Is Phonological Consonant Epenthesis Possible? A Series of Artificial Grammar Learning Experiments

Rebecca L. Morley

Abstract

Consonant epenthesis is typically assumed to be part of the basic repertoire of phonological grammars. This implies that there exists some set of linguistic data that entails epenthesis as the best analysis. However, a series of artificial grammar learning experiments found no evidence that learners ever selected an epenthesis analysis. Instead, phonetic and morphological biases were revealed, along with individual variation in how learners generalized and regularized their input. These results, in combination with previous work, suggest that synchronic consonant epenthesis may only emerge very rarely, from a gradual accumulation of changes over time. It is argued that the theoretical status of epenthesis must be reconsidered in light of these results, and that investigation of the sufficient learning conditions, and the diachronic developments necessary to produce those conditions, are of central importance to synchronic theory generally.

1 Introduction

Epenthesis is defined as insertion of a segment that has no correspondent in the relevant lexical, or underlying, form. There are various types of epenthesis that can be defined in terms of either the insertion environment, the features of the epenthesized segment, or both. The focus of this paper is on consonant epenthesis and, more specifically, default consonant epenthesis that results in markedness reduction (e.g. Prince and Smolensky (1993/2004)). Consonant epenthesis intervocalically meets these requirements; it transforms a vowel-initial syllable to a less-marked consonant-initial syllable. See the example in (1).

(1) Misantla Totonac ((MacKay, 1999))

/laa+a̰ʔḭʃki/→	[laaʔa̰ʔḭʃki]
	3PL.OBJ-lend
	'lend them'
/ta+a̰n/→	[taʔa̰n]
	3PL.SUBJ-go
	'they go'
/naa+ṵtun/→	[naaʔṵtun]
	also-they
	'they also'

That this operation is a natural phonological process is implicitly built into modern generative theory, either as a high-valued re-write rule (by the diagnostic of Chomsky and Halle (1968), or as instantiated in a universal constraint *against* epenthesis – one that is crucially violable. Thus, given exposure to the right kind of linguistic data, the learner should be able to acquire the epenthesis grammar. Determining the necessary conditions for that acquisition, however, is a non-trivial problem.

There are several sources of ambiguity in linguistic analysis. For example, in the alternations provided in (1), the analysis "insert a consonant after a low vowel" predicts the same (correct) outcome as the more general analysis "insert a consonant after any vowel". Also consistent with the available data, is the deletion of an underlying /ʔ/ in the environment before a consonant. With the collection of more data some of these ambiguities might be resolved. However, in practice, ambiguity cannot be completely eradicated (for the learner, if not the linguist) because most real-world data are messy and incomplete.

Lexical exceptions have a long history in phonological theory. Yet the phonological learning literature contains a conspicuous gap regarding the acquisition of forms that represent exceptions to a general pattern (but see Yang(2005, 2011) for a metric of morphological productivity). This lack becomes clear when we allow for the analysis that there is no regular pattern at all. This paper is an experimental investigation of the ambiguity that is at the heart of linguistic analysis. Specifically, I ask whether learners favor certain analyses over others (epenthesis, deletion, or suppletive allomorphy), and what properties of the input push them toward one analysis or another (cf. Moreton and Pater (2012))[1].

1.1 Overview

The results of 12 artificial grammar learning experiments are reported in this paper. The conditions are divided up into three thematic sets. In all conditions, the training data to which participants are exposed consist of a small set of nonce words, and the morphophonological alternations that arise when those words are inflected in the plural. The surface forms are ambiguous with respect to the underlying forms from which they were generated, and thus there are multiple possible analyses consistent with the data. After training, participants are tested with novel forms that they hear in the singular. They are then prompted to produce the plural by speaking into a microphone. An example of two training pairs is given in (2). Possible learner analyses consistent with this input follow in (3 a)-(3 d). Epenthesized segments are indicated in bold.

(2)

Sg.	Pl.	Sg.	Pl.
[skibe]	[skibejək]	[ɹatu]	[ɹatuwək]

(3) (a) Epenthesis Analysis (Phonological)
/ɹatu+ək/→[ɹatu**w**ək] /skibe+ək/→[skibe**j**ək]

(b) Epenthesis Analysis (Phonetic)
/ɹatu+ək/→[ɹatu$^{\mathbf{w}}$ək] /skibe+ək/→[skibe$^{\mathbf{j}}$ək]

(c) Suppletive Allomorphy Analysis
/ɹatu+wək/→[ɹatuwək] /skibe+jək/→[skibejək]

(d) Deletion Analysis
/ɹatuw+ək/→[ɹatuwək] /skibej+ək/→[skibejək]

Participants might conclude that there is a unique plural suffix, and a process of epenthesis conditioned by vowel context, as depicted in (3 a). This analysis competes with the deletion analysis, in which the glide belongs to the word stem, and is deleted in the singular (word-finally) (3 d). When the potential epenthetic segment is a glide homorganic with the stem-final vowel, a third possibility arises, what I will call phonetic epenthesis. This is depicted in (3 b), with the superscript glide to distinguish it from phonological epenthesis. The morphological, or suppletive allomorphy, analysis in (3 c) assigns the ambiguous glides to the suffix; the correct allomorph must be memorized for each stem.

The analysis chosen by the participants was assessed using held-out test items that provided a non-ambiguous environment for perceiving glides. For the example above, these consisted of consonant-final stems like [daɹum]. However, all analyses except suppletive allomorphy predict the same plural form: [daɹumək]. Thus, it can only be determined whether learners did or did not adopt the suppletive allomorphy analysis. The phonetic epenthesis analysis will not be available for all conditions, but the deletion analysis

[1]Some of these results have been previously reported in Morley (2011).

will always be a competitor, either making the same predictions as the phonological epenthesis analysis, or making no predictions for the relevant test items. Because this holds over all conditions, discussion of the deletion alternative will be reserved for Section 3.3, where the analytic problem is discussed in more detail.

There are two additional analytical issues that will also be discussed separately, but will be briefly described here. The predictions given for the analyses above assume idealized, error-free learning. Errors, or inconsistencies, present a scenario where it is much more difficult to assess the generating grammar. If participants produce some forms that are consistent with a rule (of epenthesis, or deletion), and some forms that are not, then all of the above analyses are possible, with the addition of an error term. In some cases it may be possible to make a decision based on the majority pattern (i.e., if there are relatively few inconsistencies). However, this is a problem, in principle, with linguistic analysis, and it also crops up in the idealized learning scenario. In the case where all data are available to the learner, the suppletive allomorphy analysis can never be ruled out. This is because it produces the correct allomorphs for all stems under perfect learning (and perfect generalization to novel stems). This issue will be discussed in Section 5.4.

As we will see from the full set of conditions, participants, in the aggregate, appeared to adopt either the suppletive allomorphy analysis or the phonetic analysis, never the phonological epenthesis analysis. This is the first main finding. The second is that learners showed a strong preference for morphological uniqueness, choosing the analysis that would result in a single plural morpheme. This was observed as a switch in the interpretation of an ambiguous glide from subsegmental to segmental across conditions. The 'Morpheme Uniqueness' pressure, however, was not strong enough to induce learners to adopt a maximally general rule – either deletion or epenthesis. These results suggest that learning true phonological generalizations might be both harder and less common than previously supposed.

The details of the experimental methodology are described in the following section. Beginning in Section 2.2, the results of each individual condition are given in context. Besides the main findings, a number of other results emerged. Learner accuracy went down as the number of allomorphs increased. Token frequency and within-category variability also affected the error rate. On average, participants chose allomorphs according to their frequency during training. However, individual participants often 'boosted' allomorphs beyond their training frequencies, i.e. regularized their input. The majority of participants who regularized chose the most frequent variant to 'boost'. But there were some participants who chose a minority allomorph instead. Variability seemed to inhibit regularization, with learners more closely matching the training frequencies. In the absence of the Morphological Uniqueness pressure a bias towards the subsegmental analysis of ambiguous glides was found. In Section 3 a global analysis of the results is given, and the body of findings is summarized in Section 4. A diachronic analysis of consonant epenthesis is given in Sections 5.1 and 5.2, and English intrusive r is discussed as a case study in Section 5.3. The paper concludes in Section 5.4 with the implications of these results for synchronic phonology.

2 Experiments

The experiments in this paper use an artificial grammar learning (AGL) paradigm. For a partial review of the phonological AGL literature, see Moreton and Pater (2012), also Peperkamp et al. (2006). The use of AGL in the study of language change allows one to simulate different stages in the diachronic evolution of a particular linguistic structure (see Culbertson and Legendre (2010), Kam and Newport (2009)). While the methodology does not eliminate native language bias, informing participants that they will be learning a new language (or a made-up language) has been shown to be effective in achieving learning of non-native language patterns. Participants can learn novel words and patterns quickly within this paradigm, and seem to implicitly acquire grammatical rules of which they are consciously unaware.

2.1 Procedure

A total of 234 participants were run in the three sets of experiments (numbers of participants by condition are given in Table A.2). All were undergraduates participating for course credit in introductory Linguistics courses at the Ohio State University. Participants were seated in front of a computer screen within a sound-attenuated booth. They listened to audio input over headphones, viewed images on the screen, and spoke into a microphone when prompted. Continuous audio was recorded in Praat (Boersma and Weenink (2009)) or Audacity(2014) over the entire experimental session.

2.1.1 Design

All participants were told that they would be hearing words in a new language, and would later be asked questions about those words. During a passive training stage, participants were exposed to auditory and visual stimuli presented in singular/plural pairs. A picture of a single object (e.g. an apple) accompanied playback of the word (for apple) in the singular, and a picture of two of the same object (e.g., two apples) accompanied playback of the word in the plural. The singular and plural were clearly related; the plural consisting of the singular plus the addition of a suffix (e.g., [ˈɹatu] and [ˈɹatuwək]). See Figure 1. At no point were participants alerted to the fact that there would be variation in the form of the plural marker, or asked about alternations. By self-report, participants were often unaware of phonetic differences in the suffix morpheme, and of the alternation ‘rule’ that they were being exposed to.

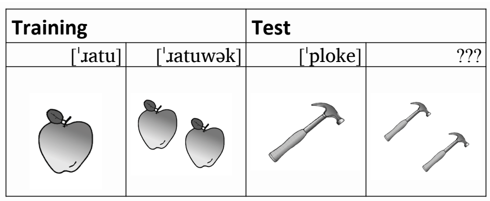


Figure 1: Training and Test Paradigm: Sequential presentation of singular and plural forms

Participants heard 12 or 18 unique singular-plural word pairs, depending on experimental condition. This set of words was repeated twice in randomized order, within each of 2 training blocks. A feedback stage occurred at the end of the first training block, and again after the second training block, immediately preceding the test block. The responses for the feedback trials were recorded but not analyzed. During feedback a familiar singular form was heard, and participants were prompted for the plural: “Now you say the plural. . . ”. Participants were instructed to speak their response into the microphone. After an interval of 5 seconds the words “The correct answer is. . . ” appeared on the screen and the plural form was played over the headphones. The procedure was the same at test, except the correct answer was not provided; additionally, participants were exposed to singular forms they had never heard before. It should be noted that although there was a distinction between stem *types* that were familiar or novel (e.g., front-vowel final vs. consonant-final stems), all test words were new words (see Table 1 for the complete list of stimuli used). In all conditions the test items were comprised of all three stem types in equal proportions. There were 6 unique words of each type: front-vowel final, back-vowel final, and consonant-final stems. Each word was repeated twice, in randomized order, for a total of 36 test items. All participants in a given condition heard the same set of words, associated with the same pictures; the order of presentation, however, was randomized. The entire experiment took roughly half an hour. No participant was run in more than one condition, therefore all comparisons reported are between-subject.

2.1.2 Stimuli

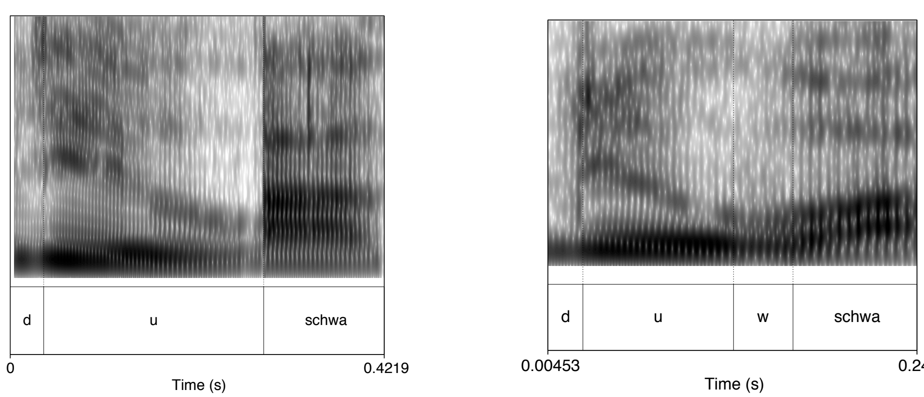
Each experiment utilized a subset of a common pool of auditory stimuli. All stimuli were recorded by a phonetically trained female speaker of American English. Singular and plural words were recorded separately. The full set of singular forms (stems) is given in Table 1. All words were stressed on the penultimate syllable of the stem, in both singular and plural forms. All plural forms were of the form singular + Xək. X was either zero, a glide homorganic to the place of the preceding vowel, an anti-homorganic glide, an obstruent (p, t, tʃ, or k), or a pause/discontinuity, depending on experiment and condition (e.g., skibeək, skibejək, skibewək, skibekək, and skibe-ək, respectively).

Table 1: Full list of stimuli across all conditions, both training and test item stems.

ˈɹatu	ˈɹɪlo	fɹaˈbomu	tʃoˈɹæno	kɹo/zo	vu
ˈhædi	ˈskibe	teˈlʌpi	gluˈdɛbe	fi	sme
ˈpiʃu/ˈgɛθu[a]	ˈhago	bəˈhʌʒu	ˈfædʒo	zo/kɹo	ˈgaidu
ˈvʌlki	ˈploke	ˈdʒimi	diˈʒaɹe	ˈθuzi	ˈʃuvi
ˈdaɹum	keˈtɛlan/d͡ʒoˈɹʌfim[a]	ˈhoʃɪn/ˈt͡ʃalum[a]	ˈɹibæz	ˈpɹɛv	ˈbihɪl
ˈtwit͡ʃo	ˈðipu	ˈmuɹo	ˈtɹɪfu	ˈmeko	ˈgɛθu/ˈpiʃu[a]
ˈsabol	ˈgenʊɹ	ˈtɹifæd	(ˈtagæf)	d͡ʒoˈɹʌfim/keˈtɛlan[a]	ˈt͡ʃalum/ˈhoʃɪn[a]

[a]items separated by / indicate substitutions: a training item used as a test item, and vice versa, for a particular condition. See Table A.2

It is difficult, if not impossible, to avoid producing a glide-like transition between adjacent vowels of a certain type. It is similarly problematic to distinguish between glides in onset position, and onset-less vowels in unstressed syllables. For example, productions of [ɹatuwək] and [ɹatuək] were auditorily and spectrographically highly similar. For this reason, a set of spliced stimuli were created to ensure the absence of glide-like material. The token [gaidu-ək], for example, was created by recording a single utterance with a long pause: [gaidu ək], then splicing out the pause, as well as the very beginning of the final vowel, to minimize glottalization. See Figure 2. The resulting stimuli exhibited a discontinuity in the spectrogram, and as a result sounded quite unnatural. However, the results show that such tokens were plausible enough to be categorized by listeners as reflexes of underlying vowel-vowel sequences. Participants developed various strategies for reproducing these unnatural words: altering vowel quality, introducing glottal stops, or significantly drawing out their articulations.



(a) spliced, in which transitions are removed: [gaidu-ək]

(b) unspliced, natural production: [gaiduək]

Figure 2: Spectrograms of Example Stimuli

All data analyzed in this paper were coded only for the unambiguous presence of a consonantal segment. That is, no attempt was made to disambiguate forms such as /ɹatuwək/ and /ɹatuək/ in participants' productions. Responses that were ambiguous were coded as being consistent with a VV production. In most cases, however, the segments were unambiguous. All consonant-final stems could be clearly distinguished in inflected form (e.g., [daɹumwək] versus [daɹumək]). The same was true of vowel-final stems with following heteroganic glides or obstruents[2].

2.2 Experiment 1: Learning Biases

The first set of experiments was designed to determine whether learners will adopt an epenthesis analysis under consistent, but incomplete, data. Table 2 provides examples of the singular and plural training stimuli for each condition; the number of participants in each condition (N); and the token counts (unique stems) for each stem type. Of three stem types, one is held out during training in each condition. The held-out stems provide a test of learner generalization. In the first two conditions the training input was completely consistent with a phonological epenthesis analysis. In the third, the pattern was partially consistent; and in the fourth, not consistent. The predictions for novel forms under each of the possible analyses in (3 a, 3 b, and 3 c) will be given by condition below.

Table 2: Experiment 1: Learning Biases: Example Training Stimuli

	Condition	N	Back Vowels	Front Vowels	Consonants	Token Counts		
1	Natural	18	ɹatuwək	skibejək		6	6	
2	Anti-Natural	18	ɹatujək	skibewək		6	6	
3	Variable0	20	ɹatuwək	hædijək		3	3	
			ɹɪlo-ək	skibe-ək		3	3	
4	Consistent-W	20	ɹatuwək		daɹumwək	6		6

2.2.1 Conditions 1 & 2: Natural and Anti-Natural Epenthesis

In the first two conditions participants heard only vowel-final stems during training. These consisted of 6 unique back-vowel final stems (three each ending in /o/ and /u/), and 6 unique front-vowel final stems (three each ending in /e/ and /i/). The singular forms were the same in both conditions. In the "Natural" condition, the plural suffix began with a glide homorganic with the stem-final vowel. In the "Anti-Natural" condition, plurals contained glides anti-homorganic with the stem-final vowel. In both conditions, training stimuli were consistent with an epenthesis analysis, in which either [w] or [j] was inserted between stem final and suffix initial vowels. However, training was also consistent with other analyses. Given this ambiguity it was expected that different participants would adopt different analyses, but that biases for one or more of the analyses would emerge.

The relevant test items were the held-out consonant-final stems. These items unambiguously show the presence or absence of a pre-suffix glide. If participants learned a phonological epenthesis pattern, then they should inflect consonant-final stems without a glide. This is also the predicted result for the phonetic epenthesis analysis. See (4). If participants instead inferred that there were two allomorphs, then they would be forced to guess about the proper inflection of consonant-final stems. I will adopt the hypothesis that participants would produce both allomorphs in equal proportions under the suppletive allomorphy analysis.

(4) Natural & Anti-Natural Conditions

	Phonological Analysis	Phonetic Analysis	Suppletive Allomorphy Analysis
C-Final Stems: [daɹum]	[daɹumək]	[daɹumək]	[daɹumwək];[daɹumjək]

[2]Because of this no test was made of inter-transcriber reliability.

2.2.2 Results

Average rate of glide production on held-out stem types are given in Table 3, by condition. The low rate of glide responses in the Natural condition seems to rule out the suppletive allomorphy analysis, but is consistent with both the phonetic and the phonological epenthesis analyses. Conversely, the result in the Anti-Natural condition is consistent with the suppletive allomorphy analysis, but not with phonological epenthesis (the phonetic analysis is not available, as the glides are anti-homorganic with the stem vowel). The failure of participants to analyze the data in the Anti-Natural condition as phonological epenthesis suggests that they also failed to do so in the Natural condition. If so, the likely conclusion is that participants in the Natural condition analyzed the glide as subsegmental - the result of coarticulation.

Table 3: Experiment 1: Average glide response on held-out stems

Condition	Glide Response %	
	mean	sd.
Natural	.119	.207
Anti-Natural	.944	.164
Variable0	0	0.00
Consistent-W	.954	.205

2.2.3 Condition 3: phonetics or phonology?

A third condition was designed in order to bias participants towards a segmental analysis of the surface glides, and in so doing, corroborate the phonetic analysis of the Natural condition data. The “Variable0” condition differed from the previous two in that two types of inflected forms were heard: phonetically natural tokens (identical with stimuli from the Natural condition) as well as spliced tokens from which all traces of excrescent glides were removed, resulting in V-V sequences (see Section 2.1.2). A subset of stems appeared with the phonetic glide, and a different subset appeared with the spliced plurals. The suffix allomorph was not predictable by stem type, but an individual stem was always inflected in the same way (both within and across participants). See Table 2. It was hypothesized that the spliced tokens would, by comparison, force a segmental interpretation of the naturally produced tokens. This would result in participants learning three (partially predictable) plural allomorphs: -/wək/, -/jək/, and -/ək/.

2.2.4 Results

As the results in Table 3 demonstrate, this was not what happened. Participants produced no glides on consonant-final stems whatsoever. Despite the clear acoustic differences between the two types of training items, participants seemed to treat them as variants of the same underlying form. This interpretation is supported by a separate orthographic version of the Variable0 condition. Participants heard auditory stimuli as before, but were instructed to write their responses, giving their best guess as to the spelling. Not a single consonant-final, or back-vowel final, stem was written with a glide in the plural[3]. Typical responses included, e.g., for the trained form [ɹatuwək], “ratuuck”, “ratuak”, and “ratoouk” (there were 2 participants who used the ’ character for all test items, e.g. “ratu’ak”). These results imply that participants were strongly biased towards the phonetic interpretation of ambiguous phones in both the Natural and Variable0 conditions.

[3]The only orthographic glides occurred with front-vowel final stems. There were 7 participants who occasionally wrote plural forms with a “y”, for a total of 22 tokens, or an average of 8.7% of their responses.

2.2.5 Condition 4: Morphological Uniqueness

Because the Variable0 condition failed to elicit a segmental glide analysis, a further condition was designed for that purpose. In the Consistent-W condition participants were trained on unambiguous /w/'s in the plural forms of consonant-final stems. Back-vowel final stem plurals were identical to those used in the Natural condition: ambiguous with respect to whether they contained a segmental or a subsegmental [w]. Front-vowel final stems were held out. See Table 2. The Phonetic Bias hypothesis (5) predicts that participants will analyze the glide in forms like [ɹatuwək] as subsegmental. Thus, participants should learn two allomorphs of the plural suffix: -/ək/ and -/wək/. Further, if participants learn that -/ək/ occurs on vowel-final stems, then there should be no -/wək/ at all with front-vowel final stems.

(5) Phonetic Bias: Surface forms ambiguous between a segmental and sub-segmental analysis are predominately analyzed as sub-segmental

However, the presence of the segmental glide in forms like [daɹumwək] may increase the likelihood that participants analyze the ambiguous forms as containing a segmental glide. This analysis would result in a single surface form for the plural suffix. If an imperative exists to limit the allomorphs of a given morpheme, then the use of -/wək/ with all stem types is predicted. This will be called the Morphological Uniqueness Hypothesis[4]:

(6) Morphological Uniqueness Hypothesis: If available, the analysis that results in a unique realization of the morpheme will be chosen

The predicted outcomes for held-out front-vowel final stems under the two hypotheses are given in (7):

(7) Consistent-W Condition

	Phonetic Bias Hypothesis	Morphological Uniqueness Hypothesis
Front-Vowel Final Stems	skibeək / skibejək	skibewək

2.2.6 Results

As Table 3 shows, participants in the Consistent-W condition produced -/wək/ on 95% of front-vowel final stems, indicating that they had largely learned a single form of the plural. A logistic regression model of proportion glide response by condition found that both the Anti-Natural and Consistent-W conditions were significantly different from the Natural condition (reference level), but not from each other. While the Variable0 condition was not significantly different from the Natural condition (each participant was treated as a single measurement of proportion glide response over all items from the held-out stem type)[5]. See Table 4. This confirms the prediction of the Morphological Uniqueness Hypothesis. The presence of the segmental /w/ completely reverses the subsegmental analysis observed in the Natural condition.

Table 4: Experiment 1: Logistic regression: proportion glide response on held-out stems by condition. A single measure of percentage glide response was calculated for each participant.

	Est	SE	Pr(>|z|)
(Intercept)	-1.99	.209	<2e-16***
Condition = Natural	(*reference level*)		
Condition = Anti-Natural	4.76	.364	<2e-16***
Condition = Variable0	-18.2	993	.985
Condition = Consistent-W	5.05	.373	<2e-16***

[4]Based on subsequent conditions, the effect seems to be limited to cases in which the number of allomorphs is reduced to 1. In cases where the number of allomorphs could be reduced from 3 to 2, for example, no bias was observed.

[5]All analyses were performed using the R software environment version 3.3.3 and the lme4 package

2.2.7 Interim Summary

In the Anti-Natural condition, participants seem to have learned two glide-initial allomorphs for the plural suffix (e.g., [daɹumwək] or [daɹumjək]). The auditorially distinct tokens in the Variable0 condition, however, resulted in a uniform subsegmental analysis. This is plausible under a model in which listeners have phonetic expectations based on their experience with spoken language (e.g., Whalen 1984, Fowler 1984). Specifically, the fact that the two acoustically distinct plural forms could be classified together is attributable to speaker knowledge that carefully articulated vowel-vowel sequences result in an intrusive pause or glottal stop; whereas rapid, or colloquial, speech often produces phonetic gliding between two vowels (otherwise termed 'intrusive', 'linking', 'excrescent', or 'transitional' segments (e.g., Gick 1999, Browman and Goldstein 1990). The results of the Consistent-W condition can be explained by a morphological uniqueness bias that is stronger than a subsegmental phonetic bias. However, the results of the Natural condition are also explainable by the morphological uniqueness bias alone. Therefore, these conditions alone cannot tell us what bias, if any, participants have for a subsegmental versus segmental interpretation of ambiguous glides.

2.3 Experiment 2: Evidence & Allomorphs

The results of Experiment 1 do not provide support for either the phonological epenthesis analysis, or the segmental analysis of ambiguous glides. Experiment 2 tests new training conditions with the goal of pushing learners towards either of these outcomes. In the Full Alternation condition, the training stimuli represent a completely consistent and exhaustive (within the world of the experiment) set of evidence for a process of dissimilatory epenthesis. See Table 5. This pattern, however, is also consistent with a deletion analysis in which underlyingly glide-initial suffixes (-/wək/, and -/jək/) lose their onsets after stem-final consonants[6]. If participants learn the conditioning environments for the three surface alternants perfectly, it cannot be determined what analysis they have adopted. Errors, however, can provide insight into what was actually learned.

Table 5: Experiment 2: Evidence & Allomorphs

	Condition	N	Back Vowels	Front Vowels	Consonants	Token Ratio		
5	Full Alternation	18	ɹatujək	skibewək	daɹumək	6	6	6
6	CTrained1	24	ɹatuwək		daɹumjək	6		6
7	CTrained2	19	ɹatujək		daɹumwək	6		6

The CTrained1 and CTrained2 conditions were designed to test two factors: the salience of the vowel/consonant natural class as a conditioning environment, and the effect of unrelated segmental glides on the analysis of ambiguous glides. It was hypothesized that the difference between consonant-final and vowel-final stems would be highly salient to learners, and that being exposed to the distinction during training would lead to high accuracy learning; this will be termed the CV Hypothesis.

(8) CV Hypothesis: Transfer rates should be low across a category boundary that is ±consonantal

The CV Hypothesis predicts higher accuracy in the CTrained1 and CTrained2 conditions than in conditions like the Anti-Natural condition, where consonant-final stems were held out. It was also hypothesized that the presence of *any* unambiguous glide during training might increase the rate of the segmental analysis for the ambiguous glide; this will be termed the Attractor Hypothesis.

[6] The analysis in which certain stem types end in underlying glides, which are deleted word-finally, is also possible. This analysis will be discussed in Section 3.3.

(9) Attractor Hypothesis: The presence of an unambiguous segment will facilitate the segmental analysis of ambiguous forms when the two segments are of the same type

There are numerous possibilities for what 'same type' might refer to; but for our purposes, segments will be considered to be of the same type if they are both glides. The Attractor Hypothesis then predicts higher rates of [skibewək] response types in the CTrained1 condition than in conditions like the Natural and Variable0. Neither CTrained1 nor CTrained2, however, is consistent with a single-morpheme analysis. It was therefore predicted that participants would learn a minimum of two allomorphs in each of those conditions.

2.3.1 Results

Fig. 3 shows the proportion of responses by allomorph[7] on consonant-final stems for each of the three conditions in Experiment 2. In all three conditions participants heard consonant-final stems during training, therefore they largely responded accurately: -/ək/ in the Full Alternation condition, -/jək/ in the CTrained1 condition, and -/wək/ in the CTrained2 condition. In all conditions, however, consonant-final stems were sometimes produced with vowel-final allomorphs. The pattern of errors in these conditions, however, can provide some insight into participant analyses. In the CTrained1 condition, 4.65% of responses contained [wək], while 16.2% of responses contained [ək]. At 8.4%, the [wək] rate was somewhat higher for front-vowel final stems (which were held out during training)[8]. However, this rate is actually lower than the segmental response rate on consonant-final stems in the Natural condition (see Table 3). These subsegmental to segmental response rates fail to support the Attractor Hypothesis.

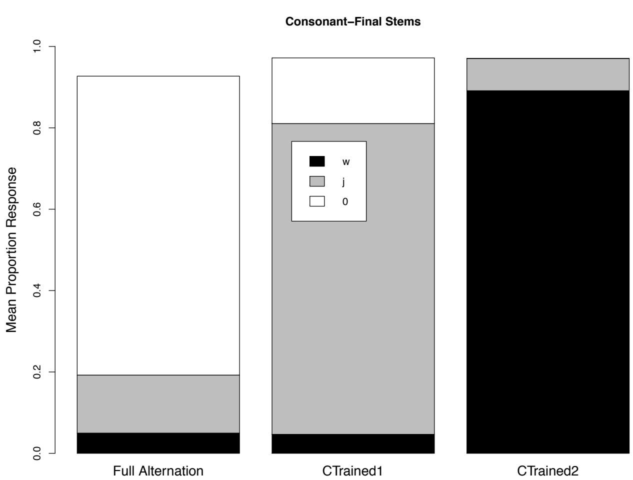


Figure 3: Experiment 2: Consonant-final stems: Mean proportion response rate by allomorph.

Despite expectations, response accuracy on trained stem types (consonant-final and back-vowel final taken together) in the Full Alternation condition was lower than in the conditions with held-out data. This means that learners were worse at learning the conditioning environments for the different allomorphs, and thus were worse at producing results consistent with an epenthesis pattern. Table 6 shows proportion correct for the three conditions, and for the two trained stem types.

[7]Errors in which participants either failed to respond, corrected their original response, or produced a form that could not be attributed to any of their training data (e.g., [ɹatumək]) were classified as 'other' responses. Such responses constituted less than 3% of total responses averaged across all conditions. They were discarded prior to calculating response accuracy.

[8]Note that the [ək] response rate cannot be accurately determined for front-vowel final stems because of its phonetic similarity to [jək].

Table 6: Experiment 2: Proportion correct by stem type: mean and standard deviation.

	Consonant-Final		Back Vowel-Final	
Conditions	mean	sd	mean	sd
Full Alternation	.788	.246	.590	.308
CTrained1	.780	.299	.753	.303
CTrained2	.920	.188	.865	.311

A logistic regression model with accuracy as the dependent variable, and fixed effects for condition (CTrained1 reference level), stem type (back-vowel reference level), and the interaction between the two, is shown in Table 7. Total proportion correct was calculated for each participant, separately for consonant-final stems, and back-vowel final stems. There was a main effect of condition: accuracy decreased significantly from the CTrained2, to the CTrained1, to the Full Alternation condition. Stem type failed to reach significance, but there was a significant interaction between stem type and condition. The decrease in accuracy in the Full Alternation condition was due to the back-vowel final stems. In other words, accuracy on consonant-final stems in the Full Alternation condition was significantly higher than predicted by the model without the interaction term.

Table 7: Logistic regression: proportion correct (one measure per participant) as a function of condition, stem type, and the interaction between condition and stem type. Condition reference level = CTrained1; Stem Type reference level=Back vowel.

	Est	SE	Pr(>|z|)
(Intercept)	1.085	.133	2.82e-16***
Condition = CTrained1	*(reference level)*		
Condition = CTrained2	1.055	.253	3.13e-05***
Condition = Full Alternation	-.711	.193	.000227***
Stem Type = Back Vowel	*(reference level)*		
Stem Type = Consonant	.366	.208	.078
Condition = CTrained2:Stem Type = C	-.077	.392	.844
Condition = Full Alternation:Stem Type = C	.607	.309	.049*

It is possible to explain the three-way difference in accuracy in terms of interference from competitor allomorphs during training: error rate increasing with increasing number of allomorphs. The Full Alternation condition contains three unambiguous allomorphs, while the CTrained2 condition contains two completely unambiguous allomorphs. They are the conditions with lowest and highest accuracy, respectively. The CTrained1 condition, intermediate in accuracy, can also be characterized as containing something more than two allomorphs, but not quite as many as three, due to the ambiguity of the surface glide. That accuracy on vowel-final stems would be affected more than on consonant-final stems is predicted under the CV Hypothesis. Front-vowel final stem allomorphs are expected to transfer more to back-vowel final stems, than to consonant-final (Full Alternation condition). And ambiguous allomorphs are associated with the back-vowel final stem, contributing some variability to that stem class (CTrained1 condition).

In order to fully test the CV Hypothesis, however, conditions in which participants were trained on both consonant-final and vowel-final stems must be compared to conditions in which consonant-final stems were held out. The Anti-Natural condition was chosen for this comparison because training consisted of two equally frequent and fully predictable allomorphs, neither of which was ambiguous. Over-all accuracy in the Anti-Natural condition was 71.7%. The results of a logistic regression with accuracy as the dependent variable show that participants achieved significantly higher accuracy in both of the CTrained conditions. See Table 8. Even though there were still cross-over errors between vowel-final and consonant-final stems, there were significantly fewer than those that occurred between vowel-final stem types in the Anti-Natural condition; the CV Hypothesis was confirmed in this case.

Table 8: Logistic Regression: proportion correct as a function of condition for all trained stem types combined.

	Est	SE	Pr(>|z|)
(Intercept)	.939	.106	<2e-16***
Condition = Anti-Natural	(*reference level*)		
Condition = CTrained1	.305	.147	.038*
Condition = CTrained2	1.33	.195	9.09e-12***

One type of participant error should be mentioned at this point, as it is relevant for determining accuracy. In some instances participants produced consonant-final forms that lacked stem-final consonants (e.g., plural of /daɹum/ as [daɹuwək]). We can't know if participants would also have dropped the consonant in the singular, as they were never asked to produce those forms. However, it was inferred that the errors arose from an occasional failure to perceive stem-final consonants. For this reason, responses like this were coded as vowel-final, rather than consonant-final. The effect was to reduce the number of consonant-final test items, and increase the number of vowel-final items. If the errors were restricted to test items then the results would be relatively unaffected. However, it is reasonable to suppose that participants also misheard a comparable number of consonant-final stems during training. This is relevant only to the Full Alternation, CTrained1, and CTrained2 conditions. The effect in those conditions would have been to introduce inconsistency in the vowel-final allomorphs. There were not a large number of these errors (at test), but they were heavily skewed toward back, rather than front, vowels. That is, stems like [genʊɹ] and [sabol] were more likely to be heard as vowel-final, than stems like [ketɛlan] and [hoʃɪn]. There were a total of 10 such responses, distributed over 6 participants, in the Full Alternation condition (accounting for 5% of all responses); a total of 11 such responses, distributed over 6 participants in the CTrained1 condition (accounting for 5% of all responses); and a total of 20 such responses, distributed over 12 participants in the CTrained2 condition (accounting for 7% of all responses). These errors may be partially responsible for the fact that accuracy on back-vowel final stems is lower than on consonant-final stems in these three conditions. However, accuracy was still higher in the CTrained conditions than the Anti-Natural; and the CTrained2 condition, with the most errors, had the highest accuracy.

2.3.2 Interim Summary

The presence of the unambiguous /j/ on consonant-final stems in the CTrained1 condition did not lead to a higher rate of segmental analysis for ambiguous [w]. Therefore, no support is found for the Attractor Hypothesis. If the additional data in the Full Alternation condition biased learners towards an epenthesis analysis, then accuracy should have been high. But over-all accuracy was lower than in conditions in which participants were trained on only one of the vowel-final stem types. Error rates were significantly higher when those errors occurred across front/back categories, than when they occurred across the consonant/vowel categories, supporting the CV Hypothesis to a certain extent. Accuracy on back-vowel final stems was consistently lower than accuracy on consonant-final stems in all three conditions. This may be due to a higher error rate across vowel-final stems than between vowel-final and consonant-final stems. It may also be due, in part, to errors during training in which certain consonant-final stem types were misperceived as vowel-final (thus increasing the variability of the vowel-final class of stems. See section 2.4).

2.4 Experiment 3: Generalization

It has often been argued that learners show strong biases towards generalizing linguistic data in “natural” ways: extending patterns only to forms that belong to the same natural class in terms of phonological features, or to a class of segments standing in an implicational relationship (e.g., Wilson (2006), Berent et al. (2009)). It has also been argued that learners are predisposed to augment, or adapt, linguistic input that is non-optimal in some way (e.g., Bickerton (1984)). More generally, there is reason to believe that there may be a point of memory load beyond which the cognitive system begins to regularize its input (Memory Load Hypothesis) (e.g., Estes (1972)). There is also support for the hypothesis that the most frequent pattern or element in a language learning paradigm will be ‘boosted’, such that it becomes the only pattern or element (e.g., Hudson Kam and Newport (2005), Culbertson and Legendre (2010)). It is possible, therefore, that inconsistent learning data might actually lead to more consistent response patterns, and, as a result, to patterns that are closer to a default epenthesis rule. Experiment 3 explores how non-uniformity, variability, and irregularity during training affect pattern learning.

2.4.1 Conditions 8-10: Frequency

Back-vowel final stems occur twice as often as front-vowel final stems in the three “HI-Freq” conditions (by both type and token counts). Because allomorphs are predictable and consistent by vowel type, back-vowel final allomorphs also occur twice as often. See Table 9. Consonant-final stems are held out in all conditions. Based on previous results, it is expected that learners will use all trained allomorphs, to some degree, on consonant-final stems. Frequency matching – participants using allomorphs in proportion to their training frequency – is the Null Hypothesis for these conditions. The Alternative Hypothesis (Frequency Boosting) is that the higher-frequency allomorph will be selected disproportionally – boosted beyond its training frequency.

Table 9: Experiment 3: Generalization: Conditions 8-10

	Condition	N	Back Vowels	Front Vowels	Consonants	Token Ratio		
8	HI-Freq-t	16	ɹatutək	skibejək		12	6	
9	HI-Freq-0	19	ɹatu-ək	skibewək		12	6	
10	HI-Freq-j	22	ɹatujək	skibe-ək		12	6	

2.4.2 Results: Continuous measures

As expected, the more frequent allomorph was selected more often than the less frequent allomorph(s), although never exclusively. See the first three bars of Fig. 4. A one-sided Wilcoxon signed rank test with an alpha level of .05 shows that the high frequency allomorph was not selected significantly more often than its baseline frequency in any of the conditions (separate test by condition; 67% baseline in all cases). See Table 10. These results support the frequency matching Null Hypothesis.

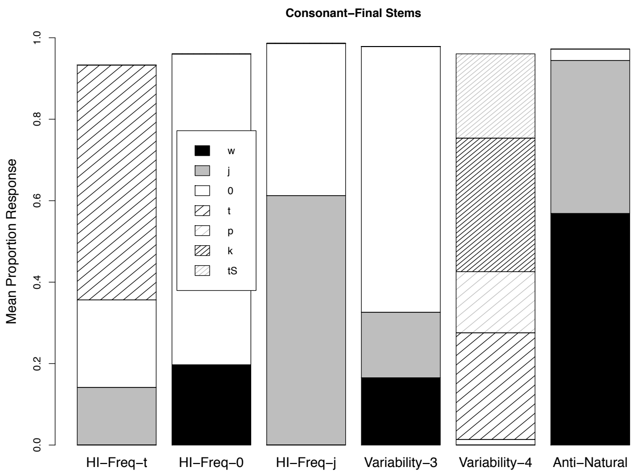


Figure 4: Experiment 3: Consonant-Final Stems: Mean proportion response by allomorph.

Table 10: Experiment 3: Consonant-Final stems. Response rates by allomorph: mean percentage response (expected response rate, based on training levels). Significant differences marked with *.

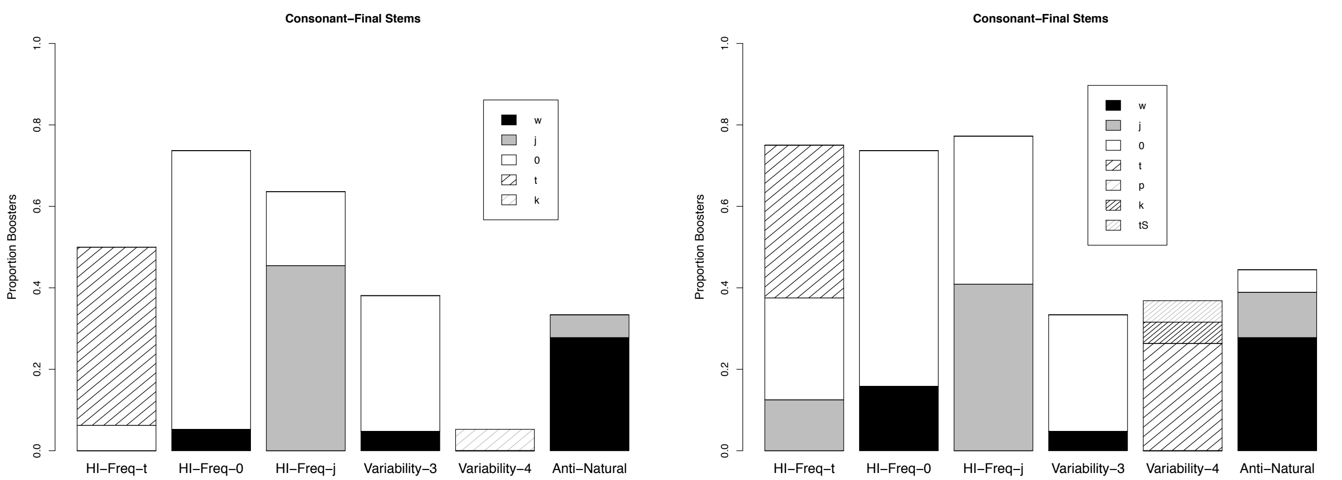
Condition	Percentage Observed (Expected)						
	-/wək/	-/jək/	-/ək/	-/tək/	-/pək/	-/kək/	-/t͡ʃək/
HI-Freq-t		14.1	21.5	57.7 (66.7)			
HI-Freq-0	19.7		76.4 (66.7)				
HI-Freq-j		61.3 (66.7)	37.4				
Variability-3	16.6	16.1	65.3 (61.1)				
Variability-4			1.4	26.2	15.0	32.8 (50)*	20.7

2.4.3 Results: Categorical Measures

Although frequency-boosting behavior was not observed in the aggregate, it may be the case that some participants frequency-matched, while others frequency-boosted. In order to quantify individual behavior with respect to the Frequency Boosting Hypothesis, two different measures are introduced and compared in this section. The first measure captures the definition of boosting as categorical behavior. It is very rare for any participant to choose the same allomorph 100% of the time, but many participants reached a high level of consistency. In Fig. 5a boosting is defined as selecting the same allomorph on at least 80% of consonant-final test items[9]. The Anti-Natural condition, with equal numbers of both vowel types and both allomorphs, and a non-overlapping distribution (see Table 2), is included to provide a comparison level. The second boosting measure requires only that a given allomorph be used at frequencies significantly greater than baseline. Fig. 5b shows the proportion of participants who meet the second boosting criterion, based on a one-sided Wilcoxon signed rank test with an alpha level of .05 (the most frequently used allomorph was determined by participant, and responses using that allomorph were coded as hits; all other

[9]For most participants this corresponded to 10, or more, out of 12 test items. However, errors in which participants produced unintelligible responses, in which they failed to inflect the stems, and in which they appeared to hear the stems as vowel-final rather than consonant-final, were excluded. Therefore, the total number of test items was occasionally fewer than 12. In the HI-Freq-t condition there was 1 booster with a response rate of 10/11. In the HI-Freq-0 condition there were 2 boosters with 10/11. In the HI-Freq-j condition there were 2 boosters with 10/10, and 1 booster with 10/11. In the Variability-3 condition there was 1 booster each with 10/10, 11/11, and 8/10. In the Anti-Natural condition there was 1 booster each with 10/10, 11/11, and 8/8. All boosters in the Variability-4 condition were calculated out of 12 test items.

responses were misses.) The proportions of boosters in each condition, and under each metric are also given in Table 11.



(a) Boosting Metric 1: participant chose given allomorph for at least 80% of test items.

(b) Boosting Metric 2: participant chose given allomorph at levels significantly greater than baseline frequency (one-sided Wilcoxon signed rank test, $\alpha = .05$)

Figure 5: Experiment 3: Consonant-Final Stems: Boosting proportions by allomorph and condition

By both metrics, the high-frequency allomorph was most likely to be boosted. However, minority allomorphs were also occasionally boosted. The three HI-Freq conditions show over-all boosting levels above the baseline of the Anti-Natural condition (33.3%). In the HI-Freq-0 and HI-Freq-j conditions the continuous measure corresponds closely to the proportion under B1, but less so in the HI-Freq-t condition. The second boosting measure (B2) shows similar behavior for the HI-Freq-0 and HI-Freq-j conditions, although the minority allomorph makes up a larger share of the total. In the HI-Freq-t condition the -/jək/ allomorph is boosted under B2, but not B1.

Table 11: Experiment 3: Consonant-Final stems. Percentage of participants, by condition, and by allomorph, who were judged to be boosters under B1, and B2, respectively (see text in 2.4.3).

Condition	% Boosted													
	-/wək/		-/jək/		-/ək/		-/tək/		-/pək/		-/kək/		-/t͡ʃək/	
	B1	B2	B1	B2	B1	B2	B1	B2	B1	B2	B1	B2	B1	B2
HI-Freq-t			0	12.5	6.25	25.0	43.8	37.5						
HI-Freq-0	5.26	15.8			68.4	57.9								
HI-Freq-j			45.5	40.9	18.2	36.4								
Variability-3	4.76	4.76			33.3	28.6								
Variability-4							0	26.3	0	0	5.26	5.26	0	5.26

2.4.4 Conditions 11-12: Variability/Irregularity

The results show that a frequency asymmetry can lead to boosting for individual participants. The Memory Load hypothesis predicts that the more allomorphs, the more trouble learners will have tracking their environments, and the more likely they will be to default to a single allomorph. A variant on this hypothesis predicts that the greater the variability in the pattern – more allomorphs *within* a single stem type – the more likely boosting will be. This will be called the Consistency Hypothesis. The "Variability" conditions contain multiple allomorphs: some more frequent than others, some unpredictable with respect to stem type, and some stem types unpredictable with respect to allomorph. See Table 12.

Table 12: Experiment 3: Generalization: Conditions 11-12

	Condition	N	Back Vowels	Front Vowels	Consonants	Token Ratio		
11	Variability-3	21	ɹatu-ək	skibe-ək		9	4	
			ɹɪlojək	hædiwək		3	2	
12	Variability-4	19	ɹatutək	telʌpitək		1	1	
			vut͡ʃək	fit͡ʃək		1	1	
			t͡ʃoɹænopək			2		
			zokək	skibekək		2	4	

2.4.5 Results

As with the previous set of conditions, there is no evidence for boosting on average. In fact, in the Variability-4 condition [kək]is actually produced at rates significantly lower than baseline (as determined by a two-sided Wilcoxon signed rank test). See Table 10. Under boosting metric B1, the categorical and continuous measures diverge considerably. Only the most frequent allomorph boosts (at low levels) in the Variability-4 condition; and only the two most frequent in the Variability-3 condition are boosted (i.e., there is no individual participant who ever boosts [jək], despite comparable continuous response rates). See Fig. 5a. Under boosting metric B2, the Variability-3 condition shows similar relative proportions of boosted allomorphs, but the total boosting proportion is now slightly below baseline. The Variability-4 condition, on the other hand, is now much closer to baseline than under B1, and three of the allomorphs undergo boosting, rather than only the most frequent. Furthermore, the highest proportion of boosting occurs for one of the minority allomorphs: [tək], and not the higher-frequency [kək]. See Fig. 5b.

2.4.6 Interim Summary

The results of this set of experiments fail to support the Consistency Hypothesis. The more variants were introduced, and the more unpredictable the distribution, the *less* likely participants were to boost any single allomorph[10]. On the other hand, there were individuals who boosted even when the allomorph distribution was uniform, (as in the Anti-Natural condition) and those numbers rose when one allomorph occurred more frequently than the others. The two boosting measures agree on these general trends. B1, however, classifies fewer participants as boosters in almost all conditions – it is generally a more conservative measure of boosting. Whereas boosting of minority allomorphs in almost all conditions is higher under B2. Both measures reflect an apparent bias against the -/kək/ allomorph, as seen in the Variability-4 condition. The failure of the -/jək/ allomorph to be boosted under either measure in the Variability-3 condition may also reflect a bias.

3 Results in the Aggregate

In this section results from conditions across the three sets of experiments are compared. These are broken down into three analyses. The first revisits the question of whether there is a preference for the subsegmental or the segmental analysis of ambiguous stimuli. The second identifies a subset of participants whose responses constitute what is defined as a ‘pre-epenthesis’ pattern. The third assesses the evidence for and against the deletion analysis.

[10] Although, it should be noted that the variability listeners experienced was within stem type, not within stem; all individual words were consistent in their inflection. Furthermore, the high-frequency allomorph was not as frequent as the high-frequency allomorphs in the HI-Freq conditions.

3.1 Biases

In both the Natural and Variable0 conditions there appeared to be an almost categorical preference for the subsegmental interpretation of ambiguous glides. However, this potential phonetic bias was confounded with a demonstrated preference for adopting a unique morpheme realization. Morphological Uniqueness does not apply in either the CTrained1 or HI-Freq-t conditions (there are two unambiguous allomorphs). Crucially, both conditions also provide an unambiguous testing environment. Consonant-final stems clearly show whether a segmental glide is present or not. Furthermore, any observed segmental glides have only one possible source: the segmental analysis of the ambiguous training forms. The training distributions for the two conditions are repeated in Table 13 for ease of reference.

In the HI-Freq-t condition front-vowel final stems, and thus ambiguous allomorphs, are less frequent. Therefore, the number of both segmental and subsegmental responses consistent with [j] are expected to be low relative to [tək] responses. However, what we are interested in here is the number of [jək] responses relative to the number of [ək] responses. In the CTrained1 condition the overall response rate is expected to be even lower because any non-[jək] responses on consonant-final stems are actually errors. However, it is still possible to compare the number of [wək] responses to the number of [ək] responses.

Table 13: Consonant-Final stems: Mean percentages of segmental vs. subsegmental analyses by condition.

		Mean Percentage of Total Responses		
Condition	Training	subsegmental (0: daɹumək)	segmental (w: daɹumwək)	segmental (j: daɹumjək)
Natural	ɹatuwək skibejək	85.7%	8%	6.2%
Variable0	ɹatuwək hædijək ɹɪlo-ək skibe-ək	100%	0	0
CTrained1	ɹatuwək daɹumjək	16.2%	4.65%	78%
HI-Freq-t	ɹatutək skibejək	21.5%	57.7%	14.1%

As Table 13 shows, subsegmental responses outnumbered segmental responses in both the CTrained1 and HI-Freq-t conditions. This difference is significant under a two-sided binomial test with an alpha level of .05 (76/111 subsegmental responses in total, $p < .0002$; although the HI-Freq-t condition by itself fails to reach significance, with 39/63 subsegmental responses, $p=.08$). The relatively high rate of segmental response in the HI-Freq-t condition, however, provides a potential explanation for the low levels of boosting in that condition compared to he HI-Freq-0 and HI-Freq-j conditions (Fig. 5). If the effect of higher rates of segmental response is to increase pattern variability, then lower generalization rates are expected (Section 2.4.4). In fact, boosting levels for the HI-Freq-t condition are closer to those found in the Variability-3 condition.

3.2 Individual pre-epenthesizers

The experiments of this paper as a whole do not support the conclusion that a rule of phonological epenthesis of any kind has been learned. Thus, they provide largely negative evidence regarding the diachronic trajectory of such a synchronic pattern. However, there was a small subset of individuals across conditions whose response patterns were *not inconsistent* with an epenthesis analysis. Pre-epenthesis patterns were defined in the following way: highly consistent use of the allomorph -ək with consonant-final stems ($\geq$ 80% of responses), and highly consistent use of a *different* allomorph with vowel-final stems ($\geq$ 80%). These patterns have the potential, over time, and possibly over generations of learners, to become a productive rule of intervocalic consonant epenthesis. In Table 14 the participants who met these criteria are listed by condition, along with the pattern their responses conform to.

In the Natural, Variable0, CTrained1, and HI-Freq-t conditions, a pre-epenthesis outcome would require a segmental analysis of the ambiguous [ɹatuwək]/[skibejək] type items. As Table 13 shows, the rate of segmental analysis was very low for those conditions. Not only that, but evidence for a segmental analysis comes from exactly those consonant-final stems that are required to surface with the bare allomorph. As a result, no pre-epenthesis patterns are found across those four conditions. In the Consistent-W condition the segmental analysis is the majority one. However, all forms are inflected in the same way, providing no alternation evidence that could lead to an epenthesis analysis. In the Anti-Natural and the Variability-4 conditions the bare allomorph -/ək/ is never heard during training, and consequently does not surface consistently with consonant-final stems at test[11].

Table 14: Proportion of Pre-Epenthesizers by condition. Gray cells indicate stem types held out during training.

Condition	Surface Pattern						Possible Epenthesis Rule	# participants/total participants
HI-Freq-0	[skibe]	[skibewək]	[ɹatu]	[ɹatuwək]	[daɹum]	[daɹumək]	$\varnothing \rightarrow$w/V+__V	8/19
HI-Freq-j	[skibe]	[skibejək]	[ɹatu]	[ɹatujək]	[daɹum]	[daɹumək]	$\varnothing \rightarrow$j/V+__V	4/22
Full Alternation	[skibe]	[skibewək]	[ɹatu]	[ɹatujək]	[daɹum]	[daɹumək]	$\varnothing \rightarrow \begin{bmatrix} \alpha front \\ +glide \end{bmatrix} /[-\alpha front]+\_\_V$	2/18

In both the HI-Freq-j and HI-Freq-0 conditions there were participants who consistently[12] produced [ək] on consonant-final stems. In the HI-Freq-0 condition, 8 of those participants were also pre-epenthesizers (with all but 1 of them consistently using [wək] with front-vowel final stems). The surface response pattern as a whole was analyzable as intervocalic /w/ epenthesis[13]. In the HI-Freq-j condition only 4 of the consonant-final boosters produced a pre-epenthesis pattern[14].

In the HI-Freq-t condition a number of participants consistently produced [ək] on consonant-final stems. However, none of those produced [tək] consistently on front-vowel final stems, and only one did so on back-vowel final stems. In the Variability-3 condition no participants reached consistency with all stem types. However, there were participants who were consistent on front-vowel stems (2 using [wək]; 7 using either [ək] or [jək]), or back-vowel stems (7 using either [ək] or [wək]), or consonant-final stems (6 using [ək]), or a combination of two stem types (8 participants total). In the Full Alternation condition participants must consistently maintain the consonant-vowel boundary in order to produce a pre-epenthesis pattern. Only two participants met this criterion.

3.3 Analytic Ambiguity

Many of the training conditions in this paper were equally consistent with an epenthesis or a deletion analysis. Take the example in (10), which could have been generated from a rule of [t] epenthesis intervocalically, or a rule of /t/ deletion word-finally (or a rule of /t/ deletion post-consonantally, if the t is analyzed as belonging to the suffix morpheme).

(10) [ɹatu]/ [ɹatutək] [daɹum]/[daɹumək]

The majority of the evidence shows that participants chose neither epenthesis nor deletion, appearing instead to learn the plural as a set of (partially conditioned) allomorphs. A reviewer suggests that this outcome might have been due to lack of evidence distinguishing between deletion and epenthesis. If learners had heard forms that biased them towards an epenthesis analysis they might have learned an epenthesis rule. Biasing data could be provided by a form that doesn't alternate, such as that in (11). Such a form presumably favors the epenthesis analysis because the root /pakut/ (and any other similar forms) would have to be marked as exceptional under a deletion analysis.

(11) [pakut]/ [pakutək]

[11] Interestingly, in both conditions there were a very small number of [ək] inflected consonant final tokens

[12] All uses of "consistently" in this section indicate a response rate of 80% or greater

[13] Based on previous results it seems likely that many of the surface [ɹatuwək]-type responses are due to underlying -/ək/ allomorphs. However, it is possible, in principle, for subsequent listeners/learners to adopt a segmental analysis.

[14] The lower number is expected given the fact that the -/ək/ allomorph is twice as frequent (as -/wək/) in the HI-Freq-0 condition, whereas it is half as frequent (as -/jək/) in the HI-Freq-j condition.

In fact, consonant-final stems like [daɹum]/[daɹumək] that fail to alternate can be considered exactly this kind of biasing evidence. The bias in this case would be against an analysis in which all stem-final consonants are deleted on the surface. There are also several conditions in this paper that can be analyzed as favoring a deletion analysis. Those conditions are collected together in Table 15. Example training data pairs are repeated for ease of reference. The final column gives the specific favored analysis for that condition. In all conditions the training data are biased towards the deletion analysis, but not always for the same reasons. For example, under the segmental analysis of the surface glides in the Variable0 and Variability-3 conditions, the epenthesis analysis would require exceptions, so the glide-deletion analysis should be preferred. The Full Alternation condition should favor glide-deletion because the epenthesis analysis requires two segments. But it should also favor deletion over suppletive allomorphy, because deletion satisfies morphological uniqueness. Deletion should also be favored in the HI-Freq-t condition. Under the subsegmental analysis epenthesis would be unpredictable; and under the segmental analysis two epenthetic segments would be required. The epenthesis analysis for the Variability-4 condition would require multiple, phonologically unpredictable epenthetic segments, so deletion should also be preferred in this condition. And in both the Hi-Freq-t and Variability-4 conditions, deletion should also be preferred over suppletive allomorphy for reasons of morphological uniqueness.

Table 15: Set of deletion-biasing conditions, with example training stimuli, and specific deletion analysis.

	Training				Deletion Analysis
Variable0	[ɹatu] [ɹatuwək]	[skibe] [skibe-ək]	[hædi] [hædijək]	[ɪɹlo] [ɪɹlo-ək]	$\{w,j\} \rightarrow \emptyset / \_\_\#$
Variability-3	[ɹatu] [ɹatu-ək]	[skibe] [skibe-ək]	[hædi] [hædiwək]	[ɪɹlo] [ɪɹlojək]	$\{w,j\} \rightarrow \emptyset / \_\_\#$
Full Alternation	[ɹatu] [ɹatujək]	[skibe] [skibewək]	[daɹum] [daɹumək]		$\{w,j\} \rightarrow \emptyset / \_\_\#$
HI-Freq-t	[ɹatu] [ɹatutək]	[skibe] [skibejək]			$t \rightarrow \emptyset / \_\_\#$ or $\{t,j\} \rightarrow \emptyset / \_\_\#$
Variability-4	[ɹatu] [ɹatutək]	[skibe] [skibekək]	[fi] [fit͡ʃək]	[t͡ʃoɹæno] [t͡ʃoɹænopək]	$\{t,k,p,t\text{ʃ}\} \rightarrow \emptyset / \_\_\#$[a]

[a]the suffix-final k is an exception to this rule

In all conditions the plural morpheme would be realized exclusively as [ək] under the deletion analysis: thus [daɹum]/[daɹumək]. The HI-Freq-t and Variability-4 conditions provide the clearest evidence against this outcome. The majority of consonant-final stems in the HI-Freq-t condition (57.7%) were inflected with [tək]; while the majority in the Variability-4 condition (32.8%) occurred with [kək] (Section 2.4.2). Additionally, all trained allomorphs appeared on at least some consonant-final stems. Not only do these results argue strongly against the deletion analysis, they also argue strongly *for* the suppletive allomorphy analysis as the preferred, or default, analysis. Morphological uniqueness should prefer either of deletion or epenthesis, over an analysis that requires 2 or 4 allomorphs. Thus, even when both the training data and the bias towards morphological uniqueness favor a particular generative analysis, participants behave as if they have learned a set of semi-predictable allomorphs. In the Variable0 condition the results are not inconsistent with a glide-deletion analysis. However, the orthographic version of this condition (Section 2.2.3) argues strongly against this interpretation.

The evidence from the remaining conditions in this section is less clear-cut. The transfer observed between vowel-final stem types in the Variability-3 and Full Alternation conditions could be attributed to errors in recalling the stem-final consonant. In the case of consonant-final stems, productions like [daɹumwək] could be attributed to a stem-final consonant cluster (/daɹumw/). Additionally, the errors described at the end of section 2.3.1 could actually be the result of "repairing" consonant-final inputs to conform to a deletion rule (rather than the result of failing to perceive coda consonants in unstressed final syllables). None of this evidence, however, is particularly strong. Furthermore I take the failure to learn a deletion (or epenthesis) analysis in the HI-Freq-t and the Variability-4 conditions as a strong argument against a deletion analysis in any of the remaining conditions. This is not definitive, of course. And, in fact, because of the ambiguity inherent in phonological analysis, it is not possible to completely rule out other analyses. The issue of analytic ambiguity within phonological theory more generally will be taken up in the final section of this paper.

4 Summary of Results

A strong bias emerged from the first set of conditions against surface allomorphy. When the training data allowed it, the majority of participants opted for an analysis in which the plural suffix was invariant. In the Natural and Variable0 conditions this meant analyzing ambiguous surface forms such as [ɹatuwək] as underlyingly /ɹatu+ək/, with an 'excrescent' glide, rather than /ɹatu+wək/, with a segmental glide. In the Consistent-W condition this meant the opposite: analyzing [ɹatuwək] as /ɹatu+wək/, rendering it consistent with forms like [daɹumwək]. While individual variation was the norm in these experiments, a pressure for morphological uniqueness imposed a level of consistency both within and across participants not seen under any other conditions.

While morphological uniqueness was confounded with a potential phonetic bias (towards the subsegmental interpretation of ambiguous glides) in the Natural and Variable0 conditions, such a bias could be inferred from the pattern of errors in the CTrained1 and HI-Freq-t conditions. In the two conditions combined, 65% of possible responses were subsegmental.

There was little to no evidence for either an epenthesis or deletion analysis in any of the experiments. In general, participant behavior was much more consistent with what would be expected from (imperfect) memorization of a set of allomorph-stem associations – a 'conjugation' analysis (Hale 1973). In fact, the multiple allomorph analysis seems actually to be preferred, as opposed to an analysis of last resort. Even when morphological uniqueness could be satisfied by adopting a deletion or epenthesis analysis participants showed no evidence of having learned such a rule.

In terms of accuracy and consistency, learners were actually worse in the full training condition than in the held-out ones. The more allomorphs they heard during training – regardless of how predictable their distribution was – the worse they did. What this means, besides the fact that participants were not learning an epenthesis rule, is that they were also failing to exploit the available natural classes as predictors of allomorph identity. In the non-variable conditions allomorphs were predictable by stem types, definable as [+consonantal] versus [-consonantal], and [+front] versus [+back] (or some other comparable configuration of features)[15]. Although consistency across the +/-consonantal natural class was higher, it was not high enough to produce epenthesis-consistent results. This is surprising under the expectation of a psychologically robust distinction between consonants and vowels. For example, speech errors seem to respect the consonant/vowel distinction: consonants swapping only with other consonants, and vowels with other vowels (e.g., MacKay 1970, Shattuck-Hufnagel 1986). However, most such errors reported are between word-initial consonants, or consonants in onset position. Data on errors with vowels appears to be restricted largely to primary stressed position (often in monosyllabic words). Natural-class categories based not only on word-final position, but on segments in unstressed syllables, are likely to be less salient to listeners. Furthermore, generalizations involving consonants may require different contexts than those involving vowels, and may depend strongly on the type of process being generalized (Finley 2011). See also Goldrick (2004) and Kapatsinski (2010) for similar cross-category transfer in artificial grammar learning experiments.

On average, frequency-matching behavior was observed in conditions with multiple allomorphs. However, categorical measures revealed that some participants 'boosted': using a single allomorph almost exclusively in their inflection of consonant-final stems. And there were more of these boosters in conditions with non-uniform frequency distributions. It may be that certain individuals have a general predilection for boosting, with approximately 33% of participants boosting one or the other of two allomorphs in the Anti-Natural condition. However, contrary to the memory load hypothesis, there were fewer boosters in conditions with more variants and less predictability.

The results also suggest that there may be individual differences in the likelihood of adopting one anal-

[15]During training participants heard at least 6 unique words, half of which ended in [i] in the singular, the other half of which ended in [e]; and another 6 unique words, half of which ended in [u] in the singular, the other half of which ended in [o]. In the conditions where consonant-final stems were heard during training, participants additionally heard 6 unique words, all of which ended in consonants in the singular, all from the set {m,n,l,ɹ,z,v}, except for one d-final word.

ysis or another (assuming the observed differences are stable over time). Individual differences provide a way for unlikely (in the aggregate) phonologization routes to arise: scenarios in which ‘phonemicizers’ and ‘boosters’ are also innovators, acting to spread a change to the rest of the population (cf. Milroy and Milroy, 1985; there is also a growing literature on individual differences and the role they may play in sound change; e.g. Yu (2013), Dimov et al. (2012)). Certain individual participants, those who boost, but also maintain category boundaries, can be categorized as producing a pattern that could become epenthetic over time. Successive incremental changes over this type of input may provide a path to phonologization (cf. Kirby (2001)).

5 Discussion & Conclusions

The focus of this paper has been on intervocalic consonant epenthesis. We began with an example of one type, what could be called a default epenthesis pattern, in which the same segment is inserted in all relevant contexts regardless of other featural differences. There is a second type, what has been termed assimilative epenthesis (cf. de Lacy (2006)), in which the epenthetic segment shares features with one or both of its flanking vowels[16]. In the next two sections I will sketch different diachronic trajectories for the two different types. But I will argue that there are certain stages that are necessary in the emergence of both as productive synchronic patterns. In light of the experimental results just presented, the conclusion will be that such patterns are difficult to learn and unlikely to be induced from ambiguous data. Section 5.3 will test those conclusions against a specific case study for which we have considerable diachronic and synchronic data: English intrusive r. The paper will end with the implications of this work for synchronic phonological theory more generally.

5.1 Epenthesis from Coarticulation

Synchronically, assimilative epenthesis can be analyzed as a process of default epenthesis, followed by assimilation to one or more vocalic features. Diachronically, however, the pattern is taken to derive from the natural coarticulation of neighboring vowels, yielding segments like glides, certain fricatives, and, in some cases, glottals (see Blevins (2008)), that can be described as either articulatorily or perceptually minimal (e.g., Steriade (2001), Clements and Hume (1995)). An example of an assimilative epenthesis pattern in Balangao is given in (12). In this language the front glide [j] is inserted after high front vowels, and the back glide [w] is inserted after back vowels. The low vowel /a/ does not participate, undergoing coalescence in hiatus contexts.

(12) Balangao (Shetler 1976)

[16] It should be noted that a third type of intervocalic epenthesis, dissimilatory epenthesis, has been reported in languages like West Greenlandic (Rischel, 1974). It is possible that alternations in which anti-homorganic glides (as in the Anti-Natural condition) occur intervocalically also arose from final deletion. However, given the paucity of data for these types of patterns I refrain from speculating about them in this paper.

/ʔalope+an/→	[ʔalopijan]	
	attach.shoulder.strap-REF.FOC	
	'attach shoulder strap to'	
/i+anpo+ju/→	[ijanpoju]	
	ASSOC.FOC-hunt-you	
	'you hunt'	
/malo+ɨn/→	[malowɨn]	
	wash.clothes-OBJ.FOC	
	'wash clothes'	
/i+bato+an+ju/→	[ibatowanju]	
	BEN.FOC-throw.rocks-BEN.FOC-you	
	'you throw rocks'	

There are a number of different permutations of this general pattern that have been described. For example, a three-way split between [j] after front vowels, [w] after back vowels, and [ʔ] after /a/ occurs in Malay. In Cupeño, epenthesis in the imperative plural inserts [j] after /i/, but either [w] or [h] after /u/. Whereas, in Abujhmaria, it is the second vowel that conditions the epenthetic segment, and /a/ sometimes patterns with front, sometimes with back vowels (See Morley (2015):Appendix for details on these and other examples). While these patterns likely have the same phonetic origin, the fact that they differ by conditioning context and by epenthetic segment argues that they have entered the phonology in the different languages in different ways. Furthermore, if a distinction can be made between a phonetic and a phonological pattern of assimilative epenthesis, there must be a mechanism by which subsegmental sounds (e.g., glides produced in the transition between two vowels) can become segmental. In several of the experiments reported here, deliberately ambiguous stimuli were used in order to determine what factors might affect learners' interpretation of such glides in one direction or another. A bias was found for the subsegmental analysis, but a strong preference for an invariant realization of the suffix allomorph was able to flip responses to the segmental analysis. It is not clear how the morphological uniqueness bias could lead to a phonological epenthesis pattern, given that it requires the absence of alternations. However, there may be a route by which an epenthesis-consistent pattern could be achieved after phonologization of one or more intervocalic glides. This is a potential avenue for future work. A second requirement for the phonologization of coarticulation-based epenthesis is pattern generalization. While an excrescent high back glide following a high back vowel may happen naturally, it is less true for a high back glide following a mid, or low, back vowel. Generalization is required either at the phonetic or the phonological stage, and probably at both. Phonological generalization is also required for non-assimilative epenthesis and, in fact, any categorical pattern that arises out of continuous change.

5.2 Epenthesis from Deletion

Default (non-assimilative) epenthesis is typically analyzed as arising from 'rule inversion' (Vennemann 1972) in which learners re-analyze the results of historic deletion (of a consonant in certain environments) as due to synchronic epenthesis (of that consonant in the complementary environments). This scenario is illustrated in (13 a-13 c). In (13 a) a hypothetical stem morpheme is shown inflected, first with a consonant-initial suffix, and then with a vowel-initial suffix. There are no phonological alternations. At some later time, (13 b), a stem-final consonant is lost in pre-consonantal position – a common historical change. This affects the first inflected form, but not the second. In the final stage, (13 c), the learner is faced with an ambiguous [t]/∅ alternation. Re-analysis, if it occurs, involves the choice of /pami/ as the new underlying form, and [t] is epenthesized when the stem is followed by a vowel-initial suffix.

(13) (a) /pamit/ /pamit+nu/→[pamitnu] /pamit+o/→[pamito]

(b) Deletion: /pamitnu/ > [paminu]

(c) Re-Analysis: /pami/ /pami+nu/→[paminu] /pami+o/→[pami**t**o]

This process seems relatively straightforward at first glance. But it has been shown that there are several necessary conditions that must hold for rule inversion to result in synchronic epenthesis. Table 16, adapted from Morley (2012), illustrates the specific case of C/Ø alternations at the stem-suffix boundary. For the re-analysis in (13 c) to be adopted, it is critical that learners analyze the second suffix as vowel-initial. Definitive evidence for this would come from non-alternating forms. This, in turn, requires that not all consonant-final roots undergo deletion – such words are represented by the form [fisem] in this example. However, while [pamito] can be analyzed as the product of [t] epenthesis between a root-final vowel and a suffix-initial vowel, the process is not consistent. No epenthesis occurs in [oruo] (or with any other historically vowel-final root). Furthermore, unless only /t/'s delete, forms like [kifuno] will result, requiring additional 'epenthetic' segments – in this case [n].

Table 16: Hypothetical Precursor Epenthesis System

historic stems	**after loss of pre-final and pre-consonant coronals**			
/pamit/	[pami]	[paminu]	[pamito]	[pamitina]
/oru/	[oru]	[orunu]	[oruo]	[oruina]
/fisem/	[fisem]	[fisemnu]	[fisemo]	[fisemina]
/kifun/	[kifu]	[kifunu]	[kifuno]	[kifunina]
historic suffixes		/nu/	/o/	/ina/

The point here is that patterns resulting from historic changes will not be consistent and unambiguous. Two types of generalization are required to transform the synchronic pattern from Table 16 into default epenthesis. The first is generalization across vowel-final stems. For the case of /t/ epenthesis, all [kifuno] type forms must be changed to [kifuto]. And all [oruo] type forms must be changed to [oruto]. Furthermore, this leveling must occur only within the synchronic vowel-final roots, and not cross over to the consonant-final ones. Transfer will introduce forms such as [fisemto], which would undermine the epenthesis analysis. The second is generalization across vowel-initial suffixes, such that the variant chosen with -/o/ is the same variant chosen with -/ina/, etc.

In the experiments reported here, learners often frequency match; they transfer across consonant-vowel boundaries; and they tend to learn sets of allomorphs. These results suggest that default synchronic epenthesis may emerge only rarely, requiring a convergence of non-typical types of learners and specific historical conditions. This conclusion is supported by at least one typological survey of intervocalic consonant epenthesis (Morley 2015). However, more work in this area is needed. In the first place, a method of sampling is reuired that can provide a good approximation for the actual rate of occurrence of epenthesis cross-linguistically. Then the proper baseline must be determined in order to assess whether the obtained counts were lower than expected, that is, lower than comparable phonological processes. This would require a much more explicit theory of how generative theory maps to typology than currently exists. However, the enterprise runs into problems much earlier. There is no universal diagnostic for establishing whether a given observed pattern is an instance of synchronic epenthesis, versus an instance of something else, such as deletion (see Lombardi (2002), de Lacy (2006), Morley (2015)). And for many sources, "epenthesis" is used merely as a descriptive label, rather than as a hypothesis about the generating grammar. Thus it is uncertain how many cases of true epenthesis are attested. If synchronic epenthesis is exceptional, as predicted, then individual cases of synchronic epenthesis can offer considerable insight into the exact conditions necessary for their emergence.

5.3 Intrusive R

English intrusive r represents a particularly well-known case of what is generally analyzed as consonant epenthesis. The phenomenon has been documented from at least the middle of the 19th century in the U.K.

(e.g., Jones (1956)), and there have been numerous analyses of the pattern within generative theory. In fact, there has been, and continues to be, disagreement about the proper phonological analysis of intrusive r. What is known is that a number of dialects of English in Britain, the U.S., New Zealand, and elsewhere experienced loss of r at some point in their history[17]; this occurred pre-consonantally both within and across words, as well as prepausally (e.g., [bɑɹ#bəl͡oʊ] > [bɑbəl͡oʊ]). However, these so-called non-rhotic varieties maintained r in inter-vocalic environments (known as "linking r"), resulting in alternations of the following kind: "bar below" [bɑbəloʊ]; "bar above" [bɑɹəbʌv]. A further subset of these dialects also developed what is known as "intrusive r": r's that subsequently appeared in contexts in which they were historically absent, but which were congruent with the existing r/Ø alternation, e.g., [sɑɹælɪs] for "saw Alice", but [sɑbɑb] for "saw Bob".

Most accounts also agree on the following facts: that both linking and intrusive r, if they appear, appear only following non-high vowels, a set which always includes [ə], and diphthongs ending in [ə], and may often include mid or low back vowels such as [ɑ], and [ɔ]. Furthermore, there do not appear to be any cases of intrusive r without accompanying linking r. There are cases, however, of linking r without intrusive r (e.g., Conservative RP, as reported by Durand (1997)), and of non-rhotic dialects which have neither, having lost r in all environments (e.g., some varieties of Southern U.S. English, as reported by Wells (1982)). While linking r is sometimes analyzed as underlying, intrusive r is more often analyzed as the result of epenthesis. The fact that innovated r's appear after vowels which are phonetically reduced, in unfamiliar place or person names, and with words that do not contain orthographic r's, comprise the most common arguments for why these r's cannot be underlying (see, for example, Rubach (2000), McCarthy (1993)).

A synchronic account of the r/Ø alternation that makes no reference to historical forms either posits a general rule of r deletion, or a general rule of r insertion. The typical argument against the deletion analysis is also orthographic – "there is obviously no r in comma" (Foulkes, 1997) – implying that no literate speaker of English could possess an underlying form of /kaməɹ/. The account of Johansson (1973) seems to represent the dominant view: that historic r deletion eventually led to rule inversion, as speakers lost track of which words were historically r-final. However, this rule inversion also requires a change in underlying representations that contradicts orthography. It is unclear why /kɑ/ would be an acceptable underlying form for car (despite the fact that there obviously *is* an "r" in "car"), while /kaməɹ/ for "comma" would not.

Furthermore, most phonological analyses of intrusive r deal exclusively with r-sandhi (innovated r at word or morpheme boundaries), or even just r-liason (at word boundaries only), but r may also be inserted within monomorphemic words, such as "daughter" as [dɑɹtəɹ] (Jackson, 1830), as well as word-finally, even when there is no pause, or no following vowel-initial word (see, e.g., Wells (1982)). These "hyper-rhotic" dialects are attributed to the stigmatization of r deletion. Speakers are aware that they "drop r's" with respect to other, possibly more standard, dialects and make an effort to "add them back in". However, this analysis implies a process of r deletion that speakers are attempting to undo, and Harris (1994) takes the over-generalization as evidence that r has become underlying in those words (e.g. /əmbrɛlə/ > /əmbrɛləɹ/).

The case for an analysis of epenthesis over deletion is thus far from unassailable. An epenthetic r would also be restricted to occur after only non-high vowels, and would be gradiently conditioned by other vowel qualities, occurring more frequently with certain vowels, and with certain words. In some cases, the epenthetic r would result in more marked forms (pre-consonantally and in certain word-final environments). Nevertheless, whether English intrusive r should be considered a case of default epenthesis or not, the pattern does illustrate an apparently uncommon, but necessary, type of learner generalization. The explanation for why such generalization occurred may be due to the phonetics of rhotic segments.

Gick (1999) analyzes r (and l) as being comprised of both a vowel gesture, and a consonant gesture. r

[17]Because the exact features of the rhotic segment that was lost vary by dialect, the convention is to use the orthographic "r" symbol in the description of this phenomenon.

loss in his framework is actually merger of the vowel gesture with the preceding vowel, with accompanying loss of the consonant gesture (also termed "r-vocalization"). The implication is that loss of such consonants may be facilitated by their dual nature. This account also provides an explanation for why r is preferentially lost before schwa, and more generally, non-high vowels. F2 and F3 values for r are typically lower than those of high/front vowels. Thus merger of r with such vowels would tend to have a lowering and backing effect, resulting in a change in vowel quality, or the addition of a schwa, e.g., [diɹ] > [di͡ə]. Generalization of r, in turn, is driven by the phonetic similarity of r and schwa. A similar explanation is offered in accounts in which r is treated as a glide, and insertion is allowed when the glide shares features with the preceding vowel (e.g., Broadbent 1991). That the results reported in this paper fail to show consistent use of contextual phonological features may be attributable to the fact that the association was not based in phonetic similarity – and in some cases was anti-similar. Both the experimental results and the case of English intrusive r suggest that phonetics may play a large role in the emergence of epenthesis, and epenthesis-like alternations. Phonetic similarity may also effectively limit the extent to which such patterns will be generalized.

5.4 Synchronic Phonology

Although current phonological theory has little to say about the likelihoods of different phonological rules, or repairs, the implicit assumption is often that all phonological operations are equally accessible to learners. However, this may not be the case, and the determination of likelihood may depend very strongly on the diachronic origins of those patterns. Epenthesis may be much less common than other phonological rules. Or it may be the case that many phonological rules are relatively uncommon (cf. the "Too Many Solutions" problem (Wilson 2000, Myers 2002)). On closer examination many apparently categorical phonological rules have been seen to show variability, both conditioned and unconditioned. Whatever the source of that variability, the end result is an explosion of possible analyses that are equally consistent with the data.

It was the apparent difficulty of the learning problem that motivated a large part of Noam Chomsky's work in both syntax and phonology (Chomsky 1965). The assumption of the degraded nature of the learner's input (filtered through performance), and the incompleteness of the evidence available to infant learners, led to the conclusion that certain properties of grammar must be innate. However, the resultant learning theory offered no guidelines for classifying inconsistent, or 'exceptional' forms. This remains true of linguistic theory generally. Because of this omission there exists no clear diagnostic for determining the best analysis of a given set of linguistic data (how many exceptions is too many? At what point does an exception change into decisive evidence for an alternative analysis?).

While no one would deny that languages change over time, and that synchronic grammars must be the result of an accumulation of diachronic changes, the far-reaching ramifications of this position are seldom explored to their full extent. Consider, once again, the example from Section 3.3, repeated here in (14) and (15).

(14) [ɹatu] /[ɹatutək] [daɹum]/[daɹumək]

The alternating forms in (14) are meant to represent a fully ambiguous scenario, consistent with both epenthesis of [t], and deletion of [t]. The additional form in (15) presumably tips the scales towards an epenthesis analysis.

(15) [pakut]/ [pakutək]

But what kind of test items could provide definitive proof that learners had adopted this analysis? The epenthesis analysis would result in novel consonant-final singular forms retaining their final consonant in the plural. Novel vowel-final forms would acquire a [t] in the plural. But the same results are expected for a suppletive allomorphy analysis. We could reverse the testing conditions, providing plurals and asking for singulars. But forms with /t/ are inherently ambiguous and listeners would have to guess under any analysis. In fact, the deletion analysis is still a possibility as well, if forms like that in (15) can be treated as exceptions. All else being equal, the analysis without exceptions is, of course, the preferred one. But when we are dealing with real language data, it is rare to find any pattern that lacks exceptions.

The conclusions in this paper are possible because there is so much supporting evidence against the generative analyses. Had the results been less unbalanced, it would have been nearly impossible to completely rule out the epenthesis or deletion analysis. Performance errors and lexical exceptions are to be expected, and, in practice, linguists seem to adopt an implicit and intuitive criterion for deciding when forms are to be classified as exceptions, versus evidence for an alternate analysis. The problem is the lack of a transparent standard for this criterion, and by extension, for selecting one linguistic analysis over another. This creates a serious falsifiability problem within linguistics, and hinders attempts to answer larger questions about language universals (see Morley (2015) for more in-depth discussion of this issue).

This paper represents part of a larger project investigting the existence of universal principles in phonological consonant epenthesis. As has just been argued, however, the undertaking goes far beyond a single phonological pattern. The true scope of the question encompasses the learning problem (this paper), diachronic change (Morley 2012), and typological methodology (Morley 2015). The conclusions at this point are the following: there is little evidence that phonological (versus phonetic) epenthesis is an accessible analysis to a learner engaged in morphological learning, and the necessary conditions under which default synchronic epenthesis could arise are very constrained. Default epenthesis is predicted to be rare, and may in fact be so, under reconsideration of the typological facts. Additionally, a general mechanism of rule inversion is unlikely to supply a simple and direct path to a categorical phonological grammar. More generally, I believe we must alter some of our core assumptions, about both the nature of synchronic grammars, and the trajectory of historic change, in order to link the two within a coherent theory of phonological competence. Exceptions must be explicitly defined and incorporated into models of learning. The burden of proof should be shifted onto the generative analysis, with suppletion, or memorized stem-allomorph pairs, as the null hypothesis. We must also devise an experimental standard for rejecting that null hypothesis. Most importantly, we need a true theory of diachronic change in which the beginning and end states, and all the states in between, are direct correspondents to states defined within synchronic theory.

References

2014. Audacity (version 2.1.2). URL `http://audacity.sourceforge.net/`.

Berent, Iris, Tracy Lennertz, Paul Smolensky, and Vered Vaknin-Nusbaum. 2009. Listeners' knowledge of phonological universals: Evidence from nasal clusters. *Phonology* 26:75–108.

Bickerton, Derek. 1984. The language bioprogram hypothesis. *Behavioral and Brain Sciences* 7:173–188.

Blevins, J. 2008. Consonant Epenthesis: Natural and Unnatural Histories. In *Linguistic universals and language change*, ed. Jeff Good, 79–106. Oxford University Press.

Boersma, P., and D. Weenink. 2009. Praat: doing phonetics by computer (Version 6.0.2.1). computer program. URL `http://www.fon.hum.uva.nl/praat/`.

Broadbent, Judith. 1991. Linking and intrusive r in English. *University College London Working Papers in Linguistics 3* 281–302.

Browman, C.P., and L.M. Goldstein. 1990. Tiers in articulatory phonology, with some implications for casual speech. In *Papers in laboratory phonology i: Between the grammar and the physics of speech*, ed. J. Kingston and M.E. Beckman, 341–376. Cambridge University Press.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. The MIT Press.

Chomsky, N., and M. Halle. 1968. *The Sound Pattern Of English*. Harper & Row.

Clements, George N, and Elizabeth V. Hume. 1995. The Internal Organization of Speech Sounds. 245–301.

Culbertson, Jennifer, and Géraldine Legendre. 2010. Investigating the evolution of agreement systems using an artificial language learning paradigm. *Pragmatic Processing Factors in Negative Island Contexts* 46.

Dimov, Svetlin, Shira Katseff, and Keith Johnson. 2012. Social and personality variables in compensation for altered auditory feedback. 185–210.

Durand, Jacques. 1997. Linking r in English: Constraints, principles and parameters, or rules? *Histoire Épistémologie Langage* 19:43–72.

Estes, W.K. 1972. Research and Theory on the Learning of Probabilities. *Journal of the American Statistical Association* 67:81–102.

Finley, Sara. 2011. Generalization to novel consonants in artificial grammar learning. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, 318–323.

Foulkes, Paul. 1997. English [r]-sandhi – a sociolinguistic perspective. *Histoire Épistémologie Langage* 19:73–96.

Fowler, C.A. 1984. Segmentation of coarticulated speech in perception. *Perception and Psychophysics* 36:359–368.

Gick, Bryan. 1999. A gesture-based account of intrusive consonants in English. *Phonology* 16:29–54.

Goldrick, M. 2004. Phonological features and phonotactic constraints in speech production. *Journal of Memory and Language* 51:586–603.

Hale, K. 1973. Deep-surface canonical disparities in relation to analysis and change: an Australian example. *Current Trends in Linguistics* 11:401–458.

Harris, John. 1994. *English sound structure*. Blackwell Oxford.

Hudson Kam, C.L., and E.L. Newport. 2005. Regularizing unpredictable variation: The roles of adult and child learners in langauage formation and change. *Language Learning and Development* 1:151–195.

Jackson, George. 1830. *Popular Errors in English Grammar, Particularly in Pronunciation, Familiarly Pointed Out: For the Use of Those Persons Who Want Either Opportunity or Inclination to Study This Science*. Effingham Wilson.

Johansson, Stig. 1973. Linking and intrusive /r/ in English: A case for a more concrete phonology. *Studia Linguistica* 27:53–68.

Jones, Daniel. 1956. *The pronunciation of English*, volume 369. Cambridge University Press.

Kam, Carla L Hudson, and Elissa L Newport. 2009. Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology* 59:30–66.

Kapatsinski, V. 2010. Velar palatalization in Russian and artificial grammar: Constraints on models of morphophonology. *Laboratory Phonology* 1:361–393.

Kirby, S. 2001. Spontaneous evolution of linguistic structure – an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation* 5:102–110.

de Lacy, P. 2006. *Markedness: Reduction and preservation in phonology*. Cambridge University Press.

Lombardi, L. 2002. Coronal epenthesis and markedness. *Phonology* 19:219–251.

MacKay, C.J. 1999. *A Grammar of Misantla Totonac*. Salt Lake City: The University of Utah Press.

MacKay, Donald G. 1970. Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia* 8:323–350.

McCarthy, John J. 1993. A case of surface constraint violation. *Constraint-based theories in multilinear phonology, special issue of Canadian Journal of Linguistics* 38:169–195.

Milroy, James, and Lesley Milroy. 1985. Linguistic change, social network and speaker innovation. *Journal of linguistics* 21:339–384.

Moreton, Elliott, and Joe Pater. 2012. Structure and Substance in Artificial-phonology Learning, Part I: Structure. *Language and Linguistics Compass* 6:686–701.

Morley, Rebecca L. 2011. Phonetics to Phonology: Learning Epenthesis. In *Proceedings of the 47th Meeting of the Chicago Linguistics Society*, volume 47, 137–151.

Morley, Rebecca L. 2012. The Emergence of Epenthesis: An Incremental Model of Grammar Change. *Language Dynamics and Change* 2:59–97.

Morley, Rebecca L. 2015. Deletion or Epenthesis? On the Falsifiability of Phonological Universals. *Lingua* 154:1–26.

Myers, S. 2002. Gaps in factorial typology: The case of voicing in consonant clusters. *ms* .

Peperkamp, S., R. Le Calvez, J.-P. Nadal, and E. Dupoux. 2006. The acquistion of allophonic rules: Statistical learning with linguistic constraints. *Cognition* 101:B31–B41.

Prince, A., and P. Smolensky. 1993/2004. *Optimality Theory*. Blackwell Publishing.

Rischel, J. 1974. *Topics in West Greenlandic Phonology*. Copenhagen: Akademisk Forlag.

Rubach, Jerzy. 2000. Glide and glottal stop insertion in Slavic languages: A DOT analysis. *Linguistic Inquiry* 31:271–317.

Shattuck-Hufnagel, Stefanie. 1986. The Representation of Phonological Information during Speech Production Planning: Evidence from Vowel Errors in Spontaneous Speech. *Phonology Yearbook* 3:117–149.

Shetler, J. 1976. *Notes on Balangao Grammar*. Language Data: Asian-Pacific Series No. 9. Huntington Beach, CA: Summer Institute of Linguistics.

Steriade, D. 2001. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. UCLA.

Vennemann, Theo. 1972. Rule inversions. *Lingua* 29:209–242.

Wells, John C. 1982. *Accents of English*, volume 1. Cambridge University Press.

Whalen, D.H. 1984. Subcategorical phonetic mismatches slow phonetic judgments. *Perception and Psychophysics* 35:49–64.

Wilson, C. 2000. Targeted Constraints: An approach to contextual neutralization in Optimality Theory. Doctoral Dissertation, Johns Hopkins, Baltimore, MD.

Wilson, C. 2006. Learning Phonology with Substantive Bias: An Experimental and Computational Study of Velar Palatalization. *Cognitive Science* 30:945–982.

Yang, C. 2005. On Productivity. *Linguistic Variation Yearbook* 5:265–302.

Yang, Charles. 2011. A statistical test for grammar. In *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*, 30–38. Association for Computational Linguistics.

Yu, Alan CL. 2013. Individual differences in socio-cognitive processing and the actuation of sound changes. In *Origins of sound change: Approaches to phonologization*, ed. A.C.L. Yu, 201–227. Oxford University Press.

Acknowledgments

This work was supported by the Targeted Investment in Excellence from the OSU College of Humanities. I would like to thank Christina Heaton, MarDez Desmond, Lark Hovey, Sara Pennington, Karen Kuhn, Dahee Kim, and others for their work in the lab. Special thanks to Emily Clem for her assistance coding and analyzing the data, as well as thinking about the results. Thanks also to Bjorn Kohnlein for his advice, and the Phonies group at OSU for input on earlier versions of this work. I would also like to thank Peter Staroverov, the anonymous reviewers, and my Associate Editor at Phonology for their work in making this a better paper.

Appendix

Table A.1: Table 1 repeated. Full list of stimuli roots for all conditions, both training and test items

i	ˈɹatu	ˈɹɪlo	fɹaˈbomu	tʃoˈɹæno	kɹo/zo	vu
ii	ˈhædi	ˈskibe	teˈlʌpi	gluˈdɛbe	fi	sme
iii	ˈpiʃu/ˈgɛθu[a]	ˈhago	bəˈhʌʒu	ˈfædʒo	zo/kɹo	ˈgaidu
iv	ˈvʌlki	ˈploke	ˈdʒimi	diˈʒaɹe	ˈθuzi	ˈʃuvi
v	ˈdaɹum	keˈtɛlan/d͡ʃoˈɹəfim[a]	ˈhoʃɪn/ˈt͡ʃalum[a]	ˈɹibæz	ˈpɹɛv	ˈbihɪl
vi	ˈtwit͡ʃo	ˈðipu	ˈmuɹo	ˈtɹɪfu	ˈmeko	ˈgɛθu/ˈpiʃu[a]
vii	ˈsabol	ˈgenʊɹ	ˈtɹifæd	(ˈtagæf)	d͡ʃoˈɹəfim/keˈtɛlan[a]	ˈt͡ʃalum/ˈhoʃɪn[a]

[a]items separated by / indicate substitutions: a training item used as a test item, and vice versa, for a particular condition. See Table A.2

Table A.2: Full List of Conditions

	Condition	N	Train	Test	Back Vowels	Front Vowels	Consonants	Token Counts			Allomorph Ratio
1	Natural	18	i & ii	iii, iv & v	ɹatuwək	skibejək		6	6		1:1[a]
2	Anti-Natural	18	i & ii	iii, iv & v	ɹatujək	skibewək		6	6		1:1
3	Variable0	20	i & ii	iii, iv & v	ɹatuwək	hædijək		3	3		2:1:1[a]
					ɹɪlo-ək	skibe-ək		3	3		
4	Consistent-W	20	i & v	iii, iv & vii	ɹatuwək		daɹumwək	6		6	1
5	Full Alternation	18	i, ii & v	iii, iv & vii	ɹatujək	skibewək	daɹumək	6	6	6	1:1:1
6	CTrained1	24	i & v	iii, iv & vii	ɹatuwək		daɹumjək	6		6	1:1
7	CTrained2	19	i & v	iii, iv & vii	ɹatujək		daɹumwək	6		6	1:1
8	HI-Freq-t	16	i, vi & ii	iii, iv & v	ɹatutək	skibejək		12	6		2:1
9	HI-Freq-0	19	i, ii, & vi	iii, iv & v	ɹatu-ək	skibewək		12	6		2:1
10	HI-Freq-j	22	i, ii, & vi	iii, iv & v	ɹatujək	skibe-ək		12	6		2:1
11	Variability-3	21	i, ii, & vi	iii, iv & v	ɹatu-ək	skibe-ək		9	4		11:4:3
					ɹɪlojək	hædiwək		3	2		
12	Variability-4	19	i & ii	iii, iv & v	ɹatutək	telʌpitək		1	1		3:1:1:1
					vut͡ʃək	fit͡ʃək		1	1		
					t͡ʃoɹænopək			2			
					zokək	skibekək		2	4		

[a]participants interpreted homorganic glide + [ək] as the bare [ək] allomorph overwhelmingly