Event-related potential evidence of abstract phonological learning in the laboratory

October 7, 2018

Abstract

The experimental study of artificial language learning has become a widely used means of investigating the predictions of theories of language learning. Although much is now known about the generalizations that learners make from various kinds of data, relatively little is known about how those generalizations are cognitively encoded. This paper presents an ERP study of brain responses to violations of lab-learned phonotactics. Novel words that violated a learned phonotactic constraint elicited a larger Late Positive Component (LPC) than novel words that satisfied it. Similar LPC's have been found for violations of natively acquired linguistic structure, as well as for violations of other types of abstract generalizations, such as musical structure. We argue that lab-learned phonotactic generalizations are represented abstractly, and similarly to natively acquired syntactic and phonological rules.

1 Introduction

When trained on an artificial language in the lab, participants form generalizations about its sound structure, similar to the generalizations that all speakers form about the sound structure of their native language. These generalizations, especially if simple, can be formed after even a very small amount of exposure (less than half an hour) to the artificial language (see Culbertson, 2012; Moreton and Pater, 2012 for reviews). Participants are consistently able to apply these generalizations to novel test words, and in some cases even to novel sounds (Cristiá and Seidl, 2008; Cristiá et al., 2013) and novel contexts (Myers and Padgett, 2014). In this paper, we use neurophysiological measurements to examine how these lab-learned phonological generalizations are cognitively encoded. In particular, we address the question of whether they are encoded in a way that is comparable to natively acquired phonological patterns.

If lab-learned phonological generalizations are learned and encoded similarly to naturalistically acquired phonological generalizations, then artificial language learning would be a valuable experimental technique for examining the qualities and limits of humans' phonological learning capacity. Researchers have already used it to examine the extent to which speakers have knowledge of phonological universals that are not instantiated in their native language (e.g. Pycha et al., 2003; Wilson, 2003; Carpenter, 2010), the relationship between phonotactics and alternations (Pater and Tessier, 2003, 2006), and the nature of biases for structural simplicity (Moreton et al., 2017; Moreton, 2008; Lai, 2012), and how humans generalize from certain phonological generalizations to others (Cristiá and Seidl, 2008; Cristiá et al., 2013; Myers and Padgett, 2014).

The conclusions of the above studies, as well as many others, rely on the assumption that artificial language learning does basically resemble natural first language acquisition. However, the process of learning a phonological generalization in the lab differs considerably from the process of acquiring a language's phonology naturalistically. The amount of input data provided in the lab, even in longer artificial language experiments, is radically different from the years of exposure that contribute to the acquisition of natural language phonology. Additionally, participants in an artificial language study are typically taught a single 'rule'. All of the items in an experiment constitute evidence about that rule, and only one rule is taught at once. In the process of natural language acquisition, not only are many phonological rules being acquired in parallel, but so is a great deal of higher-level grammatical structure. Another potential difference between the two situations lies with the learner: the participants in lab-learning experiments are typically adults, and the acquisition of one's native phonology happens in childhood. It is possible that quite different learning processes are made use of by children and adults, and that very different cognitive representations of generalizations result (though see Birdsong, 2006 for a critical overview of the 'critical period' literature).

Since the process of learning phonology in the lab bears a certain resemblance to the process of learning a second language as an adult, we turn to the L2 learning literature in order to understand what alternative strategies adults might use in the very early stages of acquiring a novel grammatical system. McLaughlin et al. (2010) review several studies comparing early stages of L2 acquisition to later stages using event-related potentials (ERPs). They argue that learners go through two distinct stages. In the first, their syntactic and phonological knowledge appears to be strongly tied to their knowledge of the lexicon, while in the second their knowledge of grammatical patterns closely resembles native grammatical knowledge. They find that for both native speakers and more proficient L2 learners, grammatical violations elicit a larger positivity than grammatically correct sequences, with the difference between conditions peaking around 600ms after the onset of the violations. This positivity peaking around 600ms is typically found in response to syntactic violations (Osterhout and Holcomb, 1992, et seq.), where it is called a P600, but a similar late positivity has been observed for phonological violations, both in the McLaughlin et al. studies, and in Domahs et al. (2009). Early L2 learners, in contrast, exhibit a larger N400 in response to grammatically illicit novel words and structures than to grammatically licit ones.

The N400 was first identified as a response to semantically incongruous, but syntactically acceptable words, such as “He took a sip from the transmitter” (Kutas and Hillyard, 1980; Kutas and Federmeier, 2011). The amplitude of the N400 has been interpreted as an index of the effort involved in lexical processing and integration with context. For example, nonwords elicit larger N400s than actual words (Rugg, 1984), and novel words (e.g., toose) preceded by rhyming strings (e.g., buice) elicit a smaller N400 than the same items preceded by unrelated strings (e.g., gock) (Coch et al., 2015). Many other factors affect the amplitude of the N400; high-cloze words elicit a smaller N400 than equally plausible low-cloze words (Kutas and Hillyard, 1984), more recently accessed words elicit a smaller N400 (van Petten et al., 1991), and both words and pseudowords with fewer lexical neighbors elicit a smaller N400 (Holcomb et al., 2002).

A late positivity has been observed for a range of syntactic violations, including agreement, phrase structure, subcategorization, and constraints on long-distance dependencies (see Gouvea et al., 2010; Morgan-Short et al., 2012 for overviews). Because in this context it usually has a peak at around 600 ms post-stimulus, it is usually referred to as a P600, though a variety of factors affect its latency, as well as its distribution on the scalp (Gouvea et al., 2010). There are a variety of proposals about the functional interpretation of this late positvity, though there is general agreement that it reflects the evaluation of an abstract structural relation. A similar late positivity (which we refer to as the LPC, or Late Positive Component) has subsequently been found for phonotactic violations: Finnish vowel harmony in McLaughlin et al. (2010), and the German sCVC constraint in Domahs et al. (2009). LPC's have also been found for violations of musical structure (Patel et al., 1998; see Carrión and Bly, 2008 for an overview), and for rule violations in arithmetic tasks (Núñez-Peña and Honrubia-Serrano, 2004). While the LPC is not language-specific, it is also an indicator of abstract structural relations in these other cognitive domains: "an index of detection for any anomaly in rule-governed sequences" (Núñez-Peña and Honrubia-Serrano, 2004, 130); [it] "reflects processes of knowledge-based structural integration" (Patel et al., 1998, 51).

Observing an N400 effect, rather than an LPC, in response to grammatical violations led McLaughlin et al. to argue that during early L2 acquisition learners memorize specific strings, or probabilistic dependencies between specific strings. They are thus able to replicate the effects of grammatical knowledge in production and comprehension. Later, learners acquire an abstract grammatical rule, and begin to show native-like processing, reflected in an LPC effect. The difference between these two stages of acquisition may be related to the difference between declarative memory (memorization of strings or probabilistic dependencies) and procedural memory (grammatical rules); see Ullman (2001, 2005); Paradis (2009) for more on the neurological grounding of this distinction.

In the current study, we examine participants' EEG responses during the process of acquiring an artificial phonological generalization in the lab. We specifically ask whether, like early L2 learners, they exhibit a larger N400 in response to novel words that violate the generalization, or whether, like later L2 learners and native speakers, they exhibit an LPC to novel words which violate the generalization.

Previous research using ERPs to examine phonotactic knowledge has focused on cases of 'perceptual repair', in which listeners have difficulty accurately perceiving sequences which are disallowed in their language (e.g. Dehaene-Lambertz et al., 2000; Breen et al., 2013). We do not expect perceptual repair to occur with lab-learned phonotactics, since both observers and violators of the pattern are legal in the participants' native language, English. Instead, we compare the processing of lab-learned phonotactics to the correctly-perceived phonotactic violations examined in McLaughlin et al. (2010) and Domahs et al. (2009), which both elicit an LPC. Based on previous research measuring the N400 and the LPC, we expect to find a smaller N400 for words presented in the training phase of an artificial language learning experiment than for words that were newly presented in the test phase. Such an N400 difference would mean that participants had successfully learned the words presented in the training, rendering them easier to process. The difference in processing trained and novel words in an artificial language is expected to parallel the difference in processing words in a known language and nonwords, demonstrated in Rugg (1984) and many subsequent studies. We might also expect to find a larger LPC for novel words that violate a generalization over the training words than for novel words that satisfy it, if the knowledge of that generalization is abstractly represented, in a way that is similar to naturalistically learned phonology (McLaughlin et al., 2010). Finally, if participants learn a generalization over the training items in a way that is similar to the early stages of L2 learning, we would expect to find a greater N400 for novel violators than for novel observers. In this case, participants would be using an analogical mechanism or other lexical search process to apply the generalization to novel words.

2 Methods

We taught 24 adult, native English speakers 16 word-object pairings by asking them to match an auditorily presented word to one of four pictures, after which they were given the correct pairing. The words each participant learned were all consistent with a phonotactic pattern. In testing, participants were asked to rate on a 4-point scale how likely it is that each word is part of the language they were learning. These words included half of the trained words (Studied), eight novel words that fit the pattern (Novel-Fit), and eight words that violated it (Novel-Violate). Testing and training blocks alternated, with a total of five each. We adopted this alternating training-testing procedure so that we could collect sufficient EEG data in the test blocks for our ERP analysis without “untraining” the participants with a too-long single test block that contains words that violate the restriction.

2.1 Participants

Participants were 24 (9 females, 15 males) native speakers of English, between the ages of 19 and 33 years. Data from one participant were excluded because she expressed explicit knowledge of the phonological generalization[1] and from one other because of excessive high-frequency noise in the EEG, likely caused by muscle tension. All participants reported being right-handed, having no neurological problems, and not taking psychoactive medication within a year of the study. Participants provided informed consent and were compensated for their time at a rate of $10/hour.

2.2 Materials

As the phonotactic patterns for our participants to learn, we selected voicing agreement and disagreement between stops in CVCV words. The patterns, and the stimulus space, were adopted from Moreton (2008; 2012). The stops were drawn from the set [d, g, t, k], and the vowels from the set [i, æ, u, ɔ]. We constructed 48 words, with the consonants in half of them agreeing in voice (e.g. [dugi], [tikɔ]), and in the other half disagreeing (e.g. [kædu], [tigæ]). Vowels were chosen to avoid patterns in terms of co-occurrence with one another, or with the consonants. From this set of items, four groups of items were created, shown in Table 1. For the two Voice-Match lists, half of the voice-agreeing items were used as training items, and half were used in testing. The items that were used in training for Group 1 were used in testing for Group 2. Half of the voice-disagreeing items were used as the novel pattern-violating items in testing for Group 1, and the other half of those items were used for Group 2. Two more lists were created for the Voice-Mismatch condition in a similar way. The 24 participants were evenly divided across Groups - six in each group. Each participant saw each item in only one role, but across participants each stimulus appeared during training, during testing, and as a pattern-violating nonword.

The 24 words in each “language” were equally likely to start with a voiced or voiceless consonant or with a velar or alveolar place of articulation. Vowel features were also controlled in each position. English words

---

[1] Of the 24 participants that contributed data to analysis, 19 reported adopting a strategy to learn the word/picture relationships. These approaches involved various types of mnemonics (e.g., [kudi] is the “current deep” where the ship sails, and [tɔki] sounds like something you would say to a little puppy). However, only five participants reported trying to come up with a system for determining which words were in the language at the time of testing; all five also stated that their attempts failed. These participants considered patterns such as “if a word starts with k, it has to end with i and if it starts with g it has to end with a.” All participants, including these five, agreed with the statement “In the end, I just guessed about the test words.” One participant did suggest that a pattern existed between the consonants in the words, and also identified the correct pattern (i.e., consonants match in voicing); data from this person were excluded from analysis.

		Agreeing		Disagreeing	
		Set 1	Set 2	Set 3	Set 4
		tɔtu kikɔ	tɔku kitɔ	tɔdu kigɔ	tɔgu kidɔ
		tɔki kuti	tɔti kuki	tɔgi kudi	tɔdi kugi
		didɔ gigæ	digɔ gidæ	ditɔ gikæ	dikɔ gitæ
		dægɔ gɔdi	dædɔ gɔgi	dækɔ gɔti	dætɔ gɔki
Voice-Match	Group 1	**Studied**	Novel-Fit	Novel-Violate	▨
	Group 2	Novel-Fit	**Studied**	▨	Novel-Violate
Voice-Mismatch	Group 3	Novel-Violate	▨	**Studied**	Novel-Fit
	Group 4	▨	Novel-Violate	Novel-Fit	**Studied**

that fit the constraints (e.g., duty and gaudy) were excluded.

The words were pronounced by a 26-year-old linguistically trained male native speaker of English. The speaker pronounced the words with stress on the initial syllable and a full vowel in the second syllable in the frame sentence “It was X said Tim.” The sentences were recorded to 32 bit / 44.1 khz digital format. The words were segmented from the sentences using the offset/onset of noise for the surrounding sibilants as criterion. The peak amplitude of all items was normalized to their mean, and a 10 ms sinusoidal fadeout was applied to the end of each recording to eliminate the effects of trimmed formants.

Words averaged 367 ms (SD = 34) in duration with the 2nd syllable beginning 136-245 ms after the first (M = 191, SD = 29). All sounds were presented over a pair of M-Audio StudioPro3 loudspeakers with EPrime software running on a PC with a Creative Audigy 2 ZS sound card. Both loudspeakers were located directly above the computer monitor. Sounds were presented at 65 dB SPL (A-weighted) as measured from the location of participants’ heads.

The objects used in training were presented as color photographs of a common concrete object (e.g., puppy, ship, shoe) that participants would be expected to describe with a single English word. Images were cropped to leave minimal background behind the objects. Pictures were then resized such that four pictures shown at the same time along with their response labels (1-4) filled the space available on the computer monitor. The image that was “correct” in training for a given aurally presented label was presented at an identical size when shown among the three distractors and when it was shown afterwards with the aural label.

2.3 Procedure

Instructions for both the training and testing tasks were given at the beginning of the session. For training, participants were told they would be learning some of the words in a made up language by matching the spoken words to pictures of the objects the words refer to. They were warned that their initial responses would be guesses, but that seeing the correct answer after every response would help them to learn the words. Participants initiated a training block by pressing any button on a response pad. Each training trial began with the appearance of a fixation cross on the computer monitor. One of the 16 training words was presented from the loudspeakers 700 - 1200 ms (randomly selected, rectangular distribution) later. The fixation cross remained on the screen for 500 ms after the offset of the word and was then replaced by pictures of 4 objects. Each object was labeled with a number (1-4) that corresponded with a button on the response pad. Participants pressed a button to indicate the meaning of the word they had heard. Immediately following any button press, the correct picture was shown in the middle of the computer monitor for 1000 ms and the word was played again. EEG data collected while pictures were shown were not analyzed and participants were free to move their eyes and blink during these presentations. In each training block, all 16 words were presented 5 times each in random order (80 trials per block).

During testing, participants were asked to rate how likely it is that each word is part of the language they were learning. They were told that some of the test words had been heard during training and were clearly part of the language. They were instructed that even though they had not learned a meaning for most of the test words, some of them were part of the same language and some were not. The experimenter shared the example that people can often tell if a word sounds like it could be Italian or Japanese even if neither of those languages is familiar. They were also encouraged to use all four response buttons rather than just button 1 (labeled “unlikely a word”) and button 4 (labeled “very likely a word”). Participants began each test block by pressing any button. At the beginning of each trial, the fixation cross was shown on the computer monitor. One of the test words was presented over the loudspeakers 700 - 1200 ms later. The fixation cross remained on the screen for 1000 - 1500 ms after the word onset and was then replaced by the response prompt "Likely a word?" with the labeled scale. A test trial ended after any response was given. In each test block, all 24 words (eight each of Studied, Novel-Fit, and Novel-Violate) were presented once in random order.

The training block - test block sequence was presented 5 times for a total of 400 training trials and 120 test trials (40 of each type). For all trials, participants were asked to refrain from blinking, moving their eyes, or moving any other part of their body, including moving a finger to press a button, whenever the fixation cross was shown on the screen. They were encouraged to make these and any other movements while the pictures (training) or response prompt (testing) were shown. Participants were asked to continue from each training block to the following test block without pause; they were encouraged to take breaks after each test block. At the end of the experiment, participants were asked if they had noticed anything about the language they learned and if they had developed any strategy to distinguish between words that were and were not in the language.

EEG was recorded continuously throughout the training and test trials (250 Hz sampling rate, 0.01-100 Hz bandpass) from 128 electrodes (EGI, Eugene OR). Scalp impedances at all electrode sites were maintained under 50 kΩs. Segments of EEG from 100 ms before to 500 ms after the onset of training words and from 100 ms before to 1000 ms after the onset of test words were examined. Trials with artifacts from muscle tension, blinks or eye-movements, or motion were excluded from analysis. EEG from remaining training trials was averaged together by each block; EEG from test trials was averaged by condition (Studied, Novel-Fit, Novel-Violate) across all blocks. The 100 ms before word onset were used as a baseline and ERPs were rereferenced to the averaged mastoid recording.

For training trials, average amplitude measurements were taken 40 - 70 ms (P1), 90 - 130 ms (N1), and 230 - 500 ms (N400) after word onset. For test trials, mean amplitude measures were made in the same P1 and N1 windows as well as 400 - 700 ms (N400 and P300) and 600 - 1000 ms (LPC) after word onset. Measurements were made at 100 of the 128 electrode sites across the scalp such that electrode position could be included as multiple factors in statistical analyses. Specifically, measurements from 4 electrodes were averaged together in a 5 (Anterior, Anterior-Central, Central, Posterior-Central, Posterior or ACP) x 5 (Left, Left-Medial, Medial, Right-Medial, Right, or LMR) grid. Data from each of the three measurements taken from training trials were entered in 5 (Block) x 5 (ACP electrode position) x 5 (LMR electrode position) repeated-measures ANOVAs. Data from test trials were entered in 3 (Word Type) x 5 (ACP electrode position) x 5 (LMR electrode position) repeated-measures ANOVAs. Greenhouse-Geisser corrected p-values (and uncorrected degrees of freedom) are reported. Significant ($p < .05$) effects of Block in the training data were followed up by comparisons of each training block with block 1. Significant ($p < .05$) interactions of Word Type and electrode position factors were followed by ANOVAs conducted on data collected at subsets of electrodes and to compare the ERPs elicited by Studied words to both types of novel words as well as Novel-Fit and Novel-Violate words.

3 Results

3.1 Training

During the first training block, participants were already well above chance performance of 25% on the word-picture matching task (M = 53.9%, SD = 2.4); in the four subsequent training blocks, performance was even better (M = 89.2%, SD = 1.8). See Figure 1

Omnibus ANOVAs were conducted over the ERP data during the training blocks. No effects of block were observed in the P1 or N1 time windows, except a marginally significant interaction between block and anteriority in the N1. In the N400 time window, there was a main effect of block across all electrodes ($F(4,92) = 2.952$, $p = .037$). The negativity was larger for block 1 (M = -1.38μV, SE = .30) than in the other four blocks (M = -.81μV, SE = .29). These results are consistent with the behavioral data showing that learning of the word meanings occurred rapidly and was maintained.

3.2 Testing

During testing, participants rated the Studied words as more likely to be in the language that they were learning (M = 3.72, SD = .18) than Novel-Fit words (M = 2.71, SD = .29) ($t(23) = 14.77$, $p < .001$).

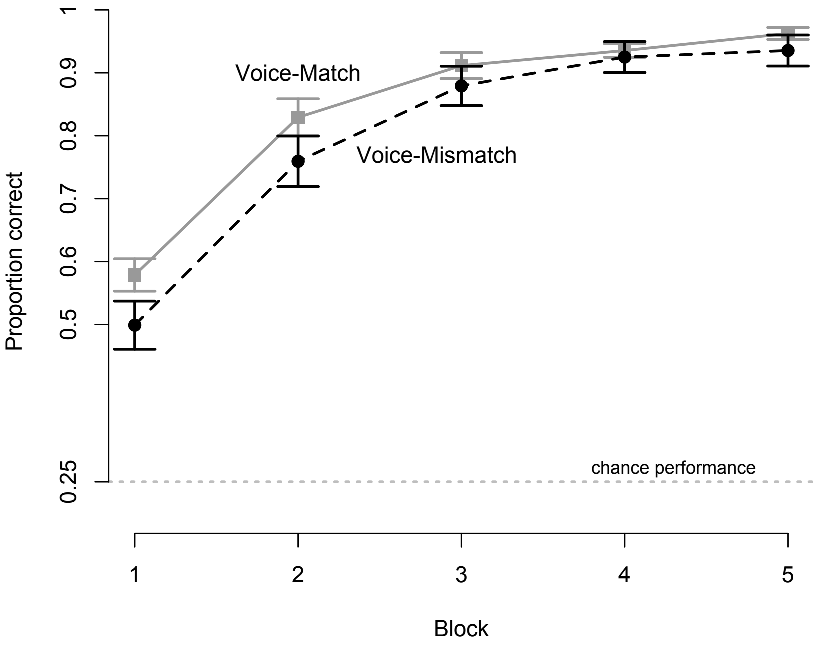

Figure 1: Proportion correct responses across blocks of training. Chance performance is at 25% because participants are choosing between four pictures while they are learning the words.

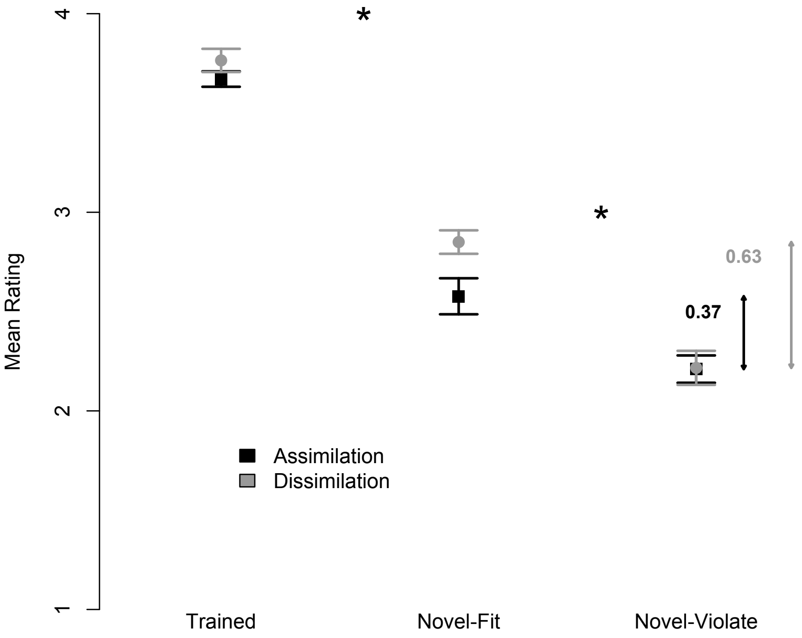

Figure 2: Rating responses during the training session. Voice-Match condition values are grey squares and Voice-Mismatch black circles. In both Voice-Match (assimilation) and Voice-Mismatch (dissimilation groups), Novel-Violate items were rated as less likely to be in the language than Novel-Fit items.

The participants rated the Novel-Fit words as more likely to be in the language than Novel-Violate (M = 2.21, SD = .26) ($t(23) = 7.98$, $p < .001$). We found no evidence that the pattern of ratings differed for the Voicing-Match and Voicing-Mismatch languages (across all types of test words, $p > .15$). These results are illustrated in Figure 2.

Results of omnibus ANOVAs conducted over each ERP time window for testing data are shown in Table 1. Over all electrodes, a significant effect of word type (Studied, Novel-Fit or Novel-Violate) was observed during the N1 window. Studied words were less negative (M = -.67μV, SE = .29) than either Novel-Fit (M = -1.27μV, SE = .30) or Novel-Violate (M = -1.11μV, SE = .29). In the N400 window, a marginally significant interaction ($p = 0.051$) between Anteriority and word type was followed up by separate ANOVAs

Table 1: ANOVA results for all four EEG time windows measured for the testing data.

Effect	P1 40-70 ms	N1 90-130 ms	N400/P300 400-700 ms	LPC 600-1000 ms
Word Type (2,46)	-	4.64^*	-	-
LeftRight x Word Type (8,184)	-	-	-	-
Anteriority x Word Type (8,184)	-	-	$2.65^\dagger$	$2.43^\dagger$
LeftRight x Anteriority x Word Type (32,736)	-	-	-	$1.70^\dagger$

† 0.1>p>0.05; * 0.05>p>0.01; **p<0.01

conducted at the five different levels of Anteriority, using a Bonferroni-corrected significance level of $\alpha = 0.05/5 = 0.01$. Over central-posterior regions, there was a significant effect of word type (F(2,46)=6.22, p = 0.004). Studied words were more positive over these electrodes (M = 1.43μV, SE = .44) than Novel-Fit (M = .45μV, SE = .39) or Novel-Violate (M = .70μV, SE = .37).

Additionally, marginally significant interactions in the LPC region between word type and Anteriority (p = .062) and between word type, Anteriority and Left-Right (p = .093) were followed up by separate ANOVAs conducted at the five different levels of Anteriority, again using a Bonferroni-corrected significance level of $\alpha = 0.05/5 = 0.01$. There was a significant effect of word type over central-posterior regions (F(2,46) = 5.10, p = .01), but no interaction between Left-Right and word type (F(8,184) = 3.01, p = .013). Over central-posterior electrodes Novel-Fit words elicited a smaller positivity (M = .85μV, SE = .49) than either Novel-Violate (M = 1.90μV, SE = .44) or Studied (M = 1.71μV, SE = .42).

The differences between conditions at the N1, N400, and LPC time windows can be seen in Figures 3 and 4, which show ERP waveforms over the whole head, comparing Novel-Fit with Studied, and Novel-Fit with Novel-Violate respectively.

4 Discussion

Both participants' increasing accuracy on training items, and the decreasing size of the N400 over the five blocks of the experiment, show that participants in fact learned the words in the training set. Additionally,

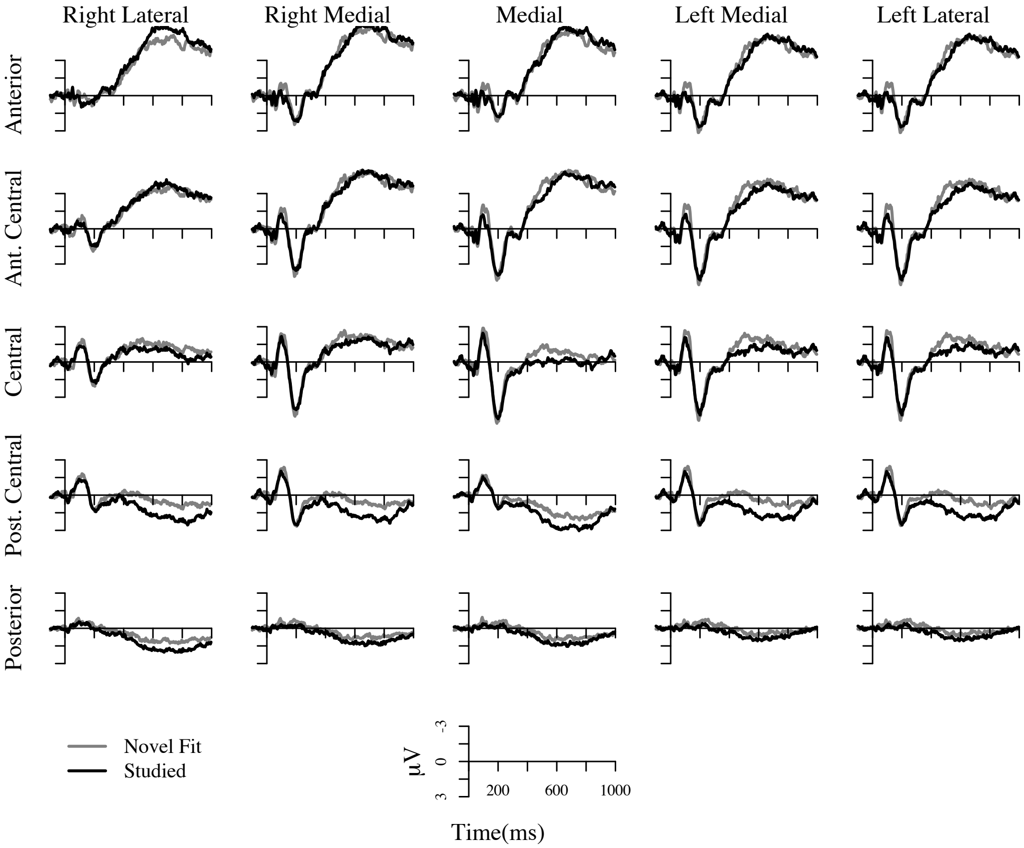

Figure 3: Novel Fit and Studied. Waveforms timelocked to the onset of the stimulus. Each one is an average over 4 electrodes, located as indicated by the row and column labels.

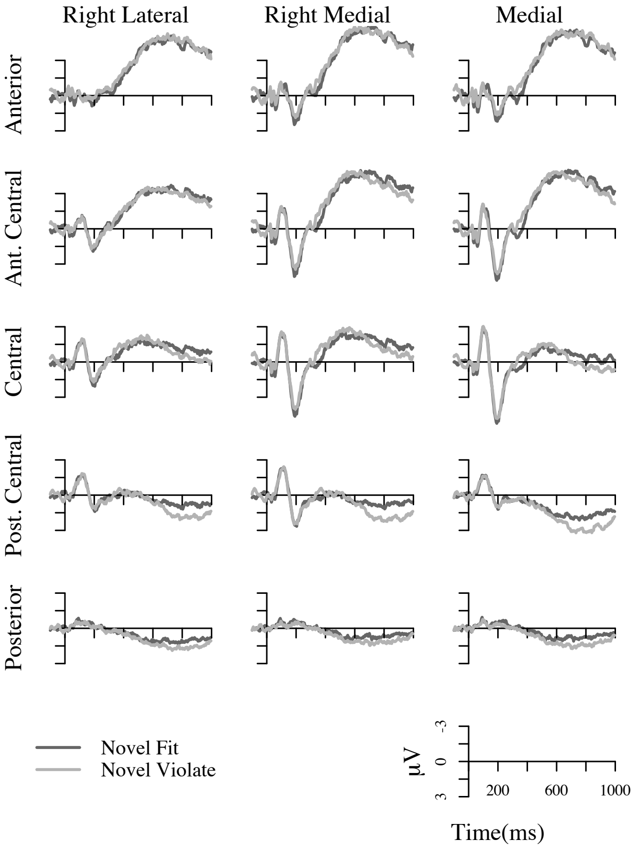

Figure 4: Novel Fit and Novel Violate. Waveforms timelocked to the onset of the stimulus. Each one is an average over 4 electrodes, located as indicated by the row and column labels.

during testing, Studied words elicited a smaller negativity in the N400 time window than novel words, whether they fit the phonological pattern or not.

In testing blocks, participants distinguished in their ratings between trained items and untrained items, but also between untrained items which fit the phonological pattern, and untrained items which did not fit the pattern. They rated Novel-Violate items worse than Novel-Fit items, and the effect was similar whether the trained pattern was an assimilation pattern or a dissimilation pattern. In testing blocks, ERP measures showed a difference between Novel-Fit and Novel-Violate items in the latest measured time window, 600-1000 ms. Novel-Violate items elicited a larger positivity than Novel-Fit items in this region. This late positive component (LPC) is identical to what has been observed in response to other violations of abstract structural relations (Carrión and Bly, 2008; Patel et al., 1998). Further, it is similar to the P600/LPC observed in response to grammatical violations in language (Osterhout and Holcomb, 1992). Our observed LPC constitutes evidence that simple artificial languages allow listeners to apply an abstract phonological rule in a manner that is similar to syntactic rule application in more proficient L2 learners (McLaughlin et al., 2010). Importantly, the effects of learning a phonological pattern in the lab are also similar to the effects of native language phonotactics on processing novel words (Domahs et al., 2009; McLaughlin et al., 2010).

In the same late time window, 600-1000 ms, a difference was also observed between Novel-Fit and Studied words, with the Studied words being more positive. We believe that this positivity is an example of a P300. P300's can be elicited by less probable items relative to more probable items (Donchin, 1981) and by items which require a response relative to items which do not (Duncan-Johnson and Donchin, 1977). In the testing session, trained items were the minority, making up one third of the stimuli. While all items in the testing session required a response, trained items required a qualitatively different type of response than untrained items (whether pattern-fitting or not), since participants only had to remember the item from training, and did not have to judge it based on its sound pattern. Both the centrally distributed N400 and the more posterior P300 are evident in the comparison of Studied and Novel-Fit items (Figure 3).

Although it has been argued that late positivities like our LPC (and the well-known P600) are instances of the P300 (e.g. Sassenhagen et al., 2014), in the present data the two components are distinct in timing as

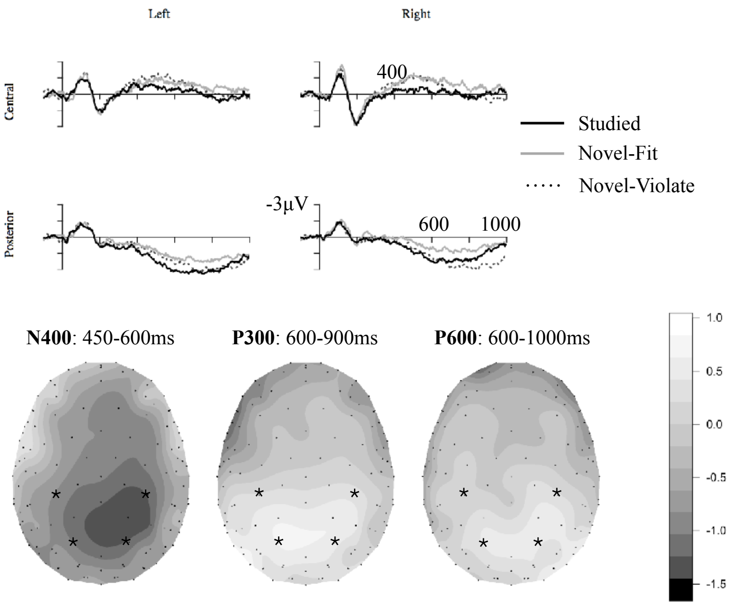

Figure 5: Studied, Novel-Fit, and Novel-Violate. Waveforms are time locked to the onset (vertical lines) of test items that had been studied (black), that were presented only during testing that fit the phonological pattern (blue), and that were presented only during testing but did not fit the phonological pattern (red). These data were measured at the four electrodes indicated (stars). Studied words elicited a smaller N400 and a larger P300. Novel-Violate items elicited a larger Late Positive Component (LPC) over posterior regions.

well as in the conditions they appear in. While the LPC is elicited by Novel-Violate items relative to Novel-Fit items (Figure 4), the P300 is elicited by Studied items relative to Novel-Fit items (Figure 3. While a sensible explanation exists for each of these differences independently, we can think of no account that unifies the Novel-Violate items and the trained items as opposed to the Novel-Fit items. Additionally, difference between Studied and Novel-Fit seems to begin earlier (these two conditions are also significantly different in the N400 window), and end earlier that the difference between Novel-Fit and Novel-Violate. The timing and distribution of all three components can be compared visually in Figure 5.

5 Conclusion

The participants in our experiment learned a dependency between the voicing of the two stop consonants of CVCV words[2]. They were exposed to a set of words obeying the restriction in the context of learning the meanings of the words, and then in testing they rated novel words that fit the restriction as more likely to belong to the language than novel words that violated it. From the study of EEG data collected during the experiment, we conclude that the phonotactic generalization is abstractly or grammatically represented, rather than the product of lexical search or analogy. The ERP response to the Novel-Violate items included a Late Positive Component (LPC), similar to that found in response to syntactic and musical harmonic structure violations. We did not find a difference in N400 amplitude between the Novel-Fit and Novel-Violate conditions, which would have been expected if differential lexical access of learned items contributed to listeners' acceptability ratings of the novel items.

The only previous neurophysiological study of the outcome of laboratory phonological learning of which we are aware is that of Wong et al. (2013), who focus on a distinction between the learning of what they call analogical and concatenative grammars, though for them analogy is a way of characterizing the knowledge of an opaque alternation (concatenation is simple addition of a suffix). Their focus is also different from ours in that they are concerned with individual differences in the learning of these two types of paradigmatic relation. The intersection of our study's concerns and theirs is a good topic for future research: are there individual differences in the learning of phonotactics in terms of a reliance on different neural subsystems?

Our results add to the broader literature on laboratory learning of language, in which there has been some previous ERP research on the outcome of morpho-syntactic acquisition in the lab. As in our study, Morgan-Short et al. (2012) show a relatively quick acquisition of an LPC. They also find a difference between implicit and explicit learning conditions, in that only implicit learning yielded an early anterior negativity (see Morgan-Short et al., 2015 on early negativities in naturalistically learned syntax). We did not find this component in our study, and it is an open question whether it will be observed in phonological violations,

---

[2] For half the participants the consonants always agreed, and for half they always disagreed. The fact that we saw no differences between the groups fits with a general lack of evidence for a difference between long-distance assimilation and dissimilation in the artificial phonology learning literature (Moreton and Pater, 2012). Some recent work indicates, however, that vowel harmony may have a learning advantage over vowel dissimilation (Rukoz, 2015).

which are less well-studied than syntax (Loui et al. (2009) find both an LPC and an early anterior negativity in responses to unexpected chords in a newly learned harmonic system). Again, the intersection between our study and this earlier work seems like a fruitful area for further work: do implicit and explicit training differentially affect phonological learning (see Moreton and Pertsova, 2015)?

Finally, a general implication of our result is that it lends support to the view that laboratory learning of phonology, while different in many ways from naturalistic acquisition, has some ecological validity, given the Domahs et al. (2009) finding discussed in the introduction that a long-distance restriction on the place of consonants in sCVC words in German also yields an LPC.

References

D. Birdsong. Age and second language acquisition and processing: A selective overview. *Language Learning*, 56(s1):9–49, 2006.

Mara Breen, John Kingston, and Lisa D. Sanders. Perceptual representations of phonotactically illegal syllables. *Attention, Perception, and Psychophysics*, 75(1):101–120, 2013. doi: 10.3758/s13414-012-0376-y.

Angela C. Carpenter. A naturalness bias in learning stress. *Phonology*, 27:345–392, 2010.

Ricardo E. Carrión and Benjamin Martin Bly. The effects of learning on event-related potential correlates of musical expectancy. *Psychophysiology*, 45:759–775, 2008.

Donna Coch, Giordana Grossi, Wendy Skendzel, and Helen J. Neville. Nonword rhyming effects in children and adults. *Journal of Cognitive Neuroscience*, 17:168–182, 2015. doi: 10.1162/0898929052880020.

Alejandrina Cristiá and Amanda Seidl. Is infants' learning of sound patterns constrained by phonological features? *Language learning and development*, 4:203–227, 2008.

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

Jennifer Culbertson. Typological universals as reflections of biased learning: evidence from artificial language learning. *Language and Linguistics Compass*, 6(5):310–329, 2012.

Ghislaine Dehaene-Lambertz, Emmanuel Dupoux, and A. Gout. Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience*, 12(4):635–647, 2000. doi: 10.1162/089892900562390.

Ulrike Domahs, W. Kehrein, J. Knaus, R. Wiese, and M. Schlesewsky. Event-related potentials reflecting the processing of phonological constraint violations. *Language and Speech*, 52(4):415–435, December 2009. doi: 10.1177/0023830909336581.

Emanuel Donchin. Surprise! surprise? *Psychophysiology*, 18(5):493–513, 1981.

Connie C. Duncan-Johnson and Emanuel Donchin. On quantifying surprise: the variation of event-related potentials with subjective probability. *Psychophysiology*, 14(5):456–467, 1977.

Ana C. Gouvea, Colin Phillips, Nina Kazanina, and David Poeppel. The linguistic processes underlying the p600. *Language and Cognitive Processes*, 25(2):149–188, 2010. doi: 10.1080/01690960902965951.

Phillip J. Holcomb, Jonathan Grainger, and Tim O'Rourke. An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience*, 14(5): 938–950, 2002.

Marta Kutas and Kara D. Federmeier. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62:621–47, 2011.

Marta Kutas and Steven A. Hillyard. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, 207(4427):203–205, January 1980. doi: 10.1126/science.7350657.

Marta Kutas and Steven A. Hillyard. Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307:161–163, 1984.

Yeeking Regine Lai. *Domain specificity in learning phonology.* PhD thesis, University of Deleware, 2012.

Psyche Loui, Elaine H. Wu, David L. Wessel, and Robert T. Knight. A generalized mechanism for perception of pitch patterns. *Journal of Neuroscience*, 29:454–459, 2009.

Judith McLaughlin, Darren Tanner, Ilona Pitkänen, Cheryl Frenck-Mestre, Kayo Inoue, Geoffrey Valentine, and Lee Osterhout. Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning*, 60(suppl. 2):123–150, December 2010. doi: 10.1111/j.1467-9922.2010.00604.x.

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

Elliott Moreton and Joe Pater. Structure and substance in artificial-phonology learning. *Language and Linguistics Compass*, 6(11):686–718, 2012.

Elliott Moreton and Katya Pertsova. Implicit and explicit phonology: What are artificial-language learners really doing? In *Handout from the Manchester Phonology Meeting*, Manchester, UK, 2015.

Elliott Moreton, Joe Pater, and Katya Pertsova. Phonological concept learning. *Cognitive science*, 41(1): 4–69, 2017.

Kara Morgan-Short, Karsten Steinhauer, Cristina Sanz, and Michael T. Ullman. Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience*, 24(4):933–947, April 2012.

Kara Morgan-Short, Mandy Faretta-Stutenberg, Laura Bartlett-Hsu, and P Rebuschat. Contributions of event-related potential research to issues in explicit and implicit second language acquisition. *Implicit and explicit learning of languages*, pages 349–84, 2015.

Scott Myers and Jaye Padgett. Domain generalisation in artificial language learning. *Phonology*, 31:399–433, 2014. doi: 10.1017/S0952675714000207.

M. Isabel Núñez-Peña and M. Luisa Honrubia-Serrano. P600 related to rule violation in an arithmetic task. *Cognitive Brain Research*, 18:130–141, 2004.

Lee Osterhout and P. J. Holcomb. Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31:785–806, 1992. doi: 10.1016/0749-596X(92)90039-Z.

Michael Paradis. *Declarative and Procedural Determinants of Second Languages*. John Benjamins, 2009.

Aniruddh Patel, Edward Gibson, Jennifer Ratner, Mireille Besson, and Phillip J. Holcomb. Processing syntactic relations in language and music: An event related potential study. *Journal of Cognitive Neuroscience*, 10:717–733, 1998. doi: 10.1162/089892998563121.

Joe Pater and Anne-Michelle Tessier. Phonotactic knowledge and the acquisition of alternations. In *Proceedings of the 15th International Congress of Phonetic Sciences.*, Barcelona, 2003.

Joe Pater and Anne-Michelle Tessier. L1 phonotactic knowledge and the L2 acquisition of alternations. In Roumyana Slabakova, Silvina Motrul, Philippe Prévost, and Lydia White, editors, *Inquiries in linguistic development: in honor of Lydia White*, pages 115–131. Benjamins, 2006.

Anne Pycha, Pawel Nowak, Eurie Shin, and Ryan Shosted. Phonological rule learning and its implications for a theory of vowel harmony. In M. Tsujimura and G. Garding, editors, *Proceedings of the 22nd West Coast Conference on Formal Linguistics (WCCFL 22)*, pages 101–114, 2003.

Michael D. Rugg. event-related potentials and the phonological processing of words and non-words. *Neuropsychologia*, 22:1984, 435-443 1984. doi: 10.1016/S0304-3940(01)02236-4.

Adriana Guevara Rukoz. The role of phonetic naturalness in shaping sound patterns: Evidence from artificial language learning and computational modeling. Master's thesis, EHESS - ENS - Université Paris Descartes, 2015.

Jona Sassenhagen, Matthias Schlesewsky, and Ina Bornkessel-Schlesewsky. The P600-as-P300 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, 137:29–39, 2014.

Michael T. Ullman. The neural basis of lexicon and grammar in first and second language: the declarative/procedural model. *Bilingualism: Language and Cognition*, 4:105–122, 2001.

Michael T. Ullman. A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In C. Sanz, editor, *Mind and context in adult second language acquisition*, pages 141–178. Georgetown University Press, 2005.

Cyma van Petten, Marta Kutas, Robert Kluender, Mark Mitchiner, and Heather McIsaac. Fractionating the word repetition effect with event related potentials. *Journal of Cognitive Neuroscience*, 3:131–150, 1991.

Colin Wilson. Experimental investigation of phonological naturalness. In G. Garding and M. Tsujimura, editors, *Proceedings of the 22nd West Coast Conference on Formal Linguistics (WCCFL 22)*, pages 533–546, Somerville, MA, 2003. Cascadilla Press.

Patrick C. M. Wong, Marc Ettlinger, and Jing Zheng. Linguistic grammar learning and DRD2-TAQ-IA polymorphism. *PloS one*, 8(5):e64983, 2013.