Learnability shapes typology: the case of the midpoint pathology[*]

Juliet Stanton, MIT – *juliets@mit.edu*

December 2014

1 Introduction

A common way to assess a proposal in phonological theory is to compare its predictions to the attested typology. In constraint-based theories, this comparison is often done by exploring the factorial typology of a proposed constraint set. When assessing a factorial typology, there are at least two important questions the analyst must ask. First, is the proposed constraint set *descriptively adequate*: does it succeed in predicting all attested languages? If it is descriptively adequate, we can then ask: is the proposed constraint set *adequately restrictive*? Does it succeed in predicting *only* the attested languages? While a lack of descriptive adequacy is typically viewed as a problem – the right theory should be able to predict all attested patterns – determining whether or not a theory is adequately restrictive is a more nuanced task.

Assessing a proposal that is not adequately restrictive – one that *overgenerates* – can be complex. This is because there are multiple reasons why a proposal might overgenerate. It is always possible that some of the predicted but unattested patterns might exist, but haven't been discovered yet. Deciding whether or not a predicted but unattested pattern is accidentally unattested largely depends on the analyst's intuition as to whether or not the predicted pattern resembles attested ones. If the unattested pattern closely resembles other attested languages, it is usually treated as accidentally missing. If the unattested pattern does not resemble any known human language, it is likely to be flagged as *pathological overgeneration*, i.e. a problematic prediction of the theory.

In constraint-based theories of phonology, the usual response to pathological overgeneration is to modify Con to eliminate it. Modifying Con to eliminate pathological overgeneration restricts the typology by modifying the learner's hypothesis space: the relevant language is unattested because the learner never has a chance to acquire it. The ultimate goal of this approach is to reach a restrictive theory of Con: one that predicts all, and only, attested patterns.

This paper explores an alternative response to pathological overgeneration. I propose that some predicted but unattested patterns, pathological or otherwise, are unattested because the required ranking is difficult to learn. The basic idea is that learnability acts as a filter on factorial typology: the larger hypothesis space available to a learner is shaped and limited by considerations of learnability (see Boersma 2003, Alderete 2008, Staubs 2014a,b). This proposal has direct implications for evaluating the predictions of a proposed constraint set. If it can be shown that a certain predicted pattern is unattested because it would be difficult for a child to acquire, there is no need to

[*]My thanks to Adam Albright, Donca Steriade, and participants at MIT's Ling-Lunch, AMP 2014, and NECPhon 8 (particularly Edward Flemming and René Kager) for helpful comments and feedback. I am also grateful to Francesca Cicileo, for help with counting words.

modify Con to exclude it. In other words, to be sure that a proposed revision to Con is necessary or even justified, we must first be certain that there is no alternative.

1.1 Our focus: midpoint systems

We will focus on a specific type of unattested stress system – those exhibiting the midpoint pathology (in the sense of Kager 2012; see Eisner 1997, Hyde 2012 for related pathologies with the same name). All midpoint systems share three characteristics; these are exemplified in (1). In short words, stress falls at or close to an edge (1a). In mid-length words, stress is drawn towards the middle (1b). In long words, stress again falls at or close to an edge (1c).

(1) Sample midpoint system

a. *Stress falls at an edge in short words*
$\boldsymbol{\acute{\sigma}}\sigma$
$\boldsymbol{\acute{\sigma}}\sigma$

b. *Stress migrates towards the midpoint in mid-length words*
$\sigma\boldsymbol{\acute{\sigma}}\sigma\sigma$
$\sigma\sigma\boldsymbol{\acute{\sigma}}\sigma\sigma$

c. *Stress falls at an edge in long words*
$\boldsymbol{\acute{\sigma}}\sigma\sigma\sigma\sigma\sigma$
$\boldsymbol{\acute{\sigma}}\sigma\sigma\sigma\sigma\sigma\sigma$

Kager (2012) shows that midpoint systems can arise when two opposite-edge contextual lapse constraints dominate all others. The pattern in (1) can be characterized by the ranking *ExtLapseL >> *ExtLapseR >> AlignL (constraint definitions in (2)).

(2) Constraints necessary to generate the pattern in (1)

a. *ExtLapseL: one * if none of the first three syllables are stressed.

b. *ExtLapseR: one * if none of the final three syllables are stressed.

c. AlignL: one * for each syllable separating stress from the left edge of the word.

In short words (1a), the domains of *ExtLapseL and *ExtLapseR overlap entirely; stressing any of the syllables in the word satisfies both constraints. The initial syllable is stressed, to maximally satisfy AlignL (3). In this tableau and the following, the domains where *ExtLapseL and *ExtLapseR can be satisfied are bracketed and subscripted with L and R, respectively[1].

(3) Initial stress in trisyllabic words

$\sigma\sigma\sigma$	*ExtLapseL	*ExtLapseR	AlignL
☞ a. $[[\boldsymbol{\acute{\sigma}}\sigma\sigma]_L]_R$			
b. $[[\sigma\boldsymbol{\acute{\sigma}}\sigma]_L]_R$			*!
c. $[[\sigma\sigma\boldsymbol{\acute{\sigma}}]_L]_R$			*!*

In mid-length words (2b), the domains of *ExtLapseL and *ExtLapseR only partially overlap. In five-syllable words, for example, satisfying *ExtLapseL and *ExtLapseR requires stressing the middle syllable (4). This "overlapping domains" effect is what causes stress to drift towards the middle of the word.

[1]When discussing systems with only one stress per word, I assume that OneStress, which assigns violations for each secondary stress, is unmoderated.

(4) Midpoint stress in five-syllable words

$\sigma\sigma\sigma\sigma\sigma$	*ExtLapseL	*ExtLapseR	AlignL
☞ a. $[\sigma\sigma[\boldsymbol{\acute{\sigma}}]_L\sigma\sigma]_R$			**
b. $[\boldsymbol{\acute{\sigma}}\sigma[\sigma]_L\sigma\sigma]_R$		*!	

In long words, the domains of *ExtLapseL and *ExtLapseR don't overlap at all. Because in long words it is not possible to satisfy both *ExtLapseL and *ExtLapseR at the same time, stress falls on the initial syllable to satisfy higher-ranked *ExtLapseL (5).

(5) Initial stress in seven-syllable words

$\sigma\sigma\sigma\sigma\sigma\sigma\sigma$	*ExtLapseL	*ExtLapseR	AlignL
☞ a. $[\boldsymbol{\acute{\sigma}}\sigma\sigma]_L\sigma[\sigma\sigma\sigma]_R$		*	
b. $[\sigma\sigma\sigma]_L\sigma[\boldsymbol{\acute{\sigma}}\sigma\sigma]_R$	*!		****

Various other combinations of contextual *Lapse and *ExtLapse give rise to other kinds of midpoint systems. For example, the system in (6) can be characterized by the ranking *LapseL >> *LapseR >> AlignL. The system in (7) can be characterized by the ranking *LapseL >> *ExtLapseR >> AlignL.

(6) *LapseL >> *LapseR >> AlignL

a. $[[\boldsymbol{\acute{\sigma}}\sigma]_L]_R$
b. $[\sigma[\boldsymbol{\acute{\sigma}}]_L\sigma]_R$
c. $[\boldsymbol{\acute{\sigma}}\sigma]_L[\sigma\sigma]_R$
d. $[\boldsymbol{\acute{\sigma}}\sigma]_L\sigma[\sigma\sigma]_R$
e. $[\boldsymbol{\acute{\sigma}}\sigma]_L\sigma\sigma[\sigma\sigma]_R$
f. $[\boldsymbol{\acute{\sigma}}\sigma]_L\sigma\sigma\sigma[\sigma\sigma]_R$

(7) *LapseL >> *ExtLapseR >> AlignL

a. $[[\boldsymbol{\acute{\sigma}}\sigma]_L]_R$
b. $[[\boldsymbol{\acute{\sigma}}\sigma]_L]\sigma_R$
c. $[\sigma[\boldsymbol{\acute{\sigma}}]_L]\sigma\sigma_R$
d. $[\boldsymbol{\acute{\sigma}}\sigma]_L[\sigma\sigma\sigma]_R$
e. $[\boldsymbol{\acute{\sigma}}\sigma]_L\sigma[\sigma\sigma\sigma]_R$
f. $[\boldsymbol{\acute{\sigma}}\sigma]_L\sigma\sigma[\sigma\sigma\sigma]_R$

1.2 Explaining the gap

As far as we know, midpoint systems do not exist. Kager (2012) claims that this is because midpoint systems are not part of the learner's hypothesis space: contextual lapse constraints (e.g. *ExtLapseL), necessary to generate midpoint systems, are not part of Con. In order to exclude midpoint systems from the typology, Kager proposes to modify Con, by removing contextual lapse constraints, as well as Gen, by incorporating weakly layered feet (see also Martínez-Paricio 2013, Martínez-Paricio & Kager 2014, Kager & Martínez-Paricio 2014). Kager's solution to the midpoint pathology has implications for foot-free theories of stress (Prince 1983, Gordon 2002). In foot-free theories, contextual lapse constraints are vital for the analysis of stress windows. If midpoint systems represent a pathological prediction of contextual lapse constraints, their existence poses a problem for foot-free theories of stress.

The alternative solution to the problem posed by the midpoint systems is to claim that they are in fact part of the learner's hypothesis space, but that they are unattested because they are difficult to learn. This paper argues that there are at least three factors that contribute to the absence of midpoint systems from the typology. First, given the assumption that each fully stratified ranking of constraints is just as probable as any other, we expect midpoint systems to be fairly rare (§2). Second, I show that midpoint systems are inherently difficult for a gradual error-driven learner, due to both the absence of certain crucial data from the learner's input as well as fundamental facts about error-driven learning (§3). Third, a typological bias against systems where stress placement depends on word length (§5) suggests that midpoint systems, along with certain other types of severely underattested systems, are dispreferred.

If this alternative succeeds, there is no need to exclude midpoint systems from the hypothesis space. This, too, has certain implications for metrical theory. If we can explain the absence of midpoint systems as resulting from considerations of learnability, then there is no need to eliminate contextual lapse constraints from Con, and no need to adopt weakly layered feet. It is important to note, though, that the point of this paper is not to deny that feet are a necessary component of metrical theory. The evidence for feet, and weakly layered feet in particular, extends beyond the arguments given by Kager (2012) (for examples of recent work, see: Vaysman 2009, Martínez-Paricio 2013, Bennett 2013, Bennett & Henderson 2013, Martínez-Paricio & Kager 2014, Kager & Martínez-Paricio 2014). Rather, the point of this paper is to examine one argument for metrical constituency, and to show that the relevant phenomenon can be explained in another way. As such, the primary goal of the paper is to explore the idea that learnability could suffice to shape this aspect of the typology. The secondary goal is to show that the midpoint pathology does not present a fatal blow for foot-free theories of stress. Further work is needed to determine whether or not there are other phenomena that do.

2 Expected frequency of midpoint systems

The plausibility of the alternative just outlined depends, in part, on how frequent we expect midpoint systems to be. It is more plausible if we expect midpoint systems to be rare; it is less plausible if we expect them to make a high percentage of the total typology. But first, how can we estimate the expected frequency of a pattern? I will make here what I believe to be the null assumption: that each total ranking of constraints is equally probable. As a single surface pattern often corresponds to more than one possible ranking, we might expect that the more rankings are consistent with a single surface pattern, the more frequent that pattern should be. In fact, Bane & Riggle 2008 show that there is a statistically significant positive correlation between ranking volume (or *r-volume*) and the frequency of quantity-insensitive stress systems, assuming Gordon's (2002) constraint set.

We can ask, now: what is the expected r-volume of midpoint systems? Kager (2012) claims that midpoint systems arise when two opposite-edge contextual lapse constraints sit at the top of the hierarchy. 322,560 (or 8.89%) rankings of Kager's anti-lapse constraint set (modeled after Gordon 2002) fit this description. But Kager's precondition alone is not sufficient to characterize when midpoint systems arise. For example, stress must be aligned to the outer edge of the window in short and long words for the "overlapping domains" effect to be visible. In (1) (below as (8)), AlignL ensures that stress falls at the left (or outer) edge of the window, in short and long words.

(8) *ExtLapseL >> *ExtLapseR >> AlignL

a. $[[\boldsymbol{\acute{\sigma}}\sigma]_L]_R$

b. $[[\boldsymbol{\acute{\sigma}}\sigma\sigma]_L]_R$

c. $[\sigma[\acute{\boldsymbol{\sigma}}\sigma]_L\sigma]_R$

d. $[\sigma\sigma[\acute{\boldsymbol{\sigma}}]_L\sigma\sigma]_R$

e. $[\acute{\boldsymbol{\sigma}}\sigma\sigma]_L[\sigma\sigma\sigma]_R$

f. $[\acute{\boldsymbol{\sigma}}\sigma\sigma]_L\sigma[\sigma\sigma\sigma]_R$

In (9), by contrast, AlignR forces stress to be aligned with the right (or inner) edge of the left-edge window. The result in (9) is not a midpoint system, but rather a system with post-peninitial stress, as in Choguita Rarámuri (Caballero 2008).

(9) *ExtLapseL >> *ExtLapseR >> AlignR

a. $[\sigma\acute{\boldsymbol{\sigma}}]_L]_R$

b. $[\sigma\sigma\acute{\boldsymbol{\sigma}}]_L]_R$

c. $[\sigma\sigma\acute{\boldsymbol{\sigma}}]_L\sigma]_R$

d. $[\sigma\sigma\acute{\boldsymbol{\sigma}}]_L\sigma\sigma]_R$

e. $[\sigma\sigma\acute{\boldsymbol{\sigma}}]_L[\sigma\sigma\sigma]_R$

f. $[\sigma\sigma\acute{\boldsymbol{\sigma}}]_L\sigma[\sigma\sigma\sigma]_R$

Excluding cases like the one in (9) and others, only 166,080 (or 4.58%) rankings of Kager's constraint set actually give rise to midpoint systems. So while the number of systems that meet Kager's precondition is quite high, the number of true midpoint systems is much lower. Translating this into expected frequency of attestation, we expect that midpoint systems should make up 4.58% of all languages with one stress per word. Assuming that Gordon's (2002) survey is representative, and that 75.75% of all languages have only a single stress per word, we then expect 3.47% of all languages to be midpoint systems. As of August 2014, 510 languages were represented in StressTyp (Goedemans & van der Hulst 2009). 3.47% of these, or 18 languages, should be midpoint systems. 18 languages is not a huge number, but the difference between 18/510 (expected) and 0/510 (attested) is extremely significant (binomial test, $p < .001$). Thus the expected rarity of midpoint systems alone is not sufficient to explain why they are absent altogether.

3 Midpoint systems are difficult to learn

Let's return now to the midpoint system in (1) (below as (10)), and take a closer look. Notice that, if we ignore the long words in (10e-f), the system in (10) could just as well be a system with antepenultimate stress[2]. The reason why it is impossible to tell that (10) is a midpoint system, from (10a-d) alone, is because it is impossible to determine the relative ranking of *ExtLapseL and *ExtLapseR. If *ExtLapseL >> *ExtLapseR, a midpoint system results (10).

(10) *ExtLapseL >> *ExtLapseR >> AlignL

a. $[[\acute{\boldsymbol{\sigma}}\sigma]_L]_R$

b. $[[\acute{\boldsymbol{\sigma}}\sigma\sigma]_L]_R$

c. $[\sigma[\acute{\boldsymbol{\sigma}}\sigma]_L\sigma]_R$

d. $[\sigma\sigma[\acute{\boldsymbol{\sigma}}]_L\sigma\sigma]_R$

e. $[\acute{\boldsymbol{\sigma}}\sigma\sigma]_L[\sigma\sigma\sigma]_R$

f. $[\acute{\boldsymbol{\sigma}}\sigma\sigma]_L\sigma[\sigma\sigma\sigma]_R$

[2](10a-d) are consistent with a number of other hypotheses, beyond (10) and (11). I do not discuss other possible hypotheses here; the point is that it is not clear from (10a-d) alone that we are dealing with a midpoint system.

If however *ExtLapseR >> *ExtLapseL, a system with antepenultimate stress results (11).

(11) *ExtLapseL >> *ExtLapseR >> AlignL

a. $[[\acute{\boldsymbol{\sigma}}\sigma]_L]_R$

b. $[[\acute{\boldsymbol{\sigma}}\sigma\sigma]_L]_R$

c. $[\sigma[\acute{\boldsymbol{\sigma}}\sigma]_L\sigma]_R$

d. $[\sigma\sigma[\acute{\boldsymbol{\sigma}}]_L\sigma\sigma]_R$

e. $[\sigma\sigma\sigma]_L[\acute{\boldsymbol{\sigma}}\sigma\sigma]_R$

f. $[\sigma\sigma\sigma]_L\sigma[\acute{\boldsymbol{\sigma}}\sigma\sigma]_R$

In order for a learner to successfully acquire the midpoint system in (10), the learner must be exposed to words of six syllables or longer. Long words are cross-linguistically rare (see §3.1); perhaps it would be difficult for a learner to distinguish between (10) and (11), as the ranking between *ExtLapseL and *ExtLapseR is only observable in words of six syllables or longer.

Before continuing, it is worth noting that this idea – that the paucity of long words might adversely affect the acquisition of a midpoint system – will only work for a subset of midpoint systems. The long-word difficulty only arises when both contextual lapse constraints are varieties of *ExtLapse. Because both *ExtLapseL and *ExtLapseR impose a trisyllabic window, the conflict between them is only apparent in six-syllable words or longer (e.g. $[\sigma\sigma\sigma]_L[\sigma\sigma\sigma]_R$), where their domains cease to overlap. The long-word difficulty does not arise when one of the contextual lapse constraints is a variety of *Lapse, for the simple reason that these constraints impose a disyllabic window. Because the windows are shorter, the conflict between contextual lapse constraints is apparent in shorter words (e.g. $[\sigma\sigma]_L[\sigma\sigma]_R$, or $[\sigma\sigma\sigma]_L[\sigma\sigma]_R$).

In §3.2, we will focus on the subset of midpoint systems where the contextual lapse constraints are *ExtLapseL and *ExtLapseR; we will call these extended midpoint systems. The claim advanced here is that extended midpoint systems are difficult to learn because the data necessary to learn them are, on average, largely absent from the learner's input. In §3.3, I show more generally that all midpoint systems are difficult to learn, even when the necessary data are available to the learner. All midpoint systems introduce a credit problem, which causes the learner to jump quickly between multiple different incorrect hypotheses before convergence. I show that both of these difficulties stem from fundamental properties of gradual error-driven learning: rankings that generate midpoint systems are inherently hard for the learner to acquire.

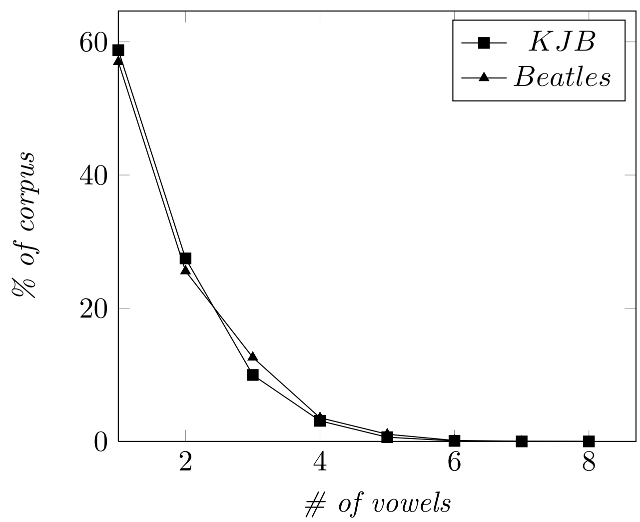
The ultimate question is, of course, whether or not a child trying to learn a midpoint system would face the same difficulties as the artificial learner investigated here. Although it is not yet known to what extent human phonological acquisition resembles the behavior of current learning algorithms, we do know that gradual error-driven learners model human phonological acquisition in certain respects, i.e. the time course of acquisition (see e.g. Boersma & Levelt 2000). It does not seem like an absurd hypothesis, then – at least, no more absurd than any other – that if a gradual error-driven learner finds midpoint systems difficult, humans will find them difficult, too.

3.1 Long words are rare

Although it has been shown for a number of lexica that short words outnumber long words (e.g. Piantadosi et al. 2011), the interest of this paper is in the distributional properties of what a learner would encounter. In order to make a meaningful crosslinguistic comparison, it is necessary to study a text corpus that is readily available for a large number of languages. To satisfy this criterion, the Bible was chosen, as it is a single text that has been translated into hundreds of languages. Counts were obtained from the book of Mark; in the rare case that Mark was unavailable, other portions of the Bible were substituted. Analysis was automated with a script that counted the number of words per number of vowels in a corpus.

The Bible has the advantage of being a text that is freely available in hundreds of languages, but using it raises several methodological questions. First, it is unclear how closely the word length distribution of Biblical language mirrors the word length distribution of everyday speech. To explore this, I compared the word length distribution of the King James Bible (790,028 words; representing Biblical language) to a selection of interviews with various former members of the Beatles[3] (90,713 words; representing everyday conversational speech). As shown below, despite the difference in both corpus size and type, their word length distributions are very similar.

Figure 1: The Beatles (Beatles) vs. the King James Bible (KJB)



At least in English, then, archaic Biblical language and more contemporary everyday speech do not appear to differ too greatly in terms of word length distribution. I have not investigated to what extend this result holds across other corpora, or in other languages.

Another methodological question surrounds the question of what counts as a vowel. In many languages, phoneme-to-orthography conversion is not one-to-one, and some graphemes can function as either a vowel or a consonant. English <y>, for example, is pronounced as the diphthong [aɪ] in the word "by", but as the glide [j] in the word "you". To avoid under-counting, English <y>, and phonemic chameleons in other languages, were always counted as vowels. More generally, if a given grapheme could function as a vowel in any context, it was counted as a vowel in all contexts; no attempt was made to account for language-specific, context-sensitive processes like glide formation. The set of vowels counted for a given language was generally determined by consulting online resources, i.e. Wikitravel's phrasebooks[4]. When this information was unavailable, the most likely set of vowels was determined by examining the distributional properties of suspect graphemes.

One final question regards the notion of wordhood. Languages often differ in which combinations of morphemes can be represented together as a single word. Even within a language, orthographic conventions are often inconsistent. Consider, for example, the pronominal clitics of French. In the imperative "Mange-les" ('eat them'), "les" is appended to the verb. In the indicative "Il les mange"

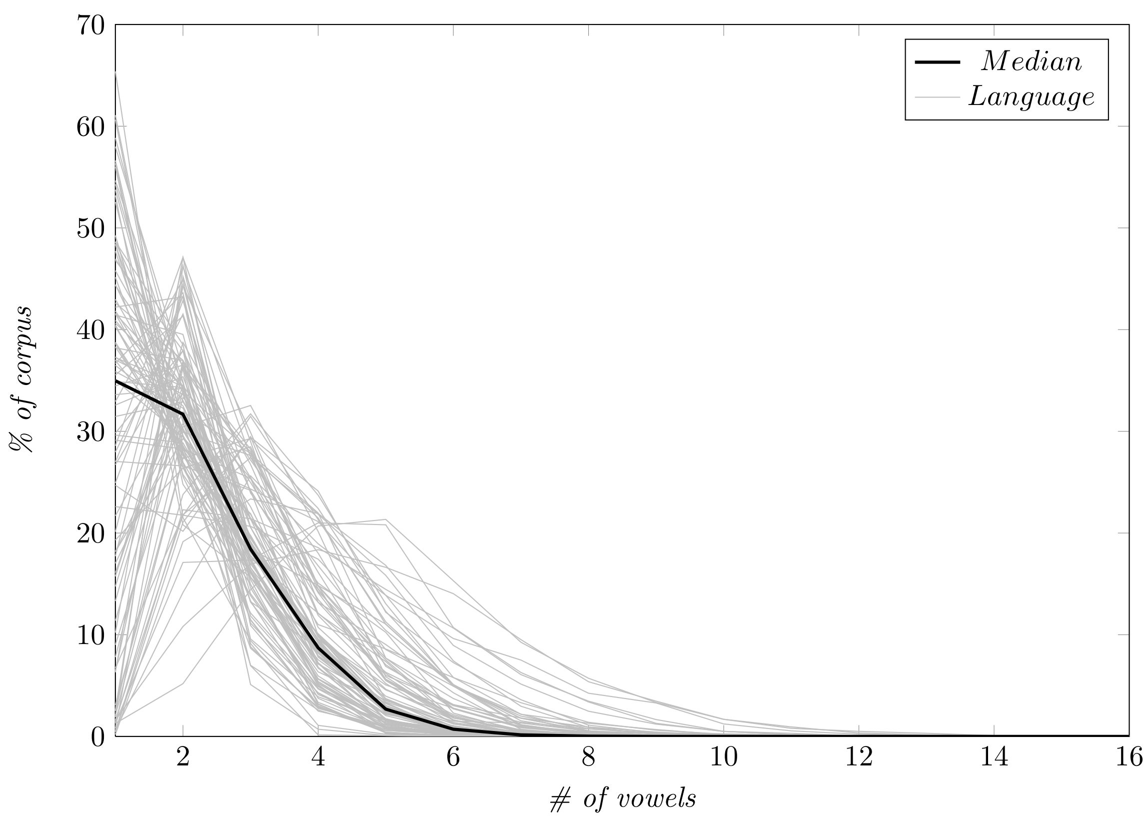
[3] All interviews dating 1970 or later, available at `http://www.beatlesinterviews.org`, were included.
[4] `http://wikitravel.org/en/List_of_phrasebooks`

('he is eating them'), however, the clitic is written as a separate word. This orthographic difference does not correspond to a difference in phonology; in neither case is the clitic prosodically independent. Although differing conventions regarding orthographic wordhood introduce a confound, I did not attempt to address it. Here, a space constitutes a word boundary.

At the time of writing, the sample consisted of data from 102 languages, selected on the basis of the availability of online resources[5]. The surveyed languages hail from 24 major language families: Afro-Asiatic (5), Algic (1), Altaic (2), Austronesian (10), Arnhem (3), Eskimo-Aleut (2), English Creole (3), French Creole (1), Indo-European (33), Iroquoian (1), Jivaroan (1), Mayan (6), Niger-Congo (8), Nilo-Saharan (1), Pama-Nyungan (10), Quechuan (1), Sepik (1), Sino-Tibetan (2), Southern Daly (1), Trans-New Guinea (2), Turkic (2), Uto-Aztecan (2), Uralic (3), and one isolate[6]. The number of words counted ranged from 6,415 (Inuktitut; Eskimo-Aleut) to 38,266 (Anindilyakwa; Pama-Nyungan), with an average of 14,402.

The investigation yielded two main results. First, the distribution of word lengths is extremely variable across languages. This is immediately visible in Figure 2, where each individual gray line represents the word length distribution of a single language. Despite this variability, however, in the large majority of languages, long words are extremely rare. The darker line in Figure 2 represents the median value for each word length; these values are also provided in (12).

Figure 2: Results of word length study



[5] I am grateful to Francesca Cicileo for her help with collecting and counting the data.

[6] See `http://web.mit.edu/juliets/www/counts.xlsx` for the list of languages, counts, and sources. All other project files are available upon request.

(12) Median values for the word count data

# of vowels	1	2	3	4	5	6	7
Median	35%	32%	18%	9%	3%	1%	0%

The important point to take away from Figure 2 is this: assuming that the median values represent approximately what the "average" learner would be exposed to, words containing six or more vowels – which I take here to be roughly equivalent to six or more syllables – make up, on average, just 1% of the input. What this means, then, is that for a learner attempting to learn the midpoint pattern in (10), evidence that *ExtLapseL >> *ExtLapseR comes from only 1% of forms present in the input. In child-directed speech, we might expect for long words to be even less frequent. Patterns where crucial rankings are only available in words of six syllables or longer might then be difficult for a child to learn.

3.2 Modeling extended midpoint systems

Given our knowledge about the distribution of word lengths, we can now ask the following question: if we give a learner realistic information about word length distribution, does it have a difficult time learning extended midpoint systems? I addressed this question by creating a learner that is informed by the results of the word length study, and comparing its performance on extended midpoint systems to its performance on superficially similar, but attested, systems. We will focus here on a small test set of quantity-insensitive languages. The system in (13) has invariant initial stress ("Initial"), the system in (14) has antepenultimate stress ("AP"), and the system in (15) is the midpoint system we have been discussing since §1 ("Midpoint").

(13) Initial (Necessary ranking: AlignL >> all)

a. $\boldsymbol{\acute{\sigma}}\sigma$

b. $\boldsymbol{\acute{\sigma}}\sigma\sigma$

c. $\boldsymbol{\acute{\sigma}}\sigma\sigma\sigma$

d. $\boldsymbol{\acute{\sigma}}\sigma\sigma\sigma\sigma$

e. $\boldsymbol{\acute{\sigma}}\sigma\sigma\sigma\sigma\sigma$

f. $\boldsymbol{\acute{\sigma}}\sigma\sigma\sigma\sigma\sigma\sigma$

(14) AP (Necessary ranking: *ExtLapseR >> AlignL)

a. $[\boldsymbol{\acute{\sigma}}\sigma]_R$

b. $[\boldsymbol{\acute{\sigma}}\sigma\sigma]_R$

c. $\sigma[\boldsymbol{\acute{\sigma}}\sigma\sigma]_R$

d. $\sigma\sigma[\boldsymbol{\acute{\sigma}}\sigma\sigma]_R$

e. $\sigma\sigma\sigma[\boldsymbol{\acute{\sigma}}\sigma\sigma]_R$

f. $\sigma\sigma\sigma\sigma[\boldsymbol{\acute{\sigma}}\sigma\sigma]_R$

(15) Midpoint (Necesssary ranking: *ExtLapseL >> *ExtLapseR >> AlignL)

a. $[[\boldsymbol{\acute{\sigma}}\sigma]_L]_R$

b. $[[\boldsymbol{\acute{\sigma}}\sigma\sigma]_L]_R$

c. $[\sigma[\boldsymbol{\acute{\sigma}}\sigma]_L\sigma]_R$

d. $[\sigma\sigma[\boldsymbol{\acute{\sigma}}]_L\sigma\sigma]_R$

e. $[\boldsymbol{\acute{\sigma}}\sigma\sigma]_L[\sigma\sigma\sigma]_R$

f. $[\boldsymbol{\acute{\sigma}}\sigma\sigma]_L\sigma[\sigma\sigma\sigma]_R$

We expect the learner's performance on the midpoint system to depend on the availability of long words. If the learner is exposed to many long words (the uncommon situation), it should find Midpoint easy to learn. If the learner is exposed to few or no long words (the more common situation), the learner should find Initial and AP easy to learn, but Midpoint more difficult.

It is worth taking a step back, now, and reminding ourselves why this is the expectation. I assume here that, in the initial state, all markedness constraints are ranked equally with respect to each other. For our purposes, learning the phonology of a language is equivalent to learning all crucial rankings associated with that language. If evidence for any given crucial ranking is rare, then the relevant language will be hard to learn. For AP (15), for example, words of four syllables or longer are required for the learner to infer that *ExtLapseR >> AlignL. This means that a learner exposed to words of only four syllables or shorter would be able to successfully learn AP. For Midpoint, however, words of six or more syllables are required for the learner to infer the relative ranking between *ExtLapseL and *ExtLapseR. If a learner is deprived of long words, we would then expect extended midpoint systems to be difficult for the learner to acquire. This point is developed further in §3.3.

3.2.1 Preliminaries: input files, and properties of the learner

Before we discuss the results, some information about the learner is necessary. The constraint set used in the simulations is Kager's (2012: 1479) anti-lapse constraint set, with WSP ("weight-to-stress principle") serving as the designated property constraint. Constraints are listed and defined in (16); given these, the patterns in (13-15) are OT-consistent, i.e. in principle learnable.

(16) Kager's (2012: 1479) anti-lapse constraint set (based on Gordon 2002)

a. *General anti-lapse constraints:*
   - i. *Lapse: a * for each sequence of two stressless syllables.
   - ii. *ExtLapse: a * for each sequence of three stressless syllables.

b. *Contextual anti-lapse constraints:*
   - i. *LapseL: a * if neither of the initial two syllables is stressed.
   - ii. *LapseR: a * if neither of the final two syllables is stressed.
   - iii. *ExtLapseL: a * if none of the initial three syllables is stressed.
   - iv. *ExtLapseR: a * if none of the final three syllables is stressed.

c. *Alignment constraints:*
   - i. AlignL: one * for each syllable separating stress from the left edge.
   - ii. AlignR: one * for each syllable separating stress from the right edge.

d. NonFinality: a * if the final syllable is stressed.

e. WSP: a * for each stressless heavy syllable.

The input files contained words of one through seven syllables. For each word length, inputs containing each possible combination of heavy and light syllables were considered; I assume here that each combination is just as frequent as any other[7]. For practical reasons, outputs considered

[7] This is a simplification: the real picture is much more complicated. To get an idea of what the relative distribution of heavy (H) and light (L) syllables looks like across languages, I investigated the distributions in Mark for five languages: Finnish, Hawaiian, Indonesian, Swahili, and Wolof. For all languages but Finnish, Ls are preferred, and words containing two Hs in a row are dispreferred. Incorporating the typologically skewed distribution of Hs and Ls into the learner does not greatly affect the results for quantity-insensitive systems, because syllables are stressed depending on their position, not their weight.

were limited to candidates with only one stress. Examples of the inputs and outputs considered for trisyllabic words are below, in (17). Heavy and light syllables are denoted by $\bar{\sigma}$ and σ, respectively.

(17) Sample input and outputs for trisyllabic words

Input:	$\bar\sigma\bar\sigma\bar\sigma$	$\bar\sigma\bar\sigma\sigma$	$\bar\sigma\sigma\bar\sigma$	$\bar\sigma\sigma\sigma$	$\sigma\bar\sigma\bar\sigma$	$\sigma\bar\sigma\sigma$	$\sigma\sigma\bar\sigma$	$\sigma\sigma\sigma$
Outputs:	$\boldsymbol{\acute{\bar\sigma}}\bar\sigma\bar\sigma$	$\boldsymbol{\acute{\bar\sigma}}\bar\sigma\sigma$	$\boldsymbol{\acute{\bar\sigma}}\sigma\bar\sigma$	$\boldsymbol{\acute{\bar\sigma}}\sigma\sigma$	$\boldsymbol{\acute\sigma}\bar\sigma\bar\sigma$	$\boldsymbol{\acute\sigma}\bar\sigma\sigma$	$\boldsymbol{\acute\sigma}\sigma\bar\sigma$	$\boldsymbol{\acute\sigma}\sigma\sigma$
	$\bar\sigma\boldsymbol{\acute{\bar\sigma}}\bar\sigma$	$\bar\sigma\boldsymbol{\acute{\bar\sigma}}\sigma$	$\bar\sigma\boldsymbol{\acute\sigma}\bar\sigma$	$\bar\sigma\boldsymbol{\acute\sigma}\sigma$	$\sigma\boldsymbol{\acute{\bar\sigma}}\bar\sigma$	$\sigma\boldsymbol{\acute{\bar\sigma}}\sigma$	$\sigma\boldsymbol{\acute\sigma}\bar\sigma$	$\sigma\boldsymbol{\acute\sigma}\sigma$
	$\bar\sigma\bar\sigma\boldsymbol{\acute{\bar\sigma}}$	$\bar\sigma\bar\sigma\boldsymbol{\acute\sigma}$	$\bar\sigma\sigma\boldsymbol{\acute{\bar\sigma}}$	$\bar\sigma\sigma\boldsymbol{\acute\sigma}$	$\sigma\bar\sigma\boldsymbol{\acute{\bar\sigma}}$	$\sigma\bar\sigma\boldsymbol{\acute\sigma}$	$\sigma\sigma\boldsymbol{\acute{\bar\sigma}}$	$\sigma\sigma\boldsymbol{\acute\sigma}$

I selected five word length distributions to feed to the model. These are provided in (18). Here, Portuguese represents our "average" language, as its distribution is closest to the median. Inuktitut represents the upper bound, as it has more long words than any other language in the study; Haitian Creole represents the lower bound, as it has very few. English and Luganda represent intermediate points along the continuum. As we move from Inuktitut to Haitian Creole, we expect that Midpoint should become harder to learn, with no comparable increase in difficulty for either Initial or AP.

(18) Word length distributions used in modeling

Distribution	**1V**	**2V**	**3V**	**4V**	**5V**	**6V**	**7V**
Inuktitut	1.5%	5.9%	16.5%	23.6%	24.4%	17.5%	10.6%
Luganda	22.7%	21.9%	20.8%	17.5%	10.5%	5.0%	1.6%
Portuguese	32.6%	35.4%	18.2%	10.0%	3.0%	0.7%	0.1%
English	56.6%	28.0%	11.5%	3.0%	0.6%	0.3%	<0.1%
H. Creole	58.0%	36.1%	5.1%	0.7%	<0.1%	0.0%	0.0%

The learner used in these simulations is Magri's (2012) convergent Gradual Learning Algorithm (GLA; see also Boersma & Hayes 2001), a gradual learner that performs slightly more demotion than it does promotion. For each input file ("language"), I ran the simulation 10 times, for a maximum of 10,000 trials per run.

3.2.2 Results: extended midpoint systems are hard to learn

The results are in (19). Each number represents the average number of trials (number of exposures to randomly selected input-output pairs) that the learner took to converge on the target grammar.

(19) Modeling results

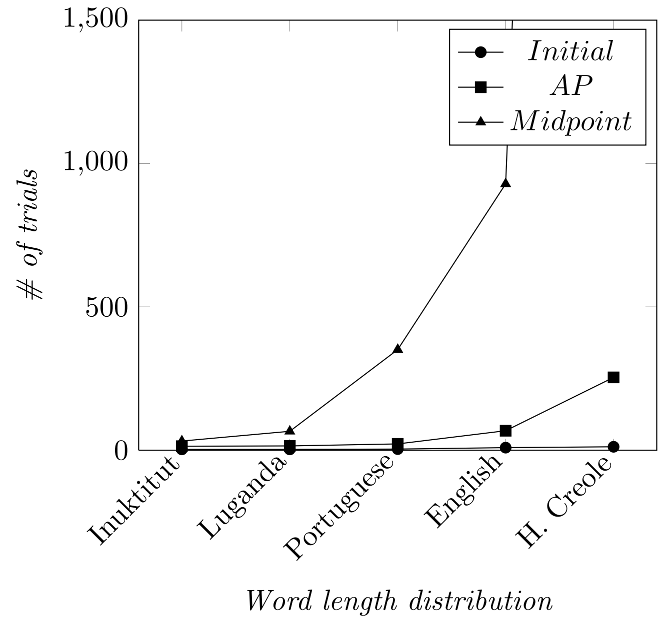
Distribution	**Initial**	**AP**	**Midpoint**
Inuktitut	3	14	32
Luganda	3	15	66
Portuguese	4	22	351
English	9	68	929
H. Creole	12	254	10,000+

Across word length distributions, Initial and AP were learned very easily. For Initial, word length distribution appears to have little effect on the learning rate. For AP, word length distribution has some effect: compare the Haitian Creole learner's average of 254 trials to the Portuguese learner's 22. Note that for AP, words of four syllables and longer are required to establish that *ExtLapseR >> AlignL; these words make up under 1% of the H. Creole learner's input. It is therefore not surprising that the H. Creole learner takes a little bit longer to learn AP, as the data demonstrating a certain crucial ranking are rare.

For Midpoint, word length distribution has a marked effect on the learning rate. As hypothesized, for the Inuktitut and Luganda learners (who were exposed to many long words), learning Midpoint does not take very long (though Midpoint is still harder for the learner than the other systems see §3.4). As the relative frequency of long words decreases, however, the number of trials needed to learn Midpoint rises sharply. For the Portuguese learner, there is a marked asymmetry between the number of trials needed to learn the attested systems (4 for Initial, 22 for AP) and the number of trials required to learn Midpoint (351). For the English learner, where the relative frequency of long words is lower, the asymmetry becomes even more pronounced (9 for Initial, 68 for AP, 929 for Midpoint). Finally, the Haitian Creole learner, who was not exposed to any words of six syllables or more, never converges on a ranking that generates Midpoint. Instead, it always converges on a ranking yielding AP.

These results are presented graphically in Figure 3. Note that the y-axis is cut off at 1,500 trials; the number of trials required to learn Midpoint for the H. Creole learner is undefined.

Figure 3: Modeling results



3.3 The effect of rare forms

The previous discussion establishes that Midpoint is difficult for the learner when long words are rare. It does not explain why this is the case. This subsection focuses on explaining why a learner exposed to only short words fails to learn a ranking that generates Midpoint. Given that the data presented to the learner are consistent with a ranking that generates Midpoint, why is this not the preferred analysis? Put differently: in the absence of overt evidence that the learner is learning a midpoint system, why is it biased against this hypothesis?

To answer this question, we'll focus on the Haitian Creole learner, as it was not exposed to any words longer than five syllables. When the Haitian Creole learner attempts to learn Midpoint, it sees only (20a-d), as words of six or more syllables are absent from its input. The forms in (20a-d) are identical to the forms that the Haitian Creole learner sees when learning AP; the two stress systems are ambiguous when a learner is not exposed to any long words.

(20) Data presented to the Haitian Creole learner for Midpoint and AP

a. $\acute{\boldsymbol{\sigma}}\sigma$

b. $\acute{\boldsymbol{\sigma}}\sigma\sigma$

c. $\sigma\acute{\boldsymbol{\sigma}}\sigma\sigma$

d. $\sigma\sigma\acute{\boldsymbol{\sigma}}\sigma\sigma$

e. $\cancel{\sigma\sigma\sigma\sigma\sigma\sigma}$

f. $\cancel{\sigma\sigma\sigma\sigma\sigma\sigma\sigma}$

Given the data in (20a-d), both AP and Midpoint are possible hypotheses. But it is clear from the results that, of these two, AP is the preferred hypothesis. Every single time the Haitian Creole learner is exposed to (20a-d), it converges on a grammar generating AP stress – regardless of whether or not that was the target grammar. To see where the bias for AP stress comes from, we will focus on the learner's error pattern in response to one run (Midpoint, Run 3). The record of the learner's errors, and its responses to them, is in (21). For convenience, all and only the updates and constraints that play a crucial role in the learner's progress are included below.

(21) Error log for Haitian Creole, Midpoint Run 3

<table>
<tr><th>Trial</th><th>Form</th><th>Guess</th><th>Correct</th></tr>
<tr><td rowspan="4">45</td><td>$\bar{\sigma}\bar{\sigma}\bar{\sigma}$</td><td>$\bar{\sigma}\acute{\bar{\boldsymbol{\sigma}}}\bar{\sigma}$</td><td>$\acute{\bar{\boldsymbol{\sigma}}}\bar{\sigma}\bar{\sigma}$</td></tr>
<tr><td colspan="3">Updates: promote AlignL</td></tr>
<tr><td colspan="3">Resulting ranking: AlignL >> *ExtLapseR, WSP</td></tr>
<tr><td colspan="3">System: Initial stress</td></tr>
<tr><td rowspan="4">71</td><td>$\bar{\sigma}\bar{\sigma}\sigma\bar{\sigma}$</td><td>$\acute{\bar{\boldsymbol{\sigma}}}\bar{\sigma}\sigma\bar{\sigma}$</td><td>$\bar{\sigma}\acute{\bar{\boldsymbol{\sigma}}}\sigma\bar{\sigma}$</td></tr>
<tr><td colspan="3">Updates: promote *ExtLapseR, demote AlignL</td></tr>
<tr><td colspan="3">Resulting ranking: AlignL >> *ExtLapseR >> WSP</td></tr>
<tr><td colspan="3">System: Initial stress</td></tr>
<tr><td rowspan="4">131</td><td>$\bar{\sigma}\bar{\sigma}\sigma\sigma$</td><td>$\acute{\bar{\boldsymbol{\sigma}}}\bar{\sigma}\sigma\sigma$</td><td>$\bar{\sigma}\acute{\bar{\boldsymbol{\sigma}}}\sigma\sigma$</td></tr>
<tr><td colspan="3">Updates: promote *ExtLapseR, demote AlignL</td></tr>
<tr><td colspan="3">Resulting ranking: *ExtLapseR >> WSP >> AlignL</td></tr>
<tr><td colspan="3">System: Right-edge trisyllabic window; quantity-sensitive</td></tr>
<tr><td rowspan="4">131</td><td>$\sigma\bar{\sigma}$</td><td>$\sigma\acute{\bar{\boldsymbol{\sigma}}}$</td><td>$\acute{\boldsymbol{\sigma}}\bar{\sigma}$</td></tr>
<tr><td colspan="3">Updates: promote AlignL, demote WSP</td></tr>
<tr><td colspan="3">Resulting ranking: *ExtLapseR >> AlignL >> WSP</td></tr>
<tr><td colspan="3">System: Antepenultimate stress</td></tr>
</table>

At trial 45, the learner encounters $\bar{\sigma}\bar{\sigma}\bar{\sigma}$, and instead of guessing the correct $\acute{\bar{\boldsymbol{\sigma}}}\bar{\sigma}\bar{\sigma}$, guesses $\bar{\sigma}\acute{\bar{\boldsymbol{\sigma}}}\bar{\sigma}$. This incorrect guess occurs because, in the initial state, all markedness constraints are ranked equally; the grammar that selects $\bar{\sigma}\acute{\bar{\boldsymbol{\sigma}}}\bar{\sigma}$ is randomly selected. In response to this error, the learner promotes AlignL (the winner-preferrer). This results in a grammar that favors initial stress. So when the learner encounters $\bar{\sigma}\bar{\sigma}\sigma\bar{\sigma}$ at trial 71, it guesses $\acute{\bar{\boldsymbol{\sigma}}}\bar{\sigma}\sigma\bar{\sigma}$, with initial stress. The correct form, however, is $\bar{\sigma}\acute{\bar{\boldsymbol{\sigma}}}\sigma\bar{\sigma}$, so the learner must adjust its weights. Because the winner (but not the loser) satisfies *ExtLapseR, *ExtLapseR is promoted; because the loser (but not the winner) fully satisfies AlignL, AlignL is demoted. But learning is gradual, and this one error is not enough to reverse the ranking between these two constraints, so the learner's grammar still generates initial stress. Thus when the learner encounters another four-syllable word, $\bar{\sigma}\bar{\sigma}\sigma\sigma$, at trial 131, it incorrectly guesses $\acute{\bar{\boldsymbol{\sigma}}}\bar{\sigma}\sigma\sigma$, instead of the correct $\bar{\sigma}\acute{\bar{\boldsymbol{\sigma}}}\sigma\sigma$. The learner again promotes *ExtLapseR and demotes AlignL. This update has the desired result of establishing the ranking *ExtLapseR >> AlignL. It also leads to an undesired result: AlignL has been demoted too much, and is now dominated by WSP, as well as all other constraints that have not moved from their initial position. The result, then, is a learner with a grammar that results in a right-edge trisyllabic window. At trial 135, however, the learner sees that satisfying AlignL is more important than satisfying WSP, and promotes the former while demoting the latter. This update allows the learner to converge on a grammar that generates antepenultimate stress, where it remains for the rest of the run. This is, with minor irrelevant differences, what happens in all runs when the learner is presented with the data in (20a-d). The learner always converges on a grammar that generates AP.

We can now return to the question of this section: if the midpoint system is a possible hypothesis given the data in (20a-d), why isn't it the chosen one? The answer is that the rankings necessary to generate a midpoint system are not directly motivated by the data. To learn the midpoint system, it's necessary for the learner to infer that *ExtLapseL >> *ExtLapseR. But as we saw in (21), while the learner makes errors that cause it to promote *ExtLapseR, it makes no errors causing it to promote *ExtLapseL. This is because the errors that would cause the learner to promote *ExtLapseL are incompatible with other aspects of the system it is learning. For example, if the learner were to encounter a four-syllable word $\sigma\sigma\sigma\sigma$ and incorrectly guess $\sigma\sigma\sigma\acute{\boldsymbol{\sigma}}$, *ExtLapseL would be promoted, as the winning candidate $\sigma\acute{\boldsymbol{\sigma}}\sigma\sigma$ satisfies *ExtLapseL, but the losing candidate $\sigma\sigma\sigma\acute{\boldsymbol{\sigma}}$ does not. But the learner never makes this error or others like it, because it learns very early on in the learning procedure that AlignL >> AlignR. The learner has no reason to guess that words should have final stress. What's important to take away from this is that AP is the preferred hypothesis, given the data in (20a-d), because it is the only hypothesis directly motivated by the data. Midpoint is not the learner's preferred hypothesis because it is not directly motivated by the data.

I return now to my assumptions about the initial state. I assume that, as all constraints involved in the simulations are markedness constraints, all are ranked equally in the initial state. Given this, another way of viewing the bias for AP stress is a bias for the *simplest possible hypothesis*, meaning here the hypothesis that is closest to the initial state. If we compare the grammars for AP stress (22) and Midpoint (23), we see that the grammar for Midpoint involves more strata.

(22) Grammar for AP stress

$$\text{*ExtLapseR} \\ \downarrow \\ \text{AlignL}$$

(23) Grammar for midpoint

$$\text{*ExtLapseL} \\ \downarrow \\ \text{*ExtLapseR} \\ \downarrow \\ \text{AlignL}$$

If all markedness constraints are unranked in the initial state, then (23) is a more significant departure. From the learner's perspective, there is no reason to assume that (23) is the correct analysis if it receives no evidence that *ExtLapseL >> *ExtLapseR must be established.

But if we make different assumptions about the initial state, does this result still hold? Will the learner still be biased against midpoint systems? Within a constraint class (e.g. M or F), proposals about biases in the initial state have typically appealed to the difference between specific and general constraints. For example, Hayes (2004) proposes that if both a specific and a general F constraint can be used to rule out a single losing candidate, the specific constraint should be selected. For example, if a learner sees that aspiration quality is preserved in prevocalic position, it should attribute the faithfulness effect to a context-sensitive version of IDENT(ASP)(i.e. IDENT(ASP)/_V). Favoring specific F constraints allows the learner to maintain a more restrictive grammar; this helps avoid overgeneration (Hayes 2004: 22).

With this logic, if learners should favor restrictive grammars, we might expect that general M constraints should be favored over specific ones (see Albright & Do 2013). But preferring general over specific M does not change the results discussed above, as neither *EXTLAPSEL nor *EXTLAPSER is more specific than the other. Even if we assume the opposite – that specific M should be favored over general M (see Do 2013: 123), there is no motivation to favor one of *EXTLAPSEL or *EXTLAPSER over the other. In fact it is difficult to envision a reason why we would ever want to favor one context-sensitive variety of *EXTLAPSE over the other, as the two constraints are completely symmetrical[8].

In short, the reason why the learner is biased to learn AP given the data in (20a-d) is because the ranking necessary to generate it is the simplest possible hypothesis, i.e. the hypothesis closest to the initial state that is compatible with all of the observed data. Even if we make different assumptions about the initial state, the learner will still be biased against Midpoint – and extended midpoint systems more generally – when it is presented with data from only short and mid-length words. Recall that there are almost certainly learners for whom long words are largely absent from the input, either because the language lacks them or because they are not present in child-directed speech. This bias against midpoint systems, given a lack of overt evidence, can then help us understand why extended midpoint systems are absent from the typology.

3.4 A credit problem

In the previous subsection, we saw that the learner is biased against acquiring Midpoint when it is deprived of long words. But there are also almost certainly learners for whom long words are present in the input, as there are languages (e.g. Inuktitut) where long words are very frequent. I show here that even when the learner has access to long words, Midpoint is still difficult to learn: the learner must encounter multiple long words before it converges on the correct grammar. This is because there is a credit problem: the updates performed in response to words of different lengths aren't consistent with one another, and often cancel each other out.

We'll focus now on the Portuguese learner, which encounters some – but not too many – long words. As a reminder, the word length distribution fed to the learner is in (24).

(24) Word length distribution for the Portuguese learner

Distribution	**1V**	**2V**	**3V**	**4V**	**5V**	**6V**	**7V**
Portuguese	56.6%	28.0%	11.5%	3.0%	0.6%	0.3%	<0.1%

In all runs, the learner converged on a grammar that generates Midpoint, but it took 351 trials on average to get there (compared to 4 for Initial, 22 for AP). We'll focus on one run (Run 4) to figure out exactly what takes it so long to converge on the target grammar (25). Up until trial

[8]It is true that, typologically speaking, more languages exhibit a right-edge window (meaning *LAPSER or *EXTLAPSER is ranked high) than a left-edge window (meaning that *LAPSEL or *EXTLAPSEL is ranked high); see Kager (2012) for details. There might then be a plausible reason to prefer right-edge contextual lapse constraints over left-edge ones. But whatever the source of this preference, favoring right-edge over left-edge constraints would just make the midpoint system under discussion even more difficult to learn.

65, the learner only sees short words, and has reached a grammar that generates antepenultimate stress through more or less the same route detailed in (21). Our interest now is on what happens when the learner is exposed to long words; as the error log is lengthy, we will discuss here only those errors occurring between trials 65 and 134.

(25) Error log for Portuguese, Midpoint Run 4

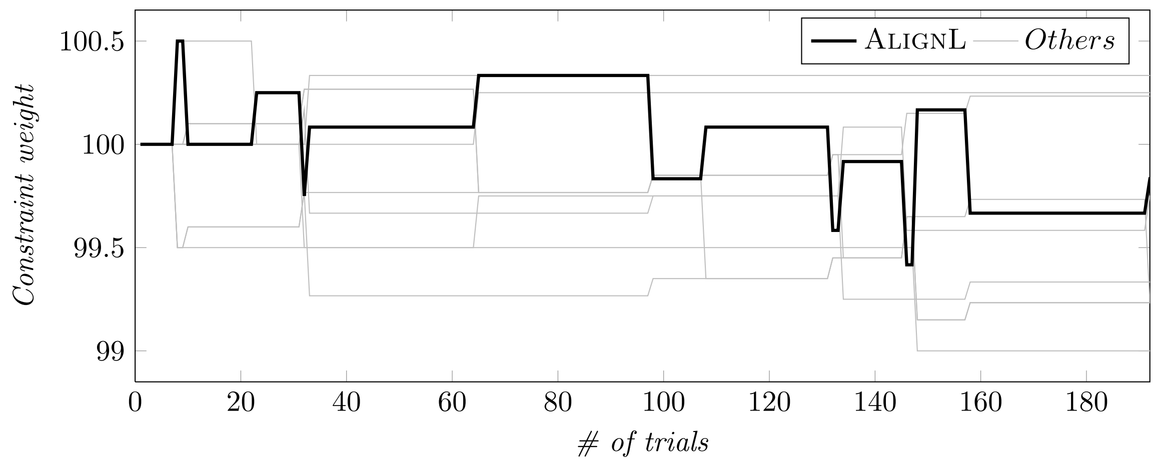
<table>
<tr><th>Trial</th><th>Form</th><th>Guess</th><th>Correct</th></tr>
<tr><td rowspan="4">65</td><td>$\bar{\sigma}\bar{\sigma}\bar{\sigma}\bar{\sigma}\sigma\sigma$</td><td>$\bar{\sigma}\bar{\sigma}\bar{\sigma}\acute{\bar{\boldsymbol{\sigma}}}\sigma\sigma$</td><td>$\acute{\bar{\boldsymbol{\sigma}}}\bar{\sigma}\bar{\sigma}\acute{\bar{\boldsymbol{\sigma}}}\sigma\sigma$</td></tr>
<tr><td colspan="3">Updates: promote *ExtLapseL, *LapseL, AlignR; demote *ExtLapseR</td></tr>
<tr><td colspan="3">Resulting ranking: AlignL >> all</td></tr>
<tr><td colspan="3">System: Initial stress</td></tr>
<tr><td rowspan="4">98</td><td>$\sigma\bar{\sigma}\sigma\bar{\sigma}$</td><td>$\acute{\boldsymbol{\sigma}}\bar{\sigma}\sigma\bar{\sigma}$</td><td>$\sigma\acute{\bar{\boldsymbol{\sigma}}}\sigma\bar{\sigma}$</td></tr>
<tr><td colspan="3">Updates: promote *Lapse; demote *AlignL</td></tr>
<tr><td colspan="3">Resulting ranking: *Lapse >> AlignL</td></tr>
<tr><td colspan="3">System: Lapse-avoiding</td></tr>
<tr><td rowspan="4">108</td><td>$\sigma\bar{\sigma}\sigma$</td><td>$\sigma\acute{\bar{\boldsymbol{\sigma}}}\sigma$</td><td>$\acute{\boldsymbol{\sigma}}\bar{\sigma}\sigma$</td></tr>
<tr><td colspan="3">Updates: promote AlignL; demote *Lapse</td></tr>
<tr><td colspan="3">Resulting ranking: AlignL >> *Lapse</td></tr>
<tr><td colspan="3">System: Initial stress</td></tr>
<tr><td rowspan="4">132</td><td>$\sigma\sigma\bar{\sigma}\sigma$</td><td>$\acute{\boldsymbol{\sigma}}\sigma\bar{\sigma}\sigma$</td><td>$\sigma\acute{\boldsymbol{\sigma}}\bar{\sigma}\sigma$</td></tr>
<tr><td colspan="3">Updates: promote *ExtLapse; demote AlignL</td></tr>
<tr><td colspan="3">Resulting ranking: *ExtLapseL >> *ExtLapse >> AlignL</td></tr>
<tr><td colspan="3">Left-edge window; extended lapse-avoiding</td></tr>
<tr><td rowspan="4">134</td><td>$\sigma\sigma\bar{\sigma}\sigma\bar{\sigma}\sigma$</td><td>$\sigma\sigma\acute{\bar{\boldsymbol{\sigma}}}\sigma\bar{\sigma}\sigma$</td><td>$\acute{\boldsymbol{\sigma}}\sigma\bar{\sigma}\sigma\bar{\sigma}\sigma$</td></tr>
<tr><td colspan="3">Updates: promote AlignL; demote *ExtLapse</td></tr>
<tr><td colspan="3">Resulting ranking: AlignL >> *ExtLapse</td></tr>
<tr><td colspan="3">Initial stress</td></tr>
<tr><td rowspan="4">146</td><td>$\sigma\sigma\sigma\bar{\sigma}\sigma$</td><td>$\acute{\boldsymbol{\sigma}}\sigma\sigma\bar{\sigma}\sigma$</td><td>$\sigma\sigma\acute{\boldsymbol{\sigma}}\bar{\sigma}\sigma$</td></tr>
<tr><td colspan="3">Updates: promote AlignR; demote AlignL</td></tr>
<tr><td colspan="3">Resulting ranking: *ExtLapseL >> AlignR >> AlignL</td></tr>
<tr><td colspan="3">Post-peninitial stress</td></tr>
</table>

At trial 65, the learner encounters six-syllable $\bar{\sigma}\bar{\sigma}\bar{\sigma}\bar{\sigma}\sigma\sigma$. As the learner has reached a ranking that generates AP, it incorrectly guesses that the output should be $\bar{\sigma}\bar{\sigma}\bar{\sigma}\acute{\bar{\boldsymbol{\sigma}}}\sigma\sigma$. Because the winner has initial stress, the learner must perform an update. But here, there is a credit problem: in this situation, *ExtLapseL, *LapseL, and AlignL are all winner-preferring. The learner, when faced with a credit problem, does not attempt to determine which of these three constraints is responsible for the winner; instead, it apportions credit evenly. By promoting AlignL and demoting *ExtLapseR (a loser-preferrer), the learner reverses the crucial ranking *ExtLapseR >> AlignL established in response to short forms, and reaches a ranking that generates initial stress. So at trial 98, when the learner encounters four-syllable $\sigma\bar{\sigma}\sigma\bar{\sigma}$, it incorrectly guesses that the form should have initial stress (*$\acute{\boldsymbol{\sigma}}\bar{\sigma}\sigma\bar{\sigma}$). In effect, the update performed in response to a long word has robbed the learner of its ability to correctly generate short words.

The update performed in response to the error at trial 98 causes the learner to reach a ranking where lapse avoidance is prized above initial stress (*Lapse >> AlignL); the update performed in response to the error at trial 108 allows the learner to reverse this ranking and return to a grammar that generates initial stress. At trial 132, the learner encounters another four-syllable form, and converges on a grammar where avoiding extended lapses is prized above initial stress, subject to the confines of a left-edge trisyllabic window (*ExtLapseL >> *ExtLapse >> AlignL). When the learner then encounters the six-syllable $\sigma\sigma\bar{\sigma}\sigma\bar{\sigma}\sigma$ at Trial 134, it guesses $\sigma\sigma\acute{\bar{\boldsymbol{\sigma}}}\sigma\bar{\sigma}\sigma$, as this form incurs the fewest violations of *ExtLapse while satisfying *ExtLapseL. As the intended winner is $\acute{\boldsymbol{\sigma}}\sigma\bar{\sigma}\sigma\bar{\sigma}\sigma$, however, the learner must update itself, and returns to a grammar that generates initial stress. At trial 146, the learner encounters five-syllable $\sigma\sigma\sigma\bar{\sigma}\sigma$, and guesses $\acute{\boldsymbol{\sigma}}\sigma\sigma\bar{\sigma}\sigma$, with initial stress. As the winner is $\sigma\sigma\acute{\boldsymbol{\sigma}}\bar{\sigma}\sigma$, it performs another update.

The learner continues to bounce back and forth between multiple incorrect hypotheses before finally converging on the target grammar at trial 192. What is immediately noticeable about the learning trajectory is that the weight of AlignL is in constant flux: it is either a winner- or loser-preferrer at each step along the path to convergence. The ranking trace in Figure 4 illustrates; the thick black line represents the weight of AlignL over time.

Figure 4: Weight of AlignL over time



The erratic behavior of AlignL, clearly visible in Figure 4, points us toward the reason why Midpoint is so difficult for the learner. Let's go back to the error in trial 65, when the learner encounters its first long word. Recall that, when the learner incorrectly guesses $\bar{\sigma}\bar{\sigma}\bar{\sigma}\acute{\bar{\boldsymbol{\sigma}}}\sigma\sigma$, there are three constraints that could be responsible for attested $\acute{\bar{\boldsymbol{\sigma}}}\bar{\sigma}\bar{\sigma}\bar{\sigma}\sigma\sigma$: *ExtLapseL, *LapseL, and AlignL. The learner is agnostic as to which constraint is responsible for the attested form, so it promotes all winner-preferrers, which results in a ranking generating initial stress. In response to this error, we could imagine a different response: the learner could look at its current weights, see that *ExtLapseR currently dominates AlignL, and refuse to reverse the ranking. This reference to previously established rankings is a property of ERC-based learners (see Brasoveanu & Prince 2011, Yanovich to appear): the updates a learner performs in response to errors are confined and informed by the crucial rankings it has already learned. ERC-based learners would not encounter the credit problem described above, as they would require the learner to retain the ranking *ExtLapseR >> AlignL. But this is not how the GLA works, and it is important to note that the GLA's lack of reference to previously established rankings is a *desirable property of the algorithm.* If for example the first word a child hears is a speech error an adult intends to produce the form $\acute{\boldsymbol{\sigma}}\sigma$, but instead produces $\sigma\acute{\boldsymbol{\sigma}}$ we do not want the child to learn that AlignR >> AlignL is an irreversible ranking. In other words, for a learner to be able to deal with noise in the input, it must be able to unlearn incorrect crucial rankings established in response to misproduced, or misperceived, forms. Recall that the learner's capability to unlearn rankings established with respect to data encountered earlier in the learning process is exactly what makes Midpoint so difficult. Thus the property of the GLA that makes Midpoint so difficult to learn is in fact an essential and important component of how the algorithm works.

Before moving on, it is important to note that the credit problem posed by midpoint systems is completely general all midpoint systems, whether they involve context-sensitive varieties of *ExtLapse or *Lapse, are difficult for the learner in this respect. Consider the midpoint system in (26) (repeated from (6)):

(26) *LapseL >> *LapseR >> AlignL

a. $[[\acute{\boldsymbol{\sigma}}\sigma]_L]_R$

b. $[\sigma[\acute{\boldsymbol{\sigma}}]_L\sigma]_R$

c. $[\acute{\boldsymbol{\sigma}}\sigma]_L[\sigma\sigma]_R$

Learning (26) does not require long words – all crucial rankings are available in (26c) – but it is still difficult for the learner, as the updates performed in response to (26b) are not consistent with those performed in response to the other forms. For illustration, see the error log[9] in (27).

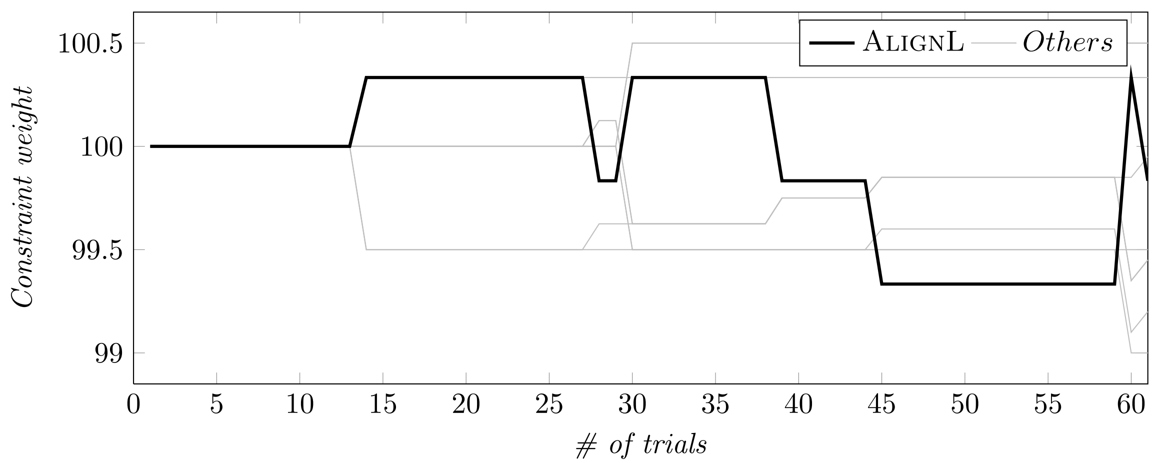
(27) Error log for (26)

Trial	Form	Guess	Correct
14	$\sigma\bar{\sigma}$	$\sigma\acute{\bar{\boldsymbol{\sigma}}}$	$\acute{\boldsymbol{\sigma}}\bar{\sigma}$
	Updates: promote AlignL		
	Resulting ranking: AlignL >> *all*		
	System: Initial stress		
28	$\bar{\sigma}\bar{\sigma}\sigma$	$\acute{\bar{\boldsymbol{\sigma}}}\bar{\sigma}\sigma$	$\bar{\sigma}\acute{\bar{\boldsymbol{\sigma}}}\sigma$
	Updates: promote *LapseR*, demote* AlignL		
	Resulting ranking: *LapseR >> AlignL		
	System: Penultimate stress		
30	$\bar{\sigma}\bar{\sigma}\bar{\sigma}\bar{\sigma}\bar{\sigma}$	$\bar{\sigma}\bar{\sigma}\bar{\sigma}\acute{\bar{\boldsymbol{\sigma}}}\bar{\sigma}$	$\acute{\bar{\boldsymbol{\sigma}}}\bar{\sigma}\bar{\sigma}\bar{\sigma}\bar{\sigma}$
	Updates: promote AlignL, *LapseL*, demote* *LapseR		
	Resulting ranking: *LapseL >> AlignL >> *LapseR		
	System: Initial stress		
39	$\sigma\sigma\bar{\sigma}$	$\acute{\boldsymbol{\sigma}}\sigma\bar{\sigma}$	$\sigma\acute{\boldsymbol{\sigma}}\bar{\sigma}$
	Updates: promote *LapseR*, demote* AlignL		
	Resulting ranking: *LapseL >> AlignL >> *LapseR		
	System: Initial stress		
45	$\sigma\bar{\sigma}\bar{\sigma}$	$\acute{\boldsymbol{\sigma}}\bar{\sigma}\bar{\sigma}$	$\sigma\acute{\bar{\boldsymbol{\sigma}}}\bar{\sigma}$
	Updates: promote *LapseR, WSP*; demote* AlignL		
	Resulting ranking: *LapseL >> *LapseR >> WSP >> AlignL		
	System: quantity-sensitive midpoint		
60	$\sigma\bar{\sigma}\sigma\sigma\bar{\sigma}$	$\acute{\boldsymbol{\sigma}}\bar{\sigma}\sigma\sigma\bar{\sigma}$	$\sigma\acute{\bar{\boldsymbol{\sigma}}}\sigma\sigma\bar{\sigma}$
	Updates: promote AlignL*; demote* WSP		
	Resulting ranking: *LapseL >> AlignL >> *LapseR >> WSP		
	System: Initial stress		
61	$\sigma\bar{\sigma}\bar{\sigma}$	$\acute{\boldsymbol{\sigma}}\bar{\sigma}\bar{\sigma}$	$\sigma\acute{\bar{\boldsymbol{\sigma}}}\bar{\sigma}$
	Updates: promote *LapseR*; demote* AlignL		
	Resulting ranking: *LapseL >> *LapseR >> AlignL >> WSP		
	System: quantity-insensitive midpoint (target)		

[9]This error log is from Run 1 of QI Lim 1. The learner has the Portuguese word length distribution.

In response to the error in trial 14 (guess: $\sigma\boldsymbol{\acute{\sigma}}$; correct: $\boldsymbol{\acute{\sigma}}\bar{\sigma}$), the learner settles on a ranking that generates initial stress; it transitions to a ranking that generates penultimate stress at trial 28, when it encounters a trisyllabic word. In trial 30, the learner incorrectly guesses that a five-syllable word should have penultimate stress; the resulting update returns the learner to a ranking that generates initial stress. This is because there are multiple constraints that could be responsible for the attested $\boldsymbol{\acute{\sigma}}\bar{\sigma}\bar{\sigma}\bar{\sigma}\bar{\sigma}$ – *ExtLapseL, *LapseL, or AlignL – the learner assigns credit equally to all. This is the same credit problem encountered in response to the extended midpoint system. In addition, the ranking trace corresponding to (27), AlignL (in bold black) displays the same erratic behavior: it is either promoted or demoted at each step of learning.

Figure 5: Weight of AlignL over time



The characteristic of midpoint systems responsible for the credit problem is the variable positioning of main stress with respect to a word edge: stress is located at a word edge in long and short words, but not others. This inconsistency causes the learner to overgeneralize: when it sees a word with initial stress, it will often update itself to a ranking that generates initial stress, regardless of whether or not it has previously seen forms that do not fit that generalization. The variable positioning of stress with respect to a word edge is a signature characteristic of midpoint systems, other unattested pathologies (§4.1.3), and only a small class of attested systems (§5).

3.5 Discussion

This section has identified two factors that make midpoint systems difficult to learn. First, a credit problem imposed by all midpoint systems often causes the learner to revise its previous hypotheses. In addition, the probable absence of long words from many learners' inputs means that data necessary to learn a subset of midpoint systems – those where both contextual lapse constraints are varieties of *ExtLapse – is not available to the learner. In the absence of this data, the learner will never converge on a midpoint system. Both difficulties arise as a direct consequence of properties of error-driven learning.

We can compare the learner's performance on the midpoint systems to its performance on AP. When long-word data are absent, the learner is biased to reach a ranking that generates AP, as AP is the simplest possible hypothesis that accounts for all of the observed data. The situation in which long-word data are absent is no different from the situation where they are present: all the long-word data do is confirm that the simplest possible hypothesis is in fact the correct one.

The observation is that midpoint systems are difficult to learn. The hypothesis is that this learnability difficulty is what leads to their unattested status. Although the discussion in this section has focused entirely on quantity-insensitive midpoint systems, the claims made here apply to quantity-sensitive midpoint systems as well. Quantity-sensitive midpoint systems tend to be even harder for the learner to acquire than their quantity-insensitive counterparts, as the position of stress within the window is not necessarily consistent with respect to a word edge.

Although the results discussed above are suggestive, we cannot yet be confident that this is the right account. What we have seen is that fundamental properties of the GLA make midpoint systems difficult to learn. What we need to show is that humans find them difficult to learn as well. This is a testable hypothesis: we can probe the behavior of human learners through artificial learning experiments, to see if the GLA models human behavior in this respect. But for now, hypothesizing that humans would find midpoint systems difficult, too, seems reasonable.

4 Exploring the predictions: attested long-word phenomena

The appeal to the infrequency of long words helps us explain why extended midpoint systems are unattested. It also predicts that all systems in which some crucial rankings are only visible in long words should be rare, or unattested altogether: systems dependent on long words should only be attested in languages with many long words. I refer to this broader prediction as the "long-word hypothesis". This section tests the long-word hypothesis by focusing on two classes of stress system in which long words appear to play a crucial role. The aim is to answer the following questions: what makes these attested patterns different than extended midpoint systems? Why should some types of long-word phenomena be attested, but not others?

I show that, in some cases, although long words appear to play a crucial role, exposure to long words is in fact not necessary to acquire the target grammar. In short, the stress pattern in longer words is *predictable* given the stress pattern of shorter words (§4.1). In other cases, there appears to be a positive correlation between the complexity of a language's stress system and its relative number of long words (§4.2). Together, these findings provide support for the long-word hypothesis.

4.1 Behavior of long words is predictable: lapses and clashes

This section discusses two types of system, those with contextually licensed lapses (§4.1.1) and clashes (§4.1.2), and shows that the stress patterns of long words are predictable given consideration of the patterns of short words. §4.1.3 acknowledges several pathologies of constraints introduced in this section, and shows that they have certain characteristics in common with midpoint systems.

4.1.1 Binary plus lapse systems

Binary plus lapse systems (name due to Gordon 2002) can be divided into two groups: those in which the lapse is located at the edge of the word (binary plus *external* lapse), and those where the lapse is located word-internally (binary plus *internal* lapse). An example of a binary plus external lapse system is Pintupi (Hansen & Hansen 1969), where odd non-final syllables are stressed (28).

(28) Pintupi (Hansen & Hansen 1969)

a. $\boldsymbol{\acute{\sigma}}\sigma$

b. $\boldsymbol{\acute{\sigma}}\sigma\sigma$

c. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma$

d. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\sigma$

e. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma$

f. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\sigma$

An example of a binary plus internal lapse system, where the lapse appears word-internally, is Garawa (Furby 1974) (29).

(29) Garawa (Furby 1974)

a. $\boldsymbol{\acute{\sigma}}\sigma$

b. $\boldsymbol{\acute{\sigma}}\sigma\sigma$

c. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma$

d. $\boldsymbol{\acute{\sigma}}\sigma\sigma\boldsymbol{\grave{\sigma}}\sigma$

e. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma$

f. $\boldsymbol{\acute{\sigma}}\sigma\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma$

The systems in (23-24) appear to be systems in which long words matter because the location of the lapse is unclear until seven-syllable words. In real Garawa (29), we find that the lapse occurs adjacent to the primary stress. It is also conceivable, however, that in unattested Garawa′, the lapse in seven-syllable words could occur between two secondary stresses. The difference between real Garawa and unattested Garawa′ (30) can only be found in seven-syllable words.

(30) Unattested Garawa′

a. $\boldsymbol{\acute{\sigma}}\sigma$

b. $\boldsymbol{\acute{\sigma}}\sigma\sigma$

c. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma$

d. $\boldsymbol{\acute{\sigma}}\sigma\sigma\boldsymbol{\grave{\sigma}}\sigma$

e. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma$

f. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\sigma\boldsymbol{\grave{\sigma}}\sigma$

But the typology of binary plus lapse systems is limited: Kager (2001) identifies two strong typological asymmetries. If the lapse is *external*, it is always right-edge adjacent, as in Pintupi (28). Kager proposes the constraint Lapse-at-End to encode this asymmetry, which licenses right-edge lapses by banning them from occurring elsewhere (31).

(31) Lapse-at-End: a * for each sequence of two stressless syllables not at the right edge.

If the lapse is internal, it is always adjacent to the primary[10] as in Garawa (29). Kager proposes the constraint Lapse-at-Peak to encode this asymmetry, which licenses peak-adjacent lapses by banning them from occurring elsewhere (32).

[10]For discussion of apparent exceptions to this generalization (Indonesian, Hawaiian, Spanish), see Kager 2001: 17, where alternative analyses are provided. In general, the existence of a language violating Kager's generalization will be difficult to find, as determining whether a language is Garawa- or Garawa′-like requires seven-syllable words, which are quite rare cross-linguistically. In addition, as many of the existing seven-syllable words are likely morphologically complex, we might expect the basic stress pattern to be obscured. Although it has been claimed that English is a counterexample (Hayes 1982), my investigation found that judgments are not clear across speakers. I asked ten native speakers of American English how they would pronounce 'Machepaconaponsuck', a seven-syllable place-name in Rhode Island. Of these, three preferred a pronunciation with an initial dactyl ($\boldsymbol{\grave{\sigma}}\sigma\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\acute{\sigma}}\sigma$), two preferred a pronunciation with a peak-adjacent lapse ($\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\sigma\boldsymbol{\acute{\sigma}}\sigma$), and three had no clear preference between these two pronunciations. The other two speakers reported different pronunciations, either placing the lapse at the end ($\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\acute{\sigma}}\sigma\sigma$) or avoiding the lapse altogether ($\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\acute{\sigma}}\sigma$). It isn't entirely clear that English is in fact an exception to Kager's generalization. Note also that the three speakers who reported the initial dactyl are phonologists, and perhaps not naïve subjects.

(32) LAPSE-AT-PEAK: a * for each sequence of two stressless syllables not adjacent to the peak

The important consequence of these constraints is that they rule out Garawa′, and other similar unattested patterns. With (31) and (32) in CON, the behavior of seven-syllable forms is entirely given consideration of the five-syllable forms. In Pintupi, five-syllable $\acute{\sigma}\sigma\grave{\sigma}\sigma\sigma$ shows that LAPSE-AT-END >> LAPSE-AT-PEAK; the seven-syllable word must then have a right-edge lapse (33).

(33) Tableau for Pintupi $\acute{\sigma}\sigma\grave{\sigma}\sigma\grave{\sigma}\sigma\sigma$

$\sigma\sigma\sigma\sigma\sigma\sigma\sigma$	LAPSE-AT-END	LAPSE-AT-PEAK
☞ a. $\acute{\sigma}\sigma\grave{\sigma}\sigma\grave{\sigma}\sigma\sigma$		*
b. $\acute{\sigma}\sigma\sigma\grave{\sigma}\sigma\grave{\sigma}\sigma$	*!	
c. $\acute{\sigma}\sigma\grave{\sigma}\sigma\sigma\grave{\sigma}\sigma$	*!	*

In Garawa, five-syllable $\acute{\sigma}\sigma\sigma\grave{\sigma}\sigma$ shows that LAPSE-AT-PEAK >> LAPSE-AT-END; the seven-syllable word must have a primary-adjacent lapse (34).

(34) Tableau for Garawa $\acute{\sigma}\sigma\sigma\grave{\sigma}\sigma\grave{\sigma}\sigma$

$\sigma\sigma\sigma\sigma\sigma\sigma\sigma$	LAPSE-AT-PEAK	LAPSE-AT-END
a. $\acute{\sigma}\sigma\grave{\sigma}\sigma\grave{\sigma}\sigma\sigma$	*!	
☞ b. $\acute{\sigma}\sigma\sigma\grave{\sigma}\sigma\grave{\sigma}\sigma$		*
c. $\acute{\sigma}\sigma\grave{\sigma}\sigma\sigma\grave{\sigma}\sigma$	*!	*

In short, binary plus lapse systems are in fact not systems in which long words are necessary to establish any crucial rankings; the stress of long words is predictable given consideration of the stress of shorter words. We might expect, then, that a learner would not face any difficulty in learning a binary plus lapse system, as exposure to long words is not necessary to reach the target grammar. This is in fact the case: a learner equipped with both LAPSE-AT-END and LAPSE-AT-PEAK takes 58 and 73 trials (on average) to learn Garawa and Pintupi, respectively[11].

4.1.2 Binary plus clash systems

In binary plus clash systems (name due to Gordon 2002), stress generally alternates in a binary fashion, but clashes arise under certain conditions. In most systems, as in Gosiute Shoshone (Miller 1996), the clash is realized at the edge (35).

(35) Gosiute Shoshone (Miller 1996)

a. $\acute{\sigma}\grave{\sigma}$

b. $\acute{\sigma}\sigma\grave{\sigma}$

c. $\acute{\sigma}\sigma\grave{\sigma}\grave{\sigma}$

d. $\acute{\sigma}\sigma\grave{\sigma}\sigma\grave{\sigma}$

e. $\acute{\sigma}\sigma\grave{\sigma}\sigma\grave{\sigma}\grave{\sigma}$

f. $\acute{\sigma}\sigma\grave{\sigma}\sigma\grave{\sigma}\sigma\grave{\sigma}$

In Southern Paiute (Sapir 1930; also Harms 1966, Wheeler 1979), clashes cannot occur at word edges because neither peripheral syllable can be stressed. The clash is realized away from the primary stress (36).

[11] The learner in these simulations differs in non-crucial ways from the one introduced in §3: the input files had to be modified to accommodate systems with more than one stress and words with more than seven syllables.

(36) Southern Paiute (Sapir 1930)

a. $\boldsymbol{\acute{\sigma}}\sigma$

b. $\sigma\boldsymbol{\acute{\sigma}}\sigma$

c. $\sigma\boldsymbol{\acute{\sigma}}\boldsymbol{\grave{\sigma}}\sigma$

d. $\sigma\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma$

e. $\sigma\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\boldsymbol{\grave{\sigma}}\sigma$

f. $\sigma\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma$

If we focus on the four-syllable words alone, the location of the clash with respect to the edge is ambiguous. For example, four-syllable $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\boldsymbol{\grave{\sigma}}$ in Gosiute Shoshone is consistent with two six-syllable forms: attested $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\boldsymbol{\grave{\sigma}}$, with a clash at the edge, and unattested *$\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}$, with a word-internal clash. As noted by Kager (2001) and van Urk (2013), however, quantity-insensitive systems with internal clashes, i.e. those separated from the edge by another stress, are unattested[12]. To capture this asymmetry, Kager (2001) proposes CLASH-AT-EDGE[13] (37), which penalizes all clashes not adjacent to a word boundary.

(37) CLASH-AT-EDGE: a * for each sequence of 2 stressless syllables not adjacent to an edge.

Both Gosiute Shoshone and Southern Paiute demonstrate another asymmetry in the typology of binary plus clash systems: in almost all systems, the clash is realized away from the primary stress[14]. To encode this asymmetry, Kager (2001) proposes *CLASH-AT-PEAK (38), which penalizes all clashes involving the primary stress.

(38) *CLASH-AT-PEAK: a * if the primary is adjacent to one or more secondaries.

Simply having these two constraints in CON is sufficient to make the six-syllable forms of Gosiute Shoshone and Southern Paiute predictable. In Gosiute Shoshone, four-syllable $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\boldsymbol{\grave{\sigma}}$ and shorter words show us that peripheral syllables must receive stress, and that *LAPSE is inviolable. Given this, it is predictable that, in six-syllable $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\boldsymbol{\grave{\sigma}}$, the clash is realized at the edge opposite the one that hosts primary stress; the alternatives are harmonically bound (39).

(39) Tableau for Gosiute Shoshone $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\boldsymbol{\grave{\sigma}}$

$\sigma\sigma\sigma\sigma\sigma\sigma$	CLASH-AT-EDGE	*CLASH-AT-PEAK
☞ a. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\boldsymbol{\grave{\sigma}}$		
b. $\boldsymbol{\acute{\sigma}}\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}$		*!
c. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}$	*!	*

Similar considerations apply for Southern Paiute. Four-syllable $\sigma\boldsymbol{\acute{\sigma}}\boldsymbol{\grave{\sigma}}\sigma$ and shorter words teach us that stressing peripheral syllables is dispreferred, and that *LAPSE is not violated. Given this, it is predictable that, in six-syllable $\sigma\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\boldsymbol{\grave{\sigma}}\sigma$, the clash should be realized at the side of the word not adjacent to the primary.

Given that the stress of long words is predictable from the stress of shorter words, we would expect a learner to converge on the target grammar fairly quickly. We expect this because exposure

[12]For sure, there are cases where word-internal clashes arise due to cyclic preservation of stem stress; for an example, see Pike (1964) on Auca. I am interested here in cases where cyclicity does not play a role.

[13]van Urk (2013: 21) notes a need to slightly revise the definition of CLASH-AT-EDGE, to be able to account for the eight-syllable forms of Southern Paiute. For simplicity, I work with the current definition.

[14]The only exception in van Urk 2013 is South Conchucos Quechua (SCQ, Hintz 2006). In SCQ $\boldsymbol{\acute{\sigma}}\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma \succ$ *$\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\boldsymbol{\grave{\sigma}}\sigma$. The pattern just shows that *CLASH-AT-PEAK is violable: CLASH-AT-EDGE >> *CLASH-AT-PEAK.

to rare long forms is not necessary to infer all crucial rankings. This is in fact the case: a learner equipped with the constraints introduced here took (on average) 69 and 77 trials to learn Southern Paiute and Gosiute Shoshone, respectively.

4.1.3 A brief excursus: on different approaches to gaps

The reader will notice that I have endorsed two different strategies for explaining gaps in the typology. For the midpoint systems in §3, I have claimed that they are absent from the typology because they are difficult to learn, not because they are outside of the learner's hypothesis space. But for the lapse- and clash-tolerating systems, I have followed Kager (2001) in attributing asymmetries in the typology to constraints on CON, i.e. constraints on the learner's hypothesis space.

My belief is that both strategies are necessary. There are reasons to think that, in addition to pruning the typology by incorporating considerations of learnability, we should also exclude certain kinds of constraints from CON. For example, it seems difficult to believe that any language would have an active dispreference for CV sequences, or that any language would exhibit a ban on even-syllabled words. Unattested systems like these, whose absence cannot obviously be attributed to any learnability-related considerations, should presumably be excluded from the hypothesis space.

This means that theories of phonological typology must take into at least two (and likely other) factors: some systems are unattested because they are not possible systems; others are unattested because they are difficult to learn. A theory must also take a stand on the relative contribution of these two factors. Should our theory of CON be more permissive, or more restrictive? How much explanatory power should we give to the role of learnability shaping phonological typology? The position argued in this paper is that learnability should play a central role in our theory of phonological typology. Our theory of CON can therefore be fairly permissive: there is no need to exclude a system from the learner's hypothesis space if we can show that it is difficult to learn.

While the midpoint systems in §3 are amenable to a learnability exclusion, this is less obviously the case for asymmetries in the typology of lapse- and clash-tolerating systems discussed above. While further investigation might reveal that these asymmetries can be completely accounted for as a consequence of learnability, perceptibility (see e.g. Lunden & Kalivoda 2013), or another extra-grammatical factor, for now, I will proceed with the assumption that the constraints introduced above – LAPSE-AT-END, CLASH-AT-EDGE, and their ilk – are part of CON.

4.1.4 Pathologies of rhythmic licensing constraints

As noted by Alber (2005) (see also Kager 2005, van Urk 2013), the rhythmic licensing constraints just discussed give rise to two pathologies. I show that these pathologies have something in common with midpoint systems: the position of main stress with respect to some edge varies as a function of syllable count. The first pathology, referred to as *licensor attraction* (Kager 2005), is characterized by a ranking in which main stress prefers to fall at the left edge (PRIMARYL >> PRIMARYR), but both LAPSE-AT-PEAK and LAPSE-AT-END dominate PRIMARYL. This results in a type of system in which primary stress usually falls at the left edge of the word (40a-c,e), but must shift to the right edge in some words to simultaneously satisfy LAPSE-AT-PEAK and LAPSE-AT-END (40d,f).

(40) Licensor attraction (LAPSE-AT-PEAK, LAPSE-AT-END >> PRIMARYL)

a. $\boldsymbol{\acute{\sigma}}\sigma$

b. $\boldsymbol{\acute{\sigma}}\sigma\sigma$

c. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma$

d. $\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\acute{\sigma}}\sigma\sigma$

e. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma$

f. $\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\acute{\sigma}}\sigma\sigma$

In even-parity words, satisfaction of PrimaryL does not induce a violation of either Lapse-at-Peak or Lapse-at-End (41).

(41) Tableau for $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma$

$\sigma\sigma\sigma\sigma$	Lapse-at-Peak	Lapse-at-End	PrimaryL
☞ a. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma$			
b. $\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\acute{\sigma}}\sigma$			*!

In odd-parity words of five syllables or more, however, satisfaction of PrimaryL entails a violation of one or both of the lapse licensing constraints. As both Lapse-at-Peak and Lapse-at-End are inviolable, main stress shifts to the right edge to license a lapse that is both peak- and right-edge adjacent (42).

(42) Tableau for $\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\acute{\sigma}}\sigma\sigma$

$\sigma\sigma\sigma\sigma\sigma$	Lapse-at-Peak	Lapse-at-End	PrimaryL
☞ a. $\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\acute{\sigma}}\sigma\sigma$			
b. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\sigma$	*!		
c. $\boldsymbol{\grave{\sigma}}\sigma\sigma\boldsymbol{\acute{\sigma}}\sigma$		*!	

A similar pathology, the *peak-shifting pathology* (name due to van Urk 2013), is associated with the constraints Clash-at-Edge and *Clash-at-Peak. In systems exhibiting this pathology, the alignment of primary stress with respect to a word edge is crucially dependent on whether or not the word contains a clash. An example of a peak-shifting language follows (43).

(43) Peak-shifting pathology (Clash-at-Edge >> *Clash-at-Peak >> PrimaryL)

a. $\boldsymbol{\acute{\sigma}}\sigma$

b. $\boldsymbol{\acute{\sigma}}\boldsymbol{\grave{\sigma}}\sigma$

c. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma$

d. $\boldsymbol{\acute{\sigma}}\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma$

e. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma$

f. $\boldsymbol{\acute{\sigma}}\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma$

In all forms but (43d,f), PrimaryL ensures that main stress lodges on the initial syllable. In longer words with an odd number of syllables, however, primary stress surfaces at the right edge of the word. In these forms, the combined forces of Clash-at-Edge and Clash-at-Peak compel a violation of PrimaryL; the primary stress is then the rightmost stress.

Kager (2005) and van Urk (2013) propose to deal with the problems of licensor attraction and peak-shifting by either redefining constraints or introducing new ones that exclude these systems from the factorial typology; see those works for more details. I would like to suggest an alternative: that the reason we do not see systems exhibiting either licensor attraction or peak-shifting is because they would be difficult to learn. The pathologies just summarized share a characteristic with midpoint systems: while main stress falls at a given edge in the majority of forms (e.g. all but (42d,f) for peak-shifting), there is a certain subset of forms where main stress falls closer to the opposite edge (43d,f). The position of main stress is dependent on syllable count, and, as in the midpoint systems, *variable* with respect to the word edge.

As shown in §3.4, the variable positioning of stress with respect to a word edge is, more or less, what makes midpoint systems difficult for the learner. The unattested status of systems exhibiting licensor attraction and peak-shifting allows us to generalize: perhaps what makes midpoint systems, peak-shifting, and licensor attraction difficult to acquire is not the variable position of *some* stress with respect to a word edge, but rather the position of *main* stress. The hypothesis, then, is that systems in which the position of main stress varies by word length are generally dispreferred.

Independent support for this hypothesis comes from a stark asymmetry in the typology of binary stress systems. In the majority of iterative binary stress systems (143/158 in StressTyp, see Staubs 2014b: 2), the placement of primary stress is correlated with the direction of iterative stress (see also Gordon 2002: 31). Systems where the primary stress is rightmost generally exhibit right-to-left iteration; systems where primary stress is leftmost generally exhibit left-to-right iteration. The result is a system in which the location of primary stress is consistent; see (44) for an example.

(44) Rightmost primary and right-to-left iteration

a. $\boldsymbol{\acute{\sigma}}\sigma$

b. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}$

c. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma$

d. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}$

e. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma$

f. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}$

In a smaller number of languages (15/158 in StressTyp, Staubs 2014b: 2), however, the placement of primary stress opposes the direction of iterative parsing. In these systems, the position of primary stress varies as a function of word parity. An example of such a system is provided in (45).

(45) Rightmost primary and right-to-left iteration

a. $\sigma\boldsymbol{\acute{\sigma}}$

b. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}$

c. $\sigma\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}$

d. $\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}$

e. $\sigma\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}$

f. $\sigma\boldsymbol{\acute{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma\boldsymbol{\grave{\sigma}}\sigma$

Staubs shows that a MaxEnt learner faces a greater difficulty in acquiring a system like (45), where the position of main stress is variable, than it does in acquiring a system like (44), where the position of main stress is fixed with respect to some edge. I suggest, then, that the absence of licensor attraction and peak-shifting, as well as midpoint systems, is symptomatic of a more general bias against systems where the position of main stress varies as a function of syllable count. Perhaps the relevant difference between the attested iterative binary systems in (44-45) and the unattested systems is that, in (44-45), it is clear from the comparison of disyllabic and trisyllabic words that the position of main stress varies as a function of word parity. For all of the unattested systems we have discussed (midpoint, licensor attraction, peak-shifting), this generalization is only clear in words of four syllables or longer. As four-syllable words are, on average, far less frequent than trisyllabic words (on average, 9% vs. 18% of the corpus in §3.1), the learner would receive more evidence that the position of main stress is variable when learning an iterative binary system than any of the unattested ones. We will return to this general point – that systems where main stress placement depends on word length are dispreferred – in §5.

4.2 Long-word phenomena in languages with long words: ternary stress systems

The second and final class of systems I will discuss in this section are languages with ternary stress (where each stress is preferably separated by two syllables, e.g. $\acute{\boldsymbol{\sigma}}\sigma\sigma\acute{\boldsymbol{\sigma}}\sigma\sigma$). Some ternary systems pose a challenge for the long-word hypothesis because they are systems in which learners must be exposed to very long words – of six syllables or longer – to be able to infer all crucial rankings. We will focus here on two ternary systems where long words matter: Chugach Alutiiq Yupik (46) (Chugach; Leer 1985) and Cayuvava (47) (Key 1961, 1967).

(46) Stress in Chugach[15] (Leer 1985)

a. $\sigma\acute{\boldsymbol{\sigma}}$

b. $\sigma\acute{\boldsymbol{\sigma}}\sigma$

c. $\sigma\acute{\boldsymbol{\sigma}}\sigma\grave{\boldsymbol{\sigma}}$

d. $\sigma\acute{\boldsymbol{\sigma}}\sigma\sigma\grave{\boldsymbol{\sigma}}$

e. $\sigma\acute{\boldsymbol{\sigma}}\sigma\sigma\grave{\boldsymbol{\sigma}}\sigma$

f. $\sigma\acute{\boldsymbol{\sigma}}\sigma\sigma\grave{\boldsymbol{\sigma}}\sigma\grave{\boldsymbol{\sigma}}$

g. $\sigma\acute{\boldsymbol{\sigma}}\sigma\sigma\grave{\boldsymbol{\sigma}}\sigma\sigma\grave{\boldsymbol{\sigma}}$

(47) Stress in Cayuvava (Key 1961, 1967)

a. $\acute{\boldsymbol{\sigma}}\sigma$

b. $\acute{\boldsymbol{\sigma}}\sigma\sigma$

c. $\sigma\acute{\boldsymbol{\sigma}}\sigma\sigma$

d. $\sigma\sigma\acute{\boldsymbol{\sigma}}\sigma\sigma$

e. $\grave{\boldsymbol{\sigma}}\sigma\sigma\acute{\boldsymbol{\sigma}}\sigma\sigma$

f. $\sigma\grave{\boldsymbol{\sigma}}\sigma\sigma\acute{\boldsymbol{\sigma}}\sigma\sigma$

g. $\sigma\sigma\grave{\boldsymbol{\sigma}}\sigma\sigma\acute{\boldsymbol{\sigma}}\sigma\sigma$

In both cases, the learner must be exposed to long words to infer all crucial rankings. In Chugach, it is not clear until words of eight syllables or longer (46g) that ternary alternation is *completely general* in this language, not just licensed at the peak. In Chugach, it isn't clear until words of six syllables or longer that there is more than one stress per word; up until (47e), the system could just as well be one with AP stress.

The fact that the Chugach and Cayuvava patterns are dependent on long words, however, does not pose a problem for the long-word hypothesis: in both languages, it is probable that long words are frequent. Although neither Chugach nor any other dialect of Yupik has an online Bible translation or accessible large text collection, in Inuktitut and Inupiatun (also Eskimo-Aleut), long words are extremely frequent. In Inuktitut, words of eight or more syllables make up 12.50% of the corpus; in Inupiatun, they make up 12.11%. We don't know for sure that the word length distribution of Chugach is similar, but another dialect of Yupik, Central Alaskan, certainly has long words (Miyaoka 2012: 132).

(48) angya-cuara-li-yu-kapigte-llru-nric-aaq-sugnarq-llru-yugnarz-aanga[16]
boat-small-make-DES-ITS-PST-NEG-CTR-INF-PST-INF-IND.3sg.1sg.
'I'm in doubt that he actually didn't really want to make me a small boat (but he did)'

[15] Leer reports all stresses as equal. I assume here that the leftmost is primary, as it is more common in quantity-insensitive systems for main stress to remain at a consistent distance from the edge.

[16] I assume that this form is a single prosodic word (see Miyaoka 2012: 70-71 for information on CAY prosody, where it appears that morphologically simple and complex words are treated alike).

Assuming that the word length distribution of Chugach resembles that of Inuktitut, we can ask if a learner can acquire the pattern in (46) if it samples long words at the rate an Inuktitut learner would encounter them. As we might expect, it is: the learner takes around 40 trials to converge.

As Cayuvava also has no Bible translation nor available text collection, it is hard to know for sure exactly what the word length distribution of this language was. But a short text in Key (1967) gives us an idea: words of six or more syllables make up 26.4% of this 76-word text (49).

(49) Word length distribution of Key's (1967) text

Syllables	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
# of words	5	0	19	21	8	10	5	4
% of total	6.9%	0%	26.4%	29.2%	11.1%	13.9%	6.9%	5.6%

A learner that samples words at the rate they are attested in (49) does not encounter any difficulty in acquiring the Cayuvava pattern (45 trials on average). This is expected, as the learner encounters long words very frequently.

4.3 Local summary

The aim of this section was to test the long-word hypothesis by investigating two classes of stress system in which long words appear to play a crucial role. The investigation yielded two findings. First, the typology of lapse- and clash-tolerating binary systems (§4.1) shows us that apparently complex patterns of stress in longer words can in fact be entirely predictable given the stress pattern in shorter words. Second, while some ternary systems (§4.2) are dependent on long words, they do not pose a problem for the long-word hypothesis, as these languages also seem to have many long words. Although it is currently unclear exactly how many long words a learner would have to encounter for a pattern dependent on long words to be learnable, we have seen here that the long-word hypothesis makes correct predictions about stress typology. We can safely point to it as a contributing factor to the unattested status of extended midpoint systems.

5 A further consideration, and implications for acquisition

Recall from the discussion in §3 that the long-word hypothesis advanced above helps us explain the absence of extended midpoint systems. Although all midpoint systems present a credit problem to the learner (§3.4), we do not know if this credit problem would be significant enough to prevent a child from learning a midpoint system. It is therefore in our best interest to continue to look for other reasons why midpoint systems, as an entire class, might be unattested. This section points toward such a reason.

5.1 On counting systems

Midpoint systems belong to a larger class of systems – counting systems – that are underattested, relative to what we would expect. Here, I define "counting system" as a one-stress system where stress falls at or close to an edge in short forms, but drifts towards the middle of the word in mid-length and longer forms. Counting systems arise when two conditions are met: (i) a general lapse constraint (*ExtLapse or *Lapse) is dominated by only a contextual lapse constraint (*LapseR, *ExtLapseL, etc.), and (ii) stress is aligned towards the outer edge of the window in short forms. An example of an unattested counting system, characterized by the ranking *ExtLapseL >> *ExtLapse >> AlignL, is below (50).

(50) Example counting system (*ExtLapseL >> *ExtLapse >> AlignL)

a. $[\boldsymbol{\acute{\sigma}}\sigma]_L$

b. $[\sigma\boldsymbol{\acute{\sigma}}\sigma]_L\sigma$

c. $[\sigma\sigma\boldsymbol{\acute{\sigma}}]_L\sigma\sigma$

d. $[\sigma\sigma\boldsymbol{\acute{\sigma}}]_L\sigma\sigma\sigma$

In short words, stress stays at the outer edge of the window to best satisfy AlignL (52).

(51) Stress falls at the outer edge in short words

$\sigma\sigma\sigma$	*ExtLapseL	*ExtLapse	AlignL
☞ a. $[\boldsymbol{\acute{\sigma}}\sigma\sigma]_L$			
b. $[\sigma\boldsymbol{\acute{\sigma}}\sigma]_L$			*!

In longer words (50c-e), stress falls at the inner edge of the window, to satisfy *ExtLapse (**??**).

(52) Stress falls at the inner edge in long words

$\sigma\sigma\sigma\sigma\sigma$	*ExtLapseL	*ExtLapse	AlignL
☞ a. $[\boldsymbol{\acute{\sigma}}\sigma\sigma]_L\sigma\sigma$		*!*	
b. $[\sigma\boldsymbol{\acute{\sigma}}\sigma]_L\sigma\sigma$			**

Counting systems differ from midpoint systems in one crucial respect: although stress migrates towards the middle of the word in mid-length words, it does not return to the word edge in longer words. And unlike midpoint systems, we expect counting systems to be fairly frequent: their r-volume is at least 30%. Assuming that 75.75% of all languages have one stress per word (Gordon 2002), we expect for counting systems to make up at least 17.2% of all languages, or at least 88 of the 510 languages in StressTyp. Despite this high predicted frequency, however, counting systems are rare: I am aware of three. In both North Kyungsang Korean (NKK; Kenstowicz & Sohn 2001) and Kimatuumbi (Odden 1996), accent falls within a right-edge disyllabic window in words of up to three syllables (53a-b), but is fixed on the penult in words of four syllables or longer (53c-e).

(53) NKK (Kenstowicz & Sohn 2001) and Kimatuumbi (Odden 1996)

a. $[\boldsymbol{\acute{\sigma}}\sigma]_R$ or $[\sigma\boldsymbol{\acute{\sigma}}]_R$

b. $\sigma[\boldsymbol{\acute{\sigma}}\sigma]_R$ or $\sigma[\sigma\boldsymbol{\acute{\sigma}}]_R$

c. $\sigma\sigma[\boldsymbol{\acute{\sigma}}\sigma]_R$

d. $\sigma\sigma\sigma[\boldsymbol{\acute{\sigma}}\sigma]_R$

e. $\sigma\sigma\sigma\sigma[\boldsymbol{\acute{\sigma}}\sigma]_R$

We can analyze this pattern with the ranking *LapseR >> *ExtLapse: speakers of NKK and Kimatuumbi prefer to avoid extended lapses, subject to the confines of a right-edge disyllabic right-edge window. In two- and three-syllable words, stress is free to fall on either syllable in the disyllabic window; neither option leads to a violation of *ExtLapse (55). (I assume that the variation between penultimate and final stress is due to a variable ranking between AlignR and NonFinality, not reflected below.)

(54) Variation in short words

$\sigma\sigma\sigma$	*LapseR	*ExtLapse
☞ a. $[\boldsymbol{\acute{\sigma}}\sigma]_R$		
☞ b. $[\sigma\boldsymbol{\acute{\sigma}}]_R$		

In words of four syllables or more, only the penult can be stressed, because stressing the final syllable incurs a violation of *ExtLapse (**??**).

(55) No variation in longer words

$\sigma\sigma\sigma\sigma$	*LapseR	*ExtLapse
☞ a. $\sigma\sigma[\boldsymbol{\acute{\sigma}}\sigma]_R$		
b. $\sigma\sigma[\sigma\boldsymbol{\acute{\sigma}}]_R$		*!

In Içuã Tupi (Abrahamson 1968), stress generally falls on the penult in words of up to four syllables (56a-c), and on the antepenult in words of five syllables or longer (56d-e).

(56) Içuã Tupi (Abrahamson 1968)

a. $[\boldsymbol{\acute{\sigma}}\sigma]_R$

b. $[\sigma\boldsymbol{\acute{\sigma}}\sigma]_R$

c. $\sigma[\sigma\boldsymbol{\acute{\sigma}}\sigma]_R$

d. $\sigma\sigma[\boldsymbol{\acute{\sigma}}\sigma\sigma]_R$

e. $\sigma\sigma\sigma[\boldsymbol{\acute{\sigma}}\sigma\sigma]_R$

The analysis of the Içuã Tupi pattern is almost identical to the analysis of NKK and Kimatumbi; the only relevant difference is that the right-edge window is three syllables long, so stress can retract to the antepenult to avoid violations of *ExtLapse[17].

5.2 A bias against counting systems

I suggest that counting systems are underattested relative to what we expect because learners disprefer systems where main stress does not occur at a fixed distance from the word edge in all words (see also Staubs 2014a, §4.1.3). A possible source of this bias comes from facts about child language acquisition. It is well-established that infants, especially young infants (6 months or younger) rely heavily on stress as a cue for word segmentation (see Morgan & Saffran 1995, as well as many others). We might imagine that, for the purposes of word segmentation, an ideal stress system would be one in which stress falls at a fixed distance from a word edge. This kind of system would be a *reliable* aid for segmentation, as a learner would be able to intuit word boundaries based on prosodic cues alone. A less ideal stress system for segmentation would be one where stress placement depends on word length. This kind of system would be an *unreliable* aid for segmentation. If the learner has not yet acquired a lexicon, she does not know what the words are, and therefore cannot possibly know that stress placement depends on their length.

[17]There is potentially an alternative analysis of the counting systems in (54) and (57). Assuming that the duration of each individual syllable decreases as the overall word length increases (Lehiste 1972, and others cited by Zhang 2002: 33), we can potentially understand these patterns as reflective of gradient restrictions on the acceptable distance between a word's rightmost stress and the word boundary. Let us redefine NonFinality as a gradient constraint: "there must be at least x milliseconds between the final peak and the word edge", where the value of x is language-dependent. It could be that, in NKK, final stress is possible in short words because the individual syllables in short words are relatively long, so a stress on the final syllable is sufficiently far away from the word's right edge. In longer words, however, the individual syllables could be shorter, so a stress on the final syllable might not be sufficiently far away from the right edge, and a violation of gradient NonFinality results. A challenge for this analysis is that the trading relation between word length and syllable duration is typically only attested in shorter words; here it would need to be the case that syllables in five-syllable Içuã Tupi words are shorter than those in four-syllable words. But if this alternative analysis succeeds, then counting systems, like midpoint systems, are unattested altogether. I am grateful to Donca Steriade for suggesting this alternative – I leave its exploration to future work.

We do not know what would happen if we attempted to teach a child such a system, but evidence from experiments on English-learning children suggest that young infants use the most frequent stress pattern in the input when segmenting the speech stream (Juszcyk et al. 1993). With this in mind, let's imagine teaching a child the midpoint pattern in (57), assuming the Portuguese word length distribution.

(57) *LAPSEL >> *LAPSER >> ALIGNL

a. $[[\boldsymbol{\acute{\sigma}}]_L]_R$ *Frequency: 33%*

b. $[[\boldsymbol{\acute{\sigma}}\sigma]_L]_R$ *Frequency: 35%*

c. $[\sigma[\boldsymbol{\acute{\sigma}}]_L\sigma]_R$ *Frequency: 18%*

d. $[\boldsymbol{\acute{\sigma}}\sigma]_L[\sigma\sigma]_R$ *Frequency: 10%*

e. $[\boldsymbol{\acute{\sigma}}\sigma]_L\sigma[\sigma\sigma]_R$ *Frequency: 3%*

f. $[\boldsymbol{\acute{\sigma}}\sigma]_L\sigma\sigma[\sigma\sigma]_R$ *Frequency: 1%*

g. $[\boldsymbol{\acute{\sigma}}\sigma]_L\sigma\sigma\sigma[\sigma\sigma]_R$ *Frequency: <1%*

In (57), 82% of all words encountered by the learner would have initial stress, while the remaining 18% would have peninitial stress. If the learner considers initial stress – the most frequent pattern – to be the most important segmentation cue, then *18% of the input would be segmented incorrectly.* For example, if presented with the hypothetical trisyllabic word [patáka], the learner would extract [táka], with initial stress, as a possible word. The result, in effect, would be that the learner acquires a regular initial stress pattern, but the wrong lexical items.

The hypothesis here is that counting systems inhibit reliable segmentation, and that for this reason learners find them difficult. But even if this is not the correct story, what is important here is that counting systems are severely underattested relative to what we might expect. As midpoint systems are a type of counting system, their absence is symptomatic of a more general dispreference against counting systems. Other kinds of counting systems (e.g. (39), (42)) are attested, to the exclusion of midpoint systems, because they have a higher r-volume (4.58% vs. 30+%), and are therefore more likely to occur in the first place.

6 Discussion and conclusions

This paper has identified three factors that could lead to the absence of midpoint systems. First, we expect midpoint systems to be fairly rare in the first place. Second, midpoint systems are inherently difficult to learn: when the data necessary to distinguish such systems from alternatives are absent, the learner prefers the alternatives; when the data necessary to learn them are present, the learner encounters a credit problem. And finally, the severe underattestation of other systems where main stress placement depends on word length suggests a more general bias against them.

If these three factors are entirely responsible for the absence of midpoint systems, then there is no need to exclude them from the learner's hypothesis space by removing the constraints that generate them from CON. Furthermore, the midpoint pathology can no longer serve as an argument that (weakly layered) feet are a necessary component of metrical theory, as there is a way to explain their absence without adopting feet. There are, however, many questions still left to answer before we can be certain that this alternative has succeeded. To some extent, the success of this proposal depends on whether or not the credit problem discussed in §3.4, as well as the bias against counting systems discussed in §5, are sufficient to explain the absence of the 75% of midpoint systems where the long-word difficulty is irrelevant. This is not something we can know without directly observing human learning. Because this question is currently unanswered, I believe that it is too early to say definitively that these three factors alone can explain the total absence of all midpoint systems from the typology. But I also believe that it is far too early to rule this alternative out.

Even if the alternative proposed in this paper ultimately fails, it is worth reminding ourselves that the only sin committed by the foot-free factorial typology in Kager (2012) is overgeneration: it predicts stress systems that look like nothing attested in human language. But the weakly layered feet model, which is argued to be superior to the foot-free model because it avoids the midpoint pathology, also predicts stress systems that do not resemble attested ones. A factorial typology for the weakly layered foot model using non-intervention constraints produces patterns like the one in (58), where two and only two differently-sized feet are aligned with a word edge (Martínez-Paricio & Kager 2014: 22). There are no known systems like (58), where each word has maximally two stresses clustering together at the word edge.

(58) Prediction of the weakly layered feet analysis

a. $(\sigma\sigma)$
b. $(\sigma\sigma\sigma)$
c. $(\sigma\sigma\sigma)\sigma$
d. $(\sigma\sigma)(\sigma\sigma\sigma)$
e. $(\sigma\sigma)(\sigma\sigma\sigma)\sigma$
f. $(\sigma\sigma)(\sigma\sigma\sigma)\sigma\sigma$
g. $(\sigma\sigma)(\sigma\sigma\sigma)\sigma\sigma\sigma$

So even though adopting weakly layered feet allows us to avoid the midpoint pathology, the more general state of affairs is this: while both foot-based and foot-free theories appear to be descriptively adequate, neither theory is adequately restrictive. It is not obvious at all why the two-foot system in (58) should be considered any less pathological than a midpoint system.

Given that both theories overgenerate, there is clearly a lot more work left to be done before we can confidently say that one theory is more restrictive, and therefore more plausible, than the other. Deciding on the right theory of stress typology will involve much more than just comparison of factorial typologies. We must also take into account factors like how easy certain systems can be learned, and whether or not the data necessary to learn them would be realistically available to a learner. This paper represents a step in that direction.

References

Abrahamson, Arne. 1968. Contrastive distribution of phoneme classes in Içuã Tupi. Anthropological Linguistics 10(6). 11–21.

Alber, Birgit. 2005. Clash, lapse and directionality. Natural Language and Linguistic Theory 23. 485–542.

Albright, Adam & Young Ah Do. 2013. Biased learning of phonological alternations. Paper presented at the 21st Manchester Phonology Meeting.

Alderete, John. 2008. Using learnability as a filter on factorial typology: A new approach to Anderson and Browne's generalization. Lingua 118. 1177–1220.

Bane, Max & Jason Riggle. 2008. Three Correlates of the Typological Frequency of Quantity-Insensitive Stress Systems. In Proceedings of the Tenth Meeting of the ACL Special Interest Group on Computational Morphology and Phonology, 29–38. Columbus, Ohio.

Bennett, Ryan. 2013. The uniqueness of metrical structure: rhythmic phonotactics in Huariapano. Phonology 30. 355–398.

Bennett, Ryan & Robert Henderson. 2013. Accent in Uspanteko. Natural Language and Linguistic Theory 31. 589–645.

Boersma, Paul. 2003. Bruce Tesar and Paul Smolensky (2000). Learnability in Optimality Theory. Cambridge, Mass.: MIT Press. Pp. vii + 140. Phonology 20. 436–446.

Boersma, Paul & Bruce Hayes. 2001. Empirical Tests of the Gradual Learning Algorithm. Linguistic Inquiry 32. 45–86.

Boersma, Paul & Clara Levelt. 2000. Gradual constraint-ranking learning algorithm predicts acquisition order. In Proceedings of child language research forum, vol. 30, 229–237.

Brasoveanu, Adrian & Alan Prince. 2011. Ranking and necessity: the Fusional Reduction Algorithm. Natural Language and Linguistic Theory 29. 3–70.

Caballero, Gabriela. 2008. Choguita Rarámuri (Tarahumara) Phonology and Morphology: UC Berkeley dissertation.

Do, Young Ah. 2013. Biased learning of phonological alternations: MIT dissertation.

Eisner, Jason. 1997. What constraints should OT allow? Paper presented at the Annual Meeting of the Linguistic Society of America, Chicago. ROA-204.

Furby, Christine. 1974. Garawa phonology Pacific Linguistics, series A. Canberra: Australian National University.

Goedemans, Rob & Harry van der Hulst. 2009. StressTyp: a database for word accentual patterns in the world's languages. In The use of databases in cross-linguistics research, 235–282. New York: Mouton de Gruyter.

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

Hansen, K. C. & L. E. Hansen. 1969. Pintupi phonology. Oceanic Linguistics 8(2). 153–170.

Harms, Robert. 1966. Stress, voice, and length in Southern Paiute. International Journal of American Linguistics 32. 228–235.

Hayes, Bruce. 1982. Extrametricality and English Stress. Linguistic Inquiry 227–276.

Hayes, Bruce. 2004. Phonological acquisition in Optimality Theory: The early stages. In René Kager, Joe Pater & Wim Zonneveld (eds.), Fixing priorities: Constraints in phonological acquisition, Cambridge: Cambridge University Press.

Hintz, Diane. 2006. Stress in South Conchucos Quechua: A phonetic and phonological study. International Journal of American Linguistics 72. 477–521.

Hyde, Brett. 2012. Alignment constraints. Natural Language and Linguistic Theory 30. 789–836.

Juszcyk, Peter, Angela D. Friederici, Jeannine M. I. Wessels, Vigdis Y. Svenkerud & Marie Jusczyk. 1993. Infants' sensitivity to the sound patterns of native language words. Journal of Memory and Language 32(3). 402–420.

Kager, René. 2001. Rhythmic directionality by positional licensing. Paper presented at the Fifth HIL Phonology Conference (HILP 5).

Kager, René. 2005. Rhythmic licensing theory: an extended typology. In Proceedings of the third international conference on phonology, 5–31. Seoul National University.

Kager, René. 2012. Stress in windows: Language typology and factorial typology. Lingua 122. 1454–1493.

Kager, René & Violeta Martínez-Paricio. 2014. Minimally recursive foot structure. Paper presented at the Annual Meeting on Phonology 2014, MIT, Cambridge, MA.

Kenstowicz, Michael & Hyang-Sook Sohn. 2001. Accentual adaptations in North Kyungsang Korean. In Michael Kenstowicz (ed.), Ken Hale: a Life in Language, 239–270. Cambridge: MIT Press.

Key, Harold H. 1961. Phonotactics of Cayuvava. International Journal of American Linguistics 27. 143–150.

Key, Harold H. 1967. Morphology of Cayuvava. The Hague: Mouton.

Leer, Jeff. 1985. Prosody in Alutiiq. In Michael Krauss (ed.), Yupik Eskimo prosodic systems: descriptive and comparative studies, 77–134. Fairbanks, AK: Alaska Native Language Center, University of Alaska.

Lehiste, Ilse. 1972. Timing of utterances and linguistic boundaries. Journal of the Acoustical Society of America 51. 2108–2124.

Lunden, Anya & Nick Kalivoda. 2013. Right edge lapses, stress cues, and final lengthening. Poster presented at the 21st Manchester Phonology Meeting.

Magri, Giorgio. 2012. Convergence of error-driven ranking algorithms. Phonology 29. 213–269.

Martínez-Paricio, Violeta. 2013. An exploration of minimal and maximal metrical feet: University of Tromsø dissertation.

Martínez-Paricio, Violeta & René Kager. 2014. Non-intervention constraints and the binary-to-ternary rhythmic continuum. Ms.

Miller, Wick. 1996. Sketch of Shoshone, a Uto-Aztecan language. In I. Goddard (ed.), Handbook of American Indian Languages, vol. 17, Washington: Smithsonian Institute.

Miyaoka, Osahito. 2012. A grammar of Central Alaskan Yupik (CAY). New York: Mouton.

Morgan, James L. & Jenny L. Saffran. 1995. Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. Child Development 66. 911–936.

Odden, David. 1996. Phonology and morphology of Kimatuumbi. Oxford: Oxford University Press.

Piantadosi, Steven T., Harry Tily & Edward Gibson. 2011. Word lengths are optimized for efficient communication. In Paul Kay (ed.), Proceedings of the National Academy of Sciences of the United States of America, vol. 108 9, 3526–3529.

Pike, Kenneth L. 1964. Stress trains in Auca. In David Abercrombie, Dennis Fry, P. A. D. McCarthy, N. C. Scott & J. L. M. Trim (eds.), In honour of Daniel Jones, 425–431. London: Longmans.

Prince, Alan S. 1983. Relating to the grid. Linguistic Inquiry 14(1). 19–100.

Sapir, Edward. 1930. Southern Paiute, a Shoshonean language. In Proceedings of the academy of arts and sciences, vol. 65, .

Staubs, Robert. 2014a. Computational modeling of learning biases in stress typology: UMass Amherst dissertation.

Staubs, Robert. 2014b. Learning and the position of primary stress. In Robert E. Santana-LaBarge (ed.), Proceedings of the 31st West Coast Conference on Formal Linguistics, 428–437.

van Urk, Coppe. 2013. A typology of clash-tolerating languages. Ms., MIT.

Vaysman, Olga. 2009. Segmental alternations and metrical theory: MIT dissertation.

Wheeler, Deirdre. 1979. A metrical analysis of stress and related processes in southern paiute and tubatulabal. In University of massachusetts working papers in linguistics, vol. 5, 145–175.

Yanovich, Igor. to appear. Handling equivalence classes of Optimality-Theoretic comparative tableaux. Journal of Language Modelling .

Zhang, Jie. 2002. The effects of duration and sonority on contour tone distribution – a typological survey and formal analysis. New York: Routledge.