Effects of probabilistic phonology on the perception of words and nonwords

Claire Moore-Cantwell[a] and Lisa Sanders[b]

[a]Department of Linguistics, Simon Fraser University, Burnaby, BC V5A 1S6 Canada;
[b]Department of Psychological and Brain Sciences, University of Massachusetts, Amherst, MA 01003, USA

ARTICLE HISTORY
Compiled August 29, 2017

ABSTRACT
Although listeners have knowledge about both categorical and probabilistic phonological patterns, it is not clear if knowledge of probabilistic patterns affects speech processing. One such pattern in English is that words ending in /i/ (e.g., rémedy) are more likely to have antepenultimate stress than penultimate stress (e.g., spaghétti). In the current study, participants extended this trend to ratings of novel words (e.g., bakati). Further, ERPs revealed that real words that violate this trend elicit an early negativity 280-380 ms after onset compared to words that observe this trend. These results indicate that probabilistic phonology interacts with lexical access. More specifically, they suggest that extra processing power is needed to recognize the stress pattern of trend-violating words because of competition between the expectations of the phonological grammar and lexical encoding of trend-violating patterns.

KEYWORDS
Speech perception; Stress; Phonology; ERP; Phonotactics

1. Introduction

The processes of perceiving and producing speech are both strongly conditioned by the listener's experience with language. Sounds and sequences of sounds which do not occur in a language may be difficult or impossible for listeners to veridically perceive (Berent, Lennertz, Smolensky, & Vaknin-Nusbaum, 2009; Breen, Kingston, & Sanders, 2013; Dehaene-Lambertz, Dupoux, & Gout, 2000; Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Moreton, 2002). Likewise, when producing novel words, speakers avoid structures which are not part of their language. Speakers strongly avoid structures which are completely absent from their language, but they also avoid structures which are rare. In general, speakers' production of rare patterns on novel forms is directly related to the statistics of the lexicon. More rare structures are produced less often, common structures more often, and structures of intermediate frequency in between (Ernestus & Baayen, 2003; B. Hayes, Zuraw, Siptár, & Londe, 2009; Zuraw, 2000). Statistical tendencies and categorical patterns have been argued to be part of the same system of phonological knowledge that speakers have about their language -

CONTACT C. Moore-Cantwell. Email: claire.moore-cantwell@sfu.ca

the phonological grammar (Coetzee & Pater, 2009; B. Hayes & Wilson, 2008; Zuraw, 2000).

Previous work has shown that categorical phonological patterns constrain both perceptual processing and lexical access (Rossi, Hartmuller, Vignotto, & Obrig, 2013; Rossi et al., 2011). Violations of such categorical patterns elicit a late positivity in ERPs that has been associated with violations of structural expectations (Osterhout & Holcomb, 1992; Patel, Gibson, Ratner, Besson, & Holcomb, 1998 et. seq.). In this paper we examine the effects on processing of a probabilistic phonological pattern in English. Because the pattern has real-word exceptions, we can, for the first time, address the interaction of grammatical and lexical knowledge in processing.

Categorical patterns can so strongly constrain processing that ill-formed structures are perceptually 'repaired'. Moreton (2002) and Breen et al. (2013) found that English-speaking listeners report hearing the ill-formed onset clusters 'dl-' and 'tl-' as well-formed 'gl-' and 'kl-'. Berent et al. (2009) found that English-speaking listeners repair 'lb-' and 'gb-' to 'ləb-' and 'gəb-'. In Japanese, which prohibits sequences of consonants[1], Dehaene-Lambertz et al. (2000); Dupoux et al. (1999) found that sequences like 'igmo' are perceptually repaired to 'igumo'. Massaro and Cohen (1983) investigated the ill-formed onset 'sr', finding that while listeners did not completely repair the sequence in perception, their categorization of ambiguous sounds was affected. Specifically, listeners were less likely to categorize a sound ambiguous between 'r' and 'l' as 'r' when it occurred after 's', than when it occurred after 'p', where both are allowed.

Breen et al. (2013) and Dehaene-Lambertz et al. (2000) investigated the timing of perceptual repair processes using ERPs. Dehaene-Lambertz et al. found a mismatch-negativity (MMN) to 'igumo' after several presentations of 'igmo' in French listeners, but not in Japanese listeners. Based on this lack of MMN for Japanese listeners, they argue that the perceptual repair process takes place very early. Rossi et al. (2011) also argue that phonological patterns affect early processing based on finding a smaller N400 in response to phonologically ill-formed pseudowords compared to phonologically well-formed pseudowords. The authors interpret this effect to mean that phonological form can be used to stop lexical access indexed by the N400 even before it begins. However, others argue that lexical access begins well before the N400 (e.g. Kretzschmar, Schlesewsky, & Staub, in press; Rayner, 1998; Reichle, Pollatsek, Fisher, & Rayner, 1998). It is important to note that even when repair occurs early in processing, some information about ill-formed sequences can be maintained. Breen et al. showed that processing of well-formed sequences (e.g., 'gl-') is influenced by presentation of repaired sounds (e.g., 'dl-' repaired to 'gl-' vs. actual 'gl-') more than a second earlier.

Not all categorical phonological patterns result in perceptual repair. Ill-formed, unrepaired sequences elicit a late positivity rather than earlier ERP effects. In German, words cannot have two identical consonants where the first is in a cluster (e.g. 'schmam' and 'plall') even though the onsets and codas are attested in other contexts(e.g., 'schmal' and 'platt'). Listeners can accurately report the ill-formed nonwords. However, they elicit a positivity over posterior regions that peaks around 1200 ms after word onset. Similarly, Pitkanen (2010) found that in Finnish speakers, unrepaired nonwords that violate Finnish vowel harmony patterns likewise elicit a posterior positivity, though in this study the effect peaked around 600 ms after stimulus onset.

In this paper, we examine the effects on processing of a probabilistic pattern, with real-word exceptions. Using a pattern of this type allows us to investigate the in-

[1] whenever the first consonant is not nasal

teraction of lexical and grammatical knowledge. Specifically, we can investigate the processing of real words which are phonologically ill-formed. We do not expect probabilistic phonological generalizations to induce perceptual repairs, since listeners must be able to accurately perceive exceptions. For the same reason, we do not expect probabilistic phonology to constrain the lexical access process, and therefore we do not expect it to modulate the size of the N400 in the way that Rossi et al. find. However, we do expect that probabilistic phonological information would be available early in processing, since categorical phonological information is.

Böcker, Bastiaansen, Vroomen, Brunia, and Gelder (1999) examined the processing of exceptional stress in Dutch, and found a larger negativity around 325 ms after word onset for phonologically ill-formed vs. phonologically well-formed stress patterns. In Dutch, initial stress prevails on two syllable words (88% of two syllable words have initial stress), so Böcker et al. compared the perception of initial and final stress on two-syllable words. However, the different stress patterns on the stimuli correspond to important physical differences such as loudness, duration, and pitch, which could also influence processing in the 325 ms range. This confound of stress pattern and phonological well-formedness, together with poor data quality (requiring a 6 Hz low-pass filter) suggest that further investigation is in order.

The ideal phonological pattern to test, then, would be one in which acoustic differences between stimuli are largely uncorrelated with phonological well-formedness. We examine a pattern within the English stress system. In longer words of English (three syllables or longer), the well-formedness of a stress pattern is modulated by the word's final vowel. For example, words ending in the vowel [i] tend to take antepenultimate stress, as in *cánnery, récipe, énemy.* Words with a final [i] and penultimate stress, like *spaghétti, bikíni* are exceptions. However, words ending in a schwa exhibit a trend in the opposite direction: words like *banána, vanílla* are common, and words like *cínema, táffeta* are exceptions. Neither the final vowel nor the stress pattern of the word is enough by itself to determine the word's phonological well-formedness; it is the combination that determines well-formedness.

The most general, and well-known, generalization about stress in English is that words tend to have stress on the first syllable (Cutler & Carter, 1987). This tendency is a property of short words of English only - one and two syllable words, which constitute 65% of the lexicon. In longer words of English, other patterns prevail[2]. Stress tends to be penultimate when the penultimate syllable is heavy ('eléctive', 'aróma'), and can be antepenultimate ('cínnamon', 'cárdamom') when the penultimate syllable is light (Chomsky & Halle, 1968; Domahs, Plag, & Carroll, 2014; Moore-Cantwell, 2016). Words with light penultimate syllables vary from word to word, taking either antepenultimate stress or penultimate stress (Pater, 2000). It is within this zone of variation that a word's final vowel has an effect.

We break up the effect of the final vowel into two phonological trends, a strong trend and a weak trend. Both were discovered through a search of the CMU pronouncing dictionary (Weide, 1994) and the SUBTLEX$_{\mathrm{US}}$ corpus (Brysbaert & New, 2009). Among words that are long enough to take antepenultimate stress (i.e., at least three syllables long), and are monomorphemic[3] and with all light syllables (to avoid other influences on stress pattern), 77% (161/208) that end in /i/ have antepenultimate stress and 70% (129/184) that end in /ə/ have penultimate stress. In addition

[2]Consider long words such as 'àbracadábra', or 'apòtheósis'. Words of this length can have a secondary stress on the first syllable, but never a primary stress. 'ábracadàbra' would not be a possible English word.

[3]Many affixes in English affect stress placement - for example, the affix '-ity' requires that stress be antepenultimate, as can be seen in the pair *sólemn* ∼ *solémnity*

to being numerically stronger among this very restricted set of words, the trend for words ending in /i/ to take antepenultimate stress holds across a larger set of almost 2000 words when morphological structure and syllable weight are ignored. In contrast, the trend for /ə/-final words to have penultimate stress only exists over the very restrictive subset of fewer than 200 words; it disappears when derivational morphology and syllable weight are ignored.

We examine the effects of these two complex trends on English speakers' processing of words and nonwords. We use a rating task to assess speakers' knowledge of the two trends, testing whether either of them are extended by speakers to novel words. Previous work on speakers' ability to extend trends in their lexicon to novel words (Ernestus & Baayen, 2003; B. Hayes et al., 2009) suggests that even complex statistical trends can be extended to pseudowords, and we expect similar effects here. Since the trends differ in strength, we can also investigate the effect of trend strength on speakers' ability to extend the trend to novel words.

We use event related potentials (ERPs) to examine the effects of these two trends on processing. We do not expect to find perceptual repair effects, since real word exceptions exist and must be correctly perceived (spaghétti, cínema). We do expect to see an effect late in processing of these phonological trends, in the form of a late positivity for novel words that violate either trend. Early in processing, we examine the interaction of lexical processing with the effects of the phonological grammar. We entertain three hypotheses: (H1) These phonological trends do not have an effect early in processing, (H2) the trends affect processing in a way that is independent from the lexical access process, and (H3) the trends interact with the lexical access process.

H1 predicts that well-formed and ill-formed nonwords should not differ early in processing, nor should well-formed and ill-formed actual words. If H2 is correct, we should see an effect of phonological well-formedness that is similar in real and novel words. We also expect real and novel words to differ in their processing, but in a way that is independent of the difference between well-forms and ill-formed items. H3 predicts that phonological well-formedness should affect the processing of real words differently from how it affects the processing of novel words.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Participants were 22 (14 females, 8 males) native speakers of English, between the ages of 18 and 30 years. All were right-handed undergraduate or graduate students at the University of Massachusetts Amherst with normal hearing, normal or corrected to normal vision, and no known neurological conditions. EEG data from one participant (female) were excluded due to artifacts caused by muscle tension and frequent blinks. Behavioral data from all 22 participants were included in analysis.

2.1.2. Stimuli

We selected 10 English words that observe the stronger trend (/i/-final antepenultimate stress), that violate the stronger trend (/i/-final penultimate stress), that observe the weaker trend (/ə/-final penultimate stress), and that violate the weaker trend (/ə/-final antepenultimate stress), shown in Table 1. All had three light syllables (monophthong vowels and no coda consonants).

In addition to the 40 real English words, we constructed 80 nonwords that paralleled the structure and stress patterns of the words. Specifically, nonwords were constructed by changing each consonant in a real word to another in the same sonority class (stops, fricatives, nasals, and liquids). For each real word, two nonwords were constructed, one with antepenultimate and one with penultimate stress (e.g. áraspa and aráspa for the word 'alaska'). Neighborhood density was calculated for the resulting nonwords using the Generalized Neighborhood Model[4] (Bailey & Hahn, 2001). Neighborhood density was similarly low for all nonwords (M = 0.0005); a neighborhood density of 1 indicates a single neighbor.

Also included in the experiment were mis-stressed versions of the real English words, e.g. bánana. These were included to balance the item set so that there were in total 80 real words and 80 nonwords, and also so that every item (real or nonword) would have a counterpart with the opposite stress pattern. In order to create these items, stress was moved to the antepenultimate syllable for penultimately stressed real words, and to the penultimate syllable for antepenultimately stressed real words.

All stimuli were transcribed into the international phonetic alphabet (IPA). A native English speaker pronounced each item in the frame Say X again. The stimuli were then isolated from the frame sentence. To ensure that the acoustic cues to stress were similar across words nonwords, we measured duration, intensity, and F0 excursion (Table 2). Along with vowel quality these are considered to capture stress cues in English (Beckman, 1986). Stimuli averaged 487 ms in length (307-652 ms, sd = 71) with the onset of the second syllable at 177 ms (70-339 ms) and the onset of the third at 347 ms (223-500 ms).

Table 1. English words used in both Experiment 1 and Experiment 2. Log frequencies from $\text{SUBTLEX}_{\text{US}}$ (Brysbaert & New, 2009) are given at the bottom.

	-[i]		-[ə]	
	Antepenultimate	Penultimate	Antepenultimate	Penultimate
	well-formed	**ill-formed**	**ill-formed**	**well-formed**
	elegy	salami	cinema	dilemma
	atrophy	finale	africa	eureka
	canopy	graffiti	nebula	lasagne
	ebony	jalopy	replica	manila
	fallacy	pastrami	retina	sierra
	cavity	bikini	spatula	sonata
	colony	safari	stamina	vanilla
	recipe	swahili	swastika	savanna
	remedy	spaghetti	taffeta	militia
	travesty	zucchini	canada	alaska
mean (sd) log(f)	*1.65(0.53)*	*1.76(0.32)*	*1.78(0.48)*	*1.72(0.46)*

2.1.3. Procedure

On each trial participants heard one of the items and were asked to rate how likely it was to be an actual word of English on a 4-point scale (1 = very unlikely, 4 = very

[4]Thanks to Adam Albright for sharing his implementation of the GNM model.

Table 2. Means and standard deviations (in parentheses) for acoustic measurements of stressed vowels in all items from Experiment 1. Note that durations do not include consonants.

Realized Stress	Parameter	Word	Nonword
Antepenultimate	Duration (ms)	97 (20)	99 (21)
	Intensity (dB)	74 (3)	75 (3)
	F0 excursion (Hz)	6.6 (4.3)	10.4 (13.2)
Penultimate	Duration (ms)	106 (24)	98 (25)
	Intensity (dB)	74 (3)	73 (3)
	F0 excursion (Hz)	17.5 (28.1)	14.9 (21.1)

likely). Listeners were told that they would hear actual and made-up words mixed together, but that some of the real words were so rare that they would not have heard them before. Two examples of very rare actual words were given: 'gálea', and 'tomálley' (both trend-violating). Participants were told that when they didnt recognize an item, they should use how it sounds to guess how likely it was to be an English word.

Participants started each trial by pressing a button on the response box at which time a fixation cross appeared on a computer monitor. Participants were asked to remain still, to avoid eye movements and blinking, and to withhold responses while the fixation cross was visible. One of the stimuli was presented over loudspeakers at an average of 65 dB SPL 100 to 1000 ms (rectangular distribution) after the onset of the fixation cross. The fixation cross was replaced with a response screen 1500 ms after the onset of the stimulus.

The first two participants heard each word, mis-stressed word, and nonword 10 times each (1600 trials). To shorten the experimental session time, the subsequent 20 participants heard each item 8 times each (1280 trials). The order of items was randomized. The entire experiment including electrode application, completion of all trials, and short (between trials) and longer (approximately every 200 trials) breaks required about 2.5 hours.

2.1.4. EEG recording and processing

Electroencephalogram (EEG) was recorded continuously by means of a 128-electrode HydroCel Geodesic Sensor Net (Electrical Geodesics Inc., Eugene, OR), at a sampling rate of 250Hz and a bandwidth of 0.01-100 Hz. For recording, the reference electrode was placed at the top of the head (Cz). The impedance for each electrode was maintained below 50kΩ throughout the experiment. A 60 Hz notch filter was applied to the EEG signal offline after recording. The continuous EEG was divided into 1600ms epochs, timelocked to (a) the beginning of the stimulus, (b) the onset of the second syllable, and (c) the onset of the third syllable. Relative to each onset, the epochs began 100ms before that onset, and continued 1500ms after. Because all effects of interest appeared timelocked to the first syllable, we do not discuss data from epochs timelocked to the second and third syllables.

Trials containing eye blinks, eye movements, or other artifacts were excluded automatically using ERPLAB. Eye blinks were detected by subtracting voltages below each eye from voltages above each eye, and eye movements were detected by subtracting voltage to the left of the eyes from voltage to the right of the eyes. Subject-specific voltage thresholds were chosen for eye blink/movement channels (between 300 and 500 μV) as well as for the whole head (between 400 and 500 μV). Any trial where the threshold was exceeded was marked as artifactual. In all conditions, between 22% and

27% of trials contained artifacts.

Data from the remaining trials were averaged separately for each participant, condition, and electrode. ERPs were then re-referenced to the average of the two mastoid recordings and baseline corrected to the average amplitude in the 100 ms before stimulus onset.

2.1.5. Analysis

A within-subjects ANOVA was conducted with mean ratings from individual subjects as the dependent variable[5]. Factors were WordStatus (Word, Nonword), FinalV (-[ə], -[i]), and Stress (antepenultimate, penultimate).

Visual inspection of grand-average ERP waveforms motivated measuring mean amplitude 280-380 ms (early difference), 400-1000 ms (N400), and 1000-1500 ms (Late Positive Component, LPC) after stimulus onset. The time windows for the N400 and LPC measurements are somewhat later than what is typically used, but appropriate for long words in which the phonological pattern information is distributed across the entire duration.

Electrode position was included in analysis by averaging data across 9 electrodes in 12 regions from anterior to posterior (Anteriority) and from left to right (LeftRight) as shown in Figure 1. These data were then subjected to repeated-measures ANOVAs using factors WordStatus (Word, Nonword), FinalV (-[ə], -[i]), Stress (antepenultimate, penultimate), Anteriority of electrode position (anterior, anterior-central, posterior-central, posterior), and LeftRight electrode position (left, medial, right). In all three time windows, an ANOVA was computed using the above five factors and their interactions. Greenhouse-Geisser corrected degrees-of-freedom were used to account for violations of sphericity; corrected p-values and uncorrected degrees-of-freedom are reported. In some cases further analyses were conducted, using Bonferroni corrections to correct for family-wise error.

2.2. Results

2.2.1. Ratings

As shown in Figure 3, ratings (1 = very unlikely to be an English word, 4 = very likely to be an English word) were higher for actual words than for nonwords (WordStatus: $F(1,21) = 1194.6$, $p < .001$). Ratings for actual words were nearly always 4 ($M = 3.5$ 4) with the exceptions of “jalopy” ($M = 2.9$) and “taffeta” ($M = 2.7$). Since subjects rated these items well outside the range of averages for the rest of the real words, they are excluded from analysis of the behavioral data.

Mis-stressed real words (like bánana) received ratings in between those of real words and those of nonwords, but there was a great deal of variability, both across participants and from item to item. Figure 2 illustrates this: real words like ‘banána’ received high ratings, clustered very strongly at the high end of the scale. Nonwords received low ratings, with means roughly normally distributed around 1.8. Mis-stressed real words on the other hand, received mean ratings distributed across the scale, but with a mean (2.8) intermediate between real words and nonwords. Because the mis-stressed words exhibit so much variability, and because they do not pattern with either real words or nonwords, we exclude them from further analyses.

[5]These mean ratings were roughly normally distributed for the nonwords, but not for real words, whose ratings were clustered at the high end of the scale

Figure 1. Data from 108 electrodes (black circles) were included in analyses by averaging across 9 electrodes in 12 regions. These regions were described as two electrode position factors: Anterior to Posterior (4 levels) and Left to Right (3 levels). These same regions of interest were used for both experiments.

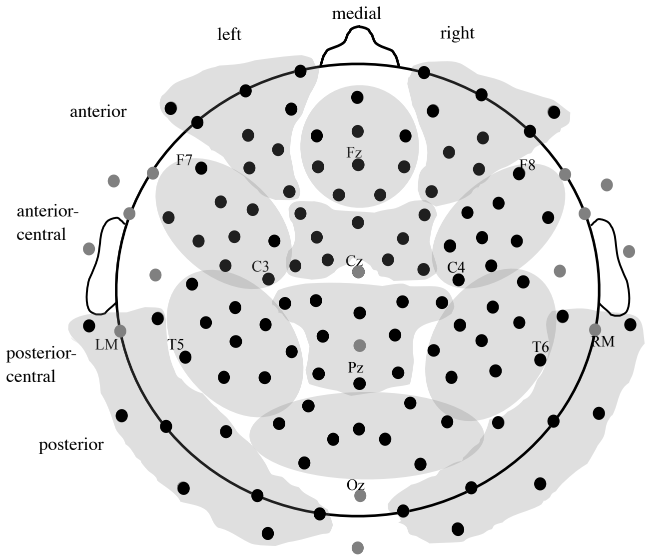


Figure 2. Distribution of mean ratings for the three major types of items. Each number in the histogram is the mean for one participant in one condition. For example, one point is participant 1's mean rating for penultimately stressed i-final nonwords.

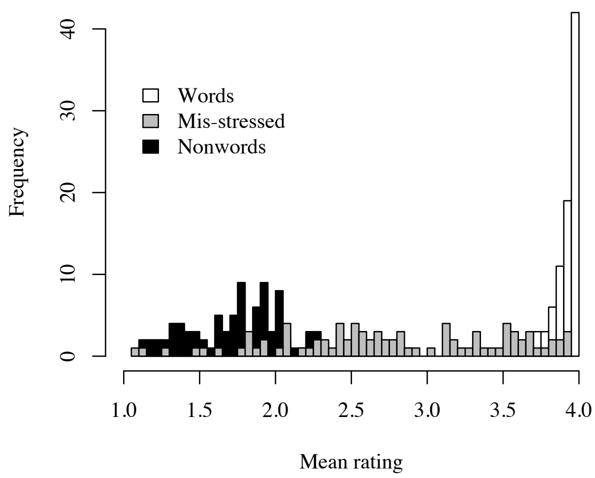
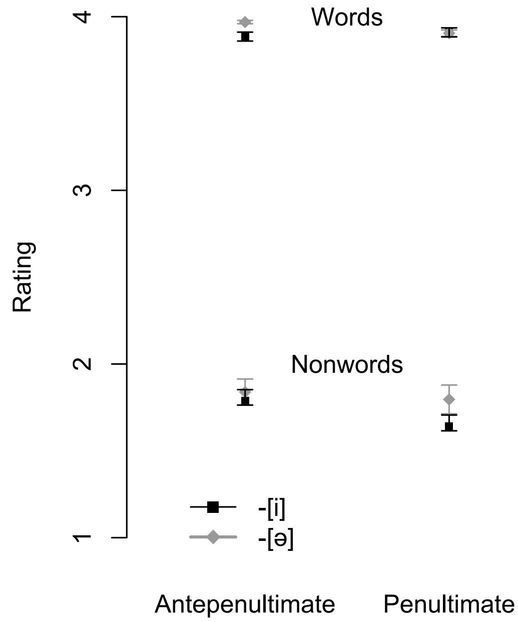


Figure 3. Mean ratings for items in Experiment 1. Error bars represent standard errors based on the distribution of individual subject means.



An omnibus ANOVA was conducted on the mean ratings from each participant (Table 3). In addition to a main effect of Word Status, there were main effects of final vowel and stress, as well as a three-way interaction between word status, final vowel, and stress. In ANOVAs conducted on words only, and on nonwords only, the interaction between final vowel and stress both reached the Bonferroni-corrected significance value of 0.025 (Words: $F(1,21)=7.7$, $p< 0.025$; Nonwords: $F(1,21)=8.1$, $p<0.025$). No main effects of final vowel or stress were found in either words or nonwords.

Ratings are graphed in Figure 3. Within antepenultimate words, ə-final items were rated a small amount (0.08) higher than i-final items, but no difference obtains in penultimately stressed items. Within nonwords, no difference obtains in antepenultimately stressed nonwords, but penultimately stressed ə-final nonwords were rated higher (by 0.15) than penultimately stressed i-final nonwords.

2.2.2. ERP data

Early effects. Results of an omnibus ANOVA analyzing the mean voltage in the 280-380 ms time window are shown in Table 4. There were no main effects of word type, final vowel, or stress, but there was a significant interaction between final vowel and stress, as well as a three-way interaction between word type, final vowel, and stress. Two ANOVAs were conducted on real words and nonwords separately. For real words, the finalV x stress interaction remained significant: $F(1,20) = 3.2$, $p<0.001$. However, for nonwords no significant interaction obtained: $F(1,20) = 0.07$, $p=0.80$. From this we conclude that the wordStatus x finalV x stress interaction in the omnibus ANOVA indicates an interaction of finalV x stress for real words only. Figure 4 shows this

Table 3. Results of an omnibus ANOVA for behavioral data from Experiment 1.

	F value
wordStatus (1,21)	1194.6^{**}
finalV (1,21)	6.4^{*}
stress (1,21)	6.3^{*}
wordStatus x finalV (1,21)	-
wordStatus x stress (1,21)	$3.5^{\dagger}$
finalV x stress (1,21)	-
wordStatus x finalV x stress (1,21)	16.0^{**}

† 0.1>p>0.05; * 0.05>p>0.01; **p<0.01

Table 4. ANOVA results for all three EEG time windows from Experiment 1. Greenhouse-geisser corrected p-values and uncorrected degrees of freedom are reported.

Effect	280-380 ms	400-1000 ms	1000-1500 ms
wordStatus (1,20)	-	44.3^{**}	$3.6^{\dagger}$
finalV (1,20)	-	-	-
stress (1,20)	-	-	-
wordStatus x finalV (1,20)	-	4.9^{*}	9.4^{**}
wordStatus x stress (1,20)	-	-	-
finalV x stress (1,20)	18.0^{**}	17.7^{**}	-
wordStatus x finalV x stress (1,20)	12.2^{**}	$3.2^{\dagger}$	$3.1^{\dagger}$
LeftRight x wordStatus (2,40)	$2.6^{\dagger}$	33.5^{**}	4.5^{*}
Anteriority x wordStatus (3,60)	-	9.2^{**}	14.5^{**}
LeftRight x finalV (2,40)	-	-	-
Anteriority x finalV (3,60)	-	-	$3.2^{\dagger}$
LeftRight x Anteriority x wordStatus (6,120)	$1.9^{\dagger}$	3.5^{*}	5.9^{**}
LeftRight x wordStatus x finalV (2,40)	-	$2.6^{\dagger}$	-
Anteriority x stress x finalV (3,60)	10.6^{**}	12.6^{**}	6.8^{**}
Anteriority x LeftRight x stress x finalV (3,60)	$2.1^{\dagger}$	-	-
LeftRight x wordStatus x stress x finalV (2,40)	$3.3^{\dagger}$	-	$2.8^{\dagger}$

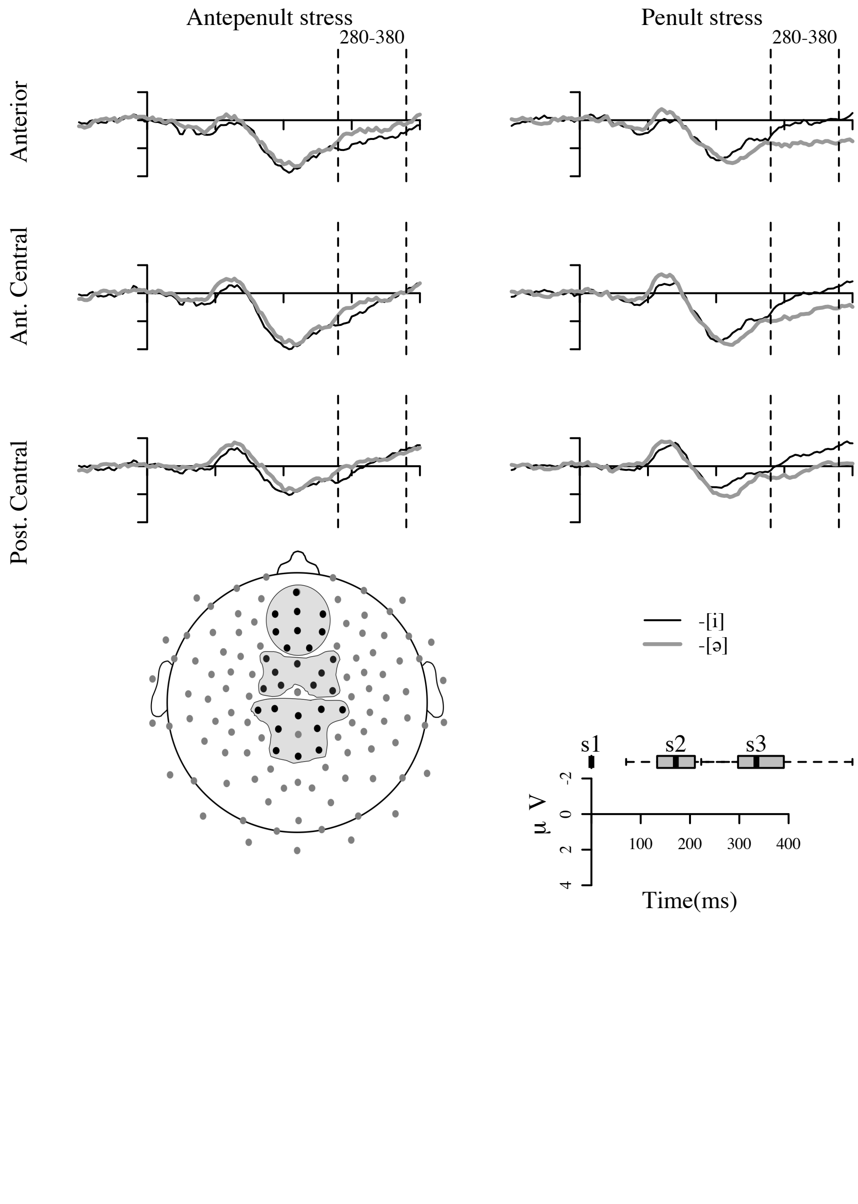
† 0.1>p>0.05; * 0.05>p>0.01; **p<0.01

interaction in more detail: ERPs for Antepenultimate and Penultimate stressed items are shown separately. In both cases, words with trend violating final vowels elicited a larger negativity than words with trend-observing final vowels. For antepenultimately stressed words, those with a final -əelicited a greater negativity than those with a final -i. For penultimately stressed words, those with a final -i elicited a larger negativity than those with a final -ə. The effect was larger for penultimately stressed items (Δ_μ=0.87μV) than for antepenultimately stressed items (Δ_μ=0.33μV), but both differences were statistically significant (Antepenultimate: t=2.06, p<0.05; Penultimate: t=4.78, p<0.01).

The omnibus ANOVA also showed a significant three-way interaction between Anteriority x FinalV x Stress. Separate ANOVAs conducted on each of the four levels of Anteriority revealed that the interaction between FinalV x Stress remained significant in all but the most posterior electrodes (α = 0.05/4 = 0.0125; Anterior electrodes: F(1,20)=32.9, p<0.001; Anterior-central electrodes: F(1,20)=14.5, p<0.0125; Posterior-central electrodes: F(1,20)=12.4, p<0.001; Posterior electrodes: F(1,20)=4.0, p=0.06). No significant interactions with LeftRight were found in the omnibus ANOVA, indicating that the effect is not lateralized.

N400. Results of an omnibus ANOVA analyzing the mean voltage in the 400-1000 ms time window are shown in Table 4. There was a main effect of word type - nonwords

Figure 4. ERPs elicited by i-final and ə-final real words with antepenultimate (left) and penultimate (right) stress in Experiment 1. Only medial non-posterior regions are shown. Onset times for each syllable are represented in the legend as boxplots. In the 280-380 ms time window, both types of trend-violating words elicited greater negativity than their trend-observing counterparts. Antepenultimately stressed ə-final words violate the weak trend, and penultimately stressed i-final words violate the strong trend.



elicited a larger N400 effect than words (Figure 5). Additionally, there were significant interactions between word type and final vowel, and between stress and final vowel, and a marginal but not significant three way interaction between word type, final vowel, and stress. The interaction of stress with final vowel is illustrated in Figure 6. Motivated by the significant interactions of left-right and anteriority with word type, as well as the interaction of Anteriority x finalV x stress, separate ANOVAs were conducted on each of the 12 regions of interest (Figure 1). Using a Bonferroni-corrected significance threshold of of 0.0042 to correct for family-wise error, we found that word type was significant in all regions of interest except for the two posterior non-medial sites (Posterior Left and Posterior Right). Additionally, the interaction of final vowel with stress was significant at anterior electrode sites, as well as the anterior-medial central region. These regions are illustrated in Figure 6.

Two separate ANOVAs were conducted on real words and nonwords using a significance threshold of 0.025. Final vowel did not reach significance in either model (Words: $F(1,20)=4.0$, $p=0.06$; Nonwords: $F(1,20)=0.42$, $p=0.05$). The interaction between final vowel and stress, as well as the three way interaction between anteriority x finalV x stress, was significant for words ($F(1,20)=15.2$, $p<0.001$; $F(3,60)=4.9$, $p<0.01$). Neither was significant for nonwords ($F(1,20)=0.95$, $p=0.34$; $F(3,60)=3.28$, $p=0.06$).

Late Positive Component (LPC). Results of an omnibus ANOVA analyzing the mean voltage in the 1000-1500 ms time window are shown in Table 4. There was a significant interaction of wordStatus and final vowel, as well as significant interactions LeftRight x wordStatus, and Anteriority x wordStatus, as well as a significant three-way interaction between LeftRight x Anteriority x wordStatus. No significant effect of final vowel, stress, or their interaction was found, but there was a significant three-way interaction between Anteriority x stress x finalV.

Motivated by the significant interactions of left-right and anteriority with word type, as well as the interaction of Anteriority x finalV x stress, separate ANOVAs were conducted on each of the 9 regions of interest (Figure 1). Using a Bonferroni-corrected significance threshold of of 0.0042 to correct for family-wise error, we found that word type was significant in just three regions, which were not continuous: Anterior left ($F(1,20)=10.7$, $p<0.004$), Anterior right ($F(1,20)=27.1$, $p<0.0001$), and Anterior-central right ($F(1,20)=13.9$, $p<0.004$). In all three regions, words were more positive than nonwords. Additionally, the stress x finalV interaction was significant in just the Anterior middle region ($F(1,20)=12.3$, $p<0.004$). All of these effects are seen over very small numbers of electrodes, and are entirely anterior - a distribution not typical of P600 effects. Additonally, while previous research (Domahs, Kehrein, Kraus, Wiese, & Schlesewsky, 2009; Pitkanen, 2010) shows a late positivity for phonologically ill-formed items, we do not expect a positivity for words compared to nonwords. We do not discuss these effects further.

Motivated by the significant interaction between word status and final vowel, two separate ANOVAs were conducted on words and on nonwords. However, final vowel did not reach significance in either words or nonwords.

Summary. In Experiment 1, participants distinguish in their ratings between i-final and ə-final penultimately stressed nonwords, preferring ə-final. This indicates that they are able to productively apply the strong trend found in the lexicon of English to novel words. Recall that this trend was for i-final words to take antepenultimate stress rather than penultimate stress - in our lexicon of English words, 77% of i-final words three syllables long and longer took antepenultimate stress. On the other hand, the weak trend for ə-final words to take penultimate stress was not reflected in participants' ratings of nonwords. Recall that this trend was weaker in that it holds only over

Figure 5. ERPs elicited by words and nonwords in Experiment 1. Words elicited a larger negativity at 400-1000 ms than nonwords in all scalp regions except for Posterior Right and Posterior Left.

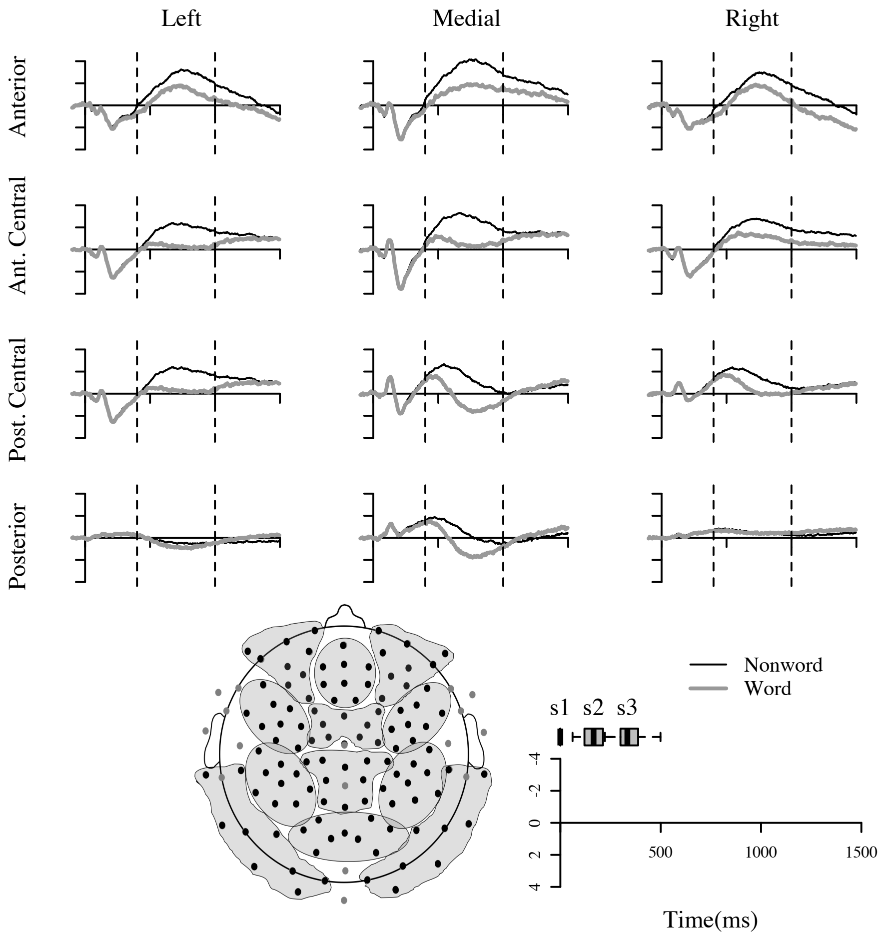
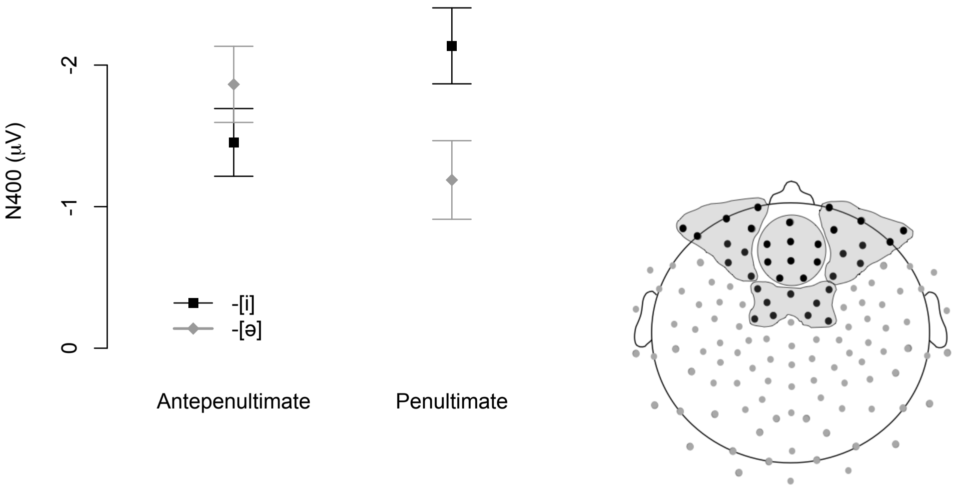


Figure 6. Mean amplitude in the N400 time window (400-1000 ms) from Experiment 1, averaged over all regions for which the stress x finalV interaction is significant - anterior and medial anterior-central regions, shown on the right. Words and nonwords are averaged together.



monomorphemic words with all light syllables, rather than over all three-syllable and longer words like the strong i-final trend.

In the ERP data, we found three effects of note. First, in the earliest time window, 280-380 ms, we found a significant interaction of stress and final vowel in words only, indicating that phonologically ill-formed real words (Penultimate i-final, Antepenultimate ə-final) elicited a larger negativity in this region than phonologically well-formed real words. Importantly, the difference betwee ill-formed and well-formed words was observed within a stress type, and cannot be attributed to a difference in stress pattern. Rather, the distinction is whether that stress pattern is well- or ill-formed given the conditioning factor of the final vowel.

Second, we found a larger N400 on nonwords compared to real words, replicating many earlier results (Rossi et al., 2013, 2011; Rugg, 1984). Finally, we found an interaction of stress and final vowel on the size of the N400, significant across both words and nonwords. Within penultimately stressed items, i-final ones elicited a larger N400 than ə-final items. Notably, we found no effects of phonological well-formedness on novel words, even though participants distinguished phonologically well- and ill-formed novel words in their ratings.

3. Experiment 2

Experiment 1 found an effect on participants' ratings of novel words based on whether or not they followed the strong trend in the lexicon for i-final words to take antepenultimate stress. However, no effects were found in the ERP measures of participants' apparent knowledge of this trend. In particular, we found no late positivity for trend-violators relative to trend-observers. A late positivity was expected, based on previous ERP work studying speakers' knowledge of the phonology of their language. One possibility is that violations of probabilistic phonology are processed differently than violations of categorical phonology, and do not elicit the same late positivity. However, in experiment 1, each nonword was repeated several times (twice per block, eight times in total). Because of this, participants may not have processed the nonwords like actual nonwords. In particular, they may have experienced priming or adaptation across repeated instances of the same nonword, which could lead to a reduction in size of the late positivity, possibly rendering it undetectable.

Experiment 2 is designed to rule out this possibility. Instead of re-using the same nonwords in each block we introduce new nonwords so that participants only hear each nonword once. If the lack of an LPC that we observed in Experiment 1 is due to the repetiton of the nonwords, then we expect an LPC to emerge in Experiment 2. However, if no LPC obtains in Experiment 2, then we can conclude that the strong i-final trend, while it is productively applied to nonwords, does not elicit the same LPC as categorical phonological patterns when violated.

3.1. Methods

3.1.1. Participants

Participants were 25 (12 females, 13 males) native speakers of English, between the ages of 18 and 30 years. Inclusion criteria were identical to those for Experiment 1 (right-handed, normal hearing, normal or corrected to normal vision, no known neurological conditions) with the addition that participants had not taken part in Experiment 1. EEG data from six participants (2 females, 4 males) were excluded due to artifacts caused by frequent blinks ($n = 3$), muscle tension that could not be resolved ($n = 1$), and low-frequency drift related to skin potentials in a too-warm room ($n = 2$). Behavioral data from all 25 participants were included in analysis.

3.1.2. Procedure

The 40 real words, their mis-stressed counterparts, and 80 nonwords from Experiment 1 were also used in Experiment 2. However, 560 new nonwords were constructed so that each nonword was only heard once during the experiment. Nonwords were counterbalanced for final vowel, so that each item appeared in all four stress by vowel conditions. Each participant saw each nonword in two stress conditions with the same vowel, but across participants each item was presented with both vowels (e.g., one participant heard /bámaki/ and /bamáki/, and another heard /bámaka/ and /bamáka/.

Lexical neighborhood density was calculated for all nonwords using the Generalized Neighborhood Model (Bailey & Hahn, 2001). Neighborhood density was similarly low for all nonwords ($M = 0.0006$); a neighborhood density of 1 indicates a single neighbor. Additionally, the Irvine Phonotactic Online Dictionary (Vaden & Hickok, n.d.) was used to calculate phoneme bigram probability ($M = 0.0034$), phoneme trigram probability ($M = 0.00022$), and phoneme positional probability ($M = 0.055$) all of which were similarly low for each prosodic form.

Items were recorded with the same procedure as in Experiment 1. Table 5 shows that the stressed syllables of all word types were comparable in duration and in F0 excursion. For both antepenultimate and penultimate stressed items, however, the mean intensity was significantly less by an average of 5-6 dB for nonwords than for real or mis-stressed words.

Table 5. Means and standard deviations (in parentheses) for acoustic measurements of stressed vowels in the items for Experiment 2. Note that durations do not include consonants. For some syllables, it was not possible to get an accurate F0 measurement, so these are excluded from the analysis (92 antepenultimately stressed items, 78 penultimately stressed items). Note that the acoustic measures for words differ from those reported for Experiment 1 - this is because these items were re-recorded for Experiment 2 so that all experimental items could be recorded together.

Realized Stress	Parameter	Word	Nonword
Antepenultimate	Duration (ms)	91 (19)	85 (27)
	Intensity (dB)	75 (3)	69 (5)
	F0 excursion (Hz)	6.1 (4.2)	5.0 (10.3)
Penultimate	Duration (ms)	107 (24)	105 (32)
	Intensity (dB)	74 (3)	67 (6)
	F0 excursion (Hz)	17.5 (28.2)	12.4 (37.9)

3.1.3. Procedure and EEG recording

Like in Experiment 1, participants heard 8 repetitions of each word and mis-stressed word. They additionally heard the 80 nonwords from Experiment 1 as well as the 560 new nonwords, for a total of 640 nonword trials and 640 word trials (1280 total)[6]. All other procedures were identical to those of Experiment 1, including measurement time windows and analyses. Stimuli averaged 442 ms in length (266-650 ms) with the onset of the second syllable at 164 ms (34-364 ms) and the onset of the third at 330 ms (171-500 ms).

As before, trials containing eye blinks, eye movements, or other artifacts were excluded automatically using ERPLAB. Participant-specific voltage thresholds were chosen for eye blink/movement channels (between 300 and 700 μV) as well as for the whole head (between 400 and 1000 μV). Any trial where the threshold was exceeded was marked as artifactual. For 18 of the 19 participants included in the EEG analysis, between 4% and 30% of trials were excluded because of artifacts. For one participant, 45% of trials had to be excluded. Six additional participants were excluded from analysis due to very high rates of artifacts.

3.1.4. Analysis

All analyses were conducted as in Experiment 1, using the same time windows for ERP measurement, and the same statistical models.

3.2. Results

3.2.1. Ratings

The pattern of ratings (Figure 7) were very similar to that found in Experiment 1. As in Experiment 1, the words "jalopy" and "taffeta" were excluded from the analysis, since their mean ratings (2.9 and 3.1, respectively) were well outside the range of mean ratings for other words in the experiment (3.4-4.0). Results of an omnibus ANOVA are shown in Table 6. Main effects of word status and stress obtained, as well as interactions between word status x final vowel, word status x stress, and a three-way interaction between word status x final vowel x stress. Motivated by these interactions, ANOVAs were also conducted on just words and on just nonwords, with α=0.025.

[6]Due to a coding error, six participants did hear some of the nonwords exactly twice throughout the course of the experiment.

Table 6. Results of an omnibus ANOVA for behavioral data from Experiment 2.

	F value
wordStatus (1,21)	1038.7^{**}
finalV (1,21)	-
stress (1,21)	7.6^{*}
wordStatus x finalV (1,21)	6.0^{**}
wordStatus x stress (1,21)	18.2^{**}
finalV x stress (1,21)	-
wordStatus x finalV x stress (1,21)	14.8^{**}

† 0.1>p>0.05; * 0.05>p>0.01; **p<0.01

As in Experiment 1, mis-stressed real words exhibited a great deal of variability, and had a mean rating (2.6) in between that of real words (3.9) and that of nonwords (1.7). As in Experiment 1, we do not include these items in our analyses.

Within words, there was a significant interaction of final vowel and stress (F(1,24)=6.8, p<0.025). As in Experiment 1, there was a small difference in ratings between antepenultimately stressed items (ə-final rated 0.12 higher than i-final items). Penultimately stressed items exhibited a much smaller difference (i-final 0.04 higher than ə-final). Within nonwords, there was a main effect of stress (F(1,24)=20.0, p<0.001). Antepenultimately stressed nonwords were rated higher than penultimately stressed nonwords by 0.1. There was also a significant interaction between stress and final vowel (F(1,24)=12.8, p<0.025). Within antepenultimately stressed words, i-final items were rated higher by 0.08, while within penultimately stressed words, i-final items were rated worse by a small margin (0.03). Put differently, ə-final items in both stress conditions were rated very similarly (antepenultimate stress was 0.04 higher than penultimate stress), whereas i-final items were rated differently in the two stress conditions (antepenultimate stress was rated 0.16 higher than penultimate stress).

3.2.2. ERP data

Early effects. Results of an omnibus ANOVA analyzing the mean voltage in the 280-380 ms time window are shown in Table 7. Unlike in Experiment 1, there was a main effect of word status, with nonwords more negative than words by an average of 0.37 μV. There were also significant interactions between left-right, anteriority, and word status. Motivated by these interactions, separate ANOVAs were conducted on each of the 12 regions of interest (Figure 1). Using a Bonferroni-corrected significance threshold of of 0.0042 to correct for family-wise error, we found that word type was significant in medial anterior and anterior-central regions, as well as right anterior and anterior central regions. Word status was not significant in left or posterior-central or posterior regions.

In the omnibus ANOVA, there was also an interaction of word status with final vowel. In two separate ANOVAs run on words and on nonwords, final vowel only reached significance (F(1,18)=9.8, p<0.025) in words, but not in nonwords.

In addition to the above analyses, two planned comparisons were conducted on the effect of final vowel within penultimately stressed words and antepenultimately stressed words. Within penultimately stressed words, i-final items elicited a larger negativity than ə-final items (t=3.5, p<0.001), by a mean of 0.55 μV. Within antepenultimately stressed items, no significant effect of final vowel was found (t=1.2, p<0.25).

N400. Results of an omnibus ANOVA analyzing the mean voltage in the 400-1000

Figure 7. Mean ratings for items in Experiment 2. Error bars represent standard errors based on the distribution of individual subject means.

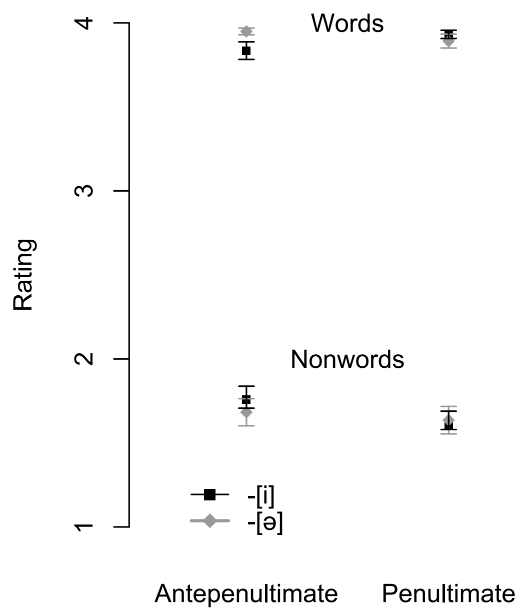


Table 7. ANOVA results for all three EEG time windows from Experiment 2. Greenhouse-geisser corrected p-values and uncorrected degrees of freedom are reported.

Effect	280-380 ms	400-1000 ms	1000-1500 ms
wordStatus (1,18)	12.5^{**}	38.8^{**}	-
finalV (1,18)	-	-	-
stress (1,18)	-	-	-
wordStatus x finalV (1,18)	9.7^{**}	6.2^{*}	-
wordStatus x stress (1,18)	-	-	-
finalV x stress (1,18)	-	-	-
wordStatus x finalV x stress (1,18)	-	-	-
LeftRight x wordStatus (2,36)	9.0^{**}	27.2^{**}	$2.9^{\dagger}$
Anteriority x wordStatus (3,54)	18.9^{**}	8.4^{**}	5.6^{*}
LeftRight x finalV (2,36)	-	$3.4^{\dagger}$	$3.3^{\dagger}$
LeftRight x stress (2,36)	$2.7^{\dagger}$	-	-
LeftRight x Anteriority x wordStatus (6,108)	2.7^{*}	7.5^{**}	-
LeftRight x stress x finalV (2,36)	-	$2.8^{\dagger}$	-
Anteriority x stress x finalV (3,54)	$2.9^{\dagger}$	-	-

† 0.1>p>0.05; * 0.05>p>0.01; **p<0.01

Figure 8. ERPs elicited by i-final and ə-final real words with antepenultimate (left) and penultimate (right) stress in Experiment 2. Only medial non-posterior regions are shown. Onset times for each syllable are represented in the legend as boxplots. In the 280-380 ms time window, i-final words elicited a larger negativity than ə-final words. The difference is larger in penultimately stressed words, and reaches significance in a contrast, while the difference in antepenultimately stressed words does not.

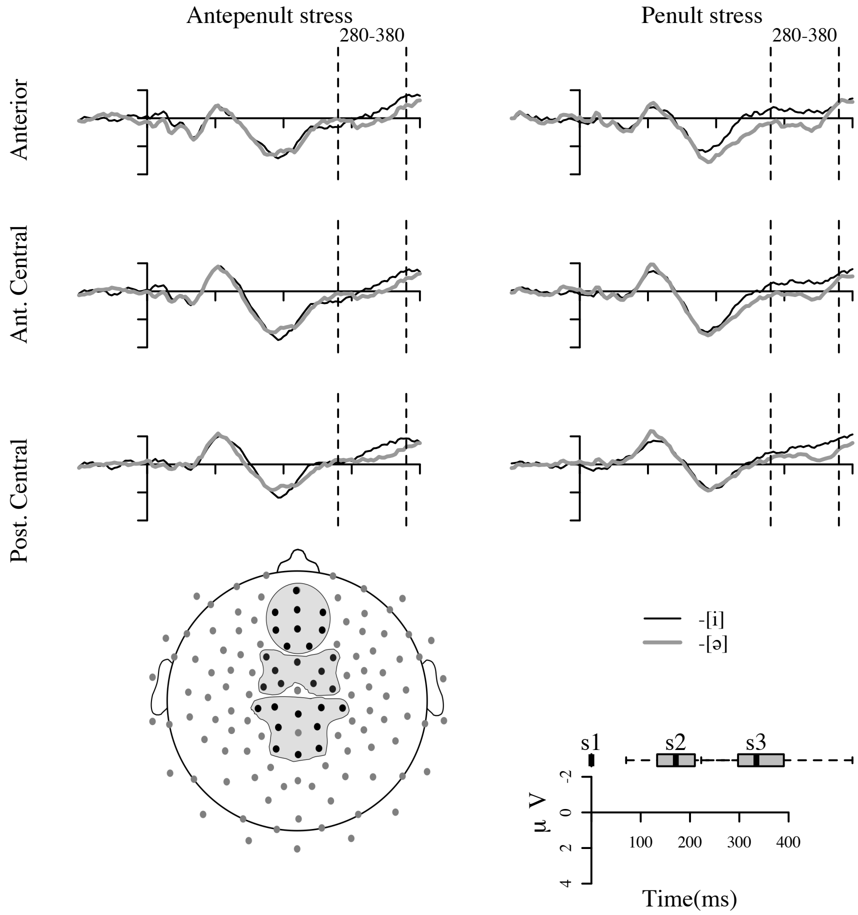
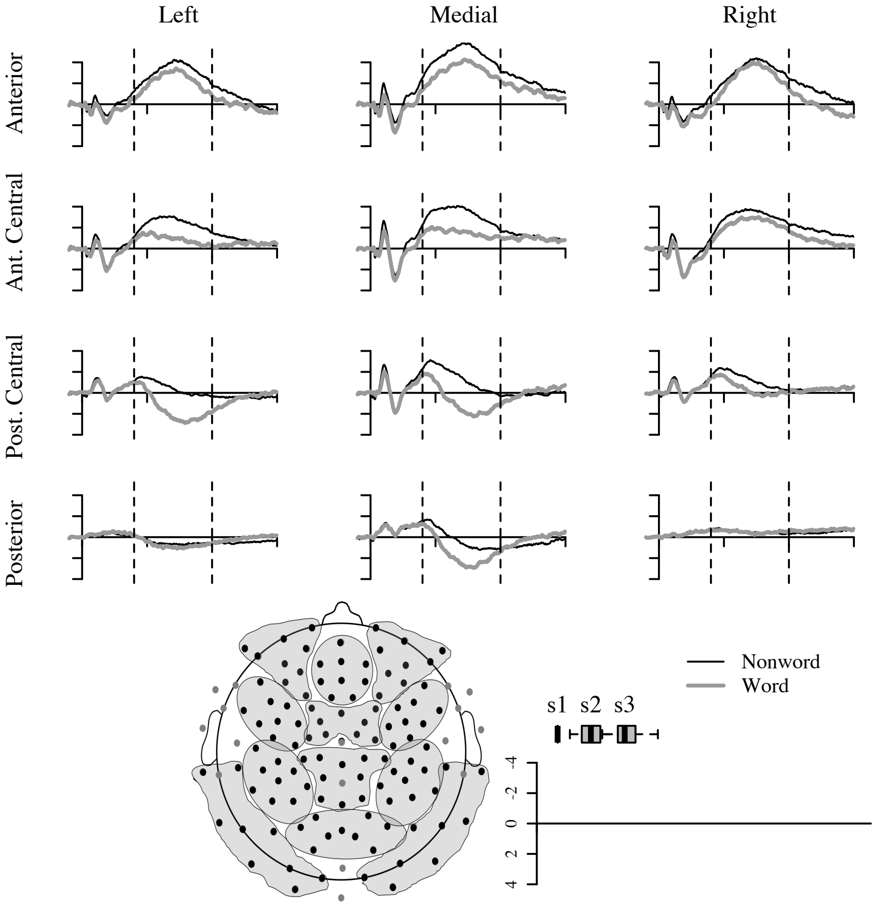


Figure 9. N400s (400-1000ms) for words and nonwords in Experiment 2. Nonwords are significantly more negative than words in all regions except for: Posterior Left, Posterior Right, Anterior Left, Anterior Right, and Anterior-central Right. Onset times for each syllable are represented in the legend as boxplots showing the distribution of onset times.



ms time window are shown in Table 7. As in Experiment 1, there was a main effect of word type - nonwords elicited a larger N400 than words (Figure 9). Additionally, there were significant interactions between word type and final vowel, and significant interactions between left-right and anteriority and word status. As in Experiment 1, separate ANOVAs were conducted on each of the 12 regions of interest (Figure 1). Using a Bonferroni-corrected significance threshold of of 0.0042 to correct for family-wise error, we found that word type was significant in all regions of interest except for the four 'corner' regions of interest, as well as the right anterior-central region (Posterior Left, Posterior Right, Anterior Left, Anterior Right, and Anterior-central Right).

Motivated by the significant interaction between word status and final vowel, two separate ANOVAs were conducted on just words and on just nonwords. No effects of final vowel reached significance in either words or nonwords. Unlike in Experiment 1, no effects final vowel, stress, or their interaction obtained.

Late Positive Component (LPC). Results of an omnibus ANOVA analyzing the mean voltage in the 1000-1500 ms time window are shown in Table 7. There were no significant main effects or interactions between the three experimental manipulations of word status, final vowel, and stress. A significant interaction between anteriority and word status was observed, but in ANOVAs conducted at each level of anteriority, word status never reached significance at α=0.0125.

4. Discussion

4.1. Ratings

These experiments tested the psychological reality and cognitive status of two probabilistic trends in the English lexicon: the tendency for [i]-final words to take antepenultimate stress, and the tendency for [ə]-final words to take penultimate stress. Only the strong i-final trend was 'productive' in the ratings. Nonwords that violated it ([bəmǽki]) were rated as less likely to be an English word than nonword observers ([bǽməki]). The weak ə-final trend was not productive: no rating difference was observed between observers ([fəmákə]) and violators ([fáməkə]). These two trends differed in terms of their degree of observance in the lexicon. The strong trend is observed over a large scope of forms - about 2000 words, including morphologically complex forms and words containing heavy syllables, and of those forms about 90% observe the trend. The weak trend is observed over about 200 words, which are morphologically simple words with no heavy syllables. Of those 200 words, about 70% observe the trend.

All three differences between the two trends (scope, percent observance, morphology and weight as conditioning factors) could contribute to differences in learnability between the two trends. The higher the scope of a trend, the more experience the learner will get with that trend, and the higher the percentage of observers of the trend, the more reliable the trend will be. The extra conditioning factors of syllable weight and morphological structure make the statement of the weak trend more complex than the statement of the strong trend ('i-final words are antepenultimately stressed' vs. 'unsuffixed ə-final words with a light penult are penultimately stressed').

The addition of the 'unsuffixed' condition is particularly problematic. A language learner would not be able to reference the set of unsuffixed things until she had acquired the relevant morphology. In this case, the relevant morphology is derivational affixes and is acquired fairly late. Tyler and Nagy (1989) find that fourth graders have not completely acquired the syntactic properties of various derivational affixes in English, and Jarmulowicz (2002) examined seven and nine year olds' knowledge of the stress-shifting nature of a selection of English derivational suffixes, finding that neither group were as good as adults at correctly judging the stress properties of a particular affix. On the other hand, children can and do begin learning generalizations about the stress pattern of their language as early as 9 months of age (Jusczyk, Cutler, & Redanz, 1993).

4.2. Early negativity

An early negativity in response to real word exceptions to the i-final trend was observed in both experiments. In Experiment 1, a similar early negativity was found for exceptions to the ə-final trend, but this negativity was not observed in the second experiment. We view this failure to replicate as evidence that the negativity observed in the first experiment was a spurious result, and we do not analyze it further. The early negativity observed across both experiments in response to exceptions to the i-final trend shows that this trend is active early during the process of lexical access. This effect is not evident for nonwords, and is early enough that information about the words final vowel is not available. For both of these reasons, this effect cannot be directly related to the perception of the acoustic string that violates the i-final trend. Both the stress pattern and the content of the final vowel are necessary information to tell whether a string is trend-observing or trend-violating. By 280-380ms post-stimulus-onset, subjects have processed acoustic information about the word's stress pattern but not about it's final vowel. This distinction between trend-observers and trend-violators may be a consequence of subjects accessing information in the lexical entry of a word, either phonological information about the final vowel (before they have actually heard it), or direct information about whether the word has a high- or low- probability shape. In nonwords, there is no stored lexical representation, so subjects cannot perceive a violation of a trend before they have heard (and processed) the entire string.

The early negative difference observed in Experiments 1 and 2 may be related to what Böcker et al. (1999) considered the exogenous portion of their N325 effect. In their study, there was some evidence for a negativity in a similar time window in response to weak-strong (12% of Dutch two-syllable words) compared to strong-weak (88%) words for words preceded by two others of the same stress pattern. The effect was far larger for weak-strong words presented after a series of three strong-weak words and when participants were engaged in a stress pattern discrimination task, but these endogenous contributions to the effect may have largely been driven by a mismatch response. In Böcker et al. (1999), it was impossible to determine if the exogenous portion of their effect was related to physical differences in the stimuli (i.e., perhaps the effect would be evident in anyone listening to weak-strong syllables compared to strong-weak syllables regardless of their language experience) or violations of the strong trend for two-syllable Dutch words to have stress on the first syllable. However, in our study the larger negativity was observed when comparing items with the same stress pattern (penultimate in Experiment 2, and across antepenultimate and penultimate in Experiment 1) and before information about the final vowel was available, and so can definitively be linked to violations of a probabilistic trend.

Our interpretation of the early negative effect is that it is related to identification of a words stress pattern in one of two ways. The first possibility is that both the early negativity and Böcker et al.'s N325 are instances of a component which is a manifestation of one stage of lexical access, namely the identification of a word's stress pattern. A greater negativity at this stage obtains if the word's stress pattern is atypical in some way. We will consider two possible mechanisms by which the greater negativity could arise. The first possibility is that the lexical encoding of stress may be different for items which violate grammatical trends than for items which do not. In the phonological literature, many such mechanisms have been proposed - for English stress specifically Becker (2009); Burzio (1994); Chomsky and Halle (1968); B. P. Hayes (1980); Pater (2005); Selkirk (1984) have all proposed some kind of extra diacritic or feature on words with exceptional stress patterns. Building on this body of work, Moore-Cantwell and Pater (2016) propose that each word's lexical representation has some weight and competes with grammatical constraints. A stress pattern which violates a lexical trend would have a higher weight (be more strongly represented) so that it can overcome the pressures of the grammar. A stress pattern which observes a lexical trend would have a lower weight, and be less strongly encoded. If this analysis is correct, then accessing the stress pattern of an exceptional word would proceed differently than accessing the stress pattern of a trend-observing word. However, it is not clear that accessing a more strongly encoded stress pattern should result in a greater negativity - rather one might imagine that accessing or recognizing a more strongly encoded stress pattern would be *less* effortful than accessing or recognizing a less strongly encoded stress pattern.

A second possible mechanism that may give rise to the observed pattern of early negativity is competition between the grammar and a word's lexically encoded stress pattern. For words whose stress pattern agrees with the probabilistic trend in the grammar, there is no competition, as the grammar's predictions and the lexical encoding of stress align. For trend-violating words however, the lexically encoded stress pattern violates the grammar's expectations. Competition between the grammar's expectations and the actual stored stress pattern of a word makes the process of correctly recognizing that stress pattern more effortful, and therefore gives rise to a greater negativity at the stage when the word's stress pattern is recognized. Although the first proposal, that differences in lexical encoding of stress give rise to the early negativity, lines up well with phonological theory, we prefer this second proposal on grounds of neurological plausibility.

4.3. N400

The N400 was much bigger for nonwords than for words - an effect replicating the large N400 for nonwords found by Rugg and Nagy (1987) and many others. Although in Experiment 1 an additional interaction of final vowel and stress was found in words, this interaction was not replicated in Experiment 2. We therefore treat it as a spurious result and do not analyze it further.

4.4. LPC

Despite the fact that the strong i-final trend is productive, it did not influence processing in the same way as categorical patterns. In particular, violations of this trend do not elicit an LPC similar to what was reported by others (Domahs et al., 2009; Pitkanen, 2010) in response to violations of categorical phonological patterns. The LPC has been linked to late, evaluative processes (Gouvea, Phillips, Kazanina, & Poeppel, 2010; Osterhout & Holcomb, 1992, et.seq.), and may not be evident for probabilistic patterns for which the listener has readily accessible exceptions. One example of an LPC, the P600, has been reported for rare syntactic sentence structures and garden-path sentences that do not actually violate grammatical rules. However, even in these cases listeners are likely to have the online experience that the syntactic structure they have extracted from the stimuli is categorically wrong during processing. With probabilistic, exceptionful patterns, the late evaluation of a pattern as deficient in some way is absent.

5. Conclusion

The strong trend for i-final words was observed in participants' ratings of nonwords, and it affected early stages of the lexical access process in real words. The ə-final trend also affected early stages of lexical access, but only in the first experiment - the effect did not replicate. We argue that this early negativity originates from the stage of lexical access at which the words stress pattern is recognized. This recognition draws upon the acoustic signal as well as stored information about a particular lexical item. The size of the negativity is modulated by the typicality of the word's stress pattern, which we argue is due to competition between the expectations of the phonological grammar and the actual lexical encoding of a trend-violating stress pattern.

Previous research on how phonological generalizations affect speech processing has examined categorical patterns only, and those can be divided into two types. Local categorical rules, such as the rule against 'tl' or 'dl' at the beginnings of words in English (Breen et al., 2013), tend to be perceptually 'repaired' relatively early in processing (Breen et al., 2013; Dehaene-Lambertz et al., 2000; Dupoux et al., 1999; Moreton, 2002). Because these violations are repaired early, there is no late evaluative process that would elicit an LPC. Nonlocal categorical rules, such as the 'spup' constraint in German, or vowel harmony in Finnish, do seem to elicit an LPC (Domahs et al., 2009; Pitkanen, 2010). Because the entire word must be evaluated in order for violations of these rules to be observed, the early perceptual repair processes that repair 'tl/dl' are not available. Instead, these structures are evaluated later in processing, and are found to violate the rule.

In this paper, we examined a probabilistic rule, which was also nonlocal in nature. With any probabilistic generalization, early perceptual repair is not an option, since extant exceptions to the generalization need to be veridically perceived. We might expect probabilistic rules to behave like nonlocal generalizations, and be subject to the same late evaluative process that yields an LPC. However, we do not find an LPC in response to violations of the probabilistic generalization examined here. Rather, the fact that real words like them exist must cause the evaluation process to deem nonwords like 'bəmǽki' acceptable. We do find that there is a cost to processing violators of probabilistic rules, however. This cost stems from the necessity of reconciling the exceptional form against the grammatical representation of the more frequent pattern. We find the effects of this conflict in the early stages of speech processing for a relatively detailed pattern, which is pertinent to only a small subset of the lexicon of English. This finding suggests that speakers' knowledge of probabilistic patterns in the sound pattern of their language is extremely detailed.

Acknowledgements

Special thanks are due to Lyn Frazier and Joe Pater for contributions to the design of the experiment, as well as to Jason Overfelt for recording the stimuli, and Austin Kopak, Hala Hasan, Maggie Ugolini, Ching Tiv, Elkhansa Elguenaoui, and Kevin Harrington for help collecting data. Thanks are also due to members of the UMass Neurocognition and Perception lab and Sound Seminar, audiences at the 22nd Manchester Phonology Meeting and LabPhon14, as well as two anonymous reviewers for useful discussion and comments.

Funding

This research was supported in part by a UMass Graduate School Dissertation Research Grant, in part by a UMass ICESL seed grant, and in part by a National Science Foundation Graduate Research Fellowship to the first author (NSF DGE-0907995).

References

Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighbors? *Journal of Memory and Language*, *44*, 568-591.

Becker, M. (2009). *Phonological trends in the lexicon: the role of constraints* (Unpublished doctoral dissertation). University of Massachusetts Amherst.

Beckman, M. E. (1986). *Stress and non-stress accent* (Vol. 7). Dordrecht, The Netherlands: Walter de Gruyter.

Berent, I., Lennertz, T., Smolensky, P., & Vaknin-Nusbaum, V. (2009). Listeners' knowledge of phonological universals: evidence from nasal clusters. *Phonology*, *26*, 75-108.

Böcker, K. B., Bastiaansen, M. C., Vroomen, J., Brunia, C. H., & Gelder, B. D. (1999). An erp correlate of metrical stress in spoken word recognition. *Psychophysiology*, *36*, 706-720.

Breen, M., Kingston, J., & Sanders, L. D. (2013). Perceptual representations of phonotactically illegal syllables. *Attention, Perception, and Psychophysics*, *75*(1), 101-120.

Brysbaert, M., & New, B. (2009, November). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977-990.

Burzio, L. (1994). *Principles of English stress* (Vol. 72). Melbourne, Australia: Cambridge University Press.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* New York, Evanston, and London: Harper and Row.

Coetzee, A. W., & Pater, J. (2009). The place of variation in phonological theory. In J. A. Goldsmith, J. Riggle, & A. C. Yu (Eds.), *Handbook of phonological theory* (2nd ed., p. 401-434). Wiley-Blackwell.

Cutler, A., & Carter, D. M. (1987, September-December). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, *2*(3-4), 133-142.

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience*, *12*(4), 635-647.

Domahs, U., Kehrein, W., Kraus, J., Wiese, R., & Schlesewsky, M. (2009, December). Event-related potentials reflecting the processing of phonological constraint violations. *Language and Speech*, *52*(4), 415-435.

Domahs, U., Plag, I., & Carroll, R. (2014). Word stress assignment in german, english and dutch: Quantity-sensitivity and extrametricality revisited. *The Journal of Comparative Germanic Linguistics*, 1–38.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in japanese: A perceptual illusion? *Journal of experimental psychology: human perception and performance*, *25*(6), 1568.

Ernestus, M., & Baayen, H. (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language*, *79*(1), 5-38.

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the p600. *Language and Cognitive Processes*, *25*(2), 149-188.

Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, *39*, 379-440.

Hayes, B., Zuraw, K., Siptár, P., & Londe, Z. (2009, December). Natural and unnatural constraints in Hungarian vowel harmony. *Language*, *85*(4), 822-863.

Hayes, B. P. (1980). *A metrical theory of stress rules* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Jarmulowicz, L. D. (2002). English derivational suffix frequency and children's stress judgments. *Brain and Language*, *81*, 192-204.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child development*, *64*(3), 675-687.

Kretzschmar, F., Schlesewsky, M., & Staub, A. (in press). Dissociating word frequency and predictability effects in reading: Evidence from co-registration of eye movements and EEG. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Massaro, D. W., & Cohen, M. M. (1983). Phonological context in speech perception. *Perception and Psychophysics*, *34*(4), 338-348.

Moore-Cantwell, C. (2016). *The representation of probabilistic phonological patterns: neurological, behavioral, and computational evidence from the English stress system* (Doctoral dissertation, University of Massachusetts Amherst). Retrieved from http://www.clairemoorecantwell.org/dissertation_CMC.pdf

Moore-Cantwell, C., & Pater, J. (2016). Gradient exceptionality in maximum entropy grammar with lexically specific constraints. *Catalan Journal of Linguistics*, *15*, 53-66. Retrieved from http://revistes.uab.cat/catJL/article/view/v15-moore-cantwell-pater/183-pdf-en

Moreton, E. (2002). Structural constraints in the perception of English stop-sonorant clusters. *Cognition*, *84*, 55-71.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785-806.

Patel, A., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event related potential study. *Journal of Cognitive Neuroscience*, *10*, 717-733.

Pater, J. (2000). Non-uniformity in english secondary stress: the role of ranked and lexically specific constraints. *Phonology*, *17*(2), 237-274.

Pater, J. (2005). Learning a stratified grammar. In A. Brugos, M. R. Clark-Cotton, & S. Ha (Eds.), *Proceedings of the 29th boston university conference on languega development* (p. 482-492). Somerville, MA: Cascadilla Press.

Pitkanen, I. (2010). *Electrophysical investigations of second language word learning, attrition, and bilingual processing.* (Unpublished doctoral dissertation). University of Washington.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372-422.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psycholinguistic Review*, *105*, 125-157.

Rossi, S., Hartmuller, T., Vignotto, M., & Obrig, H. (2013). Electrophysiological evidence for modulation of lexical processing after repetitive exposure to foreign phonotactic rules. *Brain and Language*, *127*, 404-414.

Rossi, S., Jurgenson, I. B., Hanulíková, A., Telkemeyer, S., Wartenburger, I., & Obrig, H. (2011). Implicit processing of phonotactic cues: Evidence from electrophysiological and vascular responses. *Journal of Cognitive Neuroscience*, *23*(7), 1752-1764.

Rugg, M. D. (1984, 435-443). event-related potentials and the phonological processing of words and non-words. *Neuropsychologia*, *22*, 1984.

Rugg, M. D., & Nagy, M. E. (1987). lexical contribution to nonword-repetition effects: evidence from event related potentials. *Memory and Cognition*, *15*(6), 473-481.

Selkirk, E. (1984). *Syntax and phonology: The relation between sound and structure.* Cambridge, Massachusetts: MIT Press.

Tyler, A., & Nagy, W. (1989). the acquisition of English derivational morphology. *Journal of Memory and Language*, *28*, 649-667.

Vaden, K., & Hickok, H. (n.d.). *Irvine phonotactic online dictionary, version 2.0.* Retrieved from http://www.ipod.com

Weide, R. L. (1994). *CMU pronouncing dictionary.* Retrieved from http://www.speech.cs.cmu.edu/cgi-bin/cmudict

Zuraw, K. (2000). *Patterned exceptions in phonology* (Doctoral dissertation, University of California, Los Angeles).