Phonological Data & Analysis Volume 4, Article 6: 1–35 (2022)
https://doi.org/10.3765/pda.v4art6.65

Received 25 July 2021; revised 10 February 2022; accepted 11 May 2022. © 2022 Megan J. Crowhurst et al.
Published by the Linguistic Society of America with permission of the authors under a CC BY 3.0 license.

"Natural" stress patterns and dependencies between edge alignment and quantity sensitivity

Megan J. Crowhurst,[a] Laura R. Faircloth,[a] Allison L. Wetterlin[b] & Linda R. Wheeldon[b*]

[a]University of Texas at Austin, USA – mcrowhurst@austin.utexas.edu
[b]University of Agder, Kristiansand, Norway

We conducted an artificial language learning experiment to study learning asymmetries that might reveal latent preferences relating to, and any dependencies between, the edge alignment and quantity sensitivity (QS) parameters in stress patterning. We used a poverty of the stimulus approach to teach American English speakers an unbounded QS stress rule (stress a single CV: syllable) and either a left- or right-aligning QI rule if only light syllables were present. Forms with two CV: syllables were withheld in the learning phase and added in the test phase, forcing participants to choose between left- and right-aligning options for the QS rule. Participants learned the left- and right-edge QI rules equally well, and also the basic QS rule. Response patterns for words with two CV: syllables suggest biases favoring a left-aligning QS rule with a left-edge QI default. Our results also suggest that a left-aligning QS pattern with a right-edge QI default was least favored. We argue that stress patterns shown to be preferred based on evidence from ease-of-learning and participants' untrained generalizations can be considered more natural than less favored opposing patterns. We suggest that cognitive biases revealed by artificial stress learning studies may have contributed to shaping stress typology.

Keywords: stress; artificial language learning; quantity sensitivity; alignment; natural sound patterns

1 Introduction

Half a century of research in metrical phonology has delivered a well-articulated typology of word stress patterns found in the world's languages along with formal theories intended to account for asymmetries in the stress typology (Liberman & Prince 1977; Hayes 1985, 1995; Prince 1983; Hammond 1984; Halle & Vergnaud 1987, among others). All post-SPE metrical theories are componential in that they model stress patterns in terms of a limited set of interacting parameters with opposing settings.[1] One such parameter is *edge alignment:* stress patterns orient to either the left or right edge of the prosodic word. A second key parameter is *quantity* (or weight) *sensitivity.* In quantity-sensitive (QS) stress patterns, word stress is sensitive to the internal structure of syllables. In QS languages, syllables with two morae are stress attractors. There are no known weight-sensitive languages in which stress is preferentially assigned to monomoraic syllables.

[*] The authors are grateful to Rajka Smiljanić for recording stimulus materials for the project, and to Scott Myers for his helpful input at various stages of the project. This research was presented at the Manchester Phonology Meeting in May, 2019. We also thank Andrew Nevins for his editorial handling of the submission and the anonymous PDA reviewers for their helpful feedback.

[1] Or constraints with opposing requirements. In this work, we refer to parameters identified in Hayes (1985), as it is convenient to think of oppositions in binary terms, and because we wish to distance ourselves from any particular formal theory (e.g., Optimality Theory, and its claims about independent constraints and constraint rankings).

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

In this article, we report the results of an artificial language learning (ALL) experiment in which we explored native American English speakers' preferences in regard to the edge alignment parameter in stress patterning, and whether the presence of a weight-sensitive component influences edge alignment preferences. This experiment belongs to a research program whose broadest objective is to discover evidence that can shed light on what it means for a stress pattern to be natural, and then relate any findings back to the stress typology. Do asymmetries in the stress typology exist because some known parameter settings are more natural than others? And relatedly, when stress parameters interact, do the effects of one more naturally take precedence over the effects of the other? Finally, how can naturalness in stress patterning be measured? Before describing the present research in detail, we first discuss what we mean by "naturalness" and provide background on the most relevant prior ALL research on stress pattern learning.

1.1 Naturalness and the problem with frequency statistics

The concept of "naturalness", applied to linguistic patterns, can be thought of in various ways. Obviously, if a stress pattern occurs in a human language, then it is natural in that sense; conversely, non-occurring patterns must not be natural. Adopting a more gradient metric, it would be reasonable to try to relate naturalness to frequency asymmetries in the stress typology. More prevalent stress patterns and structures could be considered more natural than patterns which occur less often. Print resources and the availability of searchable, annotated databases provide some basis for judging the prevalence of many characteristics of stress patterns (e.g. Gordon 2016).[2] With high-quality quality statistical information about frequency, we might ask, for example: if left-aligned stress patterns are more widely attested than right-edge patterns (see below), is left-alignment the more natural option?

Unfortunately, frequency cannot be assumed as a proxy for naturalness with full confidence, given a lack of comprehensive frequency counts for a wide range of stress patterns. Some useful information about edge alignment in quantity-insensitive (QI) stress patterns is provided in works such as Hyman (1977), Gordon (2002), and van der Hulst & Goedemans (2009; the online version of StressTyp in 2009) and aggregated in Gordon (2016). Figure 1 (from Gordon 2016: 177, displayed with Gordon's caption), reveals asymmetries among languages reported to have a single fixed stress positioned at or near one edge of the word or the other. While the frequency information from the three surveys differs, perhaps due to variation in sampling, averaging over these sources suggests that a single fixed stress on an initial syllable is most common, followed by stress on a final or penultimate syllable. In contrast, languages with a single fixed stress on a peninitial or antepenultimate syllable are scarce.

Strikingly, given their theoretical prominence, binary iterative stress patterns may be much less common than non-iterative patterns: Gordon (2002) reports that in a sample of 262 languages with predictable, QI word stress, 167 languages (63.7%) had a single, fixed stress and only 52 (19.8%) had patterns in which more than one stress was assigned. Of the latter, 38 languages had alternating stress and 14 had a pattern in which stress occurs on a syllable at/near each edge in words that are long enough. Figure 2 (from Gordon 2016:179, with Gordon's caption), reveals asymmetries for four binary alternating stress patterns, based on frequency counts in Gordon (2002; 38 languages) and Goedemans (2010; 171 languages). Stress on odd syllables counting from the left is most common, followed by stress on even syllables counting from the right. These distributions are formally analyzed in terms of syllabic trochees assigned iteratively beginning at the left or right edge, respectively. Both of these patterns are much more common than the mirror image patterns (stress on even syllables counted from the left, or odd syllables counted from the right). Theoretically, phonologists interpret these asymmetries to mean that trochaic feet are more common than iambic feet in QI stress systems. From a language-processing perspective, it is reasonable to think that the greater prevalence of stress patterns with stress on word-initial syllables may

[2] Examples of searchable databases with information about stress include The World Atlas of Language Structures Online (https://wals.info/) and resources such as CELEX which can be accessed through the Linguistics Data Consortium (https://www.ldc.upenn.edu/).

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

be related to the usefulness of stress cues in signaling beginnings of new words in speech, a point by Cutler & Carter (1987), who found positive evidence for this proposal in a psycholinguistic study with English speakers.

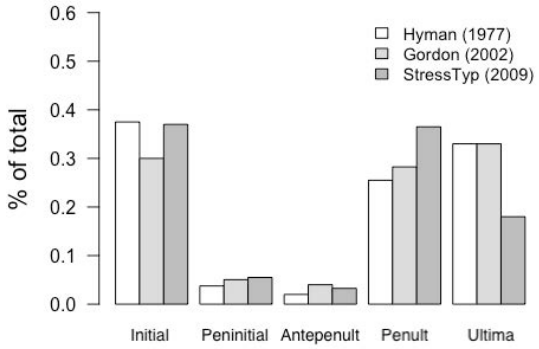


Figure 1: Relative frequency (expressed as percentage of the total cases) of different stress locations in three surveys of stress

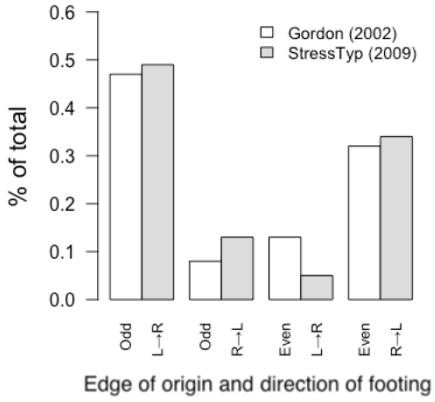


Figure 2: Relative frequency (expressed as percentage of the total cases) of different binary stress patterns in two surveys of stress

Compiled information for quantity-sensitive (QS) stress based on high-quality surveys is less available, but details from WALS (https://www.eva.mpg.de/lingua/research/tool.php), cited in Gordon (2016: 207), suggest a different balance: a sample of 48 languages with iterative QS stress patterns was evenly split between stress iterating from left-to-right vs. from right-to-left. In the same sample of QS systems, interestingly, regardless of the origin of footing (left vs. right), the syllable with primary stress tended to be positioned at the right edge (36 of 48 languages). This asymmetry hints at a dependency between weight sensitivity and alignment in stress patterning that deserves a closer look, and it provides a rationale for the study described in this article.

If it is reliable, the finding that binary alternating stress patterns are rarer than patterns with a single fixed stress might seem counterintuitive to phonologists; historically, theoretical discussions have tended to emphasize iterative stress patterns, as these have provided fertile ground for understanding the constraints (or parameters) which underpin the stress typology. But there is also reason to question the reliability of the frequency information we have. As Gordon (2016) points out, iterative stress patterns may well be underreported; some primary sources may describe main word stress and fail to mention secondary stress. Conversely, alternating secondary stress might in some cases be overreported (see Newlin-Łukowicz 2012, for Polish), given that humans tend to perceive sound sequences rhythmically, even in the absence of

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

physical cues marking rhythmic contrasts (e.g. Bolton 1894). Recent reports of serious flaws in published, well-known descriptions of stress patterns (see Tabain et al. 2014; de Lacy 2014) highlight the importance of rigorously confirming impressionistic descriptions of stress patterns, which have often been provided by researchers who are not native speakers of the languages they have described.

The problem is clear: if we lack sufficient information about the prevalence of a wide range of stress patterns, and if we have low confidence on the information we do have, then ideas about natural stress patterning cannot rely on frequency statistics alone.

1.2 Naturalness and cognitive preferences

We present as our working assumption that a stress pattern (or one of its components) is "natural" to the extent that it is optimized by at least one aspect of human biology. Patterns that are grounded in the mechanics of speech articulation, or which facilitate speech perception, have long been considered more natural than patterns which are not grounded in these ways (e.g. Diehl & Kluender 1989; Archangeli & Pulleyblank 1994; Lindblom 2000). Many biologically grounded sound patterns can be shown to be preferred by humans.

But speech perception and production are not the only human systems for which sound patterns can be optimized and, accordingly, associated with some kind of asymmetry (or general preference) in nature. Some general preferences are based in other aspects of human cognition, for example humans' natural tendencies to group objects (visual or auditory) according to Gestalt principles. We expect some general preferences or biases found among speakers of diverse languages to be cognitively based and, as such, they should be quantifiable independently of noise introduced by native language experience. We further suggest that cognitive preferences for particular components of stress patterns have helped to shape the stress typology. We are not the first to think along these lines. Moreton (2008) draws a distinction between analytic (cognitive) bias and channel bias in considering influences on the evolution of phonological patterns. Hayes (1995) proposed that perceptual biases described by the Iambic/Trochaic Law (ITL) are the cognitive precursors for asymmetries in his foot inventory. The ITL asserts that listeners associate more intense sounds with group onsets and longer sounds with group endings, and there is some evidence to support these claims.[3] The burgeoning psycholinguistic literature on ITL effects attests to the value of studying prosodic patterns in laboratory settings where inferences about study participants' preferences are based on measuring their responses in cognitively based tasks.

In the core psycholinguistic research on ITL effects among human adults, inferences about perceptual grouping biases have been based on participants' responses when they were given two-alternative forced choice or recall-based tasks. Another way in which cognitive biases can be studied is by measuring participants' performance in artificial language learning (ALL) experiments (e.g. Wilson 2003; Finley & Badecker 2009; Culbertson 2012). If natural stress patterns are those which are preferred by humans (are cognitively grounded, per our working assumption), then learning biases can be interpreted as indicators of cognitively based preferences.

ALL studies of stress acquisition can produce two kinds of information that can be useful. First, learning biases can be inferred based on the ease with which humans learn features of unfamiliar stress patterns presented to them in an exposure or training phase of an ALL study. Ease-of-learning can be quantified as an accuracy score based on participants' decisions in a post-exposure/training test phase in which they indicate whether or not they think new vocabulary items belong to the artificial language (AL).

[3] Psycholinguistic studies of ITL effects among speakers of diverse languages have assembled a nuanced picture. There is convincing evidence that humans prefer sound groupings in which louder sounds come first regardless of their language background. On the other hand, there does not seem to be a universal preference for groupings with longer sounds in final position and where this preference has been found, it is weaker than the intensity-based effect and more sensitive to contextual influences. A review of the recent literature on ITL and related effects is found in Crowhurst (2019).

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

This method tests the ability to match novel forms to trained patterns. The second kind of information is provided by ALL studies that add a poverty of the stimulus component. In such studies, participants are trained and tested as just described, but crucial data bearing on a component of the pattern are withheld during exposure. During the test phase, new forms are introduced that require participants to extend what they have learned in making an untrained generalization. The poverty of the stimulus method can be useful in revealing latent selection biases that are not reflected by an accuracy score in a strict exposure-and-matching study. The study we describe in section 2 provides both types of information.

2. The experiment: Overview and general hypotheses

2.1 Background

Given the paradigm's promise, surprisingly little ALL research has explored humans' learning of stress patterns. Some research has modelled the learnability of stress patterns computationally (e.g. Heinz 2009; Stanton 2016; Staubs 2014), for example, by implementing a gradual learning algorithm (GLA) for learning a constraint hierarchy given a data set and a set of weighted constraints. Such studies can be very useful, especially when results from the machine learner are considered together with other information. In one useful contribution, Stanton (2016) explores the midpoint pathology, which refers to an odd set of predicted but non-occurring patterns that are generated when antilapse constraints are highly ranked (Kager 2012). Stanton argues that midpoint systems are unattested not because metrical theory doesn't allow them, but because they are hard to learn, for two reasons. The first is that long forms needed to learn most midpoint systems are low-frequency and therefore unavailable to human learners. The second is that information about stress location is inconsistent in midpoint systems. In one such system, for example, a stress window shifts from the right to the left edge, as words get longer. Machine learning studies can be valuable as a proxy for human learners when investigating unattested patterns, and given that it is difficult to study the acquisition of a stress pattern with any precision in the L1 context.

In the two key human ALL studies, Carpenter (2010, 2016), compared the ability of monolingual American English and Laurentian (Canadian) French speakers to learn "natural" (attested) and "unnatural" (unreported and not predicted under standard theoretical assumptions) stress patterns. In Carpenter (2010), participants were taught quality-sensitive stress rules: in the natural condition, nonce words presented stress on the leftmost of any syllables containing a low vowel, else on the initial syllable when only high vowels were present. In the unnatural condition, the pattern was inverted so that syllables with high vowels were stress attractors. In Carpenter (2016), the stress rules were quantity-sensitive: the special, stress-attracting syllables were closed CVC syllables in the natural condition and open CV syllables in the unnatural condition. In both studies, Carpenter found that participants in both language groups learned the natural pattern more successfully than the unnatural pattern, although there were between-group differences (the English speakers achieved higher accuracy scores and with less training than the French speakers). Carpenter's findings are valuable for establishing that an ALL paradigm can be used to study stress pattern learning, and because they suggest that humans *can* learn patterns that do and don't occur, but may be better at learning attested patterns.

The stress rules taught to participants in Carpenter (2010, 2016) exemplify a "default-to-same" edge stress pattern, one of four related unbounded stress patterns in a mini-typology that reflects different settings of the edge alignment and quantity-sensitivity parameters.[4] This typology is schematically represented in Figure 3. In the pattern Carpenter taught to her participants, shown in Figure (3a), the "special" and default rules are left-aligning; stress falls on the leftmost syllable with a stress-attracting property, and when no special syllables are present, stress defaults to the same (left) edge. The mirror-image, right-aligning default-to-same edge stress pattern is represented in Figure (3d). In

[4] In the early research on this typology, quality sensitivity was treated as a species of weight sensitivity (Kiparsky 1973).

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

default-to-same edge patterns, alignment settings for the special and general rules do not compete. Other languages are reported to have a "default-to-opposite" stress pattern with conflicting edge alignment: in the default case, stress falls on the peripheral syllable at the opposite word edge when no stress attractors are present.

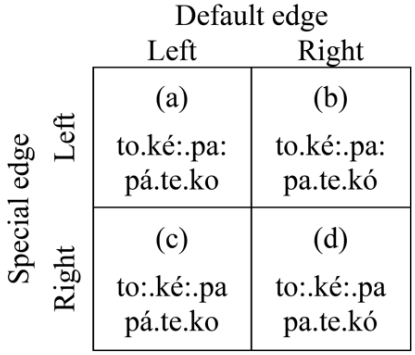


Figure 3: Edge-based stress assignment in default-to-same edge (patterns (a) and (d)) and default-to-opposite edge (patterns (b) and (c)) stress systems. Special syllables are marked with vowel length.

While all of the patterns in Figure 3 are reported to occur in languages (see Figure 10 in our concluding discussion), the existence of default-to-opposite patterns (Figure 3b and 3c) has been questioned (Gordon 2000). Such challenges raise the issue of whether the dependencies between quantity sensitivity and edge alignment that are implicit in default-to-opposite patterns exist. Other information discussed in section 1.1 suggests the possibility of dependencies between quantity sensitivity and edge alignment in iterative QS stress systems.

In our research, we were interested in studying ease-of-learning in relation to edge-alignment oppositions; given the reported prevalence of word-initial stress in QI systems (Figures 1 and 2), we wanted to test whether a left-aligning pattern of prominence might be easier to learn than a right-aligning pattern. In addition, given the possibility of typological dependencies between quantity sensitivity and edge alignment in both bounded and unbounded stress systems, we were interested in discovering whether edge-based asymmetries might be influenced by syllable weight in a study in which the learning target was an unbounded pattern of prominence.[5]

We used a poverty of the stimulus approach to study the following two questions. First, when English speakers are taught a default edge-aligning QI prominence rule and a QS prominence rule that is ambiguous with respect to alignment, which QS rule will they infer when the provision of critical new forms in a post-exposure test phase requires them to choose between aligning edges?[6] Second, are there dependencies between the trained edge for the QI rule and the edge participants infer for the "updated" QS rule?

In this study, adult native English speakers in two learning groups were exposed to a set of trisyllabic nonce words with a single prominent syllable. A subset of these items (the QI set) contained only CV syllables (light syllables, or L); another subset (the QS set) contained a single syllable with a long vowel, CV:, in initial, medial or final position (the heavy syllable or H).[7] One syllable in every trisyllable was

[5] These goals sharply distinguish our research from previous studies. Carpenter (2010, 2016) did not test the learnability of different edge alignment options.

[6] We will replace the term 'stress' with 'prominence' in discussing the ALL study, our stimuli, and what our participants might have been learning. Some care is needed in the choice of terminology given that while we were interested in obtaining evidence that might bear on the naturalness of stress patterns, we cannot be certain that participants in our study associated the patterns they were learning with stress, even though it was explained to them what stress was. What participants in our study learned were rules governing the distribution of prominence cues in the artificial language, where the cues to prominence were duration, higher pitch and increased intensity. We return to this issue in the concluding discussion.

[7] We use the traditional terms "heavy" and "light" for CV: and CV syllables for convenience, without assuming that participants necessarily formed these categories based on the AL data.

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

louder and higher pitched than the others; this we refer to as the prominent syllable (transcribed with an acute accent). Participants were to learn a QS and a QI generalization about the distribution of prominent syllables. The QS rule was that the prominent syllable must be the heavy syllable, if there was one (e.g. *pá:teko*, *paté:ko* or *patekó:*). The QI rule was that when no heavy syllables were present, prominence would fall on the short-voweled syllable in word-initial position (*kétopa,* for a group trained on the left edge), or the final syllable (*ketopá* for a group trained on the right edge). Words with two heavy syllables (the "hold out" set) were withheld during exposure and then added in the subsequent test phase.

The experiment had two phases. In Phase 1, the exposure phase, participants listened to and pronounced lists of 12 words, one by one. Each list was followed by review trials in which a two-alternative forced choice task tested participants' recall for items in the preceding list. In review trials, participants heard "decision pairs" consisting of a trisyllable from the preceding list (the target), and a foil, the same trisyllable with prominence on a different syllable. Participants were tasked with identifying the target. Phase 2 was a test in which participants responded to similar decision pairs based on novel trisyllables. This time, the target was defined as the member of each decision pair that belonged to the AL. In addition to new stimuli representing the exposure patterns, Phase 2 also included decision pairs for a Novel Patterns condition, which consisted of the trisyllables with two Hs (withheld during exposure). The Novel Patterns stimuli required participants to "update" the QS rule by deciding whether the leftmost or the rightmost heavy syllable should be the prominent one. This required participants to make an untrained generalization about edge alignment for the QS set and in these cases, the question was whether they would prefer prominence on the H closest to the trained edge (a default-to-same pattern), or on the one furthest from the trained edge (a default-to-opposite pattern).

2.2 Hypotheses

The study's design allowed for a variety of different outcomes, and competing hypotheses were possible. We expected participants in both groups to learn the trained default edge when no long vowel was present, but we anticipated a learning asymmetry that favored the left edge, for two reasons. First, if the typological asymmetry favoring prominence on initial syllables in QI stress systems (see Figures 1 and 2) reflects a general analytic bias, then we might expect to find this asymmetry reflected in participants' decisions in the ALL study. The second reason was that in participants' native language, English, initial stress predominates in nouns and adjectives with the structure LLL; words like *Pamela, trinity* are more common than medially stressed words like *banana* and *vanilla*. Also, while the stimuli used in the study were trisyllabic, we note that initial stress is also highly prevalent among English bisyllables, whereas bisyllables with stress only on the second syllable are less common (Cutler & Carter 1987). On the other hand, stress on the last syllable of an LLL form is not an English pattern. Both considerations led us to expect performance in the group with the right-edge default stress pattern to be weaker.[8]

We expected both participant groups to learn the basic QS rule. As long as they did, there were competing hypotheses as to what they might decide when presented with the hold-out set (trisyllables with two Hs) in the test phase. One possibility was that response patterns might reveal a bias favoring prominence on the rightmost H, whatever the trained default edge had been, given the asymmetry favoring right-edge primary stress in QS systems reported by Gordon (2016: 207). In this case, assuming that participants learned the trained edge in the LLL condition, a preference for prominence on the rightmost H would be consistent with a default-to-opposite pattern in the left-edge group and a default-to-same pattern in the right-edge group. Alternatively, if Gordon (2000) is right to question whether true default-to-opposite patterns occur naturally, then we might expect a bias favoring prominence on the H closest to the trained edge in the LLL condition.

A final possibility relates to the phonological observation that information-rich and/or prominent phonological structures are often restricted to initial positions in languages, a species of positional

[8] A more detailed discussion of stress asymmetries in the English noun class is deferred to section 5.

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

prominence. In languages with vowel harmony, for example, the full range of vowel contrasts is often restricted to word-initial syllables (e.g. Turkish, Clements & Sezer 1982; Shona, Beckman 1997; Yaka, Hyman 1998). Cutler & Carter (1987) report that the proportion of stress-initial lexical words in running English speech is high, and they argue that English speakers tend to infer a word boundary before a stressed syllable. Stressed long vowels are salient, information-rich segments; they can encode information both about phonological length contrasts and surface stress contrasts. If there is a tendency to prefer salient, information-rich syllables to occur earlier rather than later in words (if not precisely at the beginning), then we might expect participants to favor pronunciations with stress on the leftmost long V when more than one is present, whatever the trained edge. That is, we might expect the responses in the left-edge group to reflect a default-to-same edge preference, and for decisions in the right-edge group to be consistent with a default-to-opposite edge preference.

We did not anticipate that interference from pre-existing biases would be as problematic when a heavy syllable was present as when there was none: English trisyllables in which two light syllables combine with an initial, medial or final heavy syllable (whether closed or open with a tense vowel or diphthong) are common (e.g. *mandible, agenda, Brigadoon*), and primary stress falls on the heavy syllable.

3 Method

3.1 Stimuli and design

Vocabulary constructed for the AL were trisyllables in which light CV and heavy CV: syllables combined an onset in the set /p t k/ and one of the vowels /a e o/, or their long counterparts. Only non-high vowels were used, to minimize unwanted perceptual effects that might be associated with height-dependent differences between vowels. To create perceptual distance between the AL and the participants' native language (English), syllables were recorded by a native speaker of Croatian. Croatian does not have unstressed vowel reduction, as English does, and our intention was to create stimuli that participants would perceive as un-English-like. Syllables were recorded individually in the Croatian carrier phrase *Reći* ___ *tri puta* 'Say ___ three times'. A token of each syllable with modal voicing throughout was excised from the carrier phrase, segmenting at the beginning of the stop closure and at the beginning of the closure for [t] in *tri*.

The study's design called for trisyllables with seven arrangements of L and H syllables. A QI pattern with short Vs, LLL, and three QS patterns with one long V, HLL, LHL and LLH, were used in both the exposure and test phases. Three QS patterns with two long Vs, HHL, HLH and LHH (the novel "hold out" set) were used only in the test phase. Syllables were concatenated into trisyllables subject to the restrictions that no CV combination was repeated and no C occurred more than twice. The occurrence of Vs was not restricted in this way. A trisyllabic combination of CV sequences was never used twice with the same pattern of long and short Vs; however, some CV combinations were used with different length patterns in different phases of the experiment (e.g. *tapoke* in the exposure phase and *tapo:ke:* in the test phase). The occurrence of Cs and Vs in different positions was balanced across items used in each phase of the study. Every trisyllable was prefixed with a 500 ms silent period. The study materials are provided in Appendix A.

In addition to the duration cue, prominence was simulated by manipulating overall intensity, and pitch.[9] Prior to concatenation, syllables were edited individually in Praat (Boersma & Weenink 2019) to normalize vowel duration and intensity. Duration was adjusted to measure 150 ms and 250 ms in short and long Vs by copying or deleting every other voicing period from the sound wave as needed, preserving the vowel's

[9] Duration, intensity and pitch are among the physical cues most commonly used to signal the stress contrast and other types of prosodic structure in languages, although the precise details of their expression vary from one to the next (see Fletcher, 2013, for a crosslinguistic review). It was desirable that prominence-signaling cues in this study should not be implemented exactly as they are in English, given that this was, after all, a type of foreign language learning study.

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

natural amplitude contour. Intensity was normalized by changing gain to 65 dB in non-prominent Vs and 68 dB in prominent Vs. After concatenation, every trisyllable was given a pitch contour in which f0 was level at 225 Hz during the initial, medial, or final third of the prominent V (depending on whether prominence was initial, medial, or final) and declined smoothly to plateau at 180 Hz in non-prominent vowels.

Two lists of twelve vocabulary items were prepared for the exposure phase. List 1 contained only QI trisyllables (LLL), prepared with word-initial and word-final prominence for the left- and right-edge groups, respectively. List 2 contained four trisyllables for each QS pattern (HLL, LHL and LLH). List 2 items, all with prominence on H, were the same for the two participant groups.

Six QI trisyllables and six QS trisyllables (balanced for pattern) were selected for use in post-list review trials, structured as "decision pairs". Each decision pair consisted of a target, a trisyllable from the preceding list, together with a foil, a competing pronunciation with prominence in a different position. As an example, the List 2 target *pekó:te* was paired with a foil *péko:te*, with prominence (incorrectly) on the first syllable. Another decision pair combined *pekó:te* with the other foil, *peko:té*, with final prominence. These decision pairs (*pekó:te – péko:te* and *pekó:te – peko:té*) were presented twice, with target and foil in counterbalanced order. In all, 24 review trials followed each list (6 target words x 2 incorrect foils x 2 orders). Including the 500 ms of silence preceding each trisyllable, the longest decision pairs measured just under 3 sec in duration.

The test phase consisted of another 168 decision pairs based on a new set of trisyllables. In a "Trained Patterns" condition, 15 new LLL trisyllables were used in 60 decision pairs (two target/foil pairs per new trisyllable, with two counterbalanced orders). Fifteen new QS combinations, balanced for H's position, were used in another 60 decision pairs. For decision pairs in a "Novel Patterns" condition, eight trisyllables were prepared for each of three new patterns, HHL, HLH and LHH. Novel Patterns decision pairs combined pronunciations in which prominence was on one H or the other, and were counterbalanced. For example, the trisyllable *ka:te:po* was presented in the decision pairs *ká:te:po – ka:té:po* and *ka:té:po – ká:te:po*. (No Novel Patterns decision pairs allowed participants to choose between prominence on H or L.) In all, there were 48 Novel Patterns decision pairs (8 trisyllables x 3 patterns x 2 orders).

3.2 Testing procedures

All procedures followed a protocol (#2017-08-0089) approved by the Institutional Review Board at the University of Texas at Austin. Consent procedures and a short background questionnaire were administered when participants arrived. Afterwards, seated in a sound-treated room in a campus phonetics laboratory, participants were told they would learn to recognize words belonging to an unfamiliar, constructed language by their sound. They were instructed to pay attention to word stress, which was exemplified using the English words *banana* and *acrobat* as examples.

Training began with List 1 (pattern LLL), which participants moved through by pressing a button to hear each word in sequence. They were instructed to pronounce every word to themselves before proceeding to the next item. List 1 was followed by the 24 List 1 review trials, then List 2 and the List 2 review trials. On review trials, participants were instructed to indicate which member of the decision pair they had heard in the preceding list (the target) by pressing a designated button on a response pad. No response feedback was provided. To maximize learning opportunities, Lists 1 and 2 and the review trials were repeated once. Participants could pause, if they wished, between lists and before advancing to the test phase. Participants were told that vocabulary they would hear in the test phase were new although similar to items they had heard during training. Their task was to indicate which member of each decision pair they thought belonged to the AL (the target).

Participants listened to stimuli through studio-quality Audio-Technica headphones connected to the 13-inch Apple Pro laptop used to run the experiment, which was controlled by SuperLab 5.0 (Cedrus Corporation). The order in which items were presented at every stage was randomized by the software. Most participants completed the experiment in 25 minutes or less.

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

3.3 Participants

Thirty-eight participants were tested in the left-edge group and 43 in the right-edge group. The threshold for inclusion in the statistical analysis was an accuracy score of 70% on the second round of review trials in the exposure phase. Twenty-five participants in the left-edge group (18 women and 7 men) and 24 in the right-edge group (17 women and 7 men) met this criterion. While this exclusion rate is high, we believe it may be due to the absence of performance feedback during training. All participants were native speakers of American English and none but one had had substantial L2 exposure prior to age 13. One participant had begun to learn Spanish in classes in elementary school. All participants had studied at least one foreign language in high school or college. All participants but one were members of the university community and were recruited through a university-wide events bulletin, announcements in undergraduate classes, and by word of mouth. Participants were paid $10 or received course credit for their time. They ranged in age from 18 to 45, with most participants aged between 18 and 25.

4 Results

4.1 Data handling and statistical procedures

The response data were cleaned by eliminating trials on which participants had pressed a non-sanctioned button, and responses with reaction times (RT) shorter than 1,000 ms or longer than 10,000 ms (measured from the beginning of the stimulus event). In fact, there were no RTs between 408 ms and 1000 ms. Most short RTs were below 100 ms, which suggested that the participant might have pressed a response button twice by accident. The upper threshold of 10,000 ms was arbitrarily determined. No stimulus event was longer than 3,000 ms and we considered that a 10,000 ms RT allowed participants 7 seconds after hearing each decision pair to respond. Of the 10,584 observations expected per the design, these exclusions left 10,509 data points for the statistical analysis, an exclusion rate of .71%. The raw data, organized by category and response type, are presented in Appendix B.

The data were analyzed in R and R Studio (R Core Team 2018; RStudio Team 2016). As a first step, exact binomial tests of proportions ($\alpha = .05$) were conducted to determine whether the proportion of positively coded responses for each type of decision pair was significantly better than would be expected by chance. These tests provided an indication of participants' success in generalizing the rules they learned in the training phase to new examples of the same patterns, and of any latent bias in the Novel Patterns condition.

Variations in participants' response patterns across the inventory of decision pair types (which differed by Pattern and Foil) were explored by fitting mixed-effects models (likelihood ratio tests) to the response data, as described in the sections devoted to the training and test results, using the *mixed* function in the *afex* package (Singmann et al. 2019), a wrapper for *lmer* which provides convenience functions for the analysis of factorial data. The dependent variable was a binomial and the interpretation of positively coded responses differed depending on the analysis (as explained below). Edge was a between-subjects factor that encoded the trained edge in the QI condition (two levels, left and right). The within-subjects factors were Pattern, with four levels in the analyses of the Trained Patterns data (LLL, HLL, LHL and LLH), and three levels in the separate, Novel Patterns analysis (HHL, HLH, and LHH). Foil coded the position of the prominent syllable in the distractors paired with targets in decision pairs (two levels, the position closest to the left edge, in foils like *péko:te* and *to:téka*; and the position closer to the right edge, in foils like *peko:té* and *to:teká*). The inclusion of Edge and Pattern in the statistical analysis was hypothesis driven (see section 2.2). Foil was included because, given that participants' task required them to select one member of each decision pair and reject the other, we were interested in the types of errors participants might make. In decision pairs for the QS patterns, in particular, if participants failed to reject a foil/distractor in which prominence incorrectly occurred at the word-edge consistent with LLL forms presented during exposure, this could be interpreted as an effect of the trained edge. Significant interactions in the mixed models

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

analysis were explored *post hoc* by making pairwise comparisons using the *emmeans* package in R (Lenth 2020).

4.2 Recall accuracy in the training phase

Participants were included in the statistical analysis for the test conditions if they scored at least 70% on the second-round review trials in the exposure phase. This score measured recall accuracy, the participants' success in identifying trisyllables they had heard in a preceding vocabulary list. Our primary purpose in analyzing these data was to test for significant group-based differences in recall accuracy as a prelude to analyzing the test data. A visual inspection of Figure 4, which represents the training results by Pattern, indicates that performance in the left- and right-edge groups was comparable.

Recall accuracy was high in both groups overall, although lower for trisyllables ending in a long vowel (pattern LLH). Exact binomial tests conducted on the recall data for both groups, combined, indicated that the true probability of success (PoS) for each pattern, averaged over Foil, was significantly greater than .5, the level of chance. (The outcomes of the binomial tests are shown in Table C1 in Appendix C.) The purpose of the statistical analysis was to test for group-based differences based on the trained edge as a precursor to analyzing the response data for the Trained Patterns test condition. To this end, a mixed effects model was fitted to the recall data using the *afex* function with Edge as a fixed effect, Participant as a random intercept and random slopes for Pattern and Foil. Edge was not significant in this model ($F(1, 47) = 2.06$, $p = .16$).

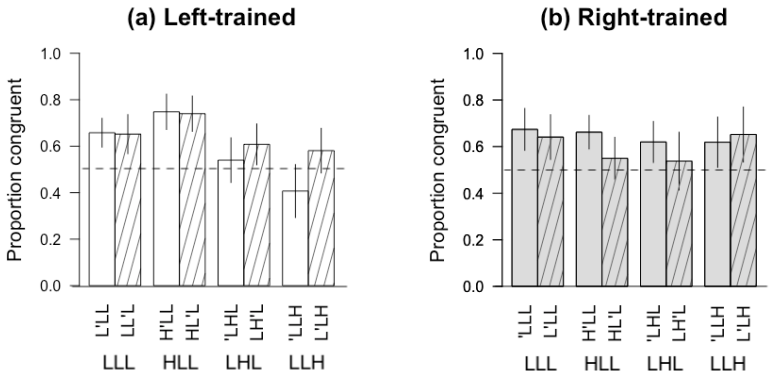


Figure 4: Proportion of correct target identifications by Pattern in review trials

4.3 Congruent identifications in the Trained Patterns test condition

Figure 5 represents the results for the Trained Patterns condition by participant group (Edge), Pattern and Foil. In this condition, the dependent variable measured *congruent responses*. Congruent responses successfully identified targets in decision pairs, but as the stimuli were new, the target was the trisyllable whose prominence position matched the exposure patterns.

A comparison of Figures 4 and 5 shows that success in the test phase was lower overall than on the review trials. We expected this, as participants were not confirming identity in test trials, but were applying learned rules to new forms. Figure 5 shows that success in the two groups was comparable for the QI pattern, but there were differences among the QS patterns: left-trained were more successful than right-trained participants in choosing the HLL target, and less successful for LLH. The proportion of congruent responses was more even across the four patterns in the right-trained group overall although there were differences by type of foil (see below).

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

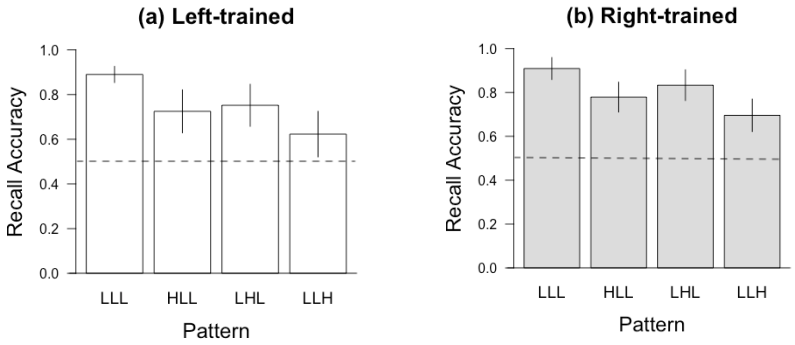


Figure 5: Congruent selections by Pattern and Foil (labelled bars) in the Trained Patterns test condition.

Exact binomial tests conducted for each category (Pattern by Foil) graphed in Figure 5 indicated that the proportion of congruent responses was significantly greater than .5 in all but four cases (see Table C2 in Appendix C). In the left-trained group, success for pattern LLH was significantly lower than chance would predict when the foil had initial prominence ('LLH). (An uptick ['] indicates prominence on the following L.) For pattern LHL, success was not significantly better than chance when the foil had initial prominence ('LHL) in the left-edge group, and final prominence (LH'L) in the right-edge group; and for pattern HLL in the right-edge group when the foil had final prominence (HL'L). These results suggest that participants were less successful in choosing the target when the foil had prominence at or closer to the trained edge.

To test the statistical strength of the differences seen in Figure 5, a mixed effects model was fitted to the Trained Patterns response data using the *afex* function in R. This model included Edge, Pattern, Foil and the interactions Edge:Pattern and Pattern:Foil in fixed effects; random slopes for Pattern and Foil; and Participant as a random intercept.[10] The output of this model, shown in Table 1, reveals that the fixed effect of Pattern was significant. The main effects of Edge and Foil were not significant, but their effects are seen in the significant values associated with the interactions Edge:Pattern, Edge:Foil and Pattern:Foil (discussed below).

The significant interactions in Table 1 were explored by making pairwise comparisons using the *emmeans* package in R (Lenth, 2020). The interpretable contrasts which were revealed to be significant are presented in Table 2. (The full set of interpretable comparisons and the code used in the statistical analysis are provided in Tables C3, C4 and C5 in Appendix C).

Table 1: Mixed model Anova table for response data in the Trained Patterns test condition (Type 3 tests, KR method; observations = 5,837)

*mixed(Accuracy ~ (1+Pattern+Foil|Participant) + (Edge*Foil*Pattern), data = Trained.Patterns)*

Edge	$F(1, 47) = 0.00$	$p = .95$
Foil	$F(1, 55.71) = 0.05$	$p = .82$
Pattern	$F(3, 45) = 5.14$	$p = .004$**
Edge:Foil	$F(1, 55.71) = 6.78$	$p = .01$*
Edge:Pattern	$F(3, 45) = 4.97$	$p = .005$**
Pattern:Foil	$F(3, 5587.64) = 6.40$	$p = .0003$***
Pattern:Foil:Edge	$F(3, 5587.64) = 1.86$	$p = .13$

[10] The model in Table 1 is the most maximal model that converged. The maximal model, with a random slope for Pattern*Foil and the three-way interaction Edge*Foil*Pattern in fixed effects, failed to converge.

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

Table 2: Pairwise comparisons for the Trained Patterns condition in the Test phase

	Contrast	Estimate	*SE*	*df*	*t*	*p*	
a. Pattern:Edge	HLL (left) – LHL (left)	0.170	0.036	46.9	4.715	0.0005	***
	HLL (left) – LLH (left)	0.250	0.058	47.0	4.286	0.0021	**
	HLL (left) – HLL (right)	0.139	0.045	47.0	3.098	0.0600	.
b. Pattern:Foil	LLH: left 'LLH vs. right L'LH	-0.104	0.032	353.5	-3.214	0.0307	*

Although the Edge:Foil interaction was significant in the omnibus model in Table 1, none of the pairwise comparisons was significant. Despite a tendency noted in Figure 5 for participants to make more mistakes when a foil had prominence at the trained left edge, in certain QS patterns, there was no statistical support for concluding that this was an effect of training.

To help with the interpretation of the significant Pattern:Edge interaction in Table 1 and the pairwise comparisons in Table 2a, the response data for Pattern in the two groups, averaged over Foil, are represented in Figure 6. The graph shows that outcomes for Pattern were more variable in the left-edge group (white bars) than in the right-edge group (grey bars). In the pairwise comparisons based on the omnibus model (Table 1), the contrasts between patterns HLL and LHL, and between HLL and LLH were significant in the left-edge group. In other words, the greater success identifying targets when H occurred at the trained edge, compared to the other QS patterns, was significant in the left-trained group. Interpreting, the trained left-edge bias in the left-edge group may have been enhanced when the first V was long. Our contention that this was an effect of the trained edge is supported by the finding that the between-group difference for pattern HLL was marginally significant (comparison HLL (left) vs. HLL (right) in Table 2). None of the by-pattern differences seen in Figure 5b for the right-trained group were significant.

The Pattern:Foil interaction can be meaningfully interpreted only for the QS patterns because target/foil combinations were different for pattern LLL in the left- and right-trained groups. Figure 7 graphs the outcomes for patterns HLL, LHL and LLH by Pattern and Foil. The only contrast that was significant in the pairwise comparison was for pattern LLH: the result in Table 2b confirms that participants were significantly less successful in choosing the target LL'H when the competitor had prominence on the initial L (i.e. 'LLH), than when it had medial prominence, L'LH. This can be seen in the rightmost bars in Figure 7. A comparison with Figure 5 indicates that this effect can be attributed to the weaker (well below chance) performance in the left-edge group.

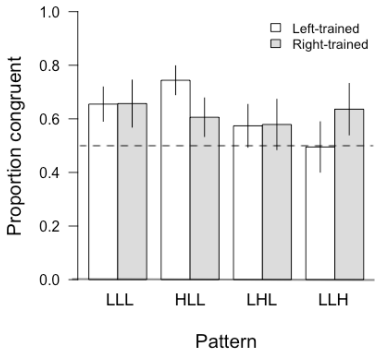


Figure 6: Congruent selections by Pattern and trained Edge in the Trained Patterns condition (white bars: left-trained; grey bars: right-trained)

While the difference between LHL and LLH in the left-edge group was not significant in the pairwise comparisons, the finding that the left-edge group was least successful with pattern LLH is still notable. Interpreting, we associate this outcome with the phenomenon of preboundary lengthening (Fletcher 2013,

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

and references cited therein): if left-trained participants discounted vowel duration in final position, they may have fallen back on the default left-edge pattern and treated the 'LLH foil as an 'LLL target. The difference between the two groups for pattern LLH, although apparent in Figure 6, was not significant in the pairwise comparisons.

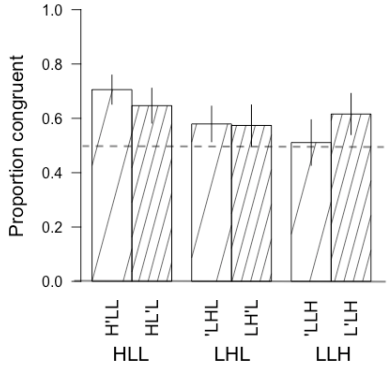


Figure 7: Congruent selections by Pattern and Foil (labelled bars) in the Trained Patterns condition for the left- and right-trained groups combined

How can the findings in the Trained Patterns test condition be interpreted? We begin by observing that the study's outcomes confirm our initial expectations: participants in both groups learned the QI rule for their group and the basic QS rule well enough to perform significantly better than chance would predict for most pattern/foil combinations (see the binomial tests in Table C2, Appendix C). We did not find a stronger effect of left-edge training for the LLL pattern; performance was comparable in the two participant groups in the QI condition. This outcome shows that English-speaking study participants were able to learn a simple prominence-based pattern that is not present in their native language (final prominence in LLL), and they performed as well as participants learning the mirror image pattern (initial prominence in LLL) which is consistent with English stress rules in some lexical classes (at least, nouns and adjectives).

While participants learned the QS rule (prominence cues should mark the heavy syllable, if one was present), the results revealed interesting effects of the trained edge on participants' success for the QS patterns. In the left-trained group, success for the HLL pattern was significantly higher than for the LHL and LLH patterns, suggesting that the effect of edge-training and syllable weight was combinative. (Success for HLL was even higher than for LLL in the left edge group, but this result was not significant.) In general, left-trained participants were more likely overall to pick the member of a decision pair with initial prominence, which resulted in more errors for patterns LHL and LLH when the foil had initial prominence. In these cases, left-trained participants failed to reject the foils 'LHL and 'LLH in favor of the L'HL and LL'H targets, and the proportion of failures in these cases was higher than would be expected by chance. Left-trained participants were particularly bad at rejecting 'LLH foils in favor of the LL'H targets. We see this not purely as an effect of the trained edge, but as related to pre-boundary lengthening: participants trained on the left edge were not primed to attend to final syllables, and we suggest that they were discounting the increased duration of final H syllables in this position. We did not find comparable differences in the right-trained group, where success for patterns LLL and LLH was comparable; if right-trained participants were discounting the duration of final H syllables, then they may simply have perceived 'LLH foils as 'LLL targets. There were other effects of the trained edge in the right-edge group in that participants' worst performance was for patterns LHL and HLL, where they failed to reject foils with prominence at the trained right edge (LH'L and HL'L) in favor of targets with prominence on H (L'HL and 'HLL) more often than in other conditions. The binomial tests in Table C2 (Appendix C) indicate that participants' success was not significantly better than chance in these cases.

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

4.4 Which heavy syllable? Inferring an edge-aligned QS rule

In the Novel Patterns condition, decision pairs offered participants the choice between prominence on the first or the second of two heavy syllables (e.g. *kó:te:pa* vs. *ko:té:pa*). While the participants' task was the same as in the Trained Patterns condition (to identify the member in each decision pair which they thought belonged to the AL), a positively coded response in the Novel Patterns condition meant the participant had selected the trisyllable with prominence on the leftmost H. Figure 8, which graphs the results by Edge and Pattern, reveals a bias favoring the leftmost H for all patterns (HHL, HLH and LHH) in the left-edge group, and no clear preference in the right-edge group.

The output of the best-fitting model for the Novel Patterns condition, which included the terms Edge, Pattern and Edge*Pattern in fixed effects, a random slope for Edge*Pattern, and Participant as a random intercept, indicated that the effect of Edge was significant ($F(1, 47) = 6.52$, $p = .01$). (The output for the full model is provided in Table C7, Appendix C). Binomial tests of proportions confirmed that the bias favoring prominence on the leftmost H in the left-edge group was significantly greater than .5 for all of the patterns HHL, HLH and LHH (see Table C6, Appendix C). Although differences across categories are apparent in Figure 8 for the left-trained group, the output of the mixed effects model indicated that the that the fixed effect of Pattern was not significant.

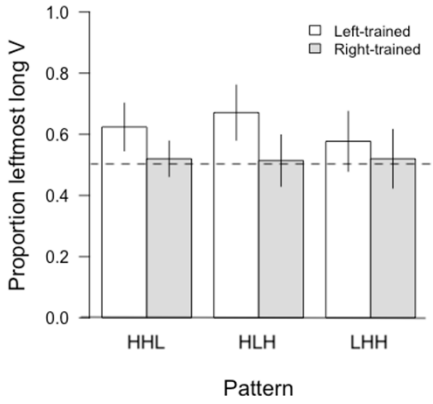


Figure 8: Proportion of selections with prominence on the leftmost long V by Pattern (x-axis) in the Novel Patterns condition in the test phase

However, the group-level outcomes do not tell the whole story. Figure 9 provides a more granular view of response patterns at the participant level, aggregating over Pattern. The x-axis represents the percentage of responses favoring prominence on the leftmost H, while the y-axis represents the number of participants in each bin. In this graph we see that all but three of the participants in the left-edge group (white bars) show a response bias favoring prominence on the leftmost H. A majority of the left-trained participants (20 of 24) cluster in the bins between 60 and 89.9%, with the 60-69.9% representing the central tendency ($M = .625$, $SD = .152$). On the other hand, the preferences of the right-trained participants (grey bars) were somewhat more distributed ($M = .518$, $SD = .140$). Half of the participants (12 of 24) showed a response bias favoring prominence on the first H, and the other half on the second H. These distributions demonstrate that in general, exposure to the left edge QI pattern predicted a left-oriented bias for the QS pattern, but right-edge experience was not strongly associated with either a left- or right-oriented bias.

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

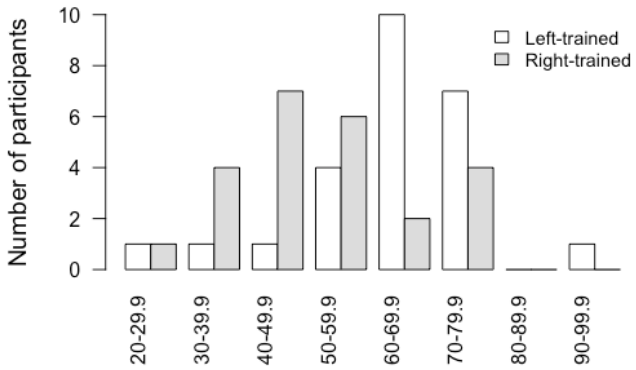


Figure 9: Percentage of leftmost H choices (x-axis) by number of participants (y-axis) in the left-trained and right-trained groups in the Novel Patterns condition

5 Discussion

5.1 Main findings

At the group level, participants in our study learned the QI rule equally well regardless of left- or right-edge exposure. We initially expected that participants might be more successful in learning the left-edge QI rule for two reasons. The first is the typological observation that initial main stress seems to be more common than final main stress in QI stress systems (see Figure 1). As noted in section 1, this asymmetry might be linked to the utility of initial stress in signaling beginnings of words in speech (Cutler & Carter, 1987).

A more compelling reason, we thought, related to participants' native experience of English. In English, initial stress predominates in LLL trisyllables (e.g., *Pamela, trinity*) whereas the mirror image pattern, with stress on the last of three light syllables, is not an English pattern. That the trained edge was learned equally well in both groups, contrary to our expectation, is an important finding: it argues that adults whose native language has a stress contrast can learn an AL with non-native characteristics, and their ability to do so is not completely constrained by L1 biases (see section 5.2, below). This should not necessarily be surprising, since adults can, after all, learn foreign languages, but it is useful to have evidence from a controlled ALL study.

We expected participants in both groups to learn the QS rule, and this expectation was broadly confirmed by the study's results. In the Trained Patterns condition, success in identifying targets was significantly above chance in most Pattern/Foil combinations. Interestingly, in light of the outcome for the LLL pattern, we observed effects of the trained edge in both participant groups for the QS patterns. The influence of trained edge was most dramatic in the left-edge group, where success was significantly higher when the heavy syllable was at the trained edge (the HLL pattern) than when it was not (the LHL and LLH patterns).

There was more variability across patterns in the left-edge group: success was significantly higher for HLL than other QS patterns. Success across patterns was more even in the right-edge group and none of the differences seen in Figures 5b, 6 and 7 for this group were significant. Our interpretation of the higher success for pattern HLL is that left-edge participants' attention to the word-initial syllable was enhanced when the vowel in that position was long. We suggested that the lower success in the left-edge group for pattern LLH may reflect participants' discounting vowel duration in final position due to their implicit expectation of pre-boundary lengthening. By contrast, performance for patterns LLL and LLH in the right-edge group was roughly comparable and as noted, differences by Pattern were not significant in the right-

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

edge group). The result for pattern LLH in the right-edge group is hard to interpret: it might suggest a greater sensitivity to final vowel duration in the right-edge group, which could also be interpreted as an effect of the trained edge. On the other hand, right-edge participants might also have discounted final vowel duration, in which case, they might not have been distinguishing LLH from LLL stimuli.

When differences by Foil were examined, success identifying targets was lower in both groups when the foil had stress on a light syllable at the trained edge. In each group, decision pairs for two patterns had targets with stress exactly at the trained edge (LLL and HLL in the left-edge group; LLL and LLH in the right-edge group). In the other two patterns (LHL and LLH in the left-edge group; LHL and HLL in the right-edge group), one target/foil combination provided an opportunity to reject the target in favor of a foil with stress on the L at the trained edge. Participants in both groups chose the target less often when this was the case. That is, for pattern LHL, success for a target like *peká:te* was lower in the left-edge group when the foil was *péka:te,* and lower in the right-edge group when the foil was *peka:té*. This effect was especially pronounced for pattern LLH in the left-edge group, where a target like *kopeté:* was rejected significantly more often when the foil had initial stress, *kópete:*.

In the Novel Patterns condition, we observed a preference for stress on the leftmost of two heavy syllables in the left-edge group for all patterns (HHL, HLH and LHH), a bias which shows the influence of the trained edge. In contrast, there was no obvious effect of the trained edge in the right-edge group; when asked to indicate whether stress on the first or the second of two long Vs was more characteristic of the AL, the outcome revealed no clear preference, overall. However, when performance was examined across participants, a clear difference between the two groups was again observed: training on the left edge predicted a left-edge bias (21 of the 24 left-trained participants), whereas right edge training did not predict a preference for either edge; participants were evenly divided between a bias favoring one edge or the other. The indeterminate outcome in the right-edge group was interesting, given that the findings in the Trained Patterns condition indicated that these participants had clearly learned both the QI and the QS rules, as well as limited evidence of a bias favoring the trained edge when a single long V was present.

5.2 Can the outcomes be explained by native language experience? A more detailed look

We expect native language experience to have an effect on L2 learning, including in ALL studies. L1 effects on learnability are interesting in their own right and worthy of future investigation. In this section, we try to pick apart the possible influence of experience with English stress patterns and the effects of exposure to AL forms in the study. Where might biases due to AL exposure be observed in our data, and which patterns are not so easily explained in these terms?

Native language characteristics can influence L2 learning at every level of the grammar, including the acquisition of prosody (e.g. Portin et al. 2008; Vainio et al. 2014; Rasier & Hiligsmann 2007), and L1 influences can be either positive or negative. While some research has shown that phonological "wordlikeness" in the L1 improves vocabulary learning in ALL studies (as in other L2 learning), other research supports a "low-wordlike" advantage, especially at the initial stages of learning, as novelty can make less word-like items more salient (Bartolotti & Marian 2017, and works cited therein). Learners are influenced by L1 stress characteristics in acquiring both novel L1 vocabulary and in straightforwardly perceptual tasks. Sulpizio & McQueen (2011) show that Italian speakers' abstract knowledge of lexical stress patterns in the native language facilitated their learning and recognition of novel L1 words. In the literature on perception, research on stress deafness has shown that speakers of languages with predictable or no word-level stress (e.g. Finnish, Hungarian, French) have more difficulty perceiving the stress contrast in nonce words than speakers of languages with unpredictable stress, such as Spanish (Peperkamp et al. 2010, and references cited there).[11]

[11] In response to Peperkamp et al. (2010, last paragraph), we suggest that the real issue is likely to be the degree to which stress is variable, not whether stress is predictable, *per se*. In a language with a QS stress pattern, such as

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

As for our participants, two different kinds of information available to native English speakers could have influenced the study's outcomes. The first would be frequency statistics about word stress position, both in speech and in the lexicon. The second would be native speakers' intuitive knowledge of English stress rules. Metrical analyses since Chomsky & Halle (1968), at least, have claimed that although English stress patterns are partially lexicalized, rules positioning primary stress in all word classes operate mostly at the right edge.

To have a sense for the types of words native English speakers actually encounter, we conducted a simple survey of primary stress in nouns in the CELEX database, using the Reelex interface, Version 0.4.5 (http://web.phonetik.uni-frankfurt.de/simplex.html).[12] Table 3 provides information for underived nouns and Table 4 for morphologically complex nouns.

Table 3: Distribution of main word stress in underived nouns in the CELEX database of English

		Words		Tokens	
	Stress position	Number	% of total	Number	% of total
Monosyllables		2,202	44.86	1,851,449	80.02
Bisyllables	Initial	1,773	36.12	378,526	16.36
	Final	187	3.81	33,251	1.44
	Subtotals	*1960*	*39.92*	*411777*	*17.80*
Trisyllables	Initial	391	7.96	38,655	1.67
	Medial	202	4.11	7,615	0.33
	Final	43	0.88	1,875	0.08
	Subtotals	*636*	*12.96*	*48145*	*2.08*
Tetrasyllables	Initial	5	0.10	37	< 0.0001
	2nd	42	0.86	1,423	0.06
	3rd	54	1.10	883	0.04
	Final	1	0.02	4	< 0.0001
	Subtotals	*102*	*2.08*	*2347*	*0.10*
Pentasyllables	3rd	9	.02	106	< 0.0001
Totals		4,909		2,313,824	

The information in Tables 3 and 4 paints a striking picture of asymmetries in word length and stress position in the data English speakers are exposed to. In Table 3, we see that monosyllables account for 80.02% and bisyllables another 17.8% of tokens in the CELEX database for monomorphemic nouns. The lexical counts (distinct words) are less extreme (44.86% for monosyllables and 39.92% for bisyllables). Trisyllables account for a paltry 2.08% of monomorphemic tokens and a larger share, 12.96% of lexical nouns in CELEX.

English, the distribution of primary and secondary stress is generally predictable but varies depending on the positions of heavy syllables. English speakers are clearly able to perceive stress contrasts.

[12] We are grateful to Henning Reetz for providing this resource.

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

Table 4: Distribution of main word stress in morphologically complex nouns in the CELEX database of English

		Types (Lexemes)		Tokens	
	Stress position	Number	% of total	Number	% of total
Monosyllables		173	0.64	23,792	3.13
Bisyllables	Initial	7,545	28.02	224,893	29.58
	Final	828	3.07	14,370	1.90
	Subtotals	*8,373*	*31.09*	*239,263*	*31.47*
Trisyllables	Initial	6,592	24.48	117,249	15.42
	Medial	2,390	8.87	154,508	20.32
	Final	732	2.72	7,318	0.96
	Subtotals	*9,714*	*36.07*	*279,075*	*36.70*
Tetrasyllables	Initial	2,073	7.70	22,263	2.93
	2nd	1,609	5.97	68,910	9.06
	3rd	1,726	6.41	75,129	9.88
	Final	217	0.81	99	0.01
	Subtotals	*5,625*	*20.89*	*166,401*	*21.88*
5+ syllables	Initial	343	1.27	1,505	0.20
	2nd	332	1.23	4,193	0.55
	3rd	780	2.90	16,914	2.22
	4th or later	1,592	5.91	29,216	3.84
	Subtotals	*3,047*	*11.31*	*51,828*	*6.82*
Totals		26,932		760,359	

The counts for monosyllables in Table 4 are dramatically lower; this is expected, given that inflectional and derivational suffixes commonly add one or more syllables. Bisyllables account for 31.47% of tokens and 31.09% of lexical words. The percentage of derived trisyllables is slightly higher, 36.70% of tokens and 36.07% of distinct words. Words of more than three syllables account for only 3% of distinct words and about .01% of tokens in Table 3. The percentages of derived words (Table 4) longer than three syllables are higher: 32.20% of distinct words and 28.7% of tokens. Counting only monomorphemic and morphologically complex nouns in CELEX, 75.27% are monomorphemes. If the CELEX counts (which are actually based on print sources) are at all representative of speech, then English speakers' surface exposure to nouns is overwhelmingly skewed toward words of fewer than three syllables.

When stress position is taken into account, we find that bisyllables overwhelmingly have initial stress: 90.46% of lexical words and 91.92% of tokens in Figure 10; and 90.1% of lexical words and 93.99% of tokens in Table 4. Table 3 shows that among the monomorphemic trisyllabic lexical words, the ratio of initial to medial stress is roughly 2:1, and only 6.8% of lexical words have final stress. Among trisyllabic tokens, primary stress predominates (80.29% of tokens), with medially stressed tokens running far behind (15.82%) and tokens with final stress a distant third (3.89%). The disparities among morphologically complex trisyllables (Table 4) are less pronounced, but here initial stress still predominates among lexical words at least, 67.86%. Among tokens, the percentage of medially stressed trisyllables is higher, 55.36%, compared to trisyllables with initial stress, 42.01%.

Counting only monomorphemic and morphologically complex nouns in CELEX, 75.27% are monomorphemes. Having no basis to infer morphological complexity (or take account of other special circumstances), we assumed that participants in our study would treat the trisyllabic

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

exposure stimuli as morphologically underived and paradigmatically regular; the figure just cited suggests that our assumption is reasonable.[13] If the CELEX counts (which are actually based on print sources) are at all representative of speech, then English speakers' "language-in-use" exposure to nouns is overwhelmingly skewed toward words of fewer than three syllables, and the predominance of word-initial main stress is clear, both in the lexical and the token count.

The initial stress bias indicated by our CELEX counts might seem to predict the success of the group in our study who learned the left edge QI rule, but it does not predict the success of the other group in learning the right edge. One way to think about this is that high-frequency patterns do not account for everything a language user learns, and a relatively small number of forms, it would seem, can provide the data needed to form a different generalization. It is generally accepted that the distribution of stress in English is mostly rule governed, although complicated by factors such as contextual quantity-sensitivity, lexical class and morphological conditioning, and the existence of subpatterns and exceptions which may be associated with loan status, for example (Chomsky & Halle, 1968; Hammond, 1984; Kager, 1989).[14] The counts in Tables 3 and 4 do not take account of differences in syllable weight and the positions of heavy syllables. In many initially stressed words, the first syllable is heavy (e.g. *fountain*), and the rule for English would be that a heavy syllables in other positions also attract main stress (for example, *giráffe*, *agénda*, *Sàskatóon*). Phonologists argue that English has a right-edge stress rule (Chomsky & Halle 1968; Hammond 1984; Kager 1989; Domahs et al. 2014). Primary word stress is restricted to a word-final trisyllabic window, and which syllable in the trisyllabic window has primary stress depends on factors including syllable quantity and extrametricality, which varies depending on lexical class and derived vs. underived status (e.g., Hayes 1982). Initial stress in a trisyllable (such as the LLL and HLL patterns in our study) is consistent with a word-final trisyllabic stress window, and indeed, we note that the right-edge location of the window can only be learned based on long words (e.g. *hàmamèlidánthemum, sùpercàlifràgilìsticsèxpiàlidócious*), which as the CELEX counts attest, are low frequency in English. It is also clear from bisyllables and words of four or more syllables in which multiple syllables are stressed (*pòntóon*, *Mìnnesóta*) that the primary stress in English is located by a right-edge rule.[15]

Given arguments that at least the English main stress rule works from the right word edge, we think that a right-edge learning advantage could have been considered as a competing hypothesis. If there are reasons to think that English speakers might be sensitive to either the right or the left word edge in stress positioning, what might our results have to tell us about the *kind* of implicit L1 knowledge that may have influenced participants' stress pattern learning in our study – knowledge of phonological rules? Or knowledge of frequency statistics relating to the distribution of stressed syllables in words (the initial stress bias indicated by the CELEX counts)? These issues cannot be completely teased apart in a study in which all stimuli were trisyllabic: since English's phonological rules allow primary stress to appear no further to the left than the antepenultimate syllable, demonstrating an influence of the phonological generalization would require observations

[13] A future study of this type could control participants' assumptions about lexical class by pairing stimuli with images of objects in the exposure phase, or presenting stimuli in a frame of the type "This is a ____". In either case, the "foreign" words would be learned as nouns.

[14] See Hammond (1999) for an excellent description of English word stress patterns and review of the literature.

[15] In trisyllables, initial main stress in HLH words (and also LLH forms like *Bígelòw*, *chíckadèe*) is the result of a stress-retracting rule (the so-called Nightingale Rule). The medial CVC syllables in *nightingale* and *Arkansas* pattern as light between stresses. Word-final CVC syllables also pattern as light; only syllables containing certain vowels and diphthongs attract stress word-finally in nouns.

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

based on stimuli longer than three syllables. An ALL study that interrogates the issues raised in this paragraph would be a useful future undertaking.

We can offer suggestions based on the findings of the current study. As noted in section 5.1, that the trained edge in the LLL pattern was learned equally well in the two participant groups is an important finding that argues against a completely constraining role for L1 biases, at least in the QI condition. Participants' success in learning the left-edge QI rule was not surprising, given the prevalence of the 'LLL pattern in English (e.g. *trinity*). However, if this dominant native pattern influenced participants' learning, then right-trained participants should not have demonstrated so strong a trained edge effect in the QI condition, given that LL'L is not allowed in English: word-final light syllables are never stressed in English. Moreover, if left-trained participants were relying on implicit L1 knowledge (whether of a right-aligning phonological rule or of more probabilistic distributional statistics), we would have expected to observe larger differences depending on stress position in the foil. Specifically, we would have expected fewer congruent decisions when the foil was L'LL than when it was LL'L; L'LL forms exist as exceptional patterns in English (e.g. *banána, vanílla*) and should compete with the 'LLL pattern, while the un-English-like LL'L pattern should have been rejected more frequently. In sum, our findings do not support the conclusion that participants' learning of the trained edge in the QI condition was influenced by L1 knowledge of stress positioning; this is contradicted by the success of the right-edge group.

We observed effects of the trained edge in the QS Trained Patterns and Novel Patterns conditions. These were most pronounced in the left-edge group, but we argue that they were also present in the right-edge group. We expected success in the QS Trained Patterns condition to be above chance regardless of trained edge because although the phonetic exponents of stress in our study differed from English, underived trisyllabic nouns with a single heavy syllable with primary stress in any position are familiar to English speakers (e.g. *góndola*, *agénda*, *Sàskatóon*). In the left-edge group, as noted, we did find a significant left-edge advantage for the QS pattern HLL compared to patterns LHL and LLH, and in the Novel Patterns condition, forms with stress on the first of the two heavy syllables were chosen significantly more often than chance.

Importantly, words in which only one of two heavy syllables attracts stress do not exist in English. However, if participants were influenced by their L1 knowledge of primary stress patterns, then the left-edge participants' bias favoring stress on the leftmost heavy syllable in the Novel Patterns condition might be expected, at least for the HLH and LHH patterns. Among underived trisyllables with a HLH pattern, stress on the first heavy syllable is the general rule, as in *níghtingàle* and *Árkansàs* (Hayes, 1982; Hammond, 1999), although forms with main stress on the second H also exist (e.g. *kàngaróo, Tìmbuktú*). While the LHH pattern seems to be less common (perhaps in part because the set of syllables patterning as heavy in final position is more restricted than in nonfinal position), main stress naturally falls on the first H (e.g. *stalágmìte*; *Nanáimò*, a place name in British Columbia).[16] However, this explanation is not available for the HHL pattern; in words of this shape, English's rules place main stress on the second H (e.g. *bàndána, Hèlsínki, Phìlémon*).

What is also not easily explained in terms of an influencing role for L1 stress knowledge is the absence of any overall bias in the right-edge group. There were no significant differences across Patterns in either the Trained or the Novel Patterns condition in the right-edge group, and in the Novel Patterns condition, there was no group-level preference for stress on either the leftmost or

[16] Note that word-final closed syllables with short/lax vowels pattern as light in English nouns. Only syllables with certain vowels and diphthongs pattern as heavy in this position. See Hayes (1982) and Hammond (1999) for detailed discussions.

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

rightmost H. Our interpretation of the outcomes we observed in the left-edge group is that an initial stress bias for which there is some evidence in English may have enhanced the effects of left-edge training in our study, even though the initial stress bias is stronger in English LLL and HLL words than in LHL words. (It should be noted that in LLH nouns in English, the initial syllable is assigned stress, whether primary, as in *chíckadèe*, or secondary, as in *Sàskatóon*.) Left-edge training in the study may have made these participants more sensitive to vowel length in initial position than elsewhere, but did nothing to increase their sensitivity to vowel length in final position where they may have discounted increased duration, which we have suggested is related to the expectation preboundary lengthening. We do not see any obvious influence of L1 stress patterns in the right-edge outcomes for the QI and QS Trained patterns, but we do see modest effects of L2 training in the study. The absence of significant differences by Pattern in the right-edge group suggests that right-edge training may have worked *against* participants' knowledge of native English stress patterns and perhaps, to some extent, against a natural tendency to discount final vowel length. Harder to explain is the greater variability at the participant level in the Novel Patterns condition in the right-edge group: as many participants preferred stress on the first as on the second of the two heavy syllables in HHL, HLH and LHH items. This outcome suggests neither a strong influence of the trained right edge, nor very clearly participants' implicit knowledge of how trisyllables with multiple heavy syllables are assigned stress in English.

Anticipating the discussion in section 6, we note that the by-participant variability in the Novel Patterns condition may reflect a more general dispreference for stress on the last of a series of heavy syllables, when there is a choice. We recognize the need for caution in raising the possibility of general analytic biases, given that only English speakers were tested in our study. On the one hand, we would argue that testing the learning of opposing characteristics of stress patterns among speakers of the same language is a good starting place for studies of this kind: given that the participants' exposure to native stress patterns was presumably comparable in both groups, it is possible to draw careful conclusions about differences based on training. However, we would emphasize that this study is a first step, and achieving a fuller understanding of general biases will require studies with speakers of languages whose prosodic characteristics differ.

Summarizing, our results point to L2 training effects and the possible influence of knowledge about native stress patterns in the left-edge group. Our interpretation of the outcomes in the left-edge group is that an initial stress bias which exists in English and left-edge training together may be responsible for a left-edge bias favoring trisyllables with stress on an initial heavy syllable and on the first of two heavy syllables. Importantly, not all of the ways in which the left edge bias was revealed in the left-edge group can be explained in terms of L1 biases, and we conclude that we have demonstrated effects of exposure to the AL pattern. We did not see clear evidence for L1 biases in the results for the right-edge group. We observed a strong effect of the trained right edge for the QI pattern and moderate effects of the trained edge in right-trained participants' biases for the QS patterns.

6 General discussion and concluding thoughts

We started out by noting that ideas about which stress patterns are most natural are often based on typological frequency statistics.[17] However, conclusions based on frequency statistics are only as solid as the typological record, and the quality of stress transcriptions in primary sources. Questions about the reliability of both kinds of information can and have been raised (e.g. de Lacy

[17] An awareness of this assumption should be part of the common ground.

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

2014; Tabain et al. 2013). The general motivation for this study, as a first step, is the idea that naturalness in sound patterning (in this case, phonological stress) can also by studied by conducting ALL experiments, by now, a well-established methodology in laboratory phonological research. In our study, participants were exposed to new patterns and tested to see both how well they recognized a form they had previously heard which was presented together with a different stress pronunciation; and how well they generalized the exposure patterns to novel forms. If participants' ability to perform these cognitive tasks is superior for one pattern, as compared with another (possibly opposing) pattern, then there may be grounds for considering the first pattern to be more natural than the second, in that sense. We believe that information from studies of this kind can potentially be used to support or challenge conclusions about natural stress patterning based on typological frequency statistics.

It is important for us to signal that we cannot know for certain that the participants in our study learned a stress pattern *per se*; the most we can do is make inferences about what they perceived based on their responses in cognitive tasks. Our participants learned a prominence-based pattern which was modelled after stress patterns found in languages, and the cues used to signal prominence were typical of stress cues documented in many languages (see references cited in the early sections of the paper). We believe that any biases observed in a study like ours, if they can be teased apart from biases related to native language experience, can shed light on basic preferences that over time, may have contributed to shaping the stress typology in human languages.

We were interested in preferences relating to the edge alignment and quantity-sensitive parameters. Typological information suggests that in QI stress systems, both iterative and noniterative main stress occurs more frequently on word-initial syllables than closer to the right edge (see Figures 1 and 2). As noted earlier, a general bias favoring word-initial stress makes sense, given its utility for the word segmentation task; some researchers have argued that English speakers associate stressed syllables with beginnings of new words (Cutler & Carter 1987). The typological left-edge asymmetry is weaker in QS stress systems. As noted in section 1.1, Gordon (2016: 207), observes a different balance in a 48-languages with iterative QS stress. The origin of footing was at the left edge in half of these languages, and at the right edge, in the other half. Regardless of the direction of footing (left vs. right), the primary stress was positioned at the right edge in 75% of the languages in the same (36 of 48 languages). This asymmetry hints at a dependency between weight sensitivity and alignment that provided a rationale for our study.

The default-to-opposite/default-to-same mini-typology that we have used as a point of reference in this study provides evidence of a different kind of dependency between edge alignment and weight-sensitivity, but in unbounded stress systems. These patterns were first introduced schematically in Figure 3. Figure 10 provides lists of languages which are reported to exemplify the patterns in this mini-typology (Hayes 1995; Gordon 2000; Walker 1997; and Heinz 2009, citing Bailey 1995).[18] These lists bring together the more straightforward cases discussed in the literature,[19] and we note again that there are questions about the reliability or consistency of the sources from which these lists are drawn. A case in point is Khalkha Mongolian, which is not included because it has been variably described as a right-orienting default-to-same edge system

[18] Some of the languages in Figure 10 have additional contingencies. For example, some (at least Kashmiri, Buriat, Classical Arabic, and Northwestern Mari) do not allow stress on a heavy syllable in final position.

[19] Other, somewhat more complicated examples in dominant/recessive accentual systems such as Abzhuy Abkhaz (Spruit, 1986) and Cupeño (Hill and Hill 1968; Crowhurst, 1994) also exist. Primary sources for the languages in Figure 10 can be found in the works cited.

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

(quadrant (c)) by Street (1963) and Walker (1997), but as a default-to-opposite edge system (quadrant (d)) by Bosson (1964) and Poppe (1970).

Default to same edge	*Default to opposite edge*
(a) Leftmost QI/Leftmost QS	(b) Rightmost QI/Leftmost QS
Amele Au Indo-European Kashmiri Lhasa Tibetan Lushootseed Maori Mordwin Murik Serbian/Croatian Yana	Kwak'wala Komi Jaz'va
(c) Rightmost QI/Rightmost QS	(d) Leftmost QI/Rightmost QS
Aguatec Mayan Golin Hill Mari Hindi (per Kelkhar) Klamath Mam Sindhi Seneca	Buriat Chuvash Classical Arabic Dongolese Nubian Huasteco Kuuku-Ya'u Meadow Mari Northwestern Mari Selkup

Figure 10: Languages with default-to-same edge and default-to-opposite edge stress systems. (In some systems, the special syllable may be characterized by some quality other than weight.)

Given the above concerns, any observations based on information from so small a sample can only be provisional: languages with these patterns are not widely attested, and some of the languages in Figure 10 are related varieties (spoken in regions of the former Soviet Union). Nonetheless, we observe that default-to-same patterns predominate over default-to-opposite patterns in this set, and the main gap in the typology is in quadrant (b): patterns with a right-edge QI and a left-edge QS rule seem to be less common.

Turning to the results of our study, it is not possible to infer a natural preference for left- or right-aligning stress in the QI context; participants learned the QI rule equally well, regardless of edge exposure. However, of the level of the system in which both the QI and QS rules operated, it is possible to say more. The decision patterns of 69.4% our participants (34 of 39) were consistent with a default-to-same edge pattern. Of the 25 participants who learned a left-edge QI rule, 22 inferred a left-edge QS pattern, when given the opportunity to do so. Only 3 of 24 participants with left-edge training inferred a default-to-opposite stress system. The picture was quite different for the participants who learned the right-edge QI rule. Twelve of these participants inferred a right-aligning, and 12 left-edge QS rule.

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

Summarizing, just over two-thirds of our participants showed patterns consistent with a default-to-same edge pattern, with the left QI/left QS option dominating. The smaller share of the participants were consistent with a default-to-opposite pattern. Interestingly, the asymmetry on this group did not turn out as one would expect, if the typological information reflected in Figure 10 were reliable: the default-to-opposite pattern most prevalent in our data was seen with participants who learned the right-edge QI rule and inferred a left-edge QS rule.[20]

Taking our results as suggestive, and without making firm claims about universality, how could we think about asymmetries in the decision behavior of our participants in language processing terms? And how can we relate the default-to-same and default-to-opposite patterns we are discussing to notions for complexity in formal terms?

Beginning with the formal, it is worth noting that default-to-same patterns are easier to model in optimality-theoretic terms (Prince & Smolensky 2004) insofar as these analyses require only simple constraints and no ambiguity in constraint rankings. The left QI/left QS pattern requires a dominant constraint **StressedL*, which punishes stress on light syllables, and requires the gradient alignment constraint *Align-Left* (assign a cost to every syllable standing between the main stressed syllable and the left word-edge) to be ranked above its mirror-image right-edge counterpart. The result of this constraint ranking is shown in Tables 5 and 6, for QI and QS forms, respectively.

Table 6: OT analysis for a QI form in a left-orienting default-to-same edge system

	*StressedL	Align-Left	Align-Right
☞ ˈLLL	*		**
LˈLL	*	*!	*
LLˈL	*	*!*	

Table 7: OT analysis for a QS form in a left-orienting default-to-same edge system

	*StressedL	Align-Left	Align-Right
ˈLLH	*!		**
LˈLH	*!	*	*
☞ LLˈH		**	

Default to opposite edge patterns are more complex to analyze in OT because they require a dependency which can be modelled as a Boolean conjunction: *StressedL is conjoined with one of the alignment constraints and the conjunction is violated only when both conjuncts are violated. Align-Left and Align-Right are ranked below conjunction. Some replication of constraints and ambiguity in their rankings is involved because whichever of the alignment constraints is conjoined with *StressedL, its mirror image must dominate in the hierarchy below. Tables 8 and 9 model the analysis for the default-to-opposite pattern which was most consistent with the decision behavior for the participants in our study, the right QI/left QS pattern.

Table 8: OT analysis for LLL in a right QI/Left QS default-to-same edge system

	*StressedL|^|Align-Right	Align-Left	Align-Right
ˈLLL	* |*!| **		**
LˈLL	* |*!| *	*	*
☞ LLˈL	* | |	**	

[20] We caution yet again that we cannot make too much of the difference between the lengths of the lists in Figure 10 and the decision patterns we observed in our study, given uncertainties about the extent to which the information in Figure 10 can be considered representative.

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

Table 9: OT analysis for HLL in a right QI/Left QS default-to-same edge system

	*StressedL|^|Align-Right	Align-Left	Align-Right
☞ ˈHLL	| |**		**
Hˈ LL	* |*!| *	*	*
HLˈ L	* | |	**!	

One of our goals at the outset was to probe for whether learning asymmetries in our ALL study would reflect dependencies between quantity sensitivity and the aligning edge. One of our main questions was this: if participants learn a QI edge-aligning pattern, and are then given a task requiring them to infer the aligning edge for a QS pattern, will they choose the same edge, or will they choose the other edge? In the first (same) case, there is no obvious dependency between the two parameters, but in the second there would be, if adding a weight-sensitive dimension changed alignment preferences. We have seen in the OT models above that it is easier to model the no-dependency (default-to-same edge) patterns, and harder to model the default-to-opposite patterns in which there are dependencies.

Asymmetries in the decision patterns of participants in our study leaned in favor of the no-dependency default-to-same edge patterns, and this was overwhelmingly the case for participants who learned the left-edge QI rule. Furthermore, for the smaller number of participants whose response asymmetries were consistent with a default-to-opposite pattern, the QS rule was in most cases (12 or 15 participants) the left-aligning one. Tendencies observed in studies such as this one are probabilistic and as such, are not so easily modelled in optimality-theoretic terms. We can, perhaps, make sense of these tendencies if we turn again to the basic word segmentation problem. A left-aligning, default-to-same edge bias fits what we have said earlier about the utility of prominence cues (stress) in signaling new words. The ability to learn a right-edge QI rule (as our participants did) confirms that humans can use prominence cues demarcatively to identify word endings (and of course, there are languages with demarcative final stress). What is interesting is that when given the opportunity to infer a QS edge aligning rule, half of our right-trained participants chose the left.

Packaging these observations in terms of the issue of dependencies, we found strong evidence for default-to-same edge patterns, with no particular dependencies between edge alignment and quantity sensitivity. The most prevalent tendency we observed was for the left-aligning default-to-same edge patterns. Where the asymmetries in our participants' decision patterns did seem to suggest a dependency (the default-to-opposite cases), the dependency favored a left-aligning QS rule. We believe that our results may suggest a dispreference for dependencies, in this type of case, but also strong evidence for positioning prominence cues closer to the beginnings of words. This outcome is not completely predicted by stress patterns in English, the participants' native language.

The focus of the current study has been on edge alignment in a set of prominence-based patterns which can be related to stress patterns occurring in languages. However, alignment is an important component of other types of phonological patterning. In one common type of tone pattern, H tone associates to a tone-bearing unit near the left edge of a verbal constituent and spreads rightward (e.g., Shona, Myers 1990). Tones do not typically align at the right and spread to the left. Mimetic palatalization in Japanese presents a segmental example of the pattern in Table 10, quadrant (d): in mimetic forms, the rightmost non-rhotic coronal C is palatalized, and if there are no coronals, then the leftmost C is palatalized (Mester & Itô, 1989). Future studies using the ALL paradigm could fruitfully investigate and compare the naturalness of edge alignment options in different phonological domains.

In conclusion, our findings indicate that adult English speakers can learn a prominence-based pattern that differs from stress in their native language, but that shared features with L1 stress patterns (in this case, aligning to the left) may make an exposure pattern easier to learn. The

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

outcomes of our study also broadly suggest that while adults can learn either a left- or right-aligning quantity-insensitive prominence-based pattern equally well, there may be a greater tendency to infer a left- than a right-aligning quantity-sensitive stress pattern when learners have the opportunity to choose. In seeking to study learning asymmetries, our approach in this study has been to teach speakers of the same native language prominence-based patterns characterized by opposing features, and we believe that there is value in this approach. We believe that ultimately, learning asymmetries observed in a study like ours, if they can be teased apart from biases related to the native language, may be a source of information about natural stress patterning. In writing of learning asymmetries, or that one stress pattern may be more readily learned in the ALL context, we make no claims about the ease with which children might learn one pattern as opposed to another in the L1 context. What we are pointing out is that asymmetries in response behavior in psycholinguistic studies can potentially reveal cognitive biases which all humans have and which, over time, may have contributed to shaping the stress typology. This study represents a beginning in a more extended program of ALL research on stress patterns, and we expect that any firm claims of connection between cognitive biases and stress patterning in human languages must be informed by the results of similar studies conducted with speakers of diverse languages.

Appendix A: Syllable combinations used in stimuli

A.1 Phase 1: Training. Trisyllables used in vocabulary lists and review trials

Trisyllables with short vowels (Type LLL)	kapoto*	pateko*	takapa*
	kepete*	pekate	tapoke
	ketako	pokepa*	tepoka*
	kotopa	potake	tokepo

	Type HLL	*Type LHL*	*Type LLH*
Trisyllables with one long vowel	ka:tepa	kope:te*	katote:
	pe:pako*	pate:ka	ketapo:*
	te:kepa	peko:te*	pokoto:
	to:kake*	topa:ko	tapoka:*

(Starred items were used in decision pairs in review trials.)

A.2 Phase 2: Test

Trisyllables with short vowels (Type LLL)	kapete	paketa	takope
	katape	papeto	tapepo
	kepopa	pekoto	tepake
	ketapo	petopa	tokape
	kopoto	potako	
	koteka		

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

	Type HLL	*Type LHL*	*Type LLH*
Trisyllables with one long vowel	ka:tote	kapo:to	kopete:
	ke:poke	pate:ko	kotepo:
	po:keto	peka:te	pateka:
	te:topa	poke:pa	pepako:
	to:teka	toka:ke	tekake:

	Type HHL	*Type HLH*	*Type LHH*
Trisyllables with two long vowels	ka:te:po	ka:topa:	kato:te:
	ko:te:pa	ke:tako:	kepe:ka:
	pa:ke:po	ko:pate:	kete:po:
	pe:ka:ko	pa:taka:	papo:to:
	po:te:ka	pe:kote:	peto:ke:
	te:po:ka	po:teke:	poke:to:
	to:ka:ke	te:peke:	tapo:ke:
	to:ko:pa	to:kepo:	teka:pe:

Appendix B: Raw data

Table B1: Observations excluded from the statistical analysis by category.

		Exposure	Test		Totals
			Trained	Novel	
Total possible		2,352	5,880	2,352	10,584
Exclusions	Wrong button	1	2	1	4
	RTs $\leq$ 408 ms	6	30	8	44
	RTs > 10,000 ms	8	11	8	27
Total available		2,337	5,837	2,335	10,509
% Exclusions		.64	.73	.72	.71

Table B2: Recall data for review trials in the exposure phase (round 2)

(a) Group A (left-trained), by Pattern and Foil. (An uptick [ˈ] marks stress on the following syllable.)

Target	ˈLLL			ˈHLL			L ˈHL			LL ˈH		
Foil	L ˈLL	LL ˈL	Tot.	H ˈLL	HL ˈL	Tot.	ˈLHL	LH ˈL	Tot.	ˈLLH	L ˈLH	Tot.
Correct	262	271	533	74	70	144	73	78	151	58	66	124
Incorrect	37	28	65	26	29	55	27	22	49	41	33	74
Total trials	299	299	598	100	99	199	100	100	200	99	99	198
Prop. corr.	.88	.91	.89	.74	.71	.72	.73	.78	.76	.59	.67	.63

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

(b) Group B (right-trained), by Pattern and Foil

Target	LL'L			'HLL			L'HL			LL'H		
Foil	'LLL	L'LL	Tot.	H'LL	HL'L	Tot.	'LHL	LH'L	Tot.	'LLH	L'LH	Tot.
Correct	260	261	521	75	73	148	77	82	159	62	71	133
Incorrect	27	25	52	20	21	41	18	13	31	32	25	57
Total trials	287	286	573	95	94	189	95	95	190	94	96	190
Prop. corr.	.91	.91	.91	.79	.78	.78	.81	.86	.84	.66	.74	.70

Table B3: Response data for the Trained Patterns condition in the test phase.

(a) Group A (left-trained), by Pattern and Foil.

Target	'LLL			'HLL			L'HL			LL'H		
Foil	L'LL	LL'L	Tot.	H'LL	HL'L	Tot.	'LHL	LH'L	Tot.	'LLH	L'LH	Tot.
Correct	493	487	980	187	185	372	135	149	284	101	145	246
Incorrect	256	260	516	63	65	128	115	96	211	147	104	251
Total trials	749	747	1496	250	250	500	250	245	495	248	249	497
Prop. corr.	.66	.65	.66	.75	.74	.74	.54	.61	.57	.41	.58	.49

(b) Group B (right-trained), by Pattern and Foil

Target	LL'L			'HLL			L'HL			LL'H		
Foil	'LLL	L'LL	Tot.	H'LL	HL'L	Tot.	'LHL	LH'L	Tot.	'LLH	L'LH	Tot.
Correct	478	456	934	158	131	289	147	128	275	148	155	303
Incorrect	232	255	487	81	107	188	90	109	199	91	83	174
Total trials	710	711	1421	239	238	477	237	237	474	239	238	477
Prop. corr.	.67	.64	.66	.66	.55	.61	.62	.54	.58	.62	.65	.64

Table B4: Response data for the Novel Patterns condition in the test phase.

	Group A (left)			Group B (right)		
Type	HHL	HLH	LHH	HHL	HLH	LHH
Leftmost heavy	249	268	228	197	196	199
Rightmost heavy	150	131	168	182	185	182
Total trials	399	399	396	379	381	381
Prop. correct	.62	.67	.58	.52	.51	.52

Appendix C: Statistical analyses

C.1 Training/Exposure phase

Table C1: Binomial tests by Pattern for review trials in the exposure phase. (Observations = 2,337)

	CI (95%)		N	PoS	p
LLL	.88	.92	1171	.90	< .00001
HLL	.71	.79	388	.75	< .00001
LHL	.75	.83	390	.79	< .00001
LLH	.61	.71	388	.66	< .00001

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

C.2 Test phase, Trained Patterns Condition

Table C2: Binomial tests by Pattern and Foil for the Trained Patterns condition in the test phase. (Observations = 5837)

	Group A (left-trained)						Group B (right-trained)					
Pattern	Foil	CI (95%)		N	PoS	p	Foil	CI (95%)		N	PoS	p
LLL	'LLL	.62	.69	749	.66	< .00001	'LLL	.64	.71	710	.67	< .00001
	L'LL	.62	.69	747	.65	< .00001	L'LL	.60	.68	711	.64	< .00001
HLL	H'LL	.69	.80	250	.75	< .00001	H'LL	.60	.72	239	.66	< .00001
	HL'L	.68	.79	250	.74	< .00001	HL'L	.48	.61	238	.55	.1358
LHL	'LHL	.48	.60	250	.54	.2294	'LHL	.55	.68	237	.62	.0003
	LH'L	.54	.67	245	.61	.0009	LH'L	.47	.60	237	.54	.2423
LLH	'LLH	.35	.47	248	.41	.0042	'LLH	.55	.68	238	.62	.0003
	L'LH	.52	.64	249	.58	.0111	L'LH	.59	.71	238	.65	< .00001

Table C3: Pairwise comparisons for Edge*Pattern

OModel = the omnibus mixed effects model whose output appears in Table 1.

Formula: emmeans(OModel, pairwise ~ Edge * Pattern, pbkrtest.limit = 5837)

Edge	*Pattern*	*emmean*	*SE*	*df*	*CI (= 0.95)*	
					lower	*upper*
Left-trained	LLL	0.655	0.038	47.0	0.579	0.731
Right-trained	LLL	0.657	0.039	47.0	0.580	0.735
Left-trained	HLL	0.744	0.031	46.9	0.681	0.807
Right-trained	HLL	0.605	0.032	47.1	0.541	0.670
Left-trained	LHL	0.574	0.043	47.0	0.488	0.661
Right-trained	LHL	0.579	0.044	47.0	0.490	0.667
Left-trained	LLH	0.494	0.047	47.0	0.400	0.589
Right-trained	LLH	0.635	0.048	47.0	0.539	0.732

	Contrasts		Estimate	SE	df	t	p	
Within subjects	Left, LLL	Left, HLL	-0.089	0.045	46.9	-1.983	0.5038	
	Left, LLL	Left, LHL	0.081	0.059	47.0	1.377	0.8630	
	Left, LLL	Left, LLH	0.161	0.057	47.0	2.833	0.1111	
	Left, HLL	**Left, LHL**	**0.170**	**0.036**	**46.9**	**4.715**	**0.0005**	***
	Left, HLL	**Left, LLH**	**0.250**	**0.058**	**47.0**	**4.286**	**0.0021**	**
	Left, LHL	Left, LLH	0.080	0.055	47.0	1.448	0.8302	
	Right, LLL	Right, HLL	0.052	0.046	47.1	1.141	0.9439	
	Right, LLL	Right, LHL	0.079	0.060	47.0	1.313	0.8894	
	Right, LLL	Right, LLH	0.022	0.058	47.0	0.382	0.9999	
	Right, HLL	Right, LHL	0.027	0.037	47.1	0.721	0.9959	
	Right, HLL	Right, LLH	-0.030	0.059	47.0	-0.507	0.9996	
	Right, LHL	Right, LLH	-0.057	0.056	47.0	-1.006	0.9712	

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

Between subjects	Left, LLL	Right, LLL	-0.002	0.054	47.0	-0.042	1.0000	
	Left, HLL	**Right, HLL**	**0.139**	**0.045**	**47.0**	**3.098**	**0.0600**	.
	Left, LHL	Right, LHL	-0.004	0.062	47.0	-0.072	1.0000	
	Left, LLH	Right, LLH	-0.141	0.067	47.0	-2.092	0.4347	

Results are averaged over the levels of Foil. Degrees-of-freedom method: Kenward-Roger
P-value adjustment: Tukey method for comparing a family of 8 estimates

Table C4: Pairwise comparisons for Pattern * Foil

*Formula: emmeans(OModel, pairwise ~ Pattern * Foil, pbkrtest.limit = 5837)*

Pattern	Foil	emmean	SE	df	CI (= 0.95) lower	upper
LLL	Left	0.666	0.028	51.3	0.610	0.722
LLL	Right	0.647	0.031	50.5	0.585	0.709
HLL	Left	0.704	0.026	81.3	0.652	0.756
HLL	Right	0.645	0.029	73.3	0.587	0.703
LHL	Left	0.580	0.036	62.0	0.508	0.651
LHL	Right	0.573	0.034	64.9	0.506	0.641
LLH	Left	0.513	0.041	57.8	0.431	0.595
LLH	Right	0.617	0.033	65.3	0.551	0.683

For pattern	Contrast	Estimate	SE	df	t	p	
HLL	Left H'LL vs. Right HL'L	0.060	0.032	351.0	1.836	0.5956	
LHL	Left 'LHL vs. Right LH'L	0.006	0.032	355.4	0.196	1.0000	
LLH	**Left 'LLH vs. Right L'LH**	**-0.104**	**0.032**	**353.5**	**-3.214**	**0.0307**	*

Results are averaged over the levels of Edge. Degrees-of-freedom method: Kenward-Roger
P-value adjustment: Tukey method for comparing a family of 8 estimates

Table C5: Pairwise comparisons for Edge * Foil

*Formula: emmeans(OModel, pairwise ~ Edge * Foil, pbkrtest.limit = 5837)*

Edge	Foil	emmean	SE	df	CI (= 0.95) lower	upper
Left-trained	Left	0.588	0.030	48.7	0.528	0.649
Right-trained	Left	0.643	0.031	48.8	0.581	0.705
Left-trained	Right	0.646	0.026	49.3	0.593	0.699
Right-trained	Right	0.595	0.027	49.3	0.541	0.649

Contrast		Estimate	SE	df	t	p
Left-trained, Left	Right, Left	-0.055	0.043	48.7	-1.278	0.5813
Left-trained, Left	Left, Right	-0.057	0.028	55.6	-2.029	0.189
Right-trained, Left	Right, Right	0.048	0.029	55.8	1.658	0.3553
Left-trained, Right	Right, Right	0.050	0.038	49.3	1.339	0.5431

Results are averaged over the levels of Pattern; Degrees-of-freedom method: Kenward-Roger
P-value adjustment: Tukey method for comparing a family of 4 estimates

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

C.3. Test phase, Novel Patterns Condition

Table C6: Outcomes of binomial tests by Pattern (α = .01) in the Novel Patterns condition in Group A (left-trained)

	CI (95%)		N	PoS	p
LLS	.57	.67	399	.62	< .00001
LSL	.62	.72	399	.67	< .00001
SLL	.53	.62	396	.58	.003

Table C7: Mixed model Anova table for the best-fitting model for the Novel Patterns condition in the Test phase

(Type 3 tests, KR method; observations = 5,837).

*mixed(Accuracy ~ (1+Pattern | Subject) + (Edge*Pattern), data = Novel.Patterns)*

Edge	$F(1, 47) = 6.52$ *	$p = 0.01$
Pattern	$F(2, 46) = 0.95$	$p = 0.39$
Edge*Pattern	$F(2, 46) = 1.34$	$p = 0.27$

References

Archangeli, Diana & Douglas Pulleyblank. 1994. *Grounded phonology*. Cambridge, MA: The MIT Press.

Bailey, Todd. 1995. *Nonmetrical constraints on stress*. Minneapolis: University of Minnesota Dissertation.

Bartolotti, James & Viorica Marian. 2017. Bilinguals' existing languages benefit vocabulary learning in a third language. *Language Learning* 67(1). 110-140. https://doi.org/10.1111/lang.12200.

Beckman, Jill. 1997. Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology* 14(1). 1–46. http://www.jstor.org/stable/4420090.

Boersma, Paul & David Weenink. 2019. Praat: Doing phonetics by computer. Computer program (Version 5.3.13). http://www.praat.org/.

Bolton, Thaddeus. 1894. Rhythm. *American Journal of Psychology* 6. 145–238. https://doi.org/10.2307/1410948.

Bosson, James E. 1964. *Modern Mongolian, a primer and reader* (Uralic and Altaic Series, Vol. 38). Bloomington, IN: Indiana University.

Carpenter, Angela C. 2010. A naturalness bias in learning stress. *Phonology* 27(3). 345–392. https://doi.org/10.1017/S0952675710000199.

Carpenter, Angela. 2016. The role of a domain-specific language mechanism in learning natural and unnatural stress. *Open Linguistics* 2. 105–131. https://doi.org/10.1515/opli-2016-0006.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. Cambridge, MA: The MIT Press.

Clements, George N. & Engin Sezer. 1982. Vowel and consonant disharmony in Turkish. In Harry van der Hulst & Norval Smith (eds.), *The structure of phonological representations*, 213–255. Berlin: De Gruyter.

Crowhurst, Megan J. 1994. Foot extrametricality and template mapping in Cupeño. *Natural Language & Linguistic Theory* 12(2). 177–201. https://doi.org/10.1007/BF00993144.

Crowhurst, Megan J. 2019. The iambic/trochaic law: Nature or nurture? *Language and Linguistics Compass* 14(1). e12360. https://doi.org/10.1111/lnc3.12360.

Culbertson, Jennifer. 2012. Typological universals as reflections of biased learning: Evidence from artificial language learning. *Language and Linguistics Compass* 6(5). 310–329. https://doi.org/10.1002/lnc3.338.

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

Cutler, Anne & David M. Carter. 1987. The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language* 2. 133–142. https://doi.org/10.1016/0885-2308(87)90004-0.

de Lacy, Paul. 2014. Evaluating evidence for stress systems. In Harry van der Hulst (ed.), *Word stress: Theoretical and typological issues*, 149–193. Cambridge, UK: Cambridge University Press.

Diehl, Randy L. & Keith R. Kluender. 1989. On the objects of speech perception. *Ecological Psychology* 1(2). 121–144.

Domahs, Ulrike, Ingo Plag & Rebecca Carroll. 2014. Word stress assignment in German, English and Dutch: Quantity-sensitivity and extrametricality revisited. *Journal of Computational German Linguistics* 17. 59–96. https://doi.org/10.1007/s10828-014-9063-9.

Finley, Sara & William Badecker. 2009. Artificial language learning and feature-based generalization. *Journal of Memory and Language* 61(3). 423–437. https://doi.org/10.1016/j.jml.2009.05.002.

Fletcher, Janet. 2013. The prosody of speech: Timing and rhythm. In William J. Hardcastle, John Laver & Fiona E. Gibbon (eds,), *The handbook of phonetic sciences*, 523–602. Hoboken, NJ: Blackwell.

Gordon, Matthew K. 2000. Re-examining default-to-opposite stress. *Proceedings of the Berkeley Linguistics Society (BLS)* 26(1). 101–112. https://doi.org/10.3765/bls.v26i1.1153.

Gordon, Matthew K. 2002. A factorial typology of quantity-insensitive stress. *Natural Language & Linguistic Theory* 20(3). 491–552. http://www.jstor.org/stable/4048055.

Gordon, Matthew K. 2016. *Phonological typology* (Oxford Survey in Phonology and Phonetics, Vol. 1). Oxford, UK: Oxford University Press.

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

Hammond, Michael. 1984. *Constraining metrical theory: A modular theory of rhythm and destressing*. Bloomington, IN: Indiana University Linguistics Club.

Hammond, Michael. 1999. *The phonology of English: A prosodic Optimality-Theoretic approach*. Oxford, UK: Oxford University Press.

Hayes, Bruce. 1985. *A metrical theory of stress rules.* New York: Garland Press.

Hayes, Bruce. 1982. Extrametricality and English stress. *Linguistic Inquiry* 13(2). 227–276. https://www.jstor.org/stable/4178274.

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

Heinz, Jeffery. 2009. On the role of locality in learning stress patterns. *Phonology* 26(2). 303–351. https://doi.org/10.1017/S0952675709990145.

Hill, Jane & Kenneth Hill. 1968. Stress in the Cupan (Uto-Aztecan) languages. *International Journal of American Linguistics* 34(4). 233–241. https://doi.org/10.1086/465023.

Hyman, Larry. 1977. On the nature of linguistic stress. In Larry Hyman (ed.), *Studies in stress and accent* (Southern California Occasional Papers in Linguistics 4), 37–82. Los Angeles: University of Southern California, Department of Linguistics.

Hyman, Larry. 1998. Positional prominence and the 'prosodic trough' in Yaka. *Phonology* 15(1). 41–75. https://www.jstor.org/stable/4420116.

Kager, Rene W. J. 1989. *A metrical theory of stress and destressing in English and Dutch.* Utrecht: University of Utrecht dissertation.

Kager, Rene W. J. 2012. Stress in windows: Language typology and factorial typology. *Lingua* 122. 1454–1493. https://doi.org/10.1016/j.lingua.2012.06.005.

Kiparsky, Paul. 1973. "Elsewhere" in phonology. In Stephen R. Anderson & Paul Kiparsky (eds.), *A Festschrift for Morris Halle*, 93–106. New York: Holt, Rinehart and Winston.

Mester, Ralf-Armin & Junko Itô. 1989. Feature predictability and underspecification: Palatal prosody in Japanese mimetics. *Language* 68(1). 258–293. https://doi.org/10.1353/lan.1992.0040.

Myers, Scott. 1990. *Tone and the structure of words in Shona*. New York: Garland Press.

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

Lenth, Russell. 2020. Package emmeans: Estimated marginal means, aka least-squares means. R package version 1.4.4. https://CRAN.R-project.org/package=emmeans.
Liberman, Mark & Alan S. Prince. 1977. On stress and linguistic rhythm. *Linguistic inquiry* 8(2). 249–336.
Lindblom, Bjorn. 2000. Developmental origins of adult phonology: The interplay between phonetic emergents and the evolutionary adaptations of sound patterns. *Phonetica* 57(2–4). 297–314.
Moreton, Elliott. 2008. Analytic bias and phonological typology. *Phonology* 25(1). 83–127.
Newlin-Łukowicz, Heather. 2012. Polish stress: Looking for phonetic evidence of a bidirectional system. *Phonology* 29(2). 271–329. https://doi.org/10.1017/S0952675712000139.
Peperkamp, Sharon, Inga Vendelin & Emmanuel Dupoux. 2010. Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics* 38(3). 422–430. https://doi.org/10.1016/j.wocn.2010.04.001.
Poppe, Nicholas. 1970. *Mongolian language handbook.* Washington, DC: Center for Applied Linguistics.
Portin, Marja, Minna Lehtonen, Gabor Harrer, Erling Wande, Jussi Niemi & Matti Laine. 2008. L1 effects on the processing of inflected nouns in L2. *Acta Psychologica* 128(3). 452–465. https://doi.org/10.1016/j.actpsy.2007.07.003.
Prince, Alan S. 1983. Relating to the grid. *Linguistic Inquiry* 14(1). 19–100. http://www.jstor.org/stable/4178311s.
Prince, Alan S. & Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Hoboken, NJ: John Wiley & Sons.
R Core Team. 2018. R: A language and environment for statistical computing. Software package, R Foundation for Statistical Computing. https://www.R-project.org/.
RStudio Team. 2016. RStudio: Integrated development for R. Software package. http://www.rstudio.com/.
Rasier, Laurent & Philippe Hiligsmann. 2007. Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Nouveaux Cahiers de Linguistique Française* 28. 41–66.
Singmann, Henrik, Ben Bolker, Jake Westfall, Frederik Aust & Mattan Ben-Shachar. 2016. afex: Analysis of factorial experiments. Software package. https://CRAN.R-project.org/package=afex.
Spruit, Arie. 1986. *Abkhaz studies.* Leiden, NL: Leiden University dissertation.
Stanton, Juliet. 2016. Learnability shapes typology: The case of the midpoint pathology. *Language* 92(4). 753–791. https://doi.org/10.1353/lan.2016.0071.
Staubs, Robert D. 2014. *Computational modeling of learning biases in stress typology*. Amherst, MA: University of Massachussetts dissertation.
Street, John C. 1962. *Khalkha structure* (Uralic and Altaic Series, Vol. 24). Bloomington, IN: Indiana University Press.
Sulpizio, Simone & James M. McQueen. 2012. Italians use abstract knowledge about lexical stress during spoken-word recognition. *Journal of Memory and Language* 66(1). 177–193. https://doi.org/10.1016/j.jml.2011.08.001.
Tabain, Marija, Janet Fletcher & Andrew Butcher. 2014. Lexical stress in Pitjantjatjara, *Journal of Phonetics* 42. 52-66. https://doi.org/10.1016/j.wocn.2013.11.005.
Tang, Kevin & Dinah Baer-Henney. 2020. Modelling L1 and the artificial language during artificial language learning. PsyArXiv preprint. https://doi.org/10.31234/osf.io/df8ru.
Vainio, Seppo, Anneli Pajunen & Jukka Hyönä. 2014. L1 and L2 word recognition in Finnish: Examining L1 effects on L2 processing of morphological complexity and morphophonological transparency. *Studies in Second Language Acquisition* 36(1). 133–162. https://doi.org/10.1017/S0272263113000478.
van der Hulst, Harry & Rob Goedemans. 2009. StressTyp: A database for word accentual patterns in the world's languages. In Martin Everaert, Simon Musgrave & Alexis Dimitriadis (eds.), *The use of databases in cross-linguistic studies*, 235–282. Berlin: De Gruyter. https://doi.org/10.1515/9783110198744.235.
Walker, Rachel. 1997. Mongolian stress, licensing and factorial typology. Manuscript. University of California, Santa Cruz.

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press

Wilson, Colin. 2003. Experimental investigation of phonological naturalness. In Gina Garding & Mimu Tsujimura (eds.), *Proceedings of the 22nd West Coast Conference on Formal Linguistics,* 533–546. Somerville, MA: Cascadilla Press.

Megan J. Crowhurst
Department of Linguistics
University of Texas, Austin
305 E. 23rd Street STOP B5100
Austin, TX 78712 USA
mcrowhurst@austin.utexas.edu

Laura R. Faircloth
laura.faircloth1693@gmail.com

Allison L. Wetterlin
Institutt for fremmedspråk og oversetting
Agder University
Kristiansand, Norway
allison.wetterlin@uia.no

Linda R. Wheeldon
Institutt for fremmedspråk og oversetting
Agder University
Kristiansand, Norway
linda.r.wheeldon@uia.no

https://doi.org/10.3765/pda.v4art6.65 Published online by Cambridge University Press