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, strong 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

In default epenthesis a consonant not present in the input, or underlying form, appears in the surface form. Consonant epenthesis is often taken to be motivated by prosodic structure constraints, occurring prevocalically in order to avoid vowel hiatus, or provide a missing syllable onset (e.g. Prince and Smolensky (2004)). Additionally, the identity of this consonant is taken to be fixed within a particular language; see example in (1).

(1) Misantla Totonac (MacKay (1999))

/laa+a̱̰ʔi̱̰ʃki/→	[laaʔa̱̰ʔi̱̰ʃki]
	3PL.OBJ-lend
	‘lend them’
/ta+a̱̰n/→	[taʔa̱̰n]
	3PL.SUBJ-go
	‘they go’
/naa+u̱̰tun/→	[naaʔu̱̰tun]
	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 (low symbol count) re-write rule, or as embodied in a 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. Language change is represented in generative theory by the loss or addition of rules on the one hand, or the re-ranking of constraints, on the other. The question of how the linguistic data come to change so as to necessitate a change to the synchronic grammar is assigned to the separate domain of historical linguistics. However, failure to consider the diachronic origins of synchronic language patterns is deeply problematic. This is because doing so effectively misrepresents the learning problem. In a real-world scenario in which change is continuous, data are messy, and learners might adopt the hypothesis that no regular pattern is present, it becomes necessary to reconsider assumptions at the very heart of linguistic theory. It is no longer possible to select the best hypothesis that achieves complete descriptive adequacy; in the first place, no hypothesis accounts for absolutely all the data; and in the second, there exist competing hypotheses that differ along dimensions that are not defined within the formal apparatus of the theory.

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. More recently, experimental work has been used to discover apparently innate biases in how learners generalize 'impoverished' linguistic input, biases that might account for observed cross-linguistic tendencies (e.g., Wilson (2006), Finley and Badecker (2008), Culbertson (2012), Berent et al. (2009), Moreton (2008)); this work has been paralleled by investigation of how learners deal with the subset/superset relationship in hypothesis testing (e.g.,Tenenbaum and Griffiths 2001, Pearl 2011). However, the proper treatment of individual, or multiple, forms that represent exceptions to a general pattern has been conspicuously missing from the phonological learning literature (see Yang 2005, 2011 for a metric to determine morphological productivity). Although it is fairly widely acknowledged that exceptions are more the rule than otherwise in linguistic data, the full ramifications of this fact have often failed to be appreciated. If certain forms can be labelled as lexical exceptions then the ambiguity inherent in phonological analysis becomes even stronger; is it segment X that becomes Y in the context of A, or is it segment Y that becomes X in all other environments? Is it [ʔ] that gets inserted before a vowel, or is it /ʔ/ that gets deleted before a consonant? This is a question not only for the linguist, but for the learner. And it is a question not only about the proper analysis of synchronic data, but about how diachronic changes are translated into new synchronic grammars. This paper presents the results of a series of experiments designed to systematically test how properties of the input data affect learning; to isolate some of the factors that could lead to the phonologization of a phonetic process; and to simulate conditions that could have arisen through certain historic sound changes[1].

1.1 From Phonetics to Phonology

It is the generally held view that most, if not all, phonological patterns arise out of existing phonetic variation (see Ohala 1981, Guy 2008, Weinreich et al. 1968, and many others). This connection is perhaps more readily apparent in certain types of patterns, such as local assimilation of voicing, nasality, or place of articulation. For consonant epenthesis, such a connection is suggested by patterns in which the identity of the epenthetic segment is conditioned by the local phonetic environment. See (2). This type of pattern is sometimes known as "assimilative epenthesis" (cf. de Lacy (2006)), in contrast to what was previously defined as default epenthesis. The two types of epenthesis are hypothesized to result from different diachronic origins. The former is taken to be one of the possible systems resulting from gestural timing/re-timing of vowel sequences, yielding segments like glides, certain fricatives, and, in some cases, glottalics, that can be described as either articulatorily or perceptually minimal (e.g., Steriade (2001), Clements and Hume (1995)); whereas the latter is taken to result from re-analysis of a diachronic process of deletion.

(2) Balangao (Shetler (1976))

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

/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’

For many consonants, features like place of articulation are most strongly cued by the transitions into a following vowel. Conversely, following consonants are apt to mask the transitional cues that signal the place of the preceding consonant and weaken the percept of that consonant in general (Fujimura et al. (1976), Repp (1977), Hura et al. (1992), etc.). This type of masking can lead to the loss of features such as nasality, voicing, or place, resulting in assimilation with the following consonant. In the extreme, the entire consonant may fail to be recovered by the listener. A diachronic change of this kind can result in a synchronic alternation between a consonant which appears in pre-vocalic contexts but is missing from pre-consonantal contexts. In theory, such an alternation could be analyzed by the learner as epenthesis due to the fact that deletion contexts mirror epenthesis contexts (undergoing ‘rule inversion’, see Vennemann (1972)). For this to occur there must, presumably, be some reason for the learner to prefer the epenthesis analysis over the deletion analysis.

Morley (2012) is an analysis of the necessary and sufficient conditions for this re-analysis to occur. Table 1 adapted from Morley (2012) illustrates the specific case of C/Ø alternations at the stem-suffix boundary. The potential epenthetic segments in this toy example are /t/ and /n/ – segments that were historically lost stem-finally. The environment for epenthesis is between the newly vowel-final stems and vowel-initial suffixes, such as -/o/. The evidence that this suffix is underlyingly vowel-initial comes from the inflected forms of historically consonant-final roots that did *not* undergo consonant loss – such words are represented by the form [fisemo] in this example. However, the resulting paradigm is 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, [kifu] appears to be a vowel-final root, yet surfaces with [n], not [t], inter-vocalically.

Table 1: 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/

Two types of generalization are required to transform the synchronic pattern from Table 1 into default epenthesis: 1) the differences between the vowel-final stems must be leveled, such that either [pamito]-type forms (and all other variants) are changed to [pamino]; or [kifuno] type forms (and all other variants) are changed to [kifuto]. At the same time, forms like [oru] must adopt an ‘epenthetic’ segment, becoming [oruto] (for a synchronic [t]-epenthesis system); this leveling must occur only within the synchronic vowel-final roots, and not cross over to the consonant-final ones to 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.

This paper tests these and other hypotheses about learner generalization over ambiguous, incomplete, unbalanced, and variable data designed to mimic different stages in the evolution of a default epenthesis pattern. These results will bear on the likelihood of the diachronic model outlined above, as well as provide information about the learnability of epenthesis as a synchronic grammatical rule. More broadly, it is expected that these experiments will provide finer-grained insight into the evolution of phonological grammars via the process of learning/acquisition.

1.2 Overview

Findings from three sets of artificial grammar learning experiments are presented, all involving a simple morphological paradigm: an uninflected singular, and a plural formed by suffixation. How much the learner's input was consistent with an epenthesis analysis varied across conditions, but competitor analyses, such as deletion, or suppletion (unpredictable allomorphy) were always possible. The first set of experiments explored the role learning biases play in the acquisition of such patterns. Given input that was consistent with both a phonetic and a phonological analysis, listeners overwhelmingly adopted the phonetic interpretation. However, an apparent pressure to minimize allomorphs could push learners almost categorically toward the phonological interpretation.

The second set of experiments tested the effect of the amount of evidence, and the number of allomorphs, heard during training. In conditions in which both vowel-final and consonant-final stems were heard during training participants made errors in which they produced consonant-final stems with allomorphs that should only have occurred with vowel-final stems, and vice versa. In other words, participants did not always maintain a distinction between a natural class of consonants and a natural class of vowels, nor did they always do so between a class of front vowels, and a class of back vowels. In fact, the best predictor of accuracy seemed to be the number of distinct allomorphs heard during training. The fewer the number of allomorphs, the better participants did. Thus, the conditions in which no stem types were held out, and the evidence for epenthesis was complete, participants performed the worst. Amount of transfer across category boundaries was also affected by token frequency and within-category variability.

The third set of experiments specifically targeted factors that might encourage, or inhibit, generalization over the input data. These factors aligned closely with those related to consistency in the second set of experiments. It was found that a disproportionately frequent allomorph tended to be selected by listeners even beyond its base frequency, as well as in contexts in which it did not appear during training. Learners thus generalized based on token frequency. Irregularity, on the other hand, where variants sometimes appeared in otherwise predictable contexts, seemed to inhibit generalization, causing learners to more closely match the training frequencies for novel test items. Likelihood of generalization was also found to vary across individuals. Not only were certain participants more likely to regularize the training paradigm by 'boosting' one allomorph, they also varied in which allomorph they chose. In conditions in which the distribution was skewed, the majority of participants who regularized chose the most frequent variant to boost. But there were some participants who chose a minority allomorph instead. In conditions in which all allomorphs appeared an equal number of times there sometimes appeared to be a general bias towards one of the variants, but this was not completely consistent across conditions. While the vast majority of participants chose the phonetic interpretation when there was ambiguity there were also occasional participants, over occasional tokens, who selected the phonemic interpretation.

The combined results of the three sets of experiments indicate that learners are very unlikely to learn epenthesis, even with completely consistent input. This is in part because learners are overwhelmingly inclined to pick the phonetic interpretation of their data when possible. However, even when the interpretation of the stimuli was not ambiguous, participants seemed to learn a morphological, or suppletion, analysis (see Hale (1973)), evidenced by their inconsistent use of the variants on novel stems. Additionally, participants did not maintain strict category boundaries based on phonological features. This outcome is problematic for an account of epenthesis emergence that relies on learner generalizations of a certain type. Examining the individual data from each experimental condition showed that a very small subset of participants did learn a consistent pattern that might qualify as a proto-epenthesis system. However, these numbers were very low, and the more of the relevant contexts learners were exposed to, the lower they got. This leaves us with a very narrow potential route to synchronic epenthesis: a small population of individuals who are prone to a phonemic analysis, likely to generalize, but less likely to merge category distinctions. Section 2 provides the full description of the experiments and the results by condition. Section 3 summarizes and synthesizes the entire set of results; and the paper concludes in Section 4 with a discussion of the diachronic model of epenthesis emergence, the synchronic typology, and the theoretical treatment of epenthesis.

2 Experiments

The experiments in this paper are conceptually part of the tradition of laboratory sound change pioneered by John Ohala and colleagues (Ohala 1981, 1983, 1993; Hombert et al. 1979, Kawasaki 1986), and continued in the work of Solé 2007, Harrington et al. 2008, Beddor 2009, and others. Methodologically, however, they belong to the class of artificial grammar learning (AGL) experiments, a paradigm that has become increasingly popular among linguists. The paradigm has been used with adults, children, and infants; to train participants on phonological, and syntactic patterns; with explicit alternations, and with conditional probabilities of occurrence. The paradigm has become quite widely used to test for differences in the learning of 'natural', or typologically common patterns, and 'unnatural', rare, or unattested patterns, as well as to detect biases in the degree and type of generalizations participants make from small samples. 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, instructing 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 (although certain types of patterns remain persistently intractable, e.g., Pallier et al. 1997, Takagi and Mann 1995). Participants have been shown to learn novel words and patterns quickly within this paradigm, and to implicitly acquire grammatical rules of which they are consciously unaware.

2.1 Procedure

A total of 244 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 which were presented as 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.

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 2 for the complete list of stimuli used). In each condition there were 18 test words, 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.

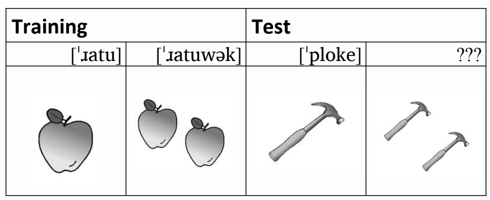


Figure 1: Training Paradigm

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 2. All words were stressed on the penultimate syllable of the stem, in both singular and plural forms. All plural forms were of the form singular + Xək. X was either 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., skibejək, skibewək, skibekək, skibe-ək), depending on experiment and condition.

The last type of stimuli was necessary for distinguishing between a phonetic versus phonological hypothesis on the part of listeners (see discussion of Experiment 1). 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. Productions of [ɹatuwək] and [ɹatuək] by the speaker of the stimuli were auditorily and spectrographically highly similar. For this reason, 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 were rather unnatural sounding. However, the experimental results show that such tokens were natural-enough sounding to be categorized by listeners as reflexes of underlying vowel-vowel sequences. Participants developed various strategies for reproducing these

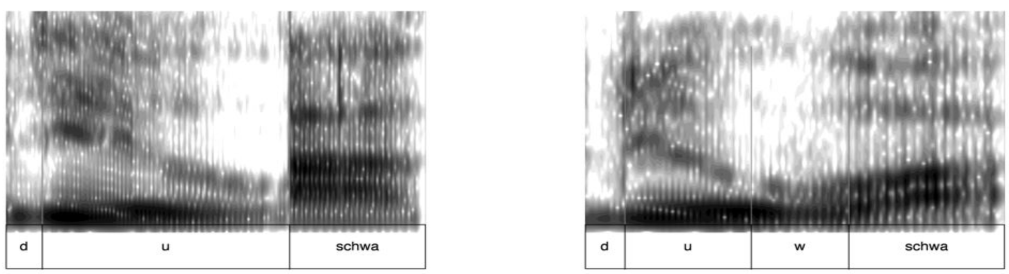
[2] the responses for the feedback trials were recorded but not analyzed

Table 2: Full list of stimuli roots across all conditions, both training and test items. Roman numerals are used to indicate the sets used in particular conditions. See A.2.

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/d͡ʒoˈɹʌfim[a]	ˈhoʃɪn/ˈt͡ʃalum[a]	ˈɹibæz	ˈpɹɛv	ˈbihɪl
vi	ˈtwit͡ʃo	ˈðipu	ˈmuɹo	ˈtɹɪfu	ˈmeko	ˈgɛθu/ˈpiʃu[a]
vii	ˈsabol	ˈgenʊɹ	ˈtɹifæd	(ˈtagæf)	d͡ʒoˈɹʌfim/keˈtɛlan[a]	ˈt͡ʃalum/ˈhoʃɪn[a]

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

unnatural words: altering vowel quality, introducing glottal stops, or significantly drawing out their articulations. Participant responses were coded by a phonetically trained listener as to whether or not they were consistent with a VV production. Since finer measurements are needed to definitively establish whether listeners intended to produce /ɹatuwək/ or /ɹatuək/, the two are not differentiated for the purpose of the analyses presented here. For test items consisting of consonant-final stems, however, the presence versus absence of a glide is quite clear (e.g., [daɹumwək] versus [daɹumək]).



(a) spliced: [gaidu-ək]   (b) unspliced: [gaiduək]

Figure 2: Spectrograms of Example Stimuli

2.1.3 Conditions

In all experiments participants were presented with a morphophonological alternation pattern that was consistent with multiple analyses. In most conditions, only a subset of the alternating forms were presented during training (e.g., vowel-final stems), while generalization to novel stem types was assessed during testing (e.g., consonant-final stems). There were a total of three classes of stem used in all experiments: back-vowel final (o/u; e.g.,[ɹatu]), front-vowel final (e/i; e.g., [skibe]), and consonant-final (e.g., [daɹum]). Table A.2 provides a reference for comparison of the full set of conditions: training items, test-items, held-out stem types, and relative token frequency of each type. There are a total of 13 conditions described in this paper. Their discussion will proceed in the following way. Subsets of conditions will be grouped under different thematic headings in order to present a series of relatively self-contained and coherent results. Since certain conditions bear on multiple questions, those conditions will appear in multiple sections. In certain cases, different aspects of the same results will be discussed in different sections. The theme of the first set of experiments is Learning Biases, sub-divided into Phonetic Primacy and Morphological Economy. The theme of the second set of experiments is Evidence & Allomorphs**.** The theme of the third set of experiments is Generalization, sub-divided into Frequency and Variability. For ease of reference the relevant portions of Table A.2 will be reproduced at the start of each section in which those conditions are discussed.

2.2 Experiment 1: Learning Biases

In the first set of experiments participants were exposed to a partial pattern from which one type of stem was missing. At test, the held-out stems provided a test of learner generalization. In two of the conditions the pattern was completely consistent with an epenthesis analysis. In the third, the pattern was partially consistent; and in the fourth, not consistent. See Table 3.

Table 3: Experiment 1

	Condition	N	Back Vowels	Front Vowels	Consonants	Token Counts		
1	Natural	18	ɹatuwək	skibejək		6	6	
2	Anti-Natural	18	ɹatujək	skibewək		6	6	
	Bi-Modal		ɹatuwək	skibejək		3	3	
3		20	vu-ək	fi-ək		3	3	
4	Consistent-W	20	ɹatuwək		daɹumwək	6		6

2.2.1 Set 1: Phonetic Primacy: Conditions 1-3

In Conditions 1-3 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 same singular items, paired with the same pictures were presented in each of the three conditions. They differed only in the formation of the plural. The plural suffix was also the same for all conditions (and all experiments described here); however, the transition from stem to suffix varied by condition. In the “Natural” condition, all plurals 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. The surface forms for both of these conditions are consistent with an epenthesis analysis. However, the linguistic information available to the participants is impoverished (in the sense of Wilson (2006)); they never hear inflected consonant-final stems. In the Natural condition, there are at least three possible analyses of the input: (i) phonetic epenthesis: the glide is the consequence of the fluent articulation of the sequence of two vowels; (ii) phonological epenthesis: the glides are full segments inserted between the stem-final vowel and the suffix-initial vowel by grammatical rule (iii) allomorphy: there are two allomorphs of the plural: -/jək/, -/wək/.

Each analysis makes predictions about how participants will inflect novel consonant-final stems: hypotheses (i) and (ii) both predict that there should be no glides in the plurals of consonant-final stems. See (3). In the case of (i) this follows from the fact that there is no articulatory reason to produce a glide between a consonant and a vowel; in the case of (ii), however, the prediction follows from the fact that there is no grammatical reason to insert a glide between a consonant and a vowel. Analysis (iii), on the other hand, predicts that consonant-final stems will surface with one or the other glide when appearing in the plural. See (4).

(3) /daɹum+ək/ $\rightarrow$ [daɹumək]

(4) /daɹum+wək/ $\rightarrow$ [daɹumwək] ; /daɹum+jək/ $\rightarrow$ [daɹumjək]

Participants overwhelmingly inflected consonant-final stems without the glide, as in (3). Because this outcome is predicted by both analysis (i) and (ii), a third condition was designed to differentiate between the two. In the "Bi-Modal" condition two types of inflected forms were presented: 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 on Stimuli above). 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). 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/.

As the results in Fig 3demonstrate, this was not what happened. Instead, participants produced even fewer glides with consonant-final stems than in the Natural condition (neither condition was significantly different from 0). This strongly implies that participants were choosing the phonetic over the phonological analysis almost without exception in these two conditions, and thus learning a single allomorph: /ək/. 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 Bi-Modal 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", "ratu-uk", "ratoouk", and "ratu'ak". No consonant-final stems were ever written with a glide[3]. The Anti-Natural condition, by contrast, does not permit a phonetic analysis. Here, participants may select a phonological epenthesis analysis or a morphological analysis. Figure 3 indicates that participants overwhelmingly chose the morphological analysis. Almost all consonant-final forms were inflected either with a [w] or a [j] in the plural. Even though the data were consistent with an epenthesis analysis participants did not select it. This result may be due to the specifics of the experimental design: the fact 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. In fact, the full set of results suggest that the conditions for acquiring an epenthesis pattern may be quite restrictive, and that a strong bias exists to attribute morphophonological alternations to allomorphy on the suffix itself.

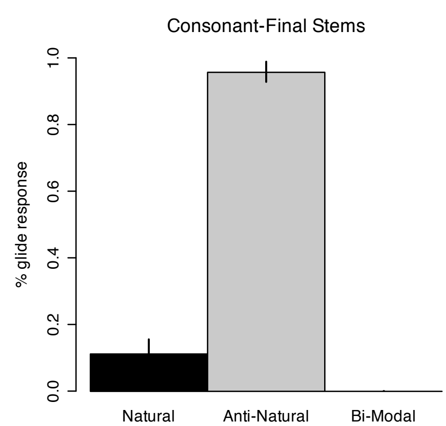


Figure 3: Proportion of glide response (-wək or -jək) on novel consonant-final stems

[3]The only orthographic glides in the entire condition came from the front-vowel final inflected forms of 3 participants, for a total of 8 tokens with "j", or 16% of responses.

2.2.2 Set 2: Morphological Economy: Condition 4

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 phonemic or a phonetic [w]. See Table 3. Front-vowel final stems were held out. Based on the results of the Natural and Bi-Modal conditions, it was expected that learners would choose the phonetic interpretation for the glide they heard in back-vowel final inflected stems. The phonetic interpretation is not possible for the consonant-final stems. Therefore, a two-allomorph analysis was predicted: -/wək/, and -/ək/.

Critical items at test were the novel front-vowel final words, where the presence versus absence of a /w/ in the plural suffix is unambiguous. Contrary to expectation, participants almost categorically produced glides in the inflected forms of front-vowel final stems, indicating that they had learned a single allomorph of the plural: -/wək/. See Fig 4. Presumably, this result is due to the presence of the [w] on inflected consonant-final stems during training which, unlike the variable forms in the Bi-Modal condition, can only be interpreted as phonemic. To further explore the conditions under which a phonological analysis can be elicited, the results from two additional conditions will be briefly discussed here.

Before proceeding further, however, the naming conventions for the remaining conditions are described. The name of the condition always begins with the number of allomorphs heard during training. Then the stem types follow, either VV for back- and front-vowel final stems; VC for back-vowel final stems and consonant-final stems (in cases in which one type of vowel-final stem was held out, it was always front-vowel final); VVC for the full set of stem types. The variable part of the allomorph (X) that goes with each stem is listed next: {0, j, w, t, etc.}, separated by dashes. The order of allomorphs is always: back-vowel final, front-vowel final, c-final. If a given stem takes multiple allomorphs those will be listed consecutively, highest frequency first, with no space between the letters. The suffix ".F" indicates that the frequency of the different allomorphs (as well as the different stem types) are not the same (in cases in which the two vowel-final stems were heard at different rates, it was always the back-vowel final that were heard more often).

In the 2VV.w-j and 2VV.t-j.F conditions participants also heard a combination of ambiguous and non-ambiguous segments during training. In the former the unambiguous glide was [j] – on consonant-final stems – , and the ambiguous glide was [w] – on back-vowel final stems. Front-vowel final stems were held out (see Table 4). In the latter the unambiguous consonant was [t] – on back-vowel final stems– , and the ambiguous consonant was [j] – on front-vowel final stems. Consonant final stems were held out (see Table 5). As Figure 4 shows, neither of these conditions induced a robust phonological analysis for the ambiguous glide: numbers of responses of front-vowel final stems with [w], and consonant final stems with [j], respectively, were not significantly different from the response levels in the Natural condition and Bi-Modal conditions, which were at floor.

2.2.3 Results & Discussion

The full set of results from conditions 1-4 is interpreted to indicate, firstly, a primacy of phonetic naturalness over phonological. 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 Bi-Modal condition, however, resulted in a phonologically unimodal response distribution. 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

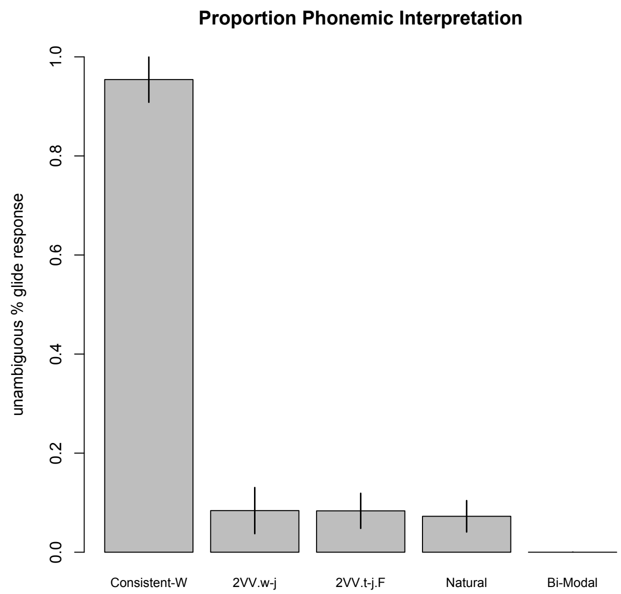


Figure 4: Novel stem-type test items: Percentage of responses containing an *unambiguous* glide (means and standard error bars). This included front-vowel final stems produced with the [wək] allomorph; back-vowel final stems produced with the [jək] allomorph; and consonant-final stems produced with either, depending on condition.

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)). Statistically, the Natural and Bi-Modal conditions do not differ from one another, but both differ significantly from the Anti-Natural condition[4] (logistic regression (glm) with condition as independent variable, $p < .05$).

In the Consistent-W condition participants also chose a single allomorph, again indicating a preference for fewer morphemes. Additionally, this result indicates that the segmental interpretation is, in fact, accessible. Listeners have been biased, potentially both by general phonetic considerations, as well as their own language-specific experience, to compensate perceptually for the glide-like acoustics that emerge predictably between vowel articulations of certain kinds. The bias towards the phonetic interpretation, while strong enough to trump phonological and acoustic factors, can be overcome. Morphological economy over-rides phonetic predictability. By comparison, the presence of an unrelated consonant in the 2VV.w-j and 2VV.t-j.F conditions does not result in a glide percept. Participants seemed to have learned glide-less allomorphs in both conditions; -/ək/ rather than -/wək/ or -/jək/, respectively. Thus it appears that a phonological analysis only results in cases where the phonemic glide has the same qualities as the ambiguous glide.

2.3 Experiment 2: Evidence & Allomorphs

Although the results of the Natural condition described above are consistent with an epenthesis analysis (a surface [w] or [j] inserted between the stem-final and suffix-initial vowels, but missing between stem-final consonants and suffix-initial vowels) it was determined that participants interpret this pattern as resulting from phonetic coarticulation and do not select a phonological epenthesis analysis. In Experiment 2 the sufficient training conditions for the adoption of the epenthesis analysis will be directly addressed.

The training stimuli in the 3VVC.j-w-0 condition represent a completely consistent and comprehensive set of evidence for a process of dissimilatory epenthesis. See Table 4. This pattern, of course, also

[4] All statistical tests were conducted using the R Software Environment Version 3.2.3

Table 4: Experiment 2. The last column provides the relative number of participants in each condition who were highly consistent over consonant-final stems.

	Condition	N	Back Vowels	Front Vowels	Consonants	Token Ratio			Allomorph	% Consistent
5	3VVC.j-w-0	18	ɹatujək	skibewək	daɹumək	6	6	6	-/ək/	50
6	2VV.w-j	24	ɹatuwək		daɹumjək	6		6	-/jək/	66.7
7	2VC.j-w	19	ɹatujək		daɹumwək	6		6	-/wək/;-/jək/	89.5; 5.26

permits of alternate analyses: in particular, (i) underlyingly consonant-initial suffixes (-/wək/, and -/jək/), accompanied by a process of consonant deletion, and (ii) three predictable allomorphs: -/wək/, -/jək/ and -/ək/. Under perfect learning, all three analyses make the same predictions, and cannot be distinguished. Under imperfect learning, however, the allomorph analysis can be differentiated from the other two. In fact, imperfect learning of individual word forms should not affect performance if a general rule of deletion or epenthesis has been learned. Whereas, by definition, imperfect learning of an allomorph analysis results in substitutions of one allomorph for another. For novel stems, of course, allomorph selection must occur via generalization. Imperfect learning is a failure to make the correct generalizations regarding the predictive environment for each allomorph and results in responses that are not fully consistent with either epenthesis or deletion.

Because learning behavior is likely to vary considerably across participants, degree of learning is assessed on a participant by participant basis. A participant was deemed to be "highly consistent" if they chose the same allomorph for a given stem type in at least 10 out of 12 instances. Table 4 gives the relative number of highly consistent participants by condition, and by allomorph. These numbers are for consonant-final stems only. In the 3VVC.j-w-0 condition only 50% of participants were consistent. By way of comparison, most participants in the 2VC.j-w condition were highly consistent, and highly consistent in selecting the correct allomorph for consonant-final stems (-/wək/). However, about 5% of participants were highly consistent in selecting the incorrect allomorph (-/jək/). In the 2VV.w-j condition high consistency was only reached for the correct allomorph (-/jək/), and there were fewer participants who showed high consistency, although more than in the 3VVC.j-w-0 condition.

The 3VVC.j-w-0 condition contains 3 unambiguous allomorphs, and the 2VC.j-w Condition contains 2 completely unambiguous allomorphs. The 2VV.w-j condition, however, contains one unambiguous allomorph (-/jək/), as well as an inflected stem type that is ambiguous with respect to the underlying form of its affix: -/wək/or -/ək/. As seen in previous conditions, participants overwhelmingly pick the phonetic interpretation of the ambiguous form, and thus the -/ək/allomorph. However, there are a small number of participants who select, in a small number of forms, the phonemic analysis (-/wək/)[5]. On average, this condition could be characterized as containing something more than 2 allomorphs, but not quite as many as 3. Under that interpretation the monotonic decrease in accuracy (measured as proportion of participants who are highly consistent) is explained as a function of the number of allomorphs: lowest accuracy in the 3VVC.j-w-0 Condition, intermediate accuracy in the 2VV.w-j condition, and highest accuracy in the 2VC.j-w condition[6].

If the additional data reinforced a predictable pattern then participants should have performed better, not worse, in conditions where they were exposed to more stem types. Instead, participants responded as though the additional allomorphs added variability, or noise, to the pattern. That is, they were not consistent about making a distinction between consonant-final stem types and vowel-final stem types, or between back-vowel final stem types, and front-vowel final stem types. The failure of participants to make a strong distinction between consonants and vowels is particularly problematic for an account of the emergence of synchronic epenthesis. Experiment 3 will look more closely at how non-uniformity, variability, and irregularity during training affect the degree to which participants generalize, or regularize,

[5]The phonemic response rate corresponded to a total of 11 -/wək/responses distributed over 5 participants.

[6]These differences are obscured by the continuous measure response data. Averaged differences in accuracy are not significant between the three conditions (logistic regression by condition, $p > .05$) although they exhibit considerable differences in variation.

their input.

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. This is the crux of Bickerton's "language bioprogram" hypothesis (1984) in which pidgins (collocations of elements from disparate languages) are transformed into creoles (fully productive independent languages) by subsequent generations of speakers (see also Singleton (1989), Senghas and Coppola (2001), 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 (Estes (1972), Gomez (2002)). Hale (1973) argues that under certain conditions learners adopt what he calls a conjugation analysis of a given linguistic pattern. That is, they memorize a set of allomorphs that correspond to a single meaning (e.g., plurality), along with the environments in which each allomorph occurs. However, high amounts of allomorphy, combined with lack of predictability (what he calls suppletion), can lead to regularization, as in the case of the Maori passive, where speakers appear to have replaced certain infrequently used forms with the commonly occurring -/tia/allomorph. 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)).

Table 5: Experiment 3

	Condition	N	Back Vowels	Front Vowels	Consonants	Token Ratio			Allomorph	% Boosted
8	2VV.t-j.F	16	ɹatutək	skibejək		12	6		-/tək/; -/ək/	43.8; 6.25
9	2VV.0-w.F	19	ɹatu-ək	skibewək		12	6		-/ək/; -/wək/	68.4; 5.26
10	2VV.j-0.F	22	ɹatujək	skibe-ək		12	6		-/jək/; /ək/	45.5; 18.2
11	3VV.0j-w0.F	21	ɹatu-ək	skibewək		9	4		-/ək/;-/wək/	33.3 ; 4.76
			vujək	fi-ək		3	2			
12	4VV.kpt͡ʃt-kt͡ʃt.F	19	ɹatutək	telʌpitək		1	1		-/tək/	5.26
			vut͡ʃək	fit͡ʃək		1	1			
			t͡ʃoɹænopək			2				
			zokək	skibekək		2	4			
13	2VVC.wt-jt-0.F	10	ɹatutək	fitək	sabolək	3	3	6	-/ək/; -/jək/; -/wək/;-/tək/	30;0;0;10[a]
			vuwək	hædijək		3	3			0;0;0;0[b]
										80;0;0;0[c]

[a]back-vowel final stems
[b]front-vowel final stems
[c]consonant-final stems

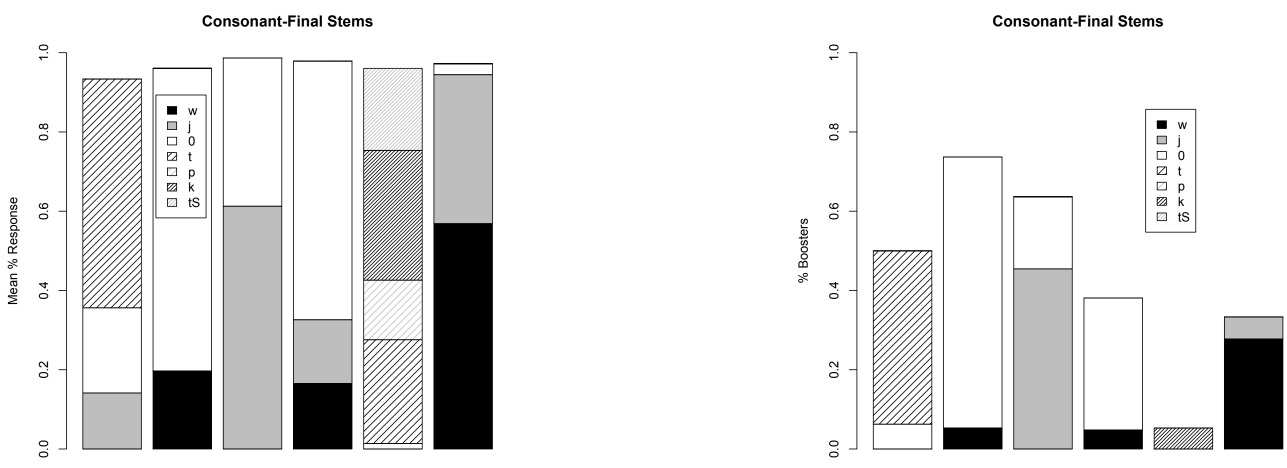
2.4.1 Frequency: Conditions 8-10

In this set of conditions no consonant-final stems are heard during training, forcing learners to generalize in some way from the vowel-final stems with which they are familiar. Additionally, the allomorphs and stem types are unequally distributed: back-vowel final stems, and back-vowel allomorphs, occur twice as often as front-vowel final stems (in both type and token counts). The last two columns of Table 5 show responses on consonant-final stems. Boosting is defined analogously with high consistency: selection of the same allomorph for at least 83% of test stems (usually 10 out of 12). The results show, as expected, that the most frequently occurring allomorph was generalized (or 'boosted') the most often in each condition.

However, minority allomorphs were also sometimes boosted. For comparison, the categorical and continuous response data are given side by side in Figure 5. The Anti-Natural condition with equal numbers of both vowel types and both allomorphs, and a non-overlapping distribution (see Table 3) is included to provide a baseline level. In all conditions the boosting levels are above baseline. In the 2VV.0-w.F and 2VV.j-0.F conditions the continuous measure corresponds closely to the categorical measure, but less so in the 2VV.t-j.F condition.

2.4.2 Variability: Conditions 11-13

This set of conditions exhibit variability in which allomorphs combine with which stem type. In the 3VV.0j-w0.F condition the -/ək/allomorph occurs with both stem types as well as occurring three times as often as the -/jək/ allomorph on back-vowel final stems, and half as often as the -/wək/ allomorph on front-vowel final stems. See Table 5. In the 4VV.kpt͡ʃt-kt͡ʃt.F condition back- and front- vowel final stems occur in equal numbers, but the four allomorphs are unequally, and overlappingly distributed. All four occur with back-vowel final stems, while only two occur with front-vowel final stems. The over-all allomorph ratio is 3:1:1:1, with -/kək/being the most frequent. In the 2VVC.wt-jt-0.F condition there are 6 tokens of each stem type, and all stem types are heard. Consonant-final stems are all consistently inflected with the bare -/ək/allomorph. But vowel-final stems are inflected with the -/tək/allomorph half the time, and with a homorganic glide half the time (because all stem-types are heard during training for this condition, the high consistency percentages are given for each stem type in the last column of Table 5). For the Variability conditions the categorical and continuous measures can be seen to diverge considerably. See Fig. 5. Only the most frequent allomorph boosts (at low levels) in the 4VV.kpt͡ʃt-kt͡ʃt.F, and only the two most frequent in the 3VV.0j-w0.F and 2VV.t-j.F conditions are boosted (i.e., there is no individual participant who ever boosts -/jək/), despite comparable continuous response rates.



(a) Continuous Measure

(b) Categorical Measure

Figure 5: Generalization to novel consonant-final stems

The 2VVC.wt-jt-0.F condition is meant to represent the outcome of a historical change in which t-final stems lost their /t/ word-finally, but retained it intervocalically. Stems ending in other consonants (l,m,ɹ) were unchanged. Vowel-final stems were also unaltered, and produced with natural coarticulation between the stem-final vowel and the suffix-initial vowel. This pattern is a potential precursor to synchronic epenthesis. What is required is that the -/tək/allomorph that occurs with originally t-final stems be generalized for use with all currently vowel-final stems (see Morley 2012). As expected, the consonant-final stems show high consistency, with only very low levels of -/tək/ and -/jək/[7] response. However, also consistent with previous results, consonant-final allomorphs were produced on vowel-final stems during testing, acting to disrupt the possible epenthesis pattern. The next section provides a quantitative analysis of transfer of allomorphs across different stem types for all conditions. This is followed by an analysis of individual participants in each condition who are possible originators of future epenthesis patterns.

2.5 Transfer Between Stem Types

Transfer from one trained stem type to another trained stem type indicates the permeability of the category boundaries, and the degree to which participants, on average, are consistent. In this section degree of transfer between vowel-final stem types and consonant-final stem types is compared across conditions 3VVC.j-w-0, 2VV.w-j, 2VC.j-w, and 2VVC.wt-jt-0.F. Degree of transfer between the two different vowel-final stem types is compared across conditions 3VV.0j-w0.F, 2VV.j-0.F, and 4VV.kpt͡ʃt-kt͡ʃt.F.

The following method is used to calculate comparable quantities under different experimental conditions. For each stem type s a class of 'other' allomorph (o) is defined as all suffix types heard on stems of the type $\neg s$ during training. Two comparison levels are then computed: the type-conditional probability ($p(o|s)$) is estimated by the proportion of s tokens heard during training with the o allomorph (usually 0) – this is the expected response level for impermeable category boundaries; the unconditioned probability ($p(o)$) is the proportion of *any* tokens heard during training with the o allomorph – this is the expected response level for completely permeable categories. The difference between observed response levels and each of the expected response levels is then determined for each condition, and each stem type pair. 'Transfer' is defined as the difference between expected and observed response rates, and is the measure that is compared across conditions. Table 6 gives the observed mean and standard deviation of the o allomorph response rate, along with the two expected response levels for the eight conditions discussed in this section.

Table 6: Percent o allomorph response rates. Trained stem types only.

	Condition	Type	o Allomorph(s)	Mean	SD	$p(o)$	$p(o\|s)$
5	3VVC.j-w-0	V to C	-/jək/, -/wək/	14.3%, 5.00%	16.4%	66.7%	0
6	2VV.w-j	Back V to C	-/ək/, -/wək/	20.8%	28.9%	50%	0
		C to Back V	-/jək/	23.8%	29.9%	50%	0
7	2VC.j-w	Back V to C	-/tək/, -/jək/	7.96%	18.8%	50%	0
		C to Back V	-/ək/,-/wək/	8.95%	21.8%	50%	0
8	3VV.0j-w0.F	Back V to Front V	-/ək/ ,-/jək/	62.5%	28.0%	77.8%	33.3%
9	2VV.t-j.F	Front V to Back V	-/ək/ ,-/jək/	3.63%	7.63%	33.3%	0
		Back V to Front V	-/tək/	12.0%	19.9%	66.7%	0
11	2VV.j-0.F	Front V to Back V	-/ək/,-/wək/	9.0%	14.6%	33.3%	0
12	4VV.kpt͡ʃt-kt͡ʃt.F	Back V to Front V	-/pək/	16.3%	14.9%	33.3%	0
13	2VVC.wt-jt-0.F	V to C	-/tək/	6.0%	13.5%	33.3%	0
		C to V[a]	-/ək/	62.8%	23.8%	66.7%	50%

[a]averaged across both vowel types

[7]These are even lower than the levels of phonemic analysis (as observed on consonant-final stems) in the 2VV.w-j condition. Only a single consonant-final token by a single speaker (~ 1% = a single token) was produced with a glide, compared to 11 tokens (~ 4% of responses = 11 tokens).

2.5.1 Results

Transfer levels from vowel-final stem types to consonant-final stem types: The transfer level of the *o* allomorph in the 2VVC.wt-jt-0.F condition (-/tək/) is not statistically distinguishable from that in the 2VC.j-w condition (-/jək/), and both are close to floor levels, while condition 2VV.w-j shows significantly more transfer than condition 2VC.j-w. The 3VVC.j-w-0 condition is significantly different from the 2VC.j-w condition, but not from the 2VV.w-j condition (logistic regression on *o* allomorph response rate by condition; $p < .05$). In terms of category permeability in the V to C direction, the conditions can be ranked in the following way: {2VV.w-j, 3VVC.j-w-0} > {2VC.j-w, 2VVC.wt-jt-0.F}. In all conditions transfer was significantly less than would be expected if there were no category distinction at all (expected level at $p(o)$). Even in the condition with the largest amount of transfer – 2VV.w-j – the response level was significantly different from the unconditioned probability level ($p < .05$: one-sample, one-sided Wilcoxon rank sum test that the mean was as high as 50%)[8].

Transfer from consonant-final to vowel-final stems: In all conditions the over-all levels of vowel-to-consonant transfer were found to be comparable to the levels of consonant-to-vowel transfer ($p > .05$: separate logistic regressions for each condition, by direction of transfer). However, a difference between directions came close to significance in the 2VVC.wt-jt-0.F condition ($p \sim .08$). This is due to the front-vowel final stems, which showed a significantly higher *o* response level than expected (Wilcoxon 1-sample Rank Sum Test; $p <.05$), and thus higher rates of transfer from consonant-to-vowel, than from vowel-to-consonant in this condition. The category permeability in the C to V direction shows the same rank ordering: {2VV.w-j, 3VVC.j-w-0} > {2VC.j-w, 2VVC.wt-jt-0.F}.

Transfer levels between vowel stem types: in the 3VV.0j-w0.F condition the response level for front-vowel final stems is significantly greater than the type-conditioned probability level (Wilcoxon 1-sample 1-sided rank sum test; $p < .05$ that the mean was as low as 1/3)[9] This is true in the 4VV.kpt͡ʃt-kt͡ʃt.F condition also (where the expected mean is 0). In the 2VV.t-j.F condition significantly higher transfer levels occur from the higher frequency back-vowel stems, to the lower frequency front-vowel stem types (logistic regression by direction of transfer; $p < .05$). However, low rates of the minority *o* allomorph still occur with the higher-frequency stems: $p < .05$ that the mean is equal to 0 in the 2VV.j-0.F condition.

As expected, as relative consistency levels increase, transfer levels decrease: 3VVC.j-w-0 to 2VV.w-j, to 2VC.j-w. More frequent allomorphs are transferred more often than less frequent ones (2VV.t-j.F), although minority allomorphs do transfer (2VV.j-0.F). Although there is almost always some degree of transfer across categories, listeners still maintain a partial distinction between stem types. No quantitative difference was found between the consonant-vowel category boundary and the front vowel–back vowel category boundary. Both appear equally permeable.

2.6 Biases

There are also indications that learners may have *a priori* 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 3VV.0j-w0.F condition as well (the -/jək/ allomorph is never boosted, although continuous measures show less of a bias). However, the bias appears to go in the opposite direction in the 3VVC.j-w-0 condition, with more than twice as much transfer of the -/jək/ over the -/wək/allomorph (Table 6). The difference between the 2VV.0-w.F and 2VV.j-0.F 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.

There may also be an asymmetry in how often the phonemic analysis is chosen based on the type of glide. Phonemic response levels were uniformly low for ambiguous tokens like [ɹatuwək]/[skibejək]. However, front-vowel contexts were somewhat more likely to lead to a phonemic analysis. As Figure 4

[8] R: wilcox.test().

[9] Back-vowel final stems may already be at ceiling.

shows, the phonemic analysis levels were very similar between the 2VV.w-j and the 2VV.t-j.F conditions. However, [ɹatuwək]-type items in the former occurred twice as often during training as did [skibejək] type items in the latter. The *proportion* of *o* allomorph responses that contained glides in the 2VV.w-j condition was 22%, versus 39.7% in the 2VV.t-j.F, a difference that was significant (logistic regression by condition; $p < .05$). If the effect of higher rates of phonemic response is to increase variability, then this difference may also account for the fact that there were so few boosters in the 2VV.t-j.F condition compared to the 2VV.0-w.F and 2VV.j-0.F conditions (Fig. 5). In fact, boosting levels for the 2VV.t-j.F condition are closer to those found in the 3VV.0j-w0.F condition which exhibited variability across three unambiguously distinct allomorphs. This argument can be extended to the results from the 2VVC.wt-jt-0.F condition. From Table 5 we see that 40% of participants were highly consistent on back-vowel final stems (30% with -/ək/, and 10% with -/tək/), but no participants were highly consistent on front-vowel final stems for any allomorph. If higher rates of phonemic analysis occurred among front-vowel final stems, leading to effectively greater variability, then consistency would be expected to be lower.

2.7 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 7 these 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.

To qualify as a pre-epenthesis outcome each allomorph must be used in a highly consistent manner (at least 10 out of 12 times with a given stem type), and with minimal transfer between vowel and consonant categories. This consistency must also result in the consonant-final stems being inflected mostly with the bare -ək allomorph, and the vowel-final stems, mostly with a consonant-initial allomorph. To meet these conditions, it is also necessary, in certain conditions, that the phonemic analysis be chosen for ambiguous forms. In the Natural, Bi-Modal, 2VV.w-j, and 2VV.t-j.F conditions, a pre-epenthesis outcome would require a phonemic analysis of the ambiguous [ɹatuwək]/[skibejək] type items. As can be seen from Fig 4, the rate of phonemic analysis is very low for these conditions. Not only that, but evidence for a phonemic analysis comes from exactly those consonant-final stems which are required to surface with the bare allomorph. Thus no pre-epenthesis patterns are found in these four conditions. In the Consistent-W condition, on the other hand, the phonemic 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 4VV.kpt͡ʃt-kt͡ʃt.F conditions the bare allomorph -/ək/is never heard during training, and consequently does not surface consistently with consonant-final stems at test[10].

Table 7: Proportion of Pre-Epenthesizers by condition. Gray cells indicate stem types held out during training.

Condition	Surface Pattern						Possible Epenthesis Rule	# participants/total participants
2VV.0-w.F	[skibe]	[skibewək]	[ɹatu]	[ɹatuwək]	[daɹum]	[daɹumək]	$\varnothing \rightarrow$w/V+__V	8/19
2VV.j-0.F	[skibe]	[skibejək]	[ɹatu]	[ɹatujək]	[daɹum]	[daɹumək]	$\varnothing \rightarrow$j/V+__V	4/22
3VVC.j-w-0	[skibe]	[skibewək]	[ɹatu]	[ɹatujək]	[daɹum]	[daɹumək]	$\varnothing \rightarrow \begin{bmatrix} \alpha front \\ +glide \end{bmatrix}$ /[$-\alpha front$]+__V	2/18

In the 2VV.j-0.F, 2VV.0-w.F, and 3VV.0j-w0.F conditions there is a subset of participants who boost the -/ək/ allomorph on consonant-final stems. In the 2VV.0-w.F condition, 8 of those participants, or about 42% of all 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

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

intervocalic /w/ epenthesis[11]. In the 2VV.j-0.F condition, only 4 of the consonant-final boosters, or about 18% of all participants, produced a pre-epenthesis pattern[12]. In the 3VV.0j-w0.F 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///-/jək/), or the back vowel stems (7 using -/ək/// -/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 conditions in which all three stem types are heard during training a pre-epenthesis pattern requires participants to consistently maintain the consonant-vowel boundary. In the 3VVC.j-w-0 condition, only two participants showed the required pattern. In the 2VVC.wt-jt-0.F condition 8 participants reached high consistency on consonant-final stems. However, none of these reached high consistency with -/tək/ on front-vowel final stems, and only one was highly consistent with -/tək/on back-vowel final stems.

3 Summary of Results

In the first set of experiments a strong phonetic bias emerged whereby listeners overwhelmingly analyzed ambiguous surface forms (e.g., [ɹatuwək]) as underlyingly /ɹatuək/, with an 'excrescent' glide, rather than /ɹatuwək/, with a phonemic glide. It was found that this bias could be completely reversed if 1) there was an identical unambiguously phonemic glide also heard during training, and 2) adopting the phonemic analysis eliminated allomorphy on the plural suffix. The presence of a different phonemic glide, or consonant, during training was not sufficient to produce a phonemic analysis for the ambiguous glide. The lack of phonemic analysis in the 2VV.0-w.F and 2VV.j-0.F conditions (see Table 5) is explained by the fact that the spliced items are not ambiguous between the presence versus absence of a homorganic glide; they are interpretable only as lacking a glide. The acoustics/phonetics only go in one direction, and morphological economy is not a strong enough force to lead to adoption of a single glide-initial allomorph in these two conditions[13].

Across the entire set of experiments participant behavior was much more consistent with a 'conjugation' analysis, than a rule-based analysis (Hale (1973)); listeners attempted to memorize which allomorphs went with which stems. In terms of consistency, learners were actually worse in the full pattern conditions than in the 'poverty of the stimulus' 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 consistently use 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). It is, perhaps, less surprising that learners would collapse the difference between the two vowel-final stem types; however, there is reason to believe that the distinction between consonants and vowels would be psychologically robust[14]. Yet transfer of

[11]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 phonemic analysis.

[12]The lower number is expected given the fact that the -/ək/allomorph is twice as frequent (as -/wək/) in the 2VV.0-w.F condition, whereas it is half as frequent (as -/jək/) in the 2VV.j-0.F condition.

[13]This account predicts that for a condition in which the second vowel-final stem types were actually ambiguous (e.g., [ɹatuwək] [skibewək]) participants should overwhelmingly interpret the pattern to result from a single glide-initial suffix. Unfortunately, this was not an originally planned condition and was not run.

[14]It is expected that English speaking participants would have explicit awareness of which sounds are classified as consonants and which as vowels. There is also evidence that speech errors 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)).

allomorphs between vowel-final and consonant-final stems was also observed, with no quantifiable difference found in the permeability of the different categories (see Goldrick (2004) and Kapatsinski (2010) for similar cross-category generalization). In most cases, however, observed response levels were significantly different from what would be predicted by a single category (i.e. front/back distinctions completely collapsed, or consonant/vowel distinctions completely collapsed), demonstrating that participants did partially retain type-conditioned statistics[15]. Larger numbers of allomorphs as well as higher frequencies of occurrence increased rates of transfer generally, and decreased the number of high-consistency learners.

Generalization, or "boosting", was also seen to depend on variability and token distribution, paralleling the consistency results. The 3VV.0j-w0.F condition shows boosting levels comparable to baseline; and the 4VV.kpt͡ʃt-kt͡ʃt.F condition, well below. Contrary to the memory load hypothesis, more variants and less predictability led to less generalization beyond the input. However, learners showed a predeliction 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), and boosted more on never before heard stem types than ones they had been trained on.

As expected, the most frequent allomorph during training is the one most often boosted, although less frequent allomorphs can also be boosted. Not all allomorphs, however, have a probability of boosting proportional to their average response rate. In the 2VV.t-j.F condition for example, the response rate for the -/jək/allomorph is roughly 14%, but this allomorph is never boosted. A similar thing occurs in the 3VV.0j-w0.F Condition. It is most striking in the 4VV.kpt͡ʃt-kt͡ʃt.F condition where there are several allomorphs with relatively high response rates, but only the -/kək/allomorph is boosted. Participants seem only to boost the two most frequent allomorphs, with additional minority allomorphs acting as a sort of persistent irregularity within a majority response pattern.

The results also revealed some unanticipated asymmetries in response patterns. The balance of evidence supports a bias against the -/jək/ allomorph when competing with the -/wək/ allomorph. Overall counts by word-form in CELEX2 (Baayen et al., 1995) are fairly close for the two glides, although /w/ is more frequent (7917 to 7355). Intervocalically, neither glide is particularly common, but /j/ is quite rare (21 unique words to 504 containing /w/'s), and the words in which it does occur tend to be infrequent themselves. This asymmetry, therefore, could derive from the context-conditioned relative frequency of the two glides in English. There was also evidence to support the conclusion that the phonemic analysis was more likely for words like [skibejək]than for words like [ɹatuwək]. Both acoustic and articulatory differences are found between the front and back glides. /j/ is typically analyzed as consisting of a single palatal gesture, but /w/ as containing both a labial and a velar gesture (e.g., Gick (2003)). Chitoran (2002) finds acoustic differences in the two types of glides in Romanian that she argues may be universal: the low F2 (approaching F1) in /w/ allowing little room to make distinctions among sounds that are both back and round, while the /j/ is more acoustically distinct than its closest vowel correspondents. This latter result could be responsible for the reported asymmetry. Because the speaker used to record the stimuli was consciously trying to produce full glides, the resulting tokens containing [j]'s may have been easier to identify as such, than the tokens containing [w]'s.

Finally, it appears that certain individuals are more likely than others to adopt a phonemic interpretation of an ambiguous token. Another individual difference emerges in the likelihood 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

[15]Note that this effect cannot be attributed to memorization of particular words heard during training since all the test items are novel words. However, it is possible that overall word similarity is driving this result, as opposed to a featural, or even segmental analysis.

this type of data may provide a path to phonologization (cf. Kirby (2001)).

4 Discussion & Conclusions

The finding that generalization depends on token frequency is consistent with expectation, as well as the hypothesis that listeners may fall back on frequently heard items when they fail to remember specific forms (e.g., Hudson Kam and Newport (2005), Hale (1973)). However, the story is somewhat complicated by the fact that the more distinct allomorphs over-all are heard in training (and thus, the more variability), the less learners generalize, exhibiting closer frequency matching both in trained categories and novel categories. There is a confound, however, in that in the highest-variability conditions the most frequent allomorph occurs less frequently, both absolutely and relatively, than the most frequent allomorph in the lower-variability conditions[16]. Thus, there is less reason to prefer one allomorph over another, or to 'boost' any given allomorph. On the other hand, an underlying preference for a single-allomorph analysis is observable in conditions in which there are two allomorphs heard during training, but one stem type is held out. In such scenarios, from about 35% to 75% of participants regularize on the novel stem type (depending on the specific allomorphs, and their relative frequencies).

There is a striking uniformity among participants in these experiments in their interpretation of surface glides as phonetic rather than phonological, and an equally striking reversal of that preference when morpheme economy plays a role. From these data alone we cannot determine whether the phonetic bias arises from experience with the native language of participants (English), or whether this is a more general result based on the way in which vowel gestures are coordinated in the absence of conflicting grammatical pressures. The 'morpheme economy' bias, on the other hand, seems likely to be a universal pressure[17], based in general learning strategies, and possibly a manifestation of a preference for paradigm uniformity (e.g., Steriade (2000), Kenstowicz (1997)).

In contrast, phonology-based biases are weak or non-existent in these experiments. The results suggest that phonological learning may be qualitatively different from phonetic and morphological learning, and may even be due to a more explicit type of learner generalization, and thus more susceptible to individual differences in learning strategies. The necessity of extensive generalization and innovation, on top of the requirements for non-typical types of learners and restrictive historic conditions leads to the prediction that true synchronic epenthesis may be quite rare. To test this prediction what is needed is a sample of languages selected so as to provide a good approximation for the actual rate of occurrence of epenthesis cross-linguistically. Additionally, the 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. According to at least one account, however, clear examples of synchronic epenthesis are considerably rarer than typically assumed (Morley (2015)). Thus, the failure to learn a rule of epenthesis in the experiments reported here may well be in line with a low observed rate of default epenthesis. As exceptional, rather

[16] This is a direct consequence of fixing the total number of unique tokens (either 12 or 18 in each condition, selected in order to have round numbers of each variant). Thus, an allomorph that is twice as frequent as a single competitor will occur on 12 unique stems, to the 6 that the lower-frequency allomorph appears on. While in a condition in which there are 3 allomorphs, the highest frequency allomorph that occurs twice as often as the rest will occur only 6 times, to the 6 times that the non-majority allomorphs occur (3 tokens of each). Therefore, the high-frequency allomorph may be less salient to learners.

[17] One might expect to see different realizations of this pressure in languages with non-concatenative morphology, or agglutinative languages with extensive morphology.

than commonplace, individual cases of synchronic epenthesis thus bear considerable scrutiny for insight into the exact conditions of their emergence.

4.1 Intrusive R

English intrusive r represents a particularly well-known case of what is generally taken to be 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])[18]. However, these so-called non-rhotic varieties maintained r in inter-vocalic environments ("linking r"), resulting in alternations of the following kind: "bar below" [bɑbəloʊ]; "bar above" [bɑɹəbʌv]. A further subset of these dialects also developed what is known as "intrusive r": r's that subsequently arose 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 (a Derby-area dialect, as reported by Foulkes (1997); Conservative RP, as reported by Durand (1997)), and of non-rhotic dialects which have neither, having lost r in all environments (some varieties of Southern U.S. English, and possibly South African English, as reported by Wells (1982) and Harris (1994)). Intrusive r alone (as opposed to intrusive and linking r together) is typically analyzed as 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 usual 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. Modern Southern English and General RP (as reported by Durand (1997), Johansson (1973)), West Yorkshire English (as reported by Broadbent (1991)), Newcastle English (as reported by Foulkes (1997)), Boston English (as reported by McCarthy (1993)), and New Zealand English (as reported by Hay and Sudbury (2005)) 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 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 stigma associated with the phenomenon; speakers attempt to correct for intrusive r, but fail to differentiate between linking (that is, historic) r, and intrusive r, leading them to sometimes drop r from the wrong contexts. Johansson takes this as evidence for r-less underlying representations, and proposes a single epenthesis rule to account both for linking and intrusive r.

The diachronic account requires a stage in which speakers had both a rule of deletion (involving linking r's) and a rule of epenthesis (intrusive r's). The epenthesis rule gradually generalizes, and in doing so increases the ambiguity around whether a given surface instantiation of the r/Ø alternation is due to deletion or epenthesis. Once speakers start to lose track of which words were historically r-final, rule inversion, in which the epenthesis rule takes over, is assumed to occur. The paradox in this account is that true rule inversion requires changes to the underlying representations of historically r-final words. Thus, the argument from orthography *against* the deletion analysis for novel r's ("there is obviously no r in comma" (Foulkes (1997)), must be abandoned to allow /kɑɹ/>/kɑ/ despite the fact that there *is* obviously

[18] 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.

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 why such generalization was possible 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 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 experiments reported in this paper fail to find consistent conditioning by vowel context in learning may be attributable to the fact that the association was not based on phonetic similarity – and in some cases was anti-similar. Both the experimental results and the case of English intrusive r suggest a primary role for phonetics in the emergence of epenthesis, and epenthesis-like alternations, a primacy that may act to limit the extent to which such patterns can be generalized.

4.2 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. Because there exists no clear diagnostic for determining the best analysis of a given set of linguistic data (given the existence of exceptions and unknown performance factors), this question cannot be satisfactorily answered. Recent experimental work addressing the question of productivity shows a range of results; some patterns productive, others apparently unproductive, with the distinction not always hinging on what is hypothesized to be phonetically or phonologically natural (e.g., Hayes et al.

(2009), Becker et al. (2011), Staroverov (2015), Zuraw (2007)). This suggests that productivity will have to be determined on a case by case basis, and may depend on details of distribution and frequency.

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. The three part methodology (typological/synchronic (Morley (2015)), diachronic (Morley (2012)), and now experimental) used to investigate the phenomenon of epenthesis is an attempt to encompass the true scope of the question about phonological, and by extension, linguistic, universals. There is little evidence that phonological (versus phonetic) epenthesis is an accessible hypothesis to a learner engaged in morphological learning. And the conditions under which synchronic epenthesis could arise seem to be very constrained. Additionally, a general mechanism of rule inversion is unlikely to supply a simple and direct path to a categorical phonological grammar. The conclusion is that we must alter some of our assumptions, about both the nature of synchronic grammars, and the trajectory of historic change. And it is likely that we will need to come up with new mechanisms in order to link the two within a coherent theory of phonological competence. The impetus for considering both diachronic and synchronic aspects of the same phonological phenomenon lies in the degree to which results from either domain can be grounded in results from the other, allowing progress to be made on questions that appear unanswerable when each domain is considered in isolation.

References

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

Baayen, R., R. Piepenbrock, and L . Gulikers. 1995. Celex2 ldc96l14. Philadelphia: Linguistic Data Consortium. URL `https://catalog.ldc.upenn.edu/LDC96L14`.

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

Becker, Michael, Nihan Ketrez, and Andrew Nevins. 2011. The surfeit of the stimulus: Analytic biases filter lexical statistics in turkish laryngeal alternations. *Language* 87:84–125.

Beddor, P.S. 2009. A coarticulatory path to sound change. *Language* 85:785–821.

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.

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.

Chitoran, Ioana. 2002. A perception-production study of romanian diphthongs and glide-vowel sequences. *Journal of the International Phonetic Association* 32:203–222.

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. 2012. Typological universals as reflections of biased learning: Evidence from artificial language learning. *Language and Linguistics Compass* 6:310–329.

Culbertson, Jennifer, and Géraldine Legendre. 2010. Investigating the evolution of agreement systems using an artificial language learning paradigm. In *Proceedings of the 39th Western Conference on Linguistics*, ed. Dina Bailey and Victoria Teliga, 46–58. Department of Lingusitics, California State University, Fresno.

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.

Finley, Sara, and William Badecker. 2008. Analytic biases for vowel harmony languages. In *WCCFL*, volume 27, 168–176.

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.

Fujimura, O., M.F. Macchi, and L.A. Streeter. 1976. Perception of stop consonants with conflicting transitional cues: a cross-linguistic study. *Language and Speech* 21:337–346.

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

Gick, Bryan. 2003. Articulatory correlates of ambisyllabicity in english glides and liquids. In *Papers in laboratory phonology vi: Constraints on phonetic interpretation.*, ed. J. Local, R. Ogden, and R. Temple, 222–236. Cambridge University Press.

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.

Gomez, R.L. 2002. Variability and detection of invariant structure. *Psychological Science* 13:431–436.

Guy, Gregory R. 2008. Variationist approaches to phonological change. In *The handbook of historical linguistics*, ed. Brian D. Joseph and Richard D. Janda, 369–400. Oxford, UK: Blackwell Publishing Ltd.

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

Harrington, Jonathan, Felicitas Kleber, and Ulrich Reubold. 2008. Compensation for coarticulation, /u/-fronting, and sound change in standard southern british: An acoustic and perceptual study. *The Journal of the Acoustical Society of America* 123:2825–2835.

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

Hay, Jennifer, and Andrea Sudbury. 2005. How rhoticity became /r/-sandhi. *Language* 799–823.

Hayes, Bruce, Péter Siptár, Kie Zuraw, and Zsuzsa Londe. 2009. Natural and unnatural constraints in hungarian vowel harmony. *Language* 85:822–863.

Hombert, Jean-Marie, John J Ohala, and William G Ewan. 1979. Phonetic explanations for the development of tones. *Language* 37–58.

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.

Hura, S.L., B. Lindblom, and R.L. Diehl. 1992. On the role of perception in shaping phonological assimilation rules. *Language and Speech* 35:59–72.

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.

Kawasaki, Haruko. 1986. Phonetic explanation for phonological universals: The case of distinctive vowel nasalization. *Experimental Phonology* 81–103.

Kenstowicz, M. 1997. Base-identity and uniform exponence: Alternatives to cyclicity. In *Current trends in phonology: Models and methods*, ed. J. Durand and B. Laks, 363–394. CNRS and University of Salford Publications.

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

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* .

Ohala, J. 1981. The listener as a source of sound change. In *Parasession on Language and Behavior: Chicago Linguistics Society*, ed. C.S. Masek, R.A. Hendrick, and M.F. Miller, 178–203. Chicago.

Ohala, J. 1993. The phonetics of sound change. In *Historical linguistics: Problems and perspectives*, ed. C. Jones, 237–278. New York, NY: Longman.

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In *The production of speech*, ed. P.F. MacNeilage, 189–216. Springer.

Pallier, Christophe, Laura Bosch, and Núria Sebastián-Gallés. 1997. A limit on behavioral plasticity in speech perception. *Cognition* 64:B9–B17.

Pearl, L.S. 2011. When unbiased probabilistic learning is not enough: Acquiring a parametric system of metrical phonology. *Language Acquisition* 18:87–120. URL `http://www.tandfonline.com/doi/abs/10.1080/10489223.2011.554261`.

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.

Repp, B.H. 1977. Dichotic competition of speech sounds: the role of acoustic stimulus structure. *Journal of Experimental Psychology: Human Perception and Performance* 3:37–50.

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.

Senghas, A., and M. Coppola. 2001. The creation of nicaraguan sign language by children: Language genesis as language acquisition. *Psychological Science* 12:323–328.

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

Singleton, J. L. 1989. Restructuring of language from impoverished input: Evidence for linguistic compensation. Doctoral Dissertation, University of Illinois at Urbana-Champaign.

Solé, Maria-Josep. 2007. Controlled and mechanical properties in speech. *Experimental Approaches to Phonology* 302–321.

Staroverov, Peter. 2015. Buriat dorsal epenthesis is not reproduced with novel morphemes. URL `http://home.uni-leipzig.de/ staroverov/papers/Staroverov`$_1$5 $-$ *productivity* $-$ *buriat* $-$ *dorsal* $-$ *zero.pdf*.

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

Steriade, D. 2000. Paradigm uniformity and the phonetics/phonology boundary. In *Papers in laboratory phonology v: Acquisition and the lexicon*, ed. M. Broe and J. Pierrehumbert, 313–332. Cambridge University Press.

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

Takagi, Naoyuki, and Virginia Mann. 1995. The limits of extended naturalistic exposure on the perceptual mastery of english /r/ and /l/ by adult japanese learners of english. *Applied Psycholinguistics* 16:379–405.

Tenenbaum, J.B., and T.L. Griffiths. 2001. Generalization, similarity and bayesian inference. *Behavioral and Brain Sciences* 24:629–641.

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

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

Weinreich, U., W. Labov, and M. Herzog. 1968. Empirical foundations for a theory of language change. In *Directions for Historical Linguistics*, ed. W. Lehmann and Y. Malkiel, 95–188. Austin: U. of Texas Press.

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.

Zuraw, Kie. 2007. The role of phonetic knowledge in phonological patterning: corpus and survey evidence from tagalog infixation. *Language* 277–316.

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 2 repeated. Full list of stimuli roots for all conditions, both training and test items

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

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

Table A.2: Full List of Conditions

	Condition	N	Train	Test	Back Vowels	Front Vowels	Consonants	Token Counts			Allomorph Ratio
1	Natural	18	i & ii	iii, iv & values	ɹatuwək	skibejək		6	6		1:1[a]
2	Anti-Natural	18	i & ii	iii, iv & v	ɹatujək	skibewək		6	6		1:1
3	Bi-Modal	20	i & ii	iii, iv & v	ɹatuwək	skibejək		3	3		2:1:1[a]
					vu-ək	fi-ək		3	3		
4	Consistent-W	20	i & v	iii, iv & vii	ɹatuwək		daɹumwək	6		6	1
5	3VVC.j-w-0	18	i, ii & v	iii, iv & vii	ɹatujək	skibewək	daɹumək	6	6	6	1:1:1
6	2VV.w-j	24	i & v	iii, iv & vii	ɹatuwək		daɹumjək	6		6	1:1
7	2VC.j-w	19	i & v	iii, iv & vii	ɹatujək		daɹumwək	6		6	1:1
8	3VV.0j-w0.F	21	i, ii, & vi	iii, iv & v	ɹatu-ək	skibewək		9	4		11:4:3
					vujək	fi-ək		3	2		
9	2VV.t-j.F	16	i, vi & ii	iii, iv & v	ɹatutək	skibejək		12	6		2:1
10	2VV.0-w.F	19	i, ii, & vi	iii, iv & variable	ɹatu-ək	skibewək		12	6		2:1
11	2VV.j-0.F	22	i, ii, & vi	iii, iv & variability	ɹatujək	skibe-ək		12	6		2:1
12	4VV.kpt͡ʃt-kt͡ʃt.F	19	i & ii	iii, iv & v	ɹatutək	telʌpitək		1	1		3:1:1:1
					vut͡ʃək	fit͡ʃək		1	1		
					t͡ʃoɹænopək			2			
					zokək	skibekək		2	4		
13[b]	2VVC.wt-jt-0.F	10	i, ii & vii (subset)	iii, iv & v	ɹatutək	fitək	sabolək	3	3	6	2:2:1:1[a]
					vuwək	hædijək		3	3		

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