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 the selection of 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 analysis 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 designated in terms of either the insertion environment, or the features of the epenthesized segment, or both. The focus of this paper is on consonant epenthesis, and more specifically, default consonant epenthesis of the kind that results in less marked surface structures (e.g. Prince and Smolensky (2004)). Consonant epenthesis intervocalically meets these requirements; it transforms a more-marked vowel-initial syllable to a less-marked onset-nucleus 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 usual, or natural, phonological process is implicitly built into current generative theory, either as a high-valued re-write rule (by the diagnostic of SPE Chomsky and Halle (1968)), or as embodied in a universal constraint banning the insertion of segments. Thus, given exposure to the right kind of linguistic data, the learner is expected to be able to acquire the epenthesis grammar. However, determining the necessary learning conditions is a non-trivial problem.

There are several sources of ambiguity in the analysis of linguistic data. For example, in the alternations provided in (1), the hypothesis “insert a consonant after a low vowel” predicts the same (correct) outcome as the more general hypothesis "insert a consonant after any vowel". Also consistent with the available data is the analysis in which it is an underlying /ʔ/ that gets deleted in the environment before a consonant. With the collection of more data some of these ambiguities might be resolvable. 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, and yet how learners might acquire forms that represent exceptions to a general pattern has been conspicuously missing from the phonological learning literature (see Yang 2005, 2011, however, for a metric to determine morphological productivity). This comes into focus most clearly when the learner is allowed to entertain the hypothesis that no regular pattern is present at all. This paper is an experimental investigation of this kind of learning scenario, and the ambiguity that is at the heart of linguistic analysis[1]. Specifically, I ask whether learners favor certain linguistic hypotheses over others (epenthesis, deletion, or suppletive allomorphy), and what properties of the input push learners toward one hypothesis or another (cf. Moreton and Pater (2012)).

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, but all are close variants of each other. The training data to which participants are exposed consist of morphophonological alternations over a small set of nonce words, corresponding to a singular and plural inflection. The alternations are ambiguous with respect to their generating grammar (or lack thereof). After training participants are tested with novel forms that they hear in the singular, and for which they provide the plural. 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) Deletion Analysis (Segmental)
/ɹatuw+ək/→[ɹatuw+ək] /skibej+ək/→[skibejək]

(c) Epenthesis Analysis (Phonetic)
/ɹatuw+ək/→[ɹatuʷək] /skibe+ək/→[skibeʲək]

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

Having heard the singular and plural forms, participants might conclude that there was a unique plural suffix, and a process of epenthesis in environments of vowel hiatus, as depicted in (3 a). This hypothesis competes with the deletion analysis, in which the glide belongs to the word stem, and is deleted in the singular (word-finally) (3 b). When glides homorganic to the stem vowel are used as the ambiguous segments a third possibility arises, namely coarticulation; what will be referred to as Phonetic Epenthesis. This is depicted in (3 c), with the superscript to distinguish it from the phonological analysis. Finally, the morphological, or suppletive allomorphy, analysis assigns the ambiguous glides to the suffix. See (3 d). The analysis chosen by the participants was determined via held-out test items that provided a non-ambiguous environment for perceiving glides. For the example above, these consisted of consonant-final stems like that in 4.

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

(4) [daɹum]

Response 1: [daɹumək]: Consistent with Phonological Epenthesis, Phonetic Epenthesis, and Deletion

Response 2: [daɹumjək]: Consistent with Suppletive Allomorphy

For the example provided here, all hypotheses except the Suppletive Allomorphy analysis predict the same outcome, namely Response 1. Thus, it can only be determined whether learners did or not adopt the Suppletive Allomorphy analysis. 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 or Deletion analysis. This is the first main finding. The second is that learners showed a strong preference for morphological uniqueness; when the Phonetic analysis resulted in a single plural morpheme (as opposed to multiple allomorphs), it was consistently adopted by learners, who treated the ambiguous surface glides as subsegmental. However, in conditions in which a single plural morpheme would result from a segmental analysis of the glide, this preference reversed. The 'Morpheme Uniqueness' pressure, however, was not strong enough to induce learners to adopt a generative analysis – either deletion or epenthesis. These results suggest that learning true phonological generalizations might be both harder and less common than previously supposed.

In the following section the details of the experimental methodology are described, and 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; more frequently heard allomorphs were more likely to be used when an error was made, and variability, i.e., unpredictability for a given stem type, increased overall 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 generalization, 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 observed. 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 the case of 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 linguistics.

2 Experiments

The experiments in this paper use the artificial grammar learning (AGL) paradigm, a technique that has become increasingly popular among linguists. For a partial review of the phonological AGL literature, see Moreton and Pater (2012), also Peperkamp et al. (2006). The use of the AGL paradigm 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 (R) (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 that they would later be asked questions about those words. What followed was a passive training stage in which 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 singular word, and a picture of two of the same object (e.g., two apples) accompanied playback of the plural word. 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 variations 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.

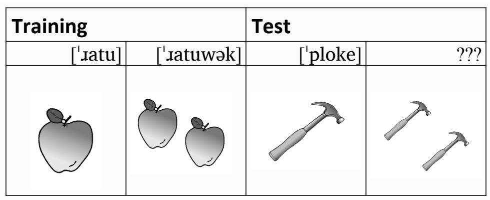


Figure 1: Training Paradigm: Sequential presentation of singular and plural forms both auditorially and pictorially

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[2]. The feedback stage provided a singular form (heard during training) and prompted the participant 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 that 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), 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. Each of these words 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 are between-subject.

2.1.2 Stimuli

The experiments each 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

[2]The responses for the feedback trials were recorded but not analyzed.

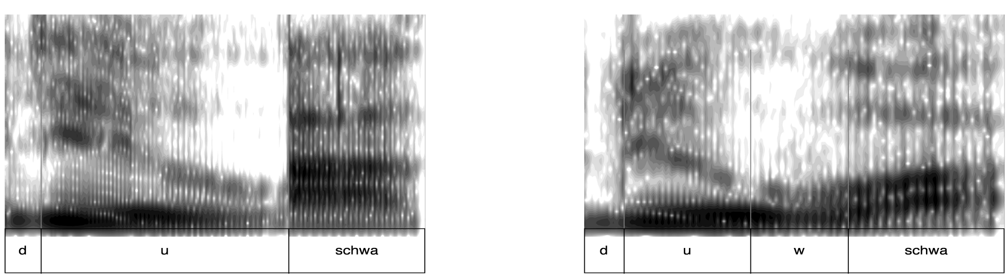
singular + Xək. X was either zero, a glide homorganic to the place of the preceding vowel, an anti-homorganic glide, an obstruent (p, t, tʃ, or k), or a pause/discontinuity (e.g., skibeək, skibejək, skibewək, skibekək, skibe-ək), depending on experiment and condition.

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	ɡluˈdɛbe	fi	sme
ˈpiʃu/ˈɡɛθu[a]	ˈhago	bəˈhʌʒu	ˈfædʒo	zo/kɹo	ˈɡaidu
ˈvʌlki	ˈploke	ˈdʒimi	diˈʒaɹe	ˈθuzi	ˈʃuvi
ˈdaɹum	keˈtɛlan/d͡ʒoˈɹʌfim[a]	ˈhoʃɪn/ˈt͡ʃalum[a]	ˈɹibæz	ˈpɹɛv	ˈbihɪl
ˈtwit͡ʃo	ˈðipu	ˈmuɹo	ˈtɹɪfu	ˈmeko	ˈɡɛθu/ˈpiʃu[a]
ˈsabol	ˈɡenʊɹ	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 some sort of 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, the discontinuity tokens were created. These tokens were spliced to avoid perceptual glides. For example, the token [gaidu-ək] 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, eliminating the natural transitional period between adjacent vowels, and as a result sounded quite unnatural. However, the experimental 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

Participant responses at test were coded only at a coarse grain. That is, no attempt was made to disambiguate forms such as /ɹatuwək/ and /ɹatuək/. They were both coded as being consistent with a VV production. In most cases, however, the productions were clear. 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[3].

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

2.2 Experiment 1: Learning Biases

The first set of experiments provides tests of the Suppletive Allomorphy, Phonology Epenthesis, and Phonetic Epenthesis Hypotheses, as exemplified in (2). The predictions for each hypothesis will be given by condition below. Table 2 provides examples of the singular and plural training stimuli for each condition in this set. The number of participants in each condition (N), and the token counts (unique stems) for each stem type are also provided. There are three stem types, of which one is held out during training. At test, the held-out stems provided a test of learner generalization. In two of the conditions the pattern was completely consistent with a phonological epenthesis analysis. In the third, the pattern was partially consistent; and in the fourth, not consistent.

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	
			ɹatuwək	hædijək		3	3	
3	Variable0	20	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 (the stems) were the same in all conditions, as was the final part of the plural suffix (-/ək/). In the “Natural” condition, plurals also contained homorganic glides at the morpheme boundary; the quality of the glide was determined by the stem-final vowel. In the “Anti-Natural” condition, all plurals contained glides anti-homorganic with the stem-final vowel. Both conditions are completely consistent with an epenthesis analysis, in which either /w/ or /j/ is inserted between adjacent vowels. However, they are 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 hypotheses would emerge.

The relevant test items are the held-out consonant-final stems. These items will unambiguously show the presence or absence of a suffix glide. If participants learned a phonological epenthesis pattern from the vowel-final training data then they would inflect consonant-final stems without the glide, as shown in (5). This is also the predicted result for the s phonetic epenthesis analysis.

(5) Phonological and Phonetic Epenthesis Analysis: [daɹumək]

If participants instead analyzed the plural as consisting of two predictable allomorphs then they would be forced to guess about the proper inflection of consonant-final stems. Many outcomes are possible, but for the sake of explicitness, we will adopt the hypothesis that participants would produce both allomorphs in equal proportions in that case (6).

(6) Suppletive Allomorphy Analysis: [daɹumjək]; [daɹumwək]

2.2.2 Results

In the Natural condition consonant-final stems were inflected with glides only 12% of the time, whereas in the Anti-Natural condition consonant-final stems were inflected with one or other of the two glides

94% of the time. See Table 3. The low rate of glide responses in the Natural condition seems to rule out the Suppletive Allomorphy Hypothesis, but it is consistent with both the Phonetic and the Phonological Epenthesis Hypotheses. Conversely, the result in the Anti-Natural condition is consistent with the Suppletive Allomorphy Hypothesis, but not with the Phonological Hypothesis (the Phonetic Hypothesis is not applicable, as the glides are anti-homorganic with the stem vowel). The input in the Anti-Natural condition was consistent with an epenthesis analysis, but participants did not select it. There are many possible reasons for this outcome; that participants were not exposed to consonant-final stems during training, or that there were not enough forms to lead to generalization, or that the exposure period was too short, etc. However, the failure to learn epenthesis in the Anti-Natural condition suggests that under the similar learning conditions of the Natural condition participants also failed to select a Phonological Epenthesis analysis. If that is the case, it leads to the conclusion that participants analyzed the glide as subsegmental - the result of coarticulation.

Table 3: Experiment 1: Results: Average glide response on Consonant-Final stems

Condition	Glide Response Ratio	
	mean	std.
Natural	.119	.207
Anti-Natural	.944	.164
Variable0	0	0.00
Consistent-W	.939	.211

2.2.3 Condition 3: phonetics or phonology?

A third condition was designed for two purposes: as an attempt to bias participants towards a segmental, rather than subsegmental analysis of the surface glides, and to corroborate the Phonetic Hypothesis as the explanation for the results of the Natural condition. 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. There was no predictability to which type of stem occurred in which form, 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. Instead, participants produced even fewer glides with consonant-final stems than in the Natural condition; in fact, they 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 also supported by a separate test of the Variable0 condition with written rather than spoken responses. Participants were instructed to give their best guess as to the spelling of the spoken words, and consistently failed to transcribe glides in pre-suffix position. 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"). No consonant-final stems were ever written with a glide[4]. These results seem to

[4]The only orthographic glides in the entire condition came from front-vowel final inflected test forms. 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.

imply that participants are strongly biased towards the phonetic interpretation of ambiguous phones, thus learning a single allomorph – /ək/ – in both the Natural and Variable0 conditions.

2.2.5 Condition 4: Morphological Uniqueness

Because the Variable0 condition failed to elicit a segmental analysis of the ambiguous glides, a further condition was designed for that purpose. In the Consistent-W condition participants were exposed to 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]. See Table 2. Front-vowel final stems were held out. If the strong bias towards the subsegmental interpretation of the glide in forms like [ɹatuwək] persists, then participants should learn two allomorphs of the plural suffix: -/ək/ and -/wək/, predictable by stem type. If participants learn that the first allomorph occurs on vowel-final stems, then the Phonetic Hypothesis predicts that there should be no -/wək/ at all with front-vowel final stems.

(7) Phonetic Hypothesis: skibeək / skibejək

However, if the presence of the segmental glide in forms like [daɹumwək] serves as some kind of attractor then there should be a non-zero number of front-vowel final stems inflected with -/wək/ [5]. Choice of -/wək/also results in a single surface form for the plural suffix in this condition. Use of -/wək/ with all stems is predicted if participants prefer to limit the allomorphs of a given morpheme. This will be called the Morphological Uniqueness Hypothesis[6]:

(8) Morphological Uniqueness Hypothesis: skibewək

2.2.6 Results

As Table 3 shows participants responding with -/wək/ for 94% of front-vowel final stems, indicating that they had largely learned a single allomorph of the plural. A logistic regression by condition (with Natural as the reference), with dependent variable proportion glide response, found the Anti-Natural and Consistent-W conditions to be significantly different from the Natural and Variable0, but not from each other[7]. See Table 4. This confirms the prediction of the Morphological Uniqueness Hypothesis. The presence of the segmental /w/ is enough to completely reverse the result seen in the Natural condition. In fact, the strong subsegmental bias seen in the Natural and Variable0 conditions could be entirely driven by morphological uniqueness. Tests of the subsegmental bias where morphological uniqueness is not applicable will be discussed in subsequent sections.

Table 4: Experiment 1: Results: Logistic regression

	Estimate	SE	Pr(>|z|)
(Intercept)	-1.994	.209	<2e-16 ***
Anti-Natural	4.761	.364	<2e-16 ***
Variable0	-18.267	992.993	.985
Consistent-W	4.721	.346	<2e-16 ***

[5]Based on the results of subsequent conditions it seems likely that if participants learned two allomorphs they would use both with novel stems. If the attractor hypothesis is correct, there should be an increase above this baseline use of -/wək/.

[6]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.

[7]All analyses were performed using the R software environment version 3.3.3 and the statistics 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]). This was in spite of the fact that they had never before heard any CC sequences (all training syllables were of the form CV). The auditorily distinct tokens in the Variable0 condition, however, resulted in a uniform subsegmental analysis. Listeners appear to treat ..Vək and ..VGək plurals as phonetic tokens of the same underlying morpheme. This is plausible under a model in which listeners have phonetic expectations based on their native language competence (e.g., Whalen, 1984, Manuel and Krakow, 1984, Fowler, 1984), as well as a predilection for limiting the surface realizations of semantically identical morphemes. Specifically, the fact that these two acoustically quite distinct phone sequences 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 (Gick 1999, Browman and Goldstein 1990)). The results of the Natural and Consistent-W conditions can also be explained by the Morphological Uniqueness Hypothesis. The Phonetic Hypothesis makes the same predictions for the Natural and Variable0 conditions, thus it cannot be determined what bias, if any, participants have towards the subsegmental interpretation of ambiguous glides.

2.3 Experiment 2: Evidence & Allomorphs

The results of Experiment 1 do not provide support either for the phonological epenthesis analysis, or for the segmental analysis of ambiguous glides. Thus, it appears that the learning conditions in those 5 conditions were not conducive to either of those analyses. The Full Alternation condition was designed to bias learners towards an epenthesis analysis. 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, of course, also allows for alternative analyses. While the Phonetic Hypothesis is not available for these stimuli, the Suppletive Allomorphy Hypothesis, with three predictable allomorphs, is still a possibility. The Deletion Hypothesis, in which underlyingly glide-initial suffixes (-/wək/, and -/jək/) lose their onsets after stem-final consonants, is also consistent with the training data. If participants learn the conditioning environments for the three surface alternants perfectly, then it cannot be determined what hypothesis they have adopted. Errors, however, can provide insight into what participants learn. However, because the Full Alternation condition exposes participants to all stem types, and all allomorphs are completely predictable, it was expected that accuracy would be high in this condition.

Table 5: Experiment 2: Evidence & Allomorphs: Example Training Stimuli

	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 vs. 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. It predicts that participants should learn a predictable pattern better when the contexts are consonant/vowel, than when they are back-vowel/front-vowel.

(9) CV Hypothesis: Error rates should be lower when one of the vowel-final stem types is held out than when the consonant-final stem type is held out

The CV Hypothesis predicts higher accuracy in the CTrained1 and CTrained2 conditions than in conditions like the Anti-Natural condition.

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 is only applicable to CTrained1, and will be termed the Attractor Hypothesis.

(10) Attractor Hypothesis: segmental response rates should be higher for an ambiguous glide in the presence of an unambiguous glide *of any kind*

The Attractor Hypothesis predicts higher rates of skibewək type responses in the CTrained1 conditions than in conditions like the Natural and Variable0. Neither CTrained1 nor CTrained2, however, is consistent with a single-morpheme hypothesis. It was therefore predicted that participants would learn a minimum of two allomorphs in those conditions.

2.3.1 Results

Fig. 3 shows the proportion of responses by allomorph on consonant-final stems for each of the three conditions in Experiment 2. The proportion of segmental analyses of the ambiguous [w] in the CTrained1 condition can be inferred from the -/wək/ responses on consonant-final stems which was 4.65% on average, corresponding to a total of 11 tokens distributed over 5 participants[8]. This response rate is actually lower than the segmental response rate in the Natural condition (see Table 3). Therefore, the Attractor Hypothesis is disconfirmed.

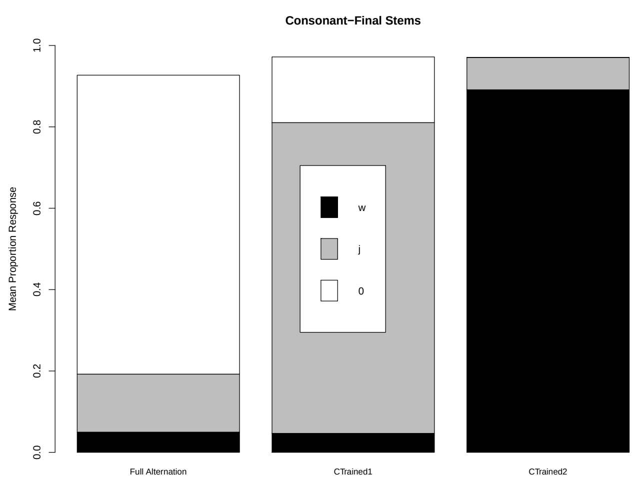


Figure 3: Results: Experiment 2: Consonant-final stems: Mean proportion response rate by allomorph. White: -/ək/; Gray: -/jək/; Black:-/wək/.

Despite expectations, response accuracy[9] on trained stem types in the Full Alternation condition was actually 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

[8]Although not shown here, the segmental response rate as evidenced by the use of [w] with front-vowel final stems averaged 8.4% of responses

[9]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.

consistent with a regular epenthesis pattern. Table 6 gives proportion correct for the three conditions, and for the two stem types: consonant-final and back-vowel final.

Table 6: Results: Experiment 2: Proportion correct by stem type, mean and standard deviation: consonant-final; back-vowel final.

	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

Accuracy ranged from 92% on consonant-final stems in the CTrained2 condition, to only 59% for back-vowel final stems in the Full Alternation condition. Four pairwise logistic regressions are reported in Table 7. The dependent variable was proportion correct, and each model was fit to a single stem type and one pair of conditions so that individual cells could be compared. The results show that accuracy on consonant-final stems is significantly higher in the CTrained2 condition than in the CTrained1 condition, but the CTrained1 condition is not significantly different from the Full Alternation condition. Accuracy on back-vowel final stems was significantly different in each pairwise comparison by condition.

Table 7: Results: Experiment 2 : Pair-wise logistic regressions for each stem type separately: Full Alternation vs. CTrained1 condition; CTrained1 vs. CTrained2 condition.

	Consonant-Final Stems			Back-Vowel Final stems		
	Est.	SE	Pr(>|z|)	Est.	SE	Pr(>|z|)
(Full Alternation)				(Full Alternation)		
(Intercept)	1.347	.180	6.65e-14***	.374	.140	.0076**
CTrained1	.105	.241	.664	.711	.193	.00023***
(CTrained1)				(CTrained1)		
(Intercept)	1.452	.160	<2e-16***	1.085	.133	2.82e-16***
CTrained2	.978	.300	.0011**	1.055	.253	3.13e-05***

It is possible to explain the three-way difference in accuracy in terms of interference from other allomorphs during training, with interference, and thus error rate, increasing with increasing number of allomorphs. The Full Alternation condition contains 3 unambiguous allomorphs, while the CTrained2 condition contains 2 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 2 allomorphs, but not quite as many as 3, due to the ambiguity of the surface glide.

To test the CV Hypothesis accuracy in the CTrained1 and CTrained2 conditions was compared to accuracy in the Anti-Natural condition (which was comparable in providing training on two stem types with predictable and equally frequent allomorphs). The results of a logistic regression comparing the three conditions (with the Anti-Natural as reference level) are shown in Table 8. The dependent variable was proportion correct over all trained stem types, and the difference was significant between each of conditions CTrained1 and CTrained2, and the Anti-Natural condition. The CTrained conditions taken together averaged 82.2% accuracy, while the Anti-Natural condition averaged 71.7%. Even though there were still errors in using vowel-final allomorphs on consonant-final stems, and vice versa, there were significantly fewer than the number of errors between vowel-final stem types in the Anti-Natural condition; the CV Hypothesis was confirmed.

Table 8: Results: CTrained1 and CTrained2 compared to the Anti-Natural condition. Logistic regression on trained stem types.

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

2.3.2 Interim Summary

The presence of the unambiguous /j/ on consonant-final stems does not lead to a higher segmental analysis of ambiguous [w]; only the presence of an unambiguous /w/ (as in the Consistent-W condition) affects segmental response rates, presumably, because it allows for a single allomorph analysis. If the additional data available in the Full Alternation condition reinforced the predictable pattern then participants should have performed better, not worse, in conditions where they were exposed to more stem types. Rather than bias learners towards an epenthesis analysis, the additional data pushed them further away from that analysis. The C-V boundary seems to be more salient to participants than the V-V boundary, as evidenced by the difference in error rates across that boundary. However, an average error rate of 18% represents a potential obstacle to the evolution of an epenthesis rule that is based on a distinction between vowel and consonant environments. The large difference between accuracy on back-vowel final stems and consonant-final stems in the Full Alternation condition suggests that there may be more interference from front-vowel final stems on back-vowel final stems, than on consonant-final stems. However, I do not have an explanation at this time for why accuracy on back-vowel final stems was consistently lower than accuracy on consonant-final stems across all three conditions.

2.4 Experiment 3: Generalization

In recent years there have been numerous studies devoted to testing how language learners generalize linguistic input under conditions of data sparsity. Although not all results are consistent, it has often been argued that participants show strong biases towards generalizing in "natural" ways, that is, 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, input that is non-optimal in some way (e.g., Bickerton (1984), Ross and Newport (1996). More broadly speaking, there is reason to believe that there may be a point of memory load beyond which the cognitive system begins to regularize or generalize its input (e.g., Estes (1972)). There is also experimental support in language learning paradigms for the hypothesis that the most frequent pattern or element 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 the degree to which participants generalize, or regularize, their input.

2.4.1 Conditions 8-10: Frequency

In the 3 HI-Freq conditions back-vowel final stems, and back-vowel allomorphs, occur twice as often as front-vowel final stems (in both type and token counts). See Table 9. Additionally, consonant-final stems are held out and allomorphs are predictable. In line with previous results, it is expected that learners will use all trained allomorphs on consonant-final stems. The null hypothesis is that participants will frequency match, using allomorphs in proportion to their training frequency. The Frequency Boosting hypothesis, on the other hand, predicts that the much higher frequency of one of the allomorphs will cause learners to select that allomorph disproportionally – to boost it beyond its training frequency.

Table 9: Experiment 3: Generalization: Example Training Stimuli

	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	
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	

2.4.2 Results: Continuous measures

The results show, as expected, that the more frequent allomorph is selected more often than the less frequent allomorph(s) on novel consonant-final stems. See the first three bars of Fig. 4. The response rates for the HI-Freq conditions are also given in the first three rows of Table 10; and these show that the high frequency allomorph was not selected more often than its baseline frequency (the responses in the HI-Freq-0 condition were not significantly different than expected by a one-sided Wilcoxon signed rank test with an alpha level of .05). On average, these results support the frequency matching null hypothesis.

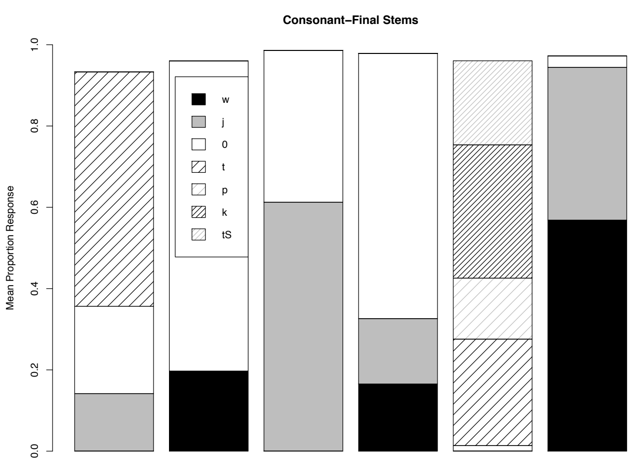


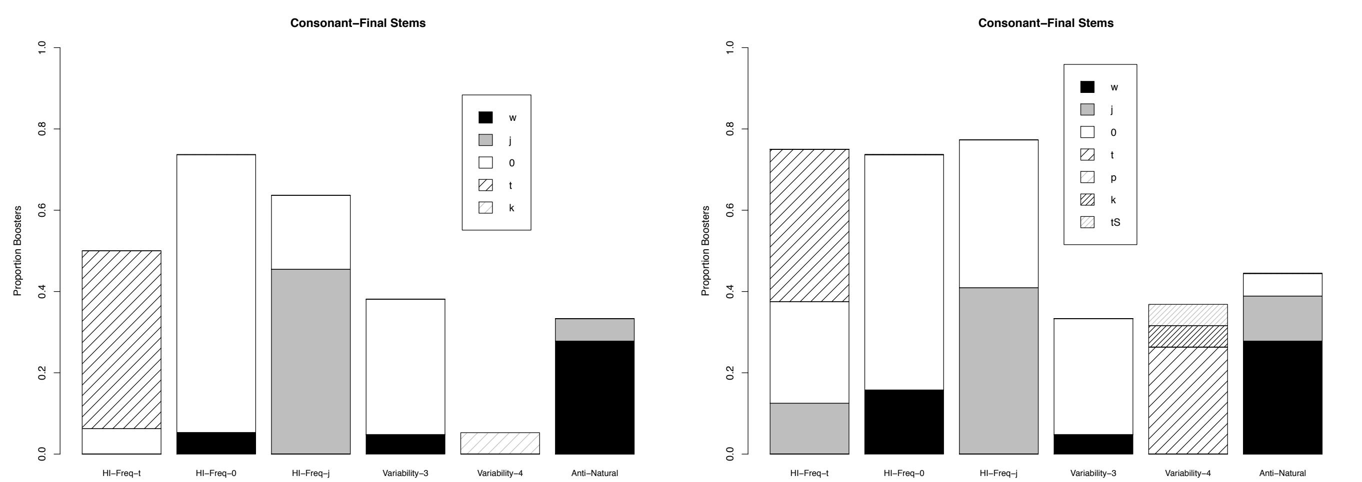
Figure 4: Results: Experiment 3: Consonant-Final Stems: Mean proportion response by allomorph: White: -/ək/; Gray: -/jək/; Black: -wək/; dark diagonal: -/tək/; dark diagonal dense: -/kək/; light diagonal: -/pək/; light diagonal dense: -/t͡ʃək/

Table 10: Results: Experiment 3: Consonant-Final stems. Response rates by allomorph: mean percentage response (expected response rates, based on training levels). Results that show a significant difference from expected (baseline) levels are marked with *. The last two columns show the percentage of participants in each condition who were judged to be boosters: showing preference for a single allomorph. These are given for each of two different metrics: B1 and B2 (see text). The percentage response values for the boosted allomorphs under B1 in each condition are bolded. Values are underlined for allomorphs that are boosted only under B2.

Condition	Percentage Observed (Expected)							% Boosted	
	-/wək/	-/jək/	-/ək/	-/tək/	-/pək/	-/kək/	-/t͡ʃək/	B1	B2
HI-Freq-t		14.1	**21.5**	**57.7 (66.7)**				6.25;43.8	12.5;**25.0**;**37.5**
HI-Freq-0	**19.7**		**76.4 (66.7)**					5.26; 68.4	**15.8;57.9**
HI-Freq-j		**61.3 (66.7)**	**37.4**					45.5; 18.2	**40.9;36.4**
Variability-3	**16.6**	16.1	**65.3 (61.1)**					4.76;33.3	**4.76;28.6**
Variability-4			1.4	26.2	15.0	**32.8 (50)***	20.7	5.26	26.3;**5.26**;5.26

2.4.3 Results: Categorical Measure

Although frequency boosting behavior was not observed in the aggregate, it is possible to ask whether variation might occur between individual learners, with some exhibiting boosting behavior. 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 at least 83% of the time (10 out of the usually 12 total test items). The second measure requires only that a given allomorph be used at frequencies significantly greater than baseline to count as boosted. 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 proportions of boosters in each condition, and under each metric are also given in the last two columns of Table 10.



(a) Results: Experiment 3: Consonant-Final Stems: Proportion of boosting participants by allomorph and by condition: Boosting Metric 1: participant chose given allomorph for at least 10 out 12 test items.

(b) Results: Experiment 3: Consonant-Final Stems: Proportion of boosting participants by allomorph and by condition: Boosting Metric 2: participant chose given allomorph at levels significantly greater than baseline frequency, where significance was calculated via a Wilcoxon signed rank test with an alpha level of .05

Figure 5: Generalization to novel consonant-final stems

On the first boosting measure (B1), the most frequent allomorph was generalized (or 'boosted') the most often in each condition. However, minority allomorphs were also sometimes boosted. 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. In all three of the HI-Freq conditions the boosting levels are above this baseline. In the HI-Freq-0 and HI-Freq-j conditions the continuous measure corresponds closely to the categorical measure, 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.

2.4.4 Conditions 11-12: Variability/Irregularity

The previous 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 also predicts that the more irregularity there is in the pattern – that is, multiple 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 in differing distributions, with some allomorphs more frequent than others, some stem types unpredictable with respect to allomorph, and some allomorphs that occur with more than one stem type. See Table 9.

2.4.5 Results

Once again, there is no evidence for boosting on average. In fact, in the Variability-4 condition -/kək/is actually produced at rates lower than baseline as determined by a two-sided Wilcoxon signed rank test. See Table 10. Under 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 Figs. 4 and 5a. Under B2, the Variability-3 condition shows similar relative proportions of boosted allomorphs, but the total boosting proportion is now 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/.

2.4.6 Interim Summary

The results of this set of experiments do not 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. On the other hand, there were individuals who boosted even when the allomorph distribution was uniform, and those numbers rose when one allomorph occurred more frequently than the others. As seen in the previous set of experiments, higher numbers of allomorphs led to lower accuracy. The two boosting measures agree on these general trends. B1, however, results in lower proportions of boosters in almost all conditions – it is generally a more conservative measure of boosting. Whereas B2 results in more boosting of minority allomorphs in almost all conditions. 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 four analyses. The first assesses the rates of subsegmental versus segmental analysis of ambiguous stimuli in conditions where the the Morphological Uniqueness effect does not apply. The second re-examines the CV Hypothesis with additional data. The third picks out a subset of participants whose responses constitute what is defined as a 'pre-epenthesis' pattern. Finally, the fourth analysis addresses the question of whether participants in all, or any, of the conditions might be learning a deletion rule rather than adopting a suppletive allomorphy analysis.

3.1 Biases

In both the Natural and Variable0 condition there appeared to be an almost categorical preference for the subsegmental analysis of ambiguous glides. However, the subsegmental analysis also allowed learners to adopt a single allomorph for the suffix. Therefore, it could not be determined if Morphological Uniqueness was solely responsible for the effect, or if a Phonetic Bias was also active. However, the CTrained1 and HI-Freq-t conditions, introduced later, can help answer this question. Both conditions include one ambiguous phone, and one unambiguous phone in training. Thus, a single-allomorph analysis is not possible, and the Phonetic Bias Hypothesis can be directly tested. 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 are unambiguous with respect to their origin – they can only derive from the segmental analysis of the ambiguous training forms. The training distributions for the two conditions are repeated in Table 11 for ease of reference.

In CTrained1, the [w] occurring with back-vowel final stems is ambiguous. Participants' analyses can be estimated, however, from their responses on consonant-final stems at test. Any consonant-final tokens at test that appear with either -/ək/ or -/wək/ are assumed to be due to transfer from back-vowel final stems. The same is true in the HI-Freq-t condition for [j]. However, in the CTrained1 any non-/jək/ responses are actually errors because consonant-final stems are heard during training. Because of this, the numbers are expected to be relatively small for both of the analyses. The numbers in the HI-Freq-t condition are also expected to be relatively low compared to -/tək/ responses, because front-vowel final stems are only half as frequent as back-vowel final stems.

Table 11: Results for relevant conditions: Mean percentage of Segmental vs. Subsegmental Analysis on Consonant-Final Stems. Segmental responses contained unambiguous glides following the stem-final consonant; subsegmental analyses unambiguously lacked glides in that position.

		Mean Percentage of Total Responses		
Condition	Training	subsegmental (0)	segmental (w)	segmental (j)
Natural	ɹatuwək skibejək	85.7%	8%	6.2%
Variable0	ɹatuwək hædijək ɹɪlo-ək skibe-ək	100%	0	0
Consistent-W	ɹatuwək daɹumwək	5.1%	94.9%	–
CTrained1	ɹatuwək daɹumjək	16.2%	4.65%	–
HI-Freq-t	ɹatutək skibejək	21.5%	–	14.1%

As Table 11 shows, the proportion of subsegmental responses was larger than the proportion of segmental responses in both conditions. This difference is significant under a two-sided binomial test with an alpha level of .05 (35 segmental vs 76 subsegmental responses in total; $p < .0002$). The segmental analysis is also more likely for the ambiguous [j] than the ambiguous [w], but this difference does not reach significance. However, if the effect of higher rates of segmental response is to increase pattern variability, then a difference between the glides may account for the fact that there were so few boosters in the HI-Freq-t condition compared to the HI-Freq-0 and HI-Freq-j conditions (Fig. 5). In fact, boosting levels for the HI-Freq-t condition are closer to those found in the Variability-3 condition (with variability across three unambiguously distinct allomorphs)[10].

3.2 Consonants vs Vowels

In Experiment 2, Section 2.3.1 a difference in accuracy was found between the CTrained conditions and the Anti-Natural condition. This difference was attributed to the difference in how participants treat the front-vowel/back-vowel natural class, versus the vowel/consonant natural class. It was hypothesized that the consonant/vowel distinction would be more salient to participants, leading to fewer errors in which they used vowel-final allomorphs on consonant-final stems, and vice versa. This section offers a more comprehensive test of that hypothesis, bringing in results from the Full Alternation and HI-Freq conditions. The full set of relevant results are given in Table **??**, and example plural training stimuli are repeated for ease of reference. The dependent variable, Transfer Proportion, was defined as responses attributable to allomorphs heard during training, but with different stem types. For example, in the CTrained1 condition this would include any consonant-final stems produced with final [wək] or [ək] in the plural.

For consistency across conditions results are given only for front-vowel final and consonant-final stems. In the Full Alternation condition both consonant-final and back-vowel final stems contribute to transfer to front-vowel final stems, just as back-vowel final and front-vowel final stems both transfer to the consonant-final stems. In all other conditions the transfer is from back-vowel final stems alone. Bolded numbers in Table **??** are transfer to front vowel-final stem types. Unbolded numbers are transfer to consonant-final stem types. Only in the CTrained2 condition was it possible to distinguish the source of the allomorphs that participants used on the held-out stem type. Transfer, or generalization, from back-vowel final stems to held-out front-vowel final stems accounted for roughly 87% participant responses.

		Held out	Trained
Conditions	Training	mean (sd) %	mean (sd) %
Full Alternation	ɹatujək **skibewək** daɹumək	–	**53.2** (35.4); 21.2 (24.6)
CTrained1	ɹatuwək daɹumjək	–	22.0 (30.0)
CTrained2	ɹatujək daɹumwək	**86.8 (27.8)**	8.0 (18.8)
HI-Freq-t	ɹatutək **skibejək**	–	**15.1 (24.2)**
HI-Freq-0	ɹatu-ək **skibewək**	–	**16.7 (24.7)**
Anti-Natural	ɹatujək **skibewək**	–	**29.2 (33.5)**

]

Three separate logistic regression models were fit to the data. In the first model data from all but the Full Alternation condition was included[11], and the independent variable was stem type (only trained stems were included in the analysis). Front-vowel stems showed average transfer proportions of 20.5% (from back-vowel final stem allomorphs), while consonant-final stems averaged 15.8% (from back-vowel final stem allomorphs). This difference was significant. See Table 12. However, there was considerable variation across the conditions, and what might be anomalously low values in the CTrained2 condition. Consonant-final and front-vowel final stems were also modeled separately, allowing relevant conditions to be directly compared to one another. The results of the consonant-final stem model shows no difference between

[10]There are indications that learners may also have segment-based biases that lead them to prefer one type of allomorph over another. In the Anti-Natural condition participants seem to prefer to generalize the -/wək/over the -/jək/allomorph (see Fig. 5). Such a bias may be operating in the Variability-3 condition as well (the -/jək/ allomorph is never boosted, although continuous measures show less of a bias). The difference between the HI-Freq-0 and HI-Freq-j conditions – with a higher minority allomorph response rate in the latter – may also be due to a bias: against the -/jək/allomorph, or in favor of the bare -/ək/ allomorph, or both. The fact that -/kək/was produced at rates significantly lower than baseline in the Variability-4 condition, and was boosted less often on both boosting measures suggests that there may also be a bias both against -/kək/and in favor of -/tək/(neither -/pək/ nor -/t͡ʃək/occurred significantly more often than expected).

[11]It was expected that transfer levels would be higher in this condition because of the two sources of transfer.

transfer in CTrained1 and the Full Alternation condition, but significantly lower rates of transfer in the CTrained2 condition. The results of the vowel-final stem model show that both HI-Freq conditions have significantly lower transfer rates than the Full Alternation condition, but do not differ from each other.

Table 12: Results across relevant conditions. Logistic Regressions: Transfer Proportion, Trained stem types only.

All conditions except Full Alternation			
Stem Type (Consonant-Final)	Est.	SE	Pr(>|z|)
(Intercept)	-1.812	.134	<2e-16***
Front-Vowel Stems	.420	.167	.012*
Consonant-Final stems only			
Condition (Full Alternation)	Est.	SE	Pr(>|z|)
(Intercept)	-1.347	.180	6.65e-14***
CTrained1	-.105	.241	.664
CTrained2	-1.082	.310	.000487***
Front-Vowel Final stems only			
Condition (HI-Freq-0)	Est.	SE	Pr(>|z|)
(Intercept)	-1.604	.178	<2e-16***
HI-Freq-t	-.336	.285	.239
Full Alternation	1.604	.229	2.47e-12***

These results are consistent in confirming the CV Hypothesis but it is not entirely clear how to interpret the full set of findings. If transfer rates are expected to be doubled in the Full Alternation condition, then transfer to consonant-final stems is roughly half of what is expected. If transfer rates are expected to be similar to other conditions (depending, perhaps, not on number of stem types but on number of tokens) then it is transfer rates for front-vowel final stems that are more than twice the expected rate. One possible explanation for higher than normal levels of transfer between vowel-final stems is the presence of consonant-final stems during training in the Full Alternation condition. The presence of the consonant-final stems could highlight the similarity between vowel-final stem types, thereby weakening the category boundary and increasing the error rate Unfortunately, there are not enough comparable data to assess this hypothesis, which must be left to future work.

3.3 Individual pre-epenthesizers

The experiments of this paper as a whole do not support the conclusion that a rule or process comparable to epenthesis has been learned. Thus they provide largely negative evidence regarding possible diachronic trajectories for the emergence of synchronic epenthesis. However, there was a small subset of individuals across conditions whose response patterns were *not inconsistent* with an underlying process of epenthesis. In Table 13 those participants are listed by condition, along with the pattern their responses conform to. This will be termed a pre-epenthesis outcome: a pattern that has the potential, over time, and possibly over generations of learners, to become a productive rule of intervocalic consonant epenthesis.

Pre-epenthesis patterns were defined in the following way: highly consistent use of each allomorph (at least 10 out of 12 times with a given stem type), with minimal transfer between vowel and consonant categories. Furthermore, the allomorph used with consonant-final stems must be the bare -ək allomorph, and the allomorph(s) used with vowel-final stems, must *not* be the bare allomorph. 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 can be seen from Table 11, the rate of segmental analysis is very low for these 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 these 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[12].

Table 13: 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 each of the HI-Freq-j, HI-Freq-0, and Variability-3 conditions there is a subset of participants who boost the -/ək/ allomorph on consonant-final stems[13]. In the HI-Freq-0 condition, 8 of those participants are also pre-epenthesizers (with all but 1 of them consistently using the -/wək/ allomorph with front-vowel final stems). The surface response pattern as a whole is analyzable as intervocalic /w/ epenthesis[14]. In the HI-Freq-j condition only 4 of the consonant-final boosters produced a pre-epenthesis pattern[15].

In the HI-Freq-t condition many participants reached high consistency on consonant-final stems. However, none of those reached high consistency with -/tək/ on front-vowel final stems, and only one was highly consistent with -/tək/on back-vowel final stems. In the Variability-3 condition no participants met the necessary conditions. However, there were participants who were highly consistent on the front vowel stems (2 using -/wək/;7 using -/ək/or -/jək/), or the back vowel stems (7 using -/ək/or -/wək/), or the consonant-final stems (6 using -/ək/), or a combination of the two (8 participants total), just none who combined high consistency on all three stem types. In the Full Alternation condition a pre-epenthesis pattern requires participants to consistently maintain the consonant-vowel boundary. Only two participants showed the required pattern.

3.4 Analytic Ambiguity

Many of the experimental conditions in this paper were equally ambiguous between an epenthesis and a deletion analysis, like the example in (11), which could have been generated from a rule of [t] epenthesis intervocalically, or a rule of /t/ deletion word-finally (or even a rule of /t/ deletion post-consonantally, if the t is analyzed as belonging to the suffix morpheme).

(11) [ɹatu] [ɹatutək] [daɹum][daɹumək]

The majority of the evidence shows that participants chose neither epenthesis nor deletion, appearing instead to learn a set of allomorphs that were all realizations of the plural suffix. A reviewer suggests that this outcome might have been due to the lack of evidence distinguishing between deletion and epenthesis. If, instead, learners heard forms that biased towards an epenthesis analysis they might have learned an epenthesis rule. Biasing data could be provided by a form that doesn't alternate. If the data in (12) were provided alongside the data in (11) it might bias learners away from a deletion analysis, and towards an epenthesis analysis. This would occur, presumably, because the deletion analysis would require marking the root /pakut/(and any other similar forms) as exceptional, whereas the epenthesis analysis would be exceptionless.

[12]Interestingly, in both these conditions there are a very small number of -ək inflected consonant final tokens

[13]By boosting measure 1.

[14]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 listener/learners to adopt a segmental analysis.

[15]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.

(12) [pakut] [pakutək]

There are, actually, several conditions in this paper that could be considered to favor one analysis over another. They are collected together in Table 14. Example training data pairs are repeated for ease of reference. The final column provides the underlying forms that are required for the hypothesized analysis. In all cases it was the deletion analysis that was favored over the epenthesis. For example, under the segmental analysis of the surface glides in the Variable0 condition the epenthesis analysis would require exceptions, so the final glide-deletion analysis should be preferred by learners. The Variability-3 condition provides exactly the same kind training data, the only difference being the relative frequencies of the various forms. The Full Alternation condition should also favor a final glide-deletion analysis because it would satisfy morphological uniqueness. Under the subsegmental analysis in the HI-Freq-t condition epenthesis would be unpredictable, but deletion would be completely consistent with the data. In fact, the deletion analysis should also be preferred under a segmental analysis because two unpredictable epenthetic segments would be required. The epenthesis analysis for the Variability-4 condition would also require multiple, phonologically unpredictable, epenthetic segments, so deletion should be preferred in this condition as well.

Table 14: Set of possible deletion-biasing conditions with example training stimuli. Underlying representations consistent with a deletion analysis are provided in the final column.

	Training				Deletion Analysis
Variable0	[ɹatu] [ɹatuwək]	[skibe] [skibe-ək]	[hædi] [hædijək]	[ɹɪlo] [ɹɪlo-ək]	/ɹatuw/ /hædij/ /ɹɪlo/
Variability-3	[ɹatu] [ɹatu-ək]	[skibe] [skibe-ək]	[hædi] [hædiwək]	[ɹɪlo] [ɹɪlojək]	/ɹatu/ /hædiw/ /ɹɪloj/
Full Alternation	[ɹatu] [ɹatujək]	[skibe] [skibewək]	[daɹum] [daɹumək]		/ɹatuj/ /skibew/ /daɹum/
HI-Freq-t	[ɹatu] [ɹatutək]	[skibe] [skibejək]			/ɹatut/ /skibe/or /ɹatut/ /skibej/
Variability-4	[ɹatu] [ɹatutək]	[skibe] [skibekək]	[fi] [fit͡ʃək]	[t͡ʃoɹæno] [t͡ʃoɹænopək]	/ɹatut/ /skibek/ /fit͡ʃ/ /t͡ʃoɹænop/

We will start with the discussion of the HI-Freq-t and Variability-4 conditions, as they provide the clearest evidence. Under the deletion analysis there is only one form of the plural suffix, namely -/ək/. Therefore, the prediction is the same in both conditions: held-out consonant-final stems should be inflected exclusively with -[ək]. As we saw in Section 2.4.2, however, 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/(and all trained allomorphs appeared on at least some consonant-final stems). Not only do these results argue strongly against the deletion hypothesis, they also argue strongly *for* the suppletive allomorphy analysis as the preferred, or default, strategy. Morphological Uniqueness should prefer either of deletion or epenthesis over an analysis that requires 2 or 4 allomorphs. And as we saw in Section 2.2 the Morphological Uniqueness pressure was strong enough to push the analysis of ambiguous forms to the extreme in either direction. 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.

The evidence from the remaining conditions in this section is less clear-cut. However, 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 /ɹatuək/, with an 'excrescent' glide, rather than /ɹatuwək/, with a segmental glide. In the Consistent-W condition this meant the opposite: analyzing [ɹatuwək] as /ɹatuwək/, rendering it consistent with forms like [daɹumwək]. While individual variation was the norm in these experiments, Morphological Uniqueness imposed a level of consistency both within and across participants not seen under any other circumstances.

While Morphological Uniqueness was confounded with the Phonetic Hypothesis in the Natural and Variable0 conditions, a preference for the subsegmental analysis of ambiguous forms could be seen in the CTrained1 and HI-Freq-t conditions. The subsegmental analysis was chosen, on average, 65% of the time.

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 a 'conjugation' analysis, than a rule-based analysis (Hale (1973)); tracking which of a set of allomorphs went with which stems. 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 generative analysis participants showed no evidence of having learned such a rule. In terms of consistency, learners were actually worse in the full pattern 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 natural classes as predictors of allomorph identity. In the non-variable conditions allomorphs were completely predictable by stem type, and stem types were definable as [+consonantal] versus [-consonantal], and [+front] versus [+back] (or some other comparable configuration of features). Although accuracy was better on average on the $\pm$consonantal natural class 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 with consonants only swapping with other consonants, and vowels with other vowels (e.g., MacKay (1970), Stemberger (1990)). However, most such errors occur at the beginnings of words – a highly salient position. It may be that distinctions, or categories, available at the beginnings of words are weaker, or non-existent, elsewhere in the word; e.g., an implicit categorization of vowel-initial versus consonant-initial words, but no clear division of vowel-final versus consonant-final words. 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.

Results averaged over participants showed frequency matching behavior: high-frequency allomorphs were not 'boosted' beyond their trained levels. However, individual participants sometimes showed generalizing behavior, picking a single allomorph with high consistency. And more boosting occurred under a non-uniform frequency distribution. Contrary to the memory load hypothesis, however, more variants and less predictability led to less generalization beyond the input. However, learners showed a general predilection for boosting; even in the absence of a majority allomorph, approximately 33% of participants boosted one or the other of two allomorphs (as in the Anti-Natural condition).

It also appears that certain individuals are more likely than others to adopt a segmental interpretation of an ambiguous token, just as certain individuals were more likely to 'boost' (or generalize). These 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), Babel et al. (2014), Dimov et al. (2012)). Certain individual participants can also be categorized as producing a pattern that could become epenthetic over time – those that boost, but also maintain category boundaries. Successive incremental changes over this type of data may provide a path to phonologization (cf. Kirby (2001)).

5 Discussion & Conclusions

The focus of this paper has been on the epenthesis of consonantal segments. More specifically, we have only looked at cases of consonant epenthesis that result in what generative theory would deem less marked, or more preferred surface structures (e.g., a syllable with an onset versus one without). This decision was made based on the hypothesis that patterns conforming to the hypothesized universals of phonological theory would be more common, and also more directly traceable to their (natural) phonetic origins. We began with an example of one type of this kind of epenthesis, what could be called a default epenthesis pattern, in which the same segment is inserted in all relevant contexts regardless of other featural differences. A second type of intervocalic epenthesis that is relevant to these results is what can be called assimilative epenthesis (cf. de Lacy (2006)), in which the epenthetic segment shares features with one or both of its flanking vowels[16]. Under the assumption that both types of pattern are attested synchronically they must be both learnable, and have a non-zero probability of arising diachronically. The next two sections will sketch out the diachronic origins of each type of epenthesis, and section 5.3 will consider a specific case study: English intrusive r. These will be analyzed in light of the experimental results which suggest that such patterns are difficult to learn and unlikely to be induced from ambiguous data. The paper will end with the ramifications of these results for synchronic phonological theory more generally.

5.1 Epenthesis from Coarticulation

Although a synchronic pattern of assimilative epenthesis may be analyzed as a default segment that assimilates to some of the features of its adjacent vowels, diachronically the pattern is taken to derive from gestural overlap, or coarticulation of vowel sequences, yielding segments like glides, certain fricatives, and, in some cases, glottalics (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 (13). In this language the front glide [j] is inserted after high front vowels, and the back glide [w] is inserted after back vowels.

(13) Balangao (Shetler (1976))

/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 observed. For example, a three-way split between [j] after front vowels, [w] after back vowels, and [ʔ] after the low vowel /a/ occurs in Malay. And Malayalam inserts [ʋ] after back vowels and [j] after front vowels (See Morley (2015) for details on these and other examples). While these patterns may all have their origin in phonetic

[16]It should be noted that a third kind of 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, where the segment deleted was either a glide or part of a diphthong. It is also possible that dissimilatory patterns have a unique diachronic trajectory. Given the paucity of data for these types of patterns I refrain from speculating about them in this paper.

coarticulation, the fact they differ by conditioning context and by epenthetic segment argues that they have entered the phonology in 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 one way 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, however, how this effect could lead to a phonological epenthesis pattern since 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 an avenue for future work. A second requirement for the phonologization of coarticulatory epenthesis is generalization, or regularization. While an excrescent high back glide following a high back vowel may happen almost inevitably, it is less true for a high back glide following a low back vowel. Generalization is required either at the phonetic or the phonological stage, and probably at both. A large degree of regularization is a requirement for non-assimilative epenthesis as well and, in fact, any categorical pattern that results from continuous change.

5.2 Epenthesis from Deletion

Default (non-assimilative) epenthesis is typically analyzed as arising from 'rule inversion' (Vennemann (1972)) in which learners analyze forms that have undergone historic deletion as underlying, and forms that have *not* undergone deletion as the result of epenthesis. This scenario is illustrated in 14 a-14 c. In (14 a) a hypothetical stem morpheme is shown inflected with a vowel-initial suffix as well as a consonant-initial suffix. There are no phonological alternations. At some later time, (14 b), a stem-final consonant is lost in those environments in which it was followed by another consonant – a common historical change. This affects the second inflected form, but not the first. In the final stage, (14 c), the learner is faced with ambiguous data. There is an alternating [t] that appears in some forms and not others. 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.

(14) (a) /pamit/ /pamit+o/→[pamito] /pamit+nu/→[pamitnu]
(b) Deletion: /pamitnu/ > [paminu]
(c) Re-Analysis: /pami/ /pami+o/→[pami**t**o] /pami+nu/→[paminu]

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 a synchronic default epenthesis pattern Morley (2012). Table 15 adapted from Morley (2012) illustrates the specific case of C/Ø alternations at the stem-suffix boundary. For the re-analysis in 14 c to be adopted, it is critical that learners analyze the first 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 [fisemo] in this example. The same argument holds for the analysis of the consonant-initial suffix, requiring historically vowel-final stems to illustrate invariance in this case. However, the resulting paradigm is still not sufficient for a synchronic epenthesis analysis. 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, which appear with different 'epenthetic' segments – in this case [n].

Table 15: Hypothetical Precursor Epenthesis System

historic stems	after loss of pre-final, 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 15 into default epenthesis: 1) the differences between the vowel-final stems must be leveled. For the case of /t/ epenthesis, all [kifuno] type forms must be changed to [kifuto]. And all [oru] 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; 2) the generalization in 1) must extend to all vowel-initial suffixes in exactly the same way, such that the variant chosen with -/o/ is the same variant chosen with -/ina/, etc.

As we have seen in the experiments reported here learners often frequency match; they transfer across consonant-vowel boundaries; and they tend to learn sets of allomorphs, with a strong preference for a unique morpheme realization. In fact, it appears as though both non-typical types of learners and specific historical conditions are required in order for default synchronic epenthesis to result. As a result it is predicted to emerge only rarely. 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 sample of languages must be agreed upon, one that would provide a good approximation for the actual rate of occurrence of epenthesis cross-linguistically. Then a proper baseline would have to 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 long before that point. 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 (2003, 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, indeed, synchronic epenthesis is exceptional, rather than commonplace, 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), Gimson (1970), Harris (1994)), 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; this occurred pre-consonantally both within and across words, as well as prepausally (e.g., [bɑɹ#bəl͡oʊ] > [bɑbəl͡oʊ])[17]. 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"

[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.

[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 r cannot be underlying in these cases (see, for example, Johansson (1973), Rubach (2000), McCarthy (1993)).

It is less straight-forward to develop an analysis that includes both types of observed r's. General RP (as reported by Durand (1997)), and Boston English (as reported by McCarthy (1993)) are among those varieties that exhibit both linking and intrusive r. Most, if not all, of these dialects show variable rates of intrusive r. In General RP, for example, the probability of intrusive r depends on the quality of the preceding vowel (occurring most frequently after [ə]), as well as speaking rate, and register. Johansson (1973) attributes some of this variability to the stigmatization of inserted r, and speakers' attempts to correct for it, sometimes incorrectly. He proposes a unified epenthesis analysis based on the following diachronic changes: First, a stage in which speakers had both a rule of deletion (involving linking r's) and a rule of epenthesis (intrusive r's). Second, gradual generalization of the epenthesis rule, due the misidentification of historic r's as intrusive. Mistakes lead to more mistakes, a they increase the ambiguity surrounding a given surface instantiation of the r/Ø alternation. Once speakers completely lose track of which words were historically r-final, rule inversion, in which the epenthesis rule takes over, is assumed to be complete. The paradox in this account is that it is subject to the same argument that is often made against the deletion analysis – namely, the evidence from orthography ("there is obviously no r in comma" (Foulkes (1997)). True rule inversion requires changes to the underlying representations of historically r-final words. Thus, for example, /kɑɹ/>/kɑ/must occur, despite the fact that there *is* obviously an "r" in "car".

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 Johansson (1973), Wells (1982), Harris (1994)). These "hyper-rhotic" dialects are also attributed to stigmatization, but in this case for r deletion, rather than r insertion. 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". This analysis implies a process of r deletion that speakers are attempting to undo, and Harris (1994) takes the over-generalization of r insertion 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 not be markedness improving (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 unlikely, but necessary, type of learner generalization. The explanation for how such generalization occurred may be found in the phonetics of rhotic segments.

Gick (1999) analyzes r (and l) as being comprised of two articulatory gestures: a vowel gesture, and a consonant gesture. r loss in his framework is actually merger of the vowel gesture with the preceding vowel, and accompanying loss of the consonant gesture (also known as "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, such that loss of the consonant-gesture of r would also result 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 (Broadbent (1991), Rubach (2000), Uffmann (2007)). 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 can 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), Steriade (2001), Myers (2002))). On closer examination many apparently categorical phonological rules have been seen to show variability and gradience. This could be the result of a later phonetic process applying to the phonological output; or evidence for an exclusively phonetic rule; an indication that there is no rule at all; or evidence for a collection of competing phonological rules.

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, Chomsky and Halle 1968). 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 in order for a given language to be learnable at all. 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 3.4, repeated here in (15) and (16).

(15) [ɹatu] [ɹatutək] [daɹum][daɹumək]

The alternating forms in 15 are meant to represent a fully ambiguous scenario, consistent with both epenthesis of [t], and deletion of [t]. The additional form in 16 presumably tips the scales towards an epenthesis analysis.

(16) [pakut] [pakutək]

But what kind of test items could provide definitive proof that learners had adopted this analysis? The epenthesis hypothesis predicts that novel consonant-final singular forms should retain their final consonant in the plural. And novel vowel-final forms should 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, if forms like that in 16 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 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 hypothesis. Performance errors and lexical exceptions are to be expected, and, in practice, linguists tend to assume some sort of 'reasonableness' criterion. The problem is the lack of a 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 linguistic universals (see Morley (2015) for a more in-depth discussion of this issue).

This paper represents part of a larger project to answer questions about universals in phonological consonant epenthesis. As has just been argued, the undertaking goes far beyond a single phonological pattern. And the true scope of the question encompasses the learning problem (this paper), diachronic change (Morley (2012)), and typological methodology (Morley (2015)). The conclusions are the following: there is little evidence that phonological (versus phonetic) epenthesis is an accessible hypothesis 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 simple association, 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/`.

Babel, Molly, Grant McGuire, Sophia Walters, and Alice Nicholls. 2014. Novelty and social preference in phonetic accommodation. *Laboratory Phonology* 5:123–150.

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 universal 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.

Gimson, Alfred Charles. 1970. *An introduction to the pronunciation of english*. Hodder Arnold.

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. 2003. Maximal words and the maori passive. *MIT Working Papers in Linguistics* 44:20–39.

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.

Manuel, Sharon Y, and RA Krakow. 1984. Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratories Status Report on Speech Research* 77:69–78.

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. 2012. The emergence of epenthesis: An incremental model of grammar change. *Language Dynamics and Change* 2:59–97.

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

Morley, R.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. 2004. *Optimality theory*. Blackwell Publishing.

Rischel, J. 1974. *Topics in west greenlandic phonology*. Copenhagen: Akademisk Forlag.

Ross, D. S., and E. L. Newport. 1996. The development of language from non-native linguistic input. In *Proceedings of the 20th annual Boston University Conference on Language Development*, ed. A. Stringfellow, D. Cahana-Amitay, E. Hughs, and A. Zukowski, 623–645. Boston: Cascadilla.

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

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

Stemberger, J.P. 1990. Wordshape errors in language production. *Cognition* 35:123–157.

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

Uffmann, Christian. 2007. Intrusive [r] and optimal epenthetic consonants. *Language Sciences* 29:451–476.

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.

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	ˈtɪɪfu	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 & values	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 & variable	ɹatu-ək	skibewək		12	6		2:1
10	HI-Freq-j	22	i, ii, & vi	iii, iv & variability	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