of the participants should not bias them toward a particular formulation of the vowel phonotactic (1). Although pitch accent placement may be foot-sensitive in Tokyo Japanese, there are no pitch-dependent segmental phonotactics in the language which might bias phonotactic learning or participant responses in the experiment. The upshot is that if Japanese speakers show a bias for (or against) the structure-dependent phonotactic (1a), this bias cannot be straightforwardly attributed to L1 phonological knowledge (see also §4.2).

[3] Stressed vowels in this experiment were probably too short to be interpreted as phonemic long vowels /V:/ by participants. The durational ratio between stressed and unstressed vowels was about 1.54, while the average durational ratio for phonemic [V:]:[V] in Japanese is roughly 2.5 (Hirata 2004).

3.3 Experimental design and results

The experiment began with a training phase in which participants listened to a set of 3σ stimuli over headphones. These stimuli all obeyed the phonotactic (1). The participants were asked to repeat each word aloud as if learning to speak a new language. There was no feedback or explicit instruction of any kind during the training phase (see Reber 1967, Saffran et al. 1997). Participants were told that the recordings were from a newly-discovered language of Papua New Guinea; after completing the experiment, they were informed that the recordings were actually synthetic speech.

3.3.1 Test phase 1 (3σ items)

In the first test phase of the experiment, participants heard a set of novel 3σ items. These test items had the same phonotactic structure as the training items (§3.1), with one important difference: half the words were 'ungrammatical' in that they contained [u] in the post-tonic syllable, violating the implicit phonotactic that governed the training items (1). Participants were asked to identify which words belonged to the language from the training phase, and which were from a new (but similar) language. The purpose of this test phase was to verify that participants learned some version of the target phonotactic (1) during training. If so, they should respond differently to items that conform to the phonotactic than to items that violate it.

Learning was assessed with a mixed-effects logistic regression over participant responses (see Jaeger 2008, Barr et al. 2013, and work cited there). Logistic regression allows us to examine which factors contributed to the grammaticality decisions that participants made during testing. Statistics were calculated in R (`http://www.r-project.org/`), using the `lme4` package (Bates et al. 2011). The base model specification included fixed effects for GRAMMATICALITY, STRESS, TRIAL, and all interactions between them. By-participant slopes for GRAM., STRESS, TRIAL, and the GRAM. x STRESS interaction were included as random effects. Trials with extreme response times were excluded from the analysis (see Bennett 2012:283, 287).

To control for possible phonotactic confounds in the training data, I used the Maximum Entropy Phonotactic Learner (Hayes and Wilson 2008) to uncover any unintended phonotactic regularities that held over the training set. The MaxEnt learner searches for phonotactic patterns by identifying statistically underrepresented strings in a given wordlist. Using the default parameter settings and the feature specifications in Bennett (2012:342-3), I had the MaxEnt learner find 40 phonotactic constraints supported by the training data. This procedure was repeated 10 times, yielding eight reliable nuisance phonotactics (Table 1). These phonotactics were added to the logistic regression as fixed effects, i.e. as possible predictors of how participants would respond to novel stimuli.[4]

(1)	*[tit]	(2)	*[sis]	(3)	*[tŭs]	(4)	*[kuk]
(5)	*[+CORONAL]+[ŭt]	(6)	*[kV̆k]	(7)	*[sV̆p]	(8)	*[CuCu#]

Table 1: Nuisance phonotactics uncovered by the MAXENT learner ([V̆] = unstressed vowel)

The final model (Table 2) was chosen by a step-down selection procedure using the log-likelihood and AIC tests to compare nested models. The highly significant effect of GRAMMATICALITY

[4]One of these phonotactics, *[CuCu#], was added independently. Even though it holds without exception in the training data, it was not uncovered by the MAXENT phonotactic learner, presumably because it is a complex constraint which assigns a proper subset of the violations assigned by the target phonotactic (1). More on this in §3.3.2.

indicates that participants distinguished grammatical and ungrammatical 3σ items, confirming that they did indeed learn some version of the basic phonotactic (1).

	Estimated β	Odds ratio: e^{β}	SE(β)	z	p-value
(Intercept)	-1.692	0.18	0.184	-9.21	$<$ .001
UNGRAMMATICAL	**1.225**	**3.40**	**0.217**	**5.63**	$<$ **.001**
*[CuCu#]	0.903	2.47	0.110	8.23	$<$ .001
*[sis]	0.815	2.26	0.235	3.46	$<$ .001
UNGRAM. X STRESS ($\acute{\sigma}_2$)	0.452	1.57	0.203	2.23	$<$ .05
UNGRAM. X TRIAL	-0.006	0.99	0.003	-2.27	$<$ .05

Table 2: Final response model for 3σ items (TEST PHASE 1). Coefficients express predicted change in likelihood of a 'No' (ungrammatical) response. Non-significant predictors are omitted to save space. D_{xy} = .62, c = 81%.

An analysis of response times yields a similar conclusion. Participants were generally slower to respond when presented with ungrammatical items (Figure 1). This also suggests too that participants were sensitive to the implicit grammaticality contrast. The effect of grammaticality on response speed comes out as statistically significant in a mixed-effects linear regression model (Table 3), computed using the same step-down procedure described above.[5]

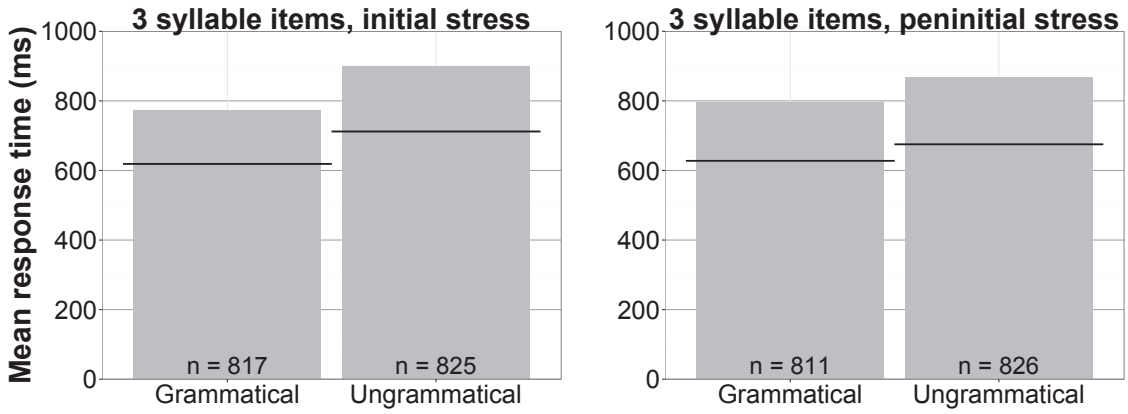


Figure 1: Mean response times across participants for 3σ test items (TEST PHASE 1). The black horizontal bars mark the median response time for each condition.

3.3.2 Test phase 2 (5σ items)

The second test phase of the experiment was designed to identify whether participants learned a structure-dependent or strictly linear version of the target phonotactic (1). In this phase participants were presented with novel *five*-syllable stimuli. Like the 3σ stimuli, these items carried stress on either the first or second syllable. Unlike the 3σ items, these 5σ items could in principle contain

[5]The initial model had fixed effects for GRAMMATICALITY, STRESS, RESPONSE, and all interactions between them; for TRIAL, TRIAL X RESP., TRIAL X GRAM.; and for the eight nuisance phonotactics in Table 1. By-participant slopes for GRAM., STRESS, TRIAL, and RESP. were included as random effects. Response times were log-transformed before being entered as dependent variables (Baayen and Milin 2010).

	Estimated β	β in ms	SE(β)	t	p-value
(Intercept)	-0.187	-104	0.091	-2.05	$< .05$
RESPONSE (NO)	0.469	364	0.068	6.85	$< .001$
UNGRAMMATICAL	**0.220**	**150**	**0.044**	**4.98**	$<$ **.001**
*[sV̌p]	0.126	82	0.062	2.02	$< .05$
*[tŭs]	0.125	81	0.062	2.01	$< .05$
UNGRAM. X RESP. (NO)	-0.310	-162	0.056	-5.57	$< .001$

Table 3: Final response time model for 3σ items (TEST PHASE 1). Coefficients express predicted increase in response time. Non-significant predictors are omitted to save space. R^2 = .40.

two different feet (2): one cued by the position of audible stress, and another 'covert' foot with no phonetic realization at all.[6]

(2) a. $(\acute{\sigma}\sigma)(\sigma\sigma)\sigma$
b. $\sigma(\acute{\sigma}\sigma)(\sigma\sigma)$

(3) a. *(CV́.Ci)(CV.<u>Cu</u>)CV
b. *CV(CV́.Ci)(CV.<u>Cu</u>)

The possibility of covert footing allows us to decouple stress from foot structure. If participants learned a foot-based version of the phonotactic (1), then the presence of two feet should entail two potential loci of ungrammaticality — even if one of those feet does not bear stress. If participants systematically respond 'Ungrammatical' to words like (3), which contain [u] *three* syllables after the stressed syllable, we would be licensed to conclude that they learned a structure-dependent version of the target phonotactic rather than a linear, stress-based one.

Half the items in this test phase were 'grammatical': [i] occurred in both the post-tonic syllable and in the weak branch of the hypothesized covert foot (2). The remaining items were split between two types of 'ungrammatical' words: those with [u] in the post-tonic syllable, and those with [u] in the weak syllable of the covert foot (3). Otherwise, the second test phase was identical to the first.

Participant responses were analyzed with another mixed-effects logistic regression, following the procedure described in §3.3.1. Table 4 shows that grammaticality violations in the post-tonic syllable had a statistically significant effect on participant responses. This suggests that participants generalized the basic ban on post-tonic [u] from 3σ forms to 5σ forms. However, grammaticality violations in the covert foot only had a near-significant effect on responses. The statistical response model therefore fails to support the hypothesis that participants learned the structure-dependent phonotactic (1a).

Unfortunately, this model is marred by a strong correlation between UNGRAM. (COVERT FT) and the nuisance phonotactic factor *[CuCu#] ($r > .80$).[7] This makes it difficult to gauge the extent to which these variables have independent effects on participant responses. If *[CuCu#] is omitted from the analysis, step-down model selection gives a final model in which UNGRAM. (OVERT FT) and UNGRAM. (COVERT FT) are both significant predictors of participant responses ($p < .05$).

Things are a bit more clear-cut when we turn to the analysis of response times. Mean response times for the second test phase are shown in Figure 2; the statistical analysis (Table 5) reveals a

[6]See McCarthy (2008), Buckley (2009), Bennett (2012) for some discussion of covert footing in natural language.

[7]The correlation arises because any test item that violates *[CuCu#] must also violate UNGRAM (COVERT FT), though not vice-versa (see (3)).

	Estimated β	Odds ratio: e^{β}	SE(β)	z	p-value
(Intercept)	-0.628	0.53	0.097	-6.51	$< .001$
*[kuk]	0.417	1.52	0.160	2.60	$< .01$
*[sis]	0.381	1.46	0.124	3.08	$< .01$
*[CuCu#]	0.360	1.43	0.149	2.42	$< .05$
UNGRAMMATICAL (OVERT FT)	**0.272**	**1.32**	**0.128**	**2.13**	**$< .05$**
UNGRAMMATICAL (COVERT FT)	**0.236**	**1.27**	**0.142**	**1.66**	**.10**
[+COR]+[ŭt]	-0.196	0.82	0.105	-1.86	0.06

Table 4: Final response model for 5σ items (TEST PHASE 2). Coefficients express predicted change in likelihood of a 'No' (ungrammatical) response. D_{xy} = .42, c = 71%.

significant slowdown in response times for test items containing [u] in either a post-tonic syllable or the weak branch of a covert foot. This finding supports the hypothesis that participants learned the structure-dependent phonotactic (1a): only the foot-based statement of the vowel restriction predicts a slowdown in response times when [u] appears three syllables after the stressed syllable, in the weak position of a phonetically unrealized, but structurally present foot.

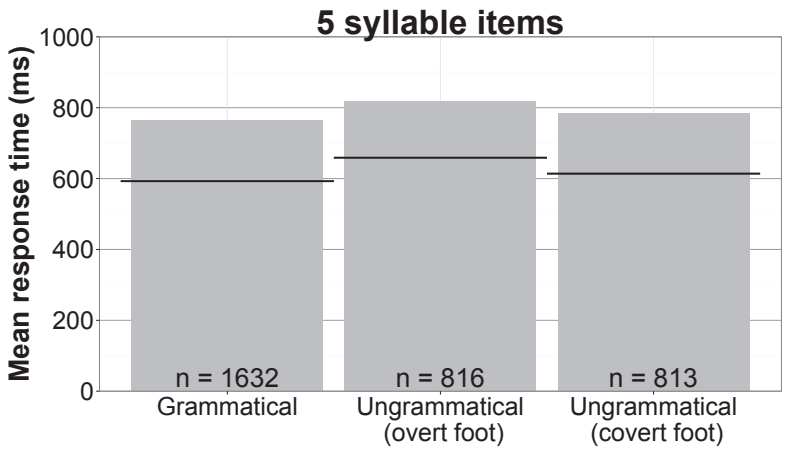


Figure 2: Mean response times across participants for 5σ test items (TEST PHASE 2). The black horizontal bars mark the median response time for each condition.

4 Discussion

The results of the second test phase suggest that participants learned a foot-based version of the target phonotactic (1) and extended it to the weak branches of covert feet in 5σ items. Grammatical sensitivity in covert feet was modest but nonetheless present, especially in the analysis of response times. These results are clearly consistent with the claim that stress is assigned on the basis of abstract foot structure. In contrast, non-metrical theories of stress fail to predict these results: if participants had learned a strictly linear version of the target phonotactic (1b), referring directly to post-tonic position, they should not have extended that restriction to vowels falling three syllables after audible stress (3). (See Newport and Aslin 2004 for related evidence that listeners are unable to learn phonotactic dependencies between non-adjacent syllables, as in a linear version of (3).)

	Estimated β	β in ms	SE(β)	t	p-value
(Intercept)	-0.138	-74	0.084	-1.66	.10
RESPONSE (NO)	0.241	157	0.050	4.83	$<$ .001
*[kV̌k]	0.133	82	0.065	2.04	$<$ .05
UNGRAMMATICAL (OVERT FT)	**0.084**	**51**	**0.041**	**2.07**	**$<$ .05**
UNGRAMMATICAL (COVERT FT)	**0.084**	**50**	**0.040**	**2.08**	**$<$ .05**
TRIAL	-0.002	-1	0.001	-3.34	$<$ .001
UNGRAM. (COVERT) X RESP. (NO)	-0.131	-71	0.062	-2.12	$<$ .05

Table 5: Final response time model for 5σ items (TEST PHASE 2). Coefficients express predicted increase in response time. Non-significant predictors are omitted to save space. R^2 = .32.

This interpretation of the results depends on two things: the existence of covert feet, and the assumption that participants imposed iterative, partially covert footing on the experimental stimuli. The cross-linguistic evidence for covert footing is mounting; much of this evidence comes from Japanese itself, which has extensive patterns of foot-conditioned morphology and phonology (§4.2). The possibility of covert iterative footing in languages with non-iterative stress is also consistent with recent claims in the formal literature (e.g. McCarthy 2003:111-2, Buckley 2009).[8]

These results provide experimental support for the abstract prosodic constituents assumed by metrical stress theory and by PHT more generally. Hierarchical prosodic structure (covert footing) accounts for generalization of the vowel phonotactic to a second locus in 5σ forms; purely linear phonology does not. Parsimony arguments against PHT are therefore irrelevant: because PHT and non-structural theories of prosody differ in their empirical coverage, Ockham's Razor simply does not apply (see also Wilson 2006).

It remains to be seen whether these findings generalize to other prosodic domains, especially at the word level and above. Still, if abstract prosodic structure exists at the level of the foot—and is psychologically real—that undercuts broad conceptual arguments against the existence of abstract prosodic constituents corresponding to domains of other sizes.

4.1 Typology, analytic bias, and prosodic universals

The main contribution of this experiment is behavioral evidence for structure-dependence in stress assignment. An important secondary finding is that participants readily imposed foot structure on novel words even when acoustic cues to stress and footing were lacking. The use of covert feet in 5σ forms thus points toward a learning bias: listeners may infer the presence of foot structure over a string of syllables whether or not such structure is signaled by actual phonetic evidence. This is an analytic bias in Moreton's (2008) terms, because it concerns the kinds of structural analyses that learners impose on linguistic data.

This apparent predilection for foot structure may even shape prosodic typology. Bennett (2012) points out that evidence for footing can be found in many languages with non-iterative stress (Irish, Russian) or without any stress at all (Kera, Awajún). In several cases foot structure seems

[8]It has been observed that listeners tend to perceive stress peaks as more evenly-spaced than they really are (Hayes 1995:31). This too may suggest that listeners impose a regular rhythmic structure on words even when none is present in the acoustic signal itself.

to have arisen *de novo* during historical development, without a clear phonetic or phonological precursor. Bennett suggests that the surprising prevalence of foot structure in such languages may have its roots in a phonological learning bias: if listeners posit metrical feet even in the absence of suprasegmental cues, then foot-conditioned grammatical patterns could arise diachronically in languages that lack stress. This mode of explanation hinges on the assumption that foot structure is *a priori* accessible to learners: if typology is influenced by a learning bias, and the learning bias in question refers to feet, then the foot must be a prosodic universal. For more discussion see Liberman (1975:§6.2), Hayes (1995:8-9), Goad and Buckley (2006), and Bennett (2012).

4.2 The role of Japanese phonology

Japanese provides a clear case of covert footing in natural language. Previous work on Japanese has uncovered robust patterns of foot-conditioned phonology (especially accent assignment) and morphology (for a review see Kubozono 2008, Itô and Mester to appear b). Many of these patterns involve trochaic feet without any independent phonetic realization (e.g. feet in unaccented words), as in the experiment described here. This raises a question: could the use of covert footing with 5σ test items reflect native-language influence from Japanese, rather than a general learning bias in favor of foot structure? This worry is certainly well-founded, as previous work has observed an effect of native language on patterns of generalization in artificial grammar studies (see Cristià et al. 2013 for references). Particularly troubling is the fact that some Japanese roots (those in the Sino-Japanese stratum) seem to obey a vowel phonotactic that restricts the quality of vowels appearing in the weak branch of a trochaic foot (see Itô and Mester to appear a). This is distressingly similar to the target phonotactic (1).

Still, there are reasons to doubt that the basic qualitative findings of this study depend on the native language of the participants. Bennett (2012) reports similar results for a parallel study with 44 English speakers in the U.S. There is little evidence for covert footing in English; this makes it unlikely that native phonological knowledge is responsible for the results of these two studies.

Whatever their source, the results reported here support the use of the foot as an analytical device in Japanese linguistics. If the response patterns of these participants reflect the influence of native Japanese phonology, then Japanese must have metrical feet. If the results reflect a general analytic bias favoring foot structure, as I have argued, then Japanese speakers must have access to metrical feet during acquisition. In either case, these findings lend credence to analyses of Japanese phonetics, phonology and morphology that rely on foot structure.

References

Baayen, R. Harald, and Petar Milin. 2010. Analyzing reaction times. *International Journal of Psychological Research* 3:12–28.

Barr, Dale, Roger Levy, Christoph Scheepers, and Harry Tily. 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language* 68:255–278.

Bates, Douglas, Martin Maechler, and Ben Bolker. 2011. *lme4: Linear mixed-effects models using S4 classes*. Version 0.999375-41, retrieved from http://CRAN.R-project.org/package=lme4.

Beckman, Mary. 1986. *Stress and non-stress accent*. Dordrecht: Foris.

Bennett, Ryan. 2012. Foot-conditioned phonotactics and prosodic constituency. Doctoral Dissertation, University of California, Santa Cruz.

Buckley, Eugene. 2009. Locality in metrical typology. *Phonology* 26:389–435.

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

Cristià, Alejandrina, Jeff Mielke, Robert Daland, and Sharon Peperkamp. 2013. Similarity in the generalization of implicitly learned sound patterns. *Laboratory Phonology* 4:259–285.

Cutler, Anne. 2005. Lexical stress. In *The handbook of speech perception*, ed. David Pisoni and Robert Remez, 264–289. Malden, MA: Blackwell.

Esper, Erwin. 1925. A technique for the experimental investigation of associative interference in artificial linguistic material. *Language* 1:1–47.

Goad, Heather, and Meaghen Buckley. 2006. Prosodic structure in child French: Evidence for the foot. *Catalan Journal of Linguistics* 5:109–142.

Goldrick, Matthew. 2011. Using psychological realism to advance phonological theory. In *The handbook of phonological theory*, ed. John Goldsmith, Alan C.L. Yu, and Jason Riggle, 631–660. Malden MA: Wiley-Blackwell.

Gordon, Matthew. 2002. A factorial typology of quantity-insensitive stress. *Natural Language and Linguistic Theory* 20:491–552.

Gouskova, Maria. 2003. Deriving economy: syncope in Optimality Theory. Doctoral Dissertation, University of Massachusetts Amherst.

Hayes, Bruce. 1995. *Metrical stress theory*. Chicago: The University of Chicago Press.

Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39:379–440.

Hillenbrand, James, Laura A. Getty, Michael J. Clark, and Kimberlee Wheeler. 1995. Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America* 97:3099–3111.

Hirata, Yukari. 2004. Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics* 32:565–589.

Hyman, Larry. 1985. *A theory of phonological weight*. Foris Publications.

Hyman, Larry. 2011. Does Gokana really have no syllables? Or: what's so great about being universal? *Phonology* 28:55–85.

Itô, Junko, and Armin Mester. 2013. Prosodic subcategories in Japanese. *Lingua* 124:20–40.

Itô, Junko, and Armin Mester. to appear a. Sino-Japanese phonology. In *The handbook of Japanese phonetics and phonology*, ed. Haruo Kubozono. Berlin: Mouton de Gruyter.

Itô, Junko, and Armin Mester. to appear b. Word formation and phonological processes. In *The handbook of Japanese phonetics and phonology*, ed. Haruo Kubozono. Berlin: Mouton de Gruyter.

Jaeger, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language* 59:434–446.

Jun, Sun-Ah. 2005. Prosodic typology. In *Prosodic typology: The phonology of intonation and phrasing*, ed. Sun-Ah Jun, 430–458. Oxford, UK: Oxford University Press.

Kubozono, Haruo. 2008. Japanese accent. In *The handbook of Japanese linguistics*, ed. Shigeru Miyagawa and Mamoru Saito, 165–191. New York: Oxford University Press.

Liberman, Mark. 1975. The intonational system of English. Doctoral Dissertation, Massachusetts Institute of Technology.

Liberman, Mark, and Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8:249–336.

McCarthy, John J. 2003. OT constraints are categorical. *Phonology* 20:75–138.

McCarthy, John J. 2008. The serial interaction of stress and syncope. *Natural Language and Linguistic Theory* 26:499–546.

Moreton, Elliott. 2008. Analytic bias and phonological typology. *Phonology* 25:83–127.

Moreton, Elliott, and Joe Pater. 2012a. Structure and substance in artificial-phonology learning, part I: structure. *Language and Linguistics Compass* 6:686–701.

Moreton, Elliott, and Joe Pater. 2012b. Structure and substance in artificial-phonology learning, part II: substance. *Language and Linguistics Compass* 6:702–718.

Nespor, Marina, and Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris Publications. Reprinted 2008 by Mouton de Gruyter.

Newport, Elissa, and Richard Aslin. 2004. Learning at a distance I. statistical learning of non-adjacent dependencies. *Cognitive psychology* 48:127–162.

Padgett, Jaye. 2014. On the origins of the prosodic word in Russian. In *Proceedings of the Speech Prosody 2014 Conference*, ed. Nick Campbell, Dafydd Gibbon, and Daniel Hirst, 368–371. Dublin, Ireland: Speech Prosody. Available online at http://www.speechprosody2014.org/.

Pierrehumbert, Janet, and Mary Beckman. 1988. *Japanese tone structure*. Cambridge, MA: MIT Press.

Prince, Alan. 1983. Relating to the grid. *Linguistic Inquiry* 14:19–100.

Prince, Alan. 1985. Improving tree theory. In *Berkeley Linguistic Society (BLS) 11*, 471–490. Berkeley, CA: Berkeley Linguistics Society.

Reber, Arthur S. 1967. Implicit learning of artificial grammars. *Journal of verbal learning and verbal behavior* 6:855–863.

Saffran, Jenny, Elissa Newport, Richard Aslin, Rachael Tunick, and Sandra Barrueco. 1997. Incidental language learning: listening (and learning) out of the corner of your ear. *Psychological Science* 8:101–105.

Samuels, Bridget. 2009. The structure of phonological theory. Doctoral Dissertation, Harvard University.

Scheer, Tobias. 2012. *Direct interface and one-channel translation*. Boston: de Gruyter.

Schiering, René, Balthasar Bickel, and Kristine Hildebrandt. 2010. The prosodic word is not universal, but emergent. *Journal of Linguistics* 46:657–709.

Seidl, Amanda. 2001. *Minimal indirect reference: a theory of the syntax-phonology interface*. Outstanding Dissertations in Linguistics. New York: Garland.

Selkirk, Elisabeth. 1980a. Prosodic domains in phonology: Sanskrit revisited. In *Juncture*, ed. Mark Aronoff and Mary-Louise Kean, Vol. 7, 107–129. Saratoga, CA: Anma Libri.

Selkirk, Elisabeth. 1980b. The role of prosodic categories in English word stress. *Linguistic Inquiry* 11:563–605.

Steriade, Donca. 1999. Alternatives to syllable-based accounts of consonantal phonotactics. In *Item order in language and speech*, ed. Osamu Fujimura, Brian Joseph, and Bohumil Palek, 205–242. Prague: Karolinum.

Sugiyama, Yukiko. 2008. The nature of Japanese accent: an experimental study. Doctoral Dissertation, SUNY Buffalo.

Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30:945–982.