TEACHING PAPA TO CHA-CHA: HOW CHANGE MAGNITUDE, TEMPORAL

CONTIGUITY, AND TASK AFFECT ALTERNATION LEARNING

by

AMY ELIZABETH SMOLEK

A DISSERTATION

DISSERTATION APPROVAL PAGE

Student: Amy Elizabeth Smolek

Title: Teaching Papa to Cha-Cha: How Change Magnitude, Temporal Contiguity, and Task Affect Alternation Learning

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Linguistics by:

Vsevolod Kapatsinski   Chairperson
Melissa Baese-Berk   Core Member
Eric Pederson   Core Member
Kaori Idemaru   Institutional Representative

and

Kate Mondloch   Interim Vice Provost and Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded December 2019.

DISSERTATION ABSTRACT

Amy Elizabeth Smolek

Doctor of Philosophy

Department of Linguistics

December 2019

Title: Teaching Papa to Cha-Cha: How Change Magnitude, Temporal Contiguity, and Task Affect Alternation Learning

In this dissertation, we investigate how speakers produce wordforms they may not have heard before. Paradigm Uniformity (PU) is the cross-linguistic bias against stem changes, particularly large changes. We propose the Perseveration Hypothesis: Motor perseveration in the production system encourages copying from related wordforms. When this conflicts with paradigmatic associations requiring a change to the base, the change may be leveled, resulting in PU. Associations are more difficult to acquire when the forms are articulatorily dissimilar, and poorly-learned associations are a lesser obstacle to the perseveratory bias, which accounts for the stronger bias against large changes.

Participants trained on a miniature artificial language with labial palatalization (p→tʃi), a large change, produce the alternation much less often than participants trained on alveolar (t→tʃi) or velar (k→tʃi) palatalization. The difficulty arises from articulatory, rather than perceptual, dissimilarity: k→tʃi and g→dʒi are learned equally well despite differing in perceptual similarity, and the bias against large changes is observed in production but not in judgment. Ratings of labial palatalization improve as much post-training as do ratings of lingual palatalization, suggesting that participants learn what

they should produce by acquiring product-oriented schemas, but are unable to acquire a paradigmatic labial-to-alveopalatal association necessary for producing the alternation.

How, then, do speakers learn to produce large changes? We propose that temporal contiguity between related forms allows speakers to notice the relationship between forms, strengthening paradigmatic associations between the chunks by which the forms differ and syntagmatic associations within these chunks. Presenting a plural immediately after the corresponding singular in training leads to more production of the exemplified pattern, whether the mapping is faithful (e.g. p→pa) or unfaithful (e.g. k→tʃa). If only one type of mapping is shown in contiguity, the pattern spreads to all inputs. Only when both types of mappings are shown in contiguity do participants learn to match inputs to the correct outputs. A simple two-layer discriminative model captures the results of the trial order manipulations, including cue availability and "chunking."

In sum, our work shows that paradigmatic associations are acquired through syntagmatic correspondence, which enables even large changes to be produced.

CURRICULUM VITAE

NAME OF AUTHOR: Amy Elizabeth Smolek

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR
Swarthmore College, Swarthmore, PA

DEGREES AWARDED:

Doctor of Philosophy, Linguistics, 2019, University of Oregon
Bachelor of Arts, Linguistics, 2011, Swarthmore College

AREAS OF SPECIAL INTEREST:

Morphophonology
Learning

PROFESSIONAL EXPERIENCE:

Graduate Employee (teaching assistant), Department of Linguistics, University of
Oregon, Eugene, 2011-2012, 2017

Graduate Employee (instructor), American English Institute, University of
Oregon, Eugene, 2012-2016

GRANTS, AWARDS, AND HONORS:

Graduate Teaching Fellowship, Linguistics, 2011 to 2017

Phi Beta Kappa, 2011

PUBLICATIONS:

Smolek, A. & Kapatsinski, V. (2019). *Syntagmatic paradigms: Learning correspondence
from contiguity.* Manuscript submitted for publication.

Smolek, A. & Kapatsinski, V. (2018). What happens to large changes? Saltation produces well-liked outputs that are hard to generate. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 9*(1), 10.

ACKNOWLEDGMENTS

To my mother, who made the cookies;
my father, who paid for the cookies;
and my brother, who let me have all the cookies:
I couldn't have done this without you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Portions of this chapter were taken from:

Smolek, A. & Kapatsinski, V. (2019). *Syntagmatic paradigms: Learning correspondence from contiguity*. Manuscript submitted for publication.

Smolek, A. & Kapatsinski, V. (2018). What happens to large changes? Saltation produces well-liked outputs that are hard to generate. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *9*(1), 10.

The acquisition of morphology includes the process of acquiring paradigm mappings, or associations between related forms (Kapatsinski, 2018b; J. P. Blevins, 2013). Everyday language use requires speakers to solve the Paradigm Cell Filling Problem (Ackerman et al., 2009): They need to be able to produce forms they may not have heard before (J. P. Blevins et al., 2017; Bonami & Beniamine, 2016; Hockett, 1967; Malouf, 2017). Diverse languages show that it is impossible for adult speakers to be exposed to every form of a known word; only 0.1% of Czech nouns are present in every inflected form in a 100 million word corpus (Malouf, 2017), and increasing the size of the corpus does not improve and may even exacerbate the problem (J. P. Blevins et al., 2017). The result is that speakers must generate novel forms of known words, and how that may happen is the topic of this dissertation.

Paradigm Uniformity (PU) is the leveling of changes across paradigmatic cells, in order to regularize the related forms of a word. The cause of PU is subject to discussion, but we propose the Perseveration Hypothesis: that weak associations are insufficient to override the perseveratory tendency in production. In order to prevent PU, there needs to

be some strategy for strengthening associations. We propose that contiguity is a vital contributor to association strength[1], and that without contiguity in language learning, there would be much more leveling of paradigms.

1.1. Paradigm Uniformity

Paradigm Uniformity is the force that militates against multiple forms of the same stem (Benua, 1997; Kenstowicz, 1996; Steriade, 2000). Especially undesirable are phonologically dissimilar allomorphs of the stem (Skoruppa et al., 2011; White, 2014). Learning of phonological patterns is biased against dissimilar sound alternations (Hayes & White, 2015; Moreton & Pater, 2012a; Peperkamp et al., 2006; Steriade, 2001/2009; White, 2013, 2014, 2017). Dissimilarity can be operationalized as an alternation "skipping over" another sound, also called saltation (Peperkamp et al., 2006; Skoruppa et al., 2011; White, 2013, 2014), or defined in terms of phonological features, articulatory gestures, or perceptual dimensions.

1.1.1. Vowel height in Canadian Raising

PU can be observed when productive phonological processes seem to fail to apply when their application would violate PU, or when they seem to overapply in order to increase PU. For example, in Canadian Raising (Joos, 1942), /aɪ/ raises to [ʌɪ] before voiceless consonants, hence [bʌɪt] but [baɪd]. However, the alternation overapplies before voiced flaps that correspond to voiceless stops in other forms of the same stem, and as a result, the paradigm is more uniform: All forms share the same vowel height, as shown in (1). The overapplication of the alternation is a form of perseveration, where an element of the base form (e.g. [ʌɪ]) is retained in the production of a related form.

---

[1] Which may be implemented via variation sets (Küntay & Slobin, 1996; Onnis et al., 2008), and in addition to consolidation (Davis & Gaskell, 2009; Kumaran et al., 2016; Lewis & Durrant, 2011; McClelland et al., 1995; §6.4.3).

1. bʌɪt   'bite'
   baɪd   'bide'
   bʌɪɾɪŋ 'biting'
   baɪɾɪŋ 'biding'

1.1.2. Saltation

There is a strong typological tendency against saltation, which is a type of large change. If a language contains X, Y, and Z sounds, such that Y is between X and Z in phonetic similarity space, then X→Z implies Y→Z, but not the opposite (shown experimentally by White, 2013, 2014; White & Sundara, 2014). For example, in White (2013, 2014), English-speaking adults exposed to p→v prefer f→v over f→f, but those exposed to b→v prefer f→f. The extension of an alternation to intermediate sounds makes sense synchronically, but less so diachronically; the rarity of saltation is implied to be due to overgeneralization of alternations to intermediate sounds, but Bybee (2008) argues that alternations lose productivity as they require greater degree of change (i.e. they become more likely to be replaced with faithful mappings). White (2013) discusses Crosswhite (2000) for diachronic vowel saltation in Russian, who shows that it is in the process of losing productivity, rather than being overextended. Most errors by children are caused by underapplying stem changes rather than overgeneralizing alternations (Do, 2013; Kerkhoff, 2007; Krajewski et al., 2011; Tomas et al., 2017), the latter of which is rare in adults as well (Bolognesi, 1998; White, 2017). Since error can seed language change (Andersen, 1973; Bybee, 2010; Bybee & Slobin, 1982; Harmon & Kapatsinski, 2017; Hudson Kam & Newport, 2009), saltation is likely to disappear through underapplication rather than extend through overapplication.

1.1.3. Explaining paradigm uniformity

Current accounts of PU are couched within the framework of Optimality Theory (Prince & Smolensky, 1993/2004; Benua, 1997; Burzio, 1996; Kenstowicz, 1996; McCarthy, 1998). Pre-OT theorists noted the importance of PU (Pinker & Prince, 1988), but claimed that its origin was morphological, not phonological. If Word = Stem + Affix, then the stem must be preserved as a consequence of the existence of morphology. However, not all aspects of the stem are likely to be retained in the output. In the Canadian Raising example in (1), the height of the vowel is retained, but the /t/ is not. Phonological theories of PU, as captured through OT, capture this asymmetry by describing retention of the vowel and retention of the consonant by distinct constraints, which can have differing weights.

Base identity constraints (Kenstowicz, 1996), also called output-output (OO) faithfulness constraints (Benua, 1997; Hayes, 2004; Kager, 1999; McCarthy, 1998), favor preserving particular features of the derivational/inflectional base. For the Canadian Raising example in (1), a high ranking of the Ident_{BA}-[low] constraint would force *biting* to retain the raised [ʌɪ] of *bite*. Base identity constraints are posited to be universal and at the top of the constraint hierarchy unless demoted by learning (Hayes, 2004; McCarthy, 1998). The initial high ranking is consistent with observations that children often level stem changes that adults in the same community reliably produce; Kerkhoff (2007) shows that the Dutch voicing alternation is less productive for Dutch children than adults, despite its high degree of regularity and phonetic naturalness, and Do (2013) shows that Korean children avoid producing stem changes using a variety of repair strategies.

Claiming that base identity constraints are innate accounts for their universality, but does not explain where they came from in the first place. What adaptive advantage do they give? The other major kinds of constraints, markedness, faithfulness, and alignment, are motivated by ease of articulation, the need to maintain lexical contrast, and the need for temporal alignment of independent structures, respectively (Kager, 1999; Prince & Smolensky, 1993/2004). PU is an independent force that shapes phonological patterns, and is separate from markedness of the output and from faithfulness to an underlying form, because some patterns can only be explained by PU (Benua, 1997; Kenstowicz, 1996), which motivates their presence in the phonological grammar. There is no universally accepted origin for PU constraints, with alternatives proposed by Kenstowicz (1996, 1998, §1.6.2), Steriade (2008) and White (2017, §1.6.3), and Moreton & Pater (2012a, §1.6.4).

We propose the Perseveration Hypothesis, summarized below in (2).

2. **The Perseveration Hypothesis:** The tendency to avoid stem changes is grounded in motor perseveration during the process of generating a novel form of a known word. When the form cannot be retrieved from memory quickly enough, activation cascades over related forms, and their activated motor representations are incorporated into the production plan under construction.

We accept the existence of a PU bias, but do not think it originates from an innate universal grammar or constraint inventory. Instead, it is grounded in two sources:

1) Paradigmatic motor perseveration (§1.2), and

2) The difficulty of learning associations between dissimilar alternants (§1.3).

It is the competing tendencies between paradigmatic perseveration and paradigmatic associations that lead to the patterns in paradigm uniformity that we find. In §1.2, we discuss prior work on perseveration. In §1.3 and §1.4, we explore the learning of

associations. §1.5 describes our view of the process that speakers undergo when attempting to construct a novel form of a known word, and §1.6 compares our theory to other work. In §1.7, we lay out the structure of the dissertation.

1.2. Perseveration

Perseveration on the base is usually beneficial. Even though we most often notice it when it is a mistake, most to-be-produced forms retain at least most of aspects of the base, and full suppletion is rare. Otherwise, there could be no "morphology" (Pinker & Prince, 1988). Perseveration on aspects of recently-produced forms that do not share a stem with the target form (in other words, mistaken perseveration) has been noted in elicited production tests and natural language. Bickel et al. (2007) show that speakers of Chintang produce prefix orderings that mirror those of a previous utterance when the following utterance is syntactically similar; Caballero (2010) demonstrates that certain morphemes are produced in non-scopal order if they had just been produced in the same order in a context where the order was scopal; and Lobben (1991) shows that Hausa plurals produced in seeming violation of the rules of the language can be explained by referencing the form produced immediately beforehand, which is usually formally similar to the anomalous form.

While perseveration may be encouraged in elicited production, the same process still applies any time a speaker produces a derivative of a known word, whereby known forms of the word are activated, providing the speaker with gestures on which to perseverate. Perseveration on morphological elements is always in the background, racing against lexical retrieval (Baayen et al., 1997), but it is usually outpaced; thus, paradigmatic perseveration is most obvious when a novel form of a known word is to be produced.

6

1.2.1. Where perseveration is visible

Paradigm Uniformity is visible when productive phonological processes fail to apply when they should (because it would violate PU), or apply when they should not (to increase PU; see Benua, 1997, for a review, and Raffelsiefen, 2005, for a critical perspective). One avenue for exploration of these processes is through "wug tests" (Berko, 1958), where participants are given a form of a word and asked to produce another form using the given as a base. In the original study, participants were shown a picture of an unfamiliar creature and told, "This is a wug." They were then shown a picture with multiple creatures and prompted with, "Now there are two of them, there are two…". For productive morphemes like the English plural -*s*, young children are able to generate the correct plural *wugs* and often able to select the correct allomorph of /z/ (*wug*[z] from *wug, fep*[s] from *fep*, and *gutch*[əz] from *gutch*).

The aim of the study of morphology is to explain real language, not the elicited production task, but we think the latter is a good approximation of the former. The primary difference is that in the lab, a single form is activated, whereas in natural language multiple forms can be activated and compete simultaneously[2] (cf. Albright, 2008), which can lead to multiple inheritance/multiple motivation, where the output retains aspects of multiple related forms (Goldberg, 1995; Umbreit, 2011).

Novel wordforms are the result of a blending process, whereby the production plan is constructed by blending the base form's production representation with the schema associated with the meaning to be conveyed (e.g. ...z#~PLURAL for the English plural).

---

[2] Rich-get-richer dynamics do favor re-use of previously-used forms and paradigmatic mappings between forms (Martin, 2007; Zipf, 1949), so a single wordform may often have a dominant influence on the constructed output because it is more frequent than competitors and/or more predictive of characteristics of the output (Albright, 2008). However, activation spreads to multiple words in parallel during lexical retrieval (Dell, 1986; Roelofs, 1992), offering them the opportunity to influence the output.

When a morphologically-related word is activated more (or earlier) than the target, aspects of its form can be retained and erroneously copied into the production plan under construction. Children show motor perseveration across domains (Dell et al., 1997; Smith et al., 1999) and are slower to retrieve and plan the target. Thus the morphological relative is often partially activated before the target is fully accessed, which allows aspects of the relative's form to influence production of the target, leveling stem changes (Do, 2013; Kerkhoff, 2007). Allowing a recently-activated form to drive production may help children learn language by encouraging imitation. Children will often repeat words used by the interlocutor, which can lead to pronoun confusion, e.g. in using "you" instead of "I" in response to "Do you want to eat?" (Clark, 1974, 1977; Rubino & Pine, 1998). Except for the rare pronoun confusion and stem leveling, perseveration in child speech is often functional, in that it allows the child to reproduce structures that would be too complex for them to produce compositionally on their own (Clark, 1977; Farrar, 1992; Rubino & Pine, 1998). This imitation of the interlocutor is akin to perseverating on aspects of the given or recently produced form in wug tests, in that in both cases, speakers incorporate a form activated by external factors (rather than internal, top-down input) into the production plan when the target is difficult to plan or retrieve, which suggests that the same mechanism lies behind both processes.

1.2.1.1. Paradigmatic perseveration in Russian

Paradigmatic perseveration seems inevitable in the context of elicited production; how could a speaker not perseverate on the form provided when there is nothing else available for memory retrieval? But its existence in natural language requires empirical support that the base form can become activated, down to the level of production units,

8

before the target form is produced, and then those activated units must be incorporated into the production plan under construction. In contrast to syntagmatic perseveration, in (3), paradigmatic perseveration has rarely been subject to scholarly attention (though see Kapatsinski, 2010).

3. Was there, like, an expl[oʊ]sion of p[oʊ]p?

Bez-initial adjectives (*bez*- meaning '-less') in Russian, like those in (4) and (5), provide evidence for paradigmatic perseveration in natural language. These adjectives always have a corresponding prepositional phrase, but often lack a *bez*-less adjective counterpart. *Bez*-initial adjectives are predictable if we assume that they are created from prepositional phrases via an adjectivizing schema like [N]*nyi#~*ADJ.MASC.SG.NOM, but not if we assume they are created from *bez*-less adjectives, which frequently do not exist.

4. *bezkrylyj*   *bez kryl'jev*   **krylyj*
   'wingless'   'without wings'   'winged'

5. *bezsmyslennyj*   *bez smysla*   **smyslennyj*
   'pointless'   'without a point'   'having a point'

The preposition *bez* and the prefix *bez*- are both underlyingly /z/, which is [z] before vowels and voiced consonants and [s] before voiceless consonants. The preposition has a constant spelling, corresponding to the underlying /z/ (written with the Cyrillic equivalent of <z>). The prefix, like all /z/-final prefixes in Russian, should be spelled the way it is pronounced, namely <s> before voiceless consonants and <z> elsewhere. Kapatsinski (2010) shows that Russian speakers struggle to correctly spell unfamiliar/novel *bez*-initial adjectives, where their errors (like *bezkreditnyi* 'creditless') stem from spelling the prefix like the corresponding prepositional phrase (*bez kredita* 'without credit'). The error rates

for familiar *bez*-adjectives, whose orthographic representations they can retrieve from memory, are very low. They do not make the same spelling mistakes with /z/-final prefixes that do not have a differently-spelled preposition (like *iz-*, *roz-*, *voz-*), and the verbal prefix *iz-* is particularly informative as to why.

Iz- 'out' has a synonymous preposition *iz* 'out of', but *iz-*initial words do not have corresponding prepositional phrases; for example, *izpisatʲ* 'to cover with writing' comes from *pisatʲ* 'to write', not from a prepositional phrase. This means that it is not the preposition *bez* that interferes with *bez-*, but rather the prepositional *phrase*, the derivational base, that interferes. This is direct evidence for paradigmatic perseveration: A base form (in this case, a prepositional phrase) is activated more and/or earlier than the target (in this case, a rare or unfamiliar *bez-*initial adjective), which allows some elements of its form to be perseverated on in the output.

1.2.2. How perseveration creates PU

Perseveration can become conventionalized (i.e. phonologized), like other processing pressures, thereby becoming part of the grammar. Conventionalized perseveration is copying, where certain parts of the input are conventionally copied into the output under construction. Following faithfulness constraints (Prince & Smolensky, 1993/2004), the strength of the pressure to copy can vary across submorphological structures and is acquired through language experience. In our proposal, the target structures are positions within prosodic templates rather than features or gestures (e.g. Copy$_{FIN}$ mandates copying of the final element of the stem, regardless of its identity). The reasons are twofold:

1) Changes do not target all instances of a unit within a base, but rather only instances in certain positions (Kapatsinski, 2013; Kapatsinski, 2017a). For example, in Korean

verbs, the voiceless obstruents [p;t;k] are voiced before vowel-initial suffixes, thus /tat-a/ 'close-INTERROGATIVE' surfaces as *tada* (Do, 2013, 2018). Only the stem-final stop undergoes the change: The initial /t/ should be retained, even though it is also followed by /a/.

2) Learners need to be able to generalize from experienced elements in a certain position to novel elements in the same position (Kapatsinski, 2017a). The English regular past tense mandates copying over the entire root, and Berko (1958) shows that children and adults are able to apply the rule to stems they have not previously encountered. Without the ability to extract copy generalizations, learners would be prone to producing garbled foms like *membled* from *mail* (Pinker & Prince, 1988; Rumelhart & McClelland, 1986).

Copying is conditioned by two factors, the meaning to be conveyed (e.g. PLURAL) and the phonological characteristics of the input (e.g. word-final [p]). For example, in hypothetical language A, Copy$_{FIN}$ is activated when the speaker intends to express the plural meaning and the input ends in [p], whereas in hypothetical language B, Copy$_{FIN}$ is inhibited in the same context and allows the speaker to replace the final [p] with a different segment. Speakers of Southern Bantu, where labial palatalization is productive (p→tʃ/__w; Braver & Bennett, 2015; Ohala, 1978), must have learned an inhibitory association from word-final [p] to Copy$_{FIN}$, like that of language B, presumably on the basis of observing many cases where [p] is not preserved when a form with a certain meaning is produced.

11

Figure 1.1 shows a partial model of a language with labial palatalization[3], like language B and Southern Bantu, with input nodes representing [blaɪp] and output nodes that are activated or inhibited when the network is asked to express the PLURAL meaning. The state of knowledge represented in Figure 1.1 can be achieved by the network experiencing singular and plural forms and, on a minority of occasions, recalling the singular when the plural is experienced, and trying to predict the plural from the recalled singular and punishing (downweighting) the connections that lead to incorrect predictions (Kapatsinski, 2018a). Exposure to enough unfaithful mappings between cells in a morphological paradigm can cause the learner to prefer not to retain units that are frequently changed or removed when producing the novel form from the known. The ability to learn a preference for anti-faithfulness (non-retention) is useful for learning stem changes, subtractive patterns (like affix stripping), and "morphological toggles" (Alderete, 2001; Kurisu, 2001). Learners must be able to derive all the forms of a paradigm, including the base: For example, sometimes the plural is retrieved before the singular (Biedermann et al., 2013), and producing the singular requires removing the plural suffix. Kurisu (2001) discusses a range of languages that employ grammatical subtraction, including Koasati, which produces plurals through rime deletion (p. 83-84), and Icelandic, which creates deverbal nouns by deleting the final vowel of the infinitive (p. 111). The variety of languages and contexts in which subtraction applies suggests it is a fairly common process and that speakers must therefore be capable of learning anti-faithfulness.

---

[3] The full network (see Kapatsinski, 2018a) includes nodes for all phonological units of the language and all meanings corresponding to cells in a morphological paradigm, but for our purposes, the simplified version suffices.
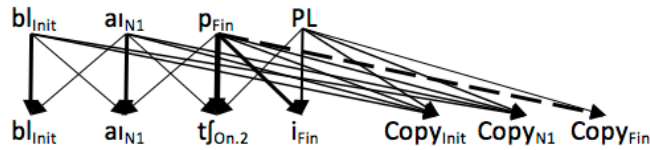
Figure 1.1. A language with phonologized labial palatalization. Width of the line shows connection strength. The dashed line shows an inhibitory connection. Word-initial onsets are always copied into the output, as are the vowel nuclei that follow them. However, a final [p] is not copied into the plural, being replaced with [tʃi][a]. Every feature of the input excites Copy$_{Init}$ and Copy$_{N1}$ because initial onsets and following vowels are always copied between paradigm cells in this language. A final [p] strongly activates output [tʃi] and inhibits copying of the consonant in the final position (itself). The plural meaning is strongly associated with the plural suffix *-i* (present whenever the plural meaning is present and not otherwise) and is more weakly associated with a preceding [tʃ] (which is overrepresented in plurals). Copying of onsets and first-syllable nuclei is associated with plurality as it is with other input features, while copying of the final consonant is associated more weakly since it does not always happen in the plural.

[a] This templatic coding scheme is sufficient for the languages presented to learners in the present thesis, where stems are monosyllabic, but would have to be extended for real languages.

Paradigm uniformity is generally seen as the tendency to preserve the *stem* of a form. Regarding cases like Polish noun case (Krajewski et al., 2011), where there is no true "stem" and generation of morphological relatives therefore requires affix switching rather than addition, we must address how the Perseveration Hypothesis restricts perseveration to the base and not afffixal elements. In generating a related form of a morphologically complex word, only the meaning of the stem is compatible with the meaning of the target word; the meaning of the affix is not, and therefore the articulatory gestures corresponding to the affix will be inhibited, making them unlikely to be incorporated into the new form.

1.3. Associations

As mentioned above, copying is often the right thing to do (e.g. *wug~wugs*). For it to be an error, there must be some other generalization that requires a change to the base that perseveration violates. A particularly important class of change-demanding

generalizations for our work is arbitrary paradigmatic mappings, also called paradigmatic

associations (Ervin, 1961), the latter of which could be considered to be the cognitive

representations of the former. Paradigmatic mappings are controversial in usage-based

and constructionist accounts (Bybee 1985, 2001; Goldberg, 2002; Kapatsinski, 2013), but

they seem to be necessary in morphology (perhaps *only* in morphology, Kapatsinski,

2018a, 2018b) because the shape of the to-be-produced form can depend on what other

forms of the same word are like, as for Genitive Plural production in Russian in (6)-(7).

6. *trop*   *tropov*
 'trope'   'tropes.GEN'

7. *tropa*   *trop*
 'path'   'paths.GEN'

The shape of the Genitive Plural is determined by whether the base (Nominative

Singular) ends in -*a*, a suffix which is absent from the Genitive Plural form itself. For

other examples, see Booij (2010), Gouskova & Becker (2014), Kapatsinski (2017b),

Nesset (2008), and Pierrehumbert (2006).

When a paradigmatic association requires a change to the base (like in

electri[k]~*electri*[s]*ity*), it conflicts with paradigmatic perseveration and obeying the

latter results in an error.

1.3.1. The Paradigm Cell Filling Problem

Some theorists argue that the Paradigm Cell Filling Problem can be solved without

referring to other forms of the word (Malouf, 2017; Thymé, 1993; Thymé et al., 1994).

Malouf (2017) achieves impressive performance on a range of morphological systems by

using the lexico-semantic features to be expressed and the preceding phonological cues as

input and generating the target incrementally left-to-right. However, we think a general

solution to the PCFP needs to incorporate paradigmatic mappings (Albright & Hayes,

2003; Bonami & Beniamine, 2016; Booij, 2010; Kapatsinski, 2010, 2018a, 2018b; Nesset, 2008; Plunkett & Juola, 1999; Rumelhart & McClelland, 1986; Westermann & Ruh, 2012), since the phonological features of other forms of the word are often the most informative cue to aspects of the target form. Without these paradigmatic mappings, learners can be misled by phonological neutralizations within target forms when they make the preceding phonological context uninformative (Becker & Gouskova, 2016; Gouskova & Becker, 2013).

Masculine diminutives in Russian serve as an illustrative example. They are formed using a set of suffxes including *-ik* and *-ok*, where *-ik* is favored by non-velars and *-ok* by velars. *Luk* 'onion' selects *-ok*, and *lutʃ* 'beam' selects *-ik*, but velars are palatalized before *-ok*, so *luk*→*lutʃ* before *-ok*, neutralizing the contrast; the diminutive forms of 'onion' and 'beam' are *lutʃok* and *lutʃik*, respectively. The neutralization of the consonant contrast means that there is nothing in the diminutive form to predict whether *-ik* or *-ok* should be chosen. When trying to generate 'little onion', a left-to-right process like that of Malouf (2017) could change [k] into [tʃ], because [k] is underattested in the postvocalic position in diminutives, but would be unable to determine whether to continue with [o] or [i], since both occur in that context.[4] In order to know what the final vowel of the diminutive should be, there needs to be some way of referencing the consonant of the non-diminutive[5]. The diminutive and non-diminutive are in an

---

[4] The model could be augmented by including inflectional class in addition to the morphosyntactic features used for specifying the paradigm cell; this would, however, grant the model prior knowledge of the structure of the lexicon that may not be warranted, if the goal is to approximate a novice learner.

[5] In rule-based models, the neutralization before *-ok* is taken to mean that the suffix must be chosen first and then trigger the consonant change. However, as long as the final consonant of the non-diminutive base remains accessible to cue the choice of the suffix after the stem is changed, then either order is acceptable. The crucial part is the availability of the paradigmatically related non-diminutive base when the suffix is being chosen.

asymmetrical implicative relationship: The form of the non-diminutive predicts the form

of the diminutive, but not vice versa.

Paradigmatic mappings from [k] to [tʃok] and [tʃ] to [tʃik] are active for speakers:

Russian adults correctly affix novel/unfamiliar nouns ending in [k] with -*ok* and almost

categorically palatalize (and almost never palatalize before -*ik*), and nouns ending with

[tʃ] are suffixed with -*ik*[6]. For other productive paradigmatic mappings, see Becker &

Gouskova (2016), Gouskova & Becker (2013), Krajewski et al (2011), and Pierrehumbert

(2006).

Cross-lingustic research shows the ubiquity of implicative relationships in

morphological grammars (Ackerman et al., 2009; Ackerman & Malouf, 2013; J. P.

Blevins, 2013; Bonami & Beniamine, 2016; Bonami & Strnadová, 2019; Finkel &

Stump, 2007; Sims & Parkers, 2016; Stump & Finkel, 2013). Not all of these

relationships are necessarily productive, but it is difficult to believe that none of them are.

While one could once question the productivity of paradigmatic mappings, and by

extension their existence in the mental grammar (Bybee, 2001; Kapatsinski, 2013;

Ramscar et al., 2013), the evidence now seems sufficient for a consensus of their

psychological reality (Ramscar et al. 2010, p. 914, 2013, p. 782 vs. J. P. Blevins et al.,

2017; Kapatsinski, 2013 vs. 2017b, 2018b).

1.3.2. The difficulty in associating dissimilar representations

Thus far, the Perseveration Hypothesis does not predict a bias against *large* changes.

Speakers perseverate on the input, but it is not yet clear why, for example, p~tʃ in

---

[6] There are some semantic influences on the choice of suffix, too (Magomedova, 2017), but they do not
override paradigmatic phonology (see also Ramscar, 2002, for English).

bup~butʃi is harder to learn than k~tʃ in *buk~butʃi*, since in both cases, Copy$_{FIN}$ must be overriden, yet there is extensive evidence that large changes are disliked more than small (Skoruppa et al., 2011; Stave et al., 2013; White, 2014). We assume that the base forms that provide material for the production of novel forms are production representations; following Articulatory Phonology, we assume they are composed of articulatory gestures/target constrictions of the vocal tract (Browman & Goldstein, 1989). Paradigmatic associations, then, require learning an association between two gestures: Learning to change a base gesture X into another gesture Y requires learning a paradigmatic association between X and Y, such that activating X also activates Y and allows X to be changed into Y (Ervin, 1961; Rumelhart & McClelland, 1986, et seq.). The associative learning literature suggests that acquiring the X$\rightarrow$Y association is easier when X is similar to Y (Rescorla & Furrow, 1977; Rescorla, 1986; regarding phonotactics, Moreton, 2008, 2012; Warker & Dell, 2006; Warker et al., 2008).

A plausible mechanism for why similarity should matter for associability comes from Kapatsinski (2011), who argues that learning an association requires modifying the synaptic connections between associated representations. If the representations are very different, they may be stored in different parts of the cortex and be separated by more and/or weaker synaptic connections. More modification to the synapses is necessary to form the association, and may require more training. This is consistent with the motor sequence learning literature, where learning an association between X and Y appears to involve changing the behavior of the cortical and subcortical areas separating X and Y (Hluštík et al., 2004).

In the present work, we extend the finding that associating dissimilar elements requires more connections from syntagmatic associations between perceptual representations to paradigmatic associations between production representations. The crucial prediction is that associations between dissimilar gestures will be harder to learn than associations between similar gestures, which makes large changes particularly difficult to learn and perform.

1.4. Learning associations

Given the importance of paradigmatic associations to our account, we must consider how they are learned. Learning paradigmatic associations seems very challenging, and has proven difficult to observe in the lab (Braine et al., 1990; Brooks et al., 1993; Frigo & McDonald, 1998; McNeill 1963, 1966, though see Seyfarth et al., 2014, and Williams, 2003, for successful examples). In natural language, the acquisition of paradigmatic mappings develops into adulthood. Many Polish adults use only some of the factors conditioning suffix choice in extant vocabulary (Dąbrowska & Sczerbinski, 2006), and Polish children productively use frequent paradigms/cells but struggle with less frequent or reliable mappings (Krajewski et al., 2011), with a similar pattern present for Korean children (Do, 2013). The protracted development in acquiring paradigmatic mappings (Dąbrowska & Sczerbinski, 2006; Do, 2013; Krajewski et al., 2011) suggests that opportunities for learning them are rare, but we nonetheless think they are crucial.

1.4.1. Early proposals

Early work on paradigmatic mappings in the 1960s (Ervin, 1961; McNeill, 1963, 1966) examined paradigms of antonymous adjectives, like *deep~shallow*, *big~little*, *large~small*. There is evidence for the existence of paradigmatic mappings in that

domain. Adults tend to produce antonyms in response to adjectives in free association tasks, whereas children tend to produce associated nouns (Brown & Berko, 1960; Ervin, 1961; Woodrow & Lowell, 1916); e.g. if cued with *shallow*, a child might produce *lake* while an adult would produce *deep*. Adults also have intuitions about which adjectives go together, with *big* associated with *little and large* with *small* (Justeson & Katz, 1991), suggesting that adults have formed paradigmatic associations between antonymous adjectives.

McNeill (1966) claims that acquiring paradigmatic associations is difficult because paradigmatic associates rarely appear together in sentences, so co-occurrence cannot be relied on to learn them, and that therefore the opportunity to learn them is absent under normal speech conditions. Nevertheless, paradigmatic associations are learned, and there have been a number of proposals regarding what allows that to happen in the absence of contiguity. McNeill (1963, 1966) and Ervin (1961) propose that the strength of a paradigmatic response depends on how often it has been incorrectly anticipated for a given stimulus, with antonyms being characterized by their ability to substitute for each other within syntactic frames. Plunkett & Juola (1999) claim that children compare the word tokens they hear to what they expect to hear and use the differences to update their beliefs. Albright & Hayes (2003) propose that speakers create rules that apply to subsets of the grammar, describing what changes apply to create a past tense from a present tense form of a word. Regardless of the particulars, these accounts all propose impressive feats, namely that whenever a learner hears a word, they either

1) generate predictions about other forms of the same word and maintain those predictions until they are encountered, or

2) retrieve other forms of the same word from memory and evaluate whether they expected to hear what they heard.

Existing computational models of morphological learning from perceptual experience assume a high level of reliability in these processes, e.g. most models of English past tense acquisition assume that the present tense is always available for predicting the corresponding past tense form (Albright & Hayes, 2003; Plunkett & Juola, 1999; Rumelhart & McClelland, 1986; Westermann & Ruh, 2012), which seems highly suspect.

1.4.2. Discriminative learning and contiguity

If memory is not fully reliable, then the temporal relationship between predictor and predicted forms should be crucial. Comparisons between separately presented stimuli present substantial difficulties elsewhere, leading to, for example, change blindness (Mitroff et al., 2004). Using one form to predict another should be easiest when the predicted closely follows the predictor. Temporal contiguity has been argued to be important for learning associations since at least Thorndike (1898, the "law of effect"). Ramscar et al. (2010) argue that order matters for what is learned about form-meaning mappings. In their experiment, each trial showed participants either a spoken form followed by the pictorial representation of meaning or a picture followed by the form. Participants discovered the most predictive semantic features when the meaning preceded the form, but not vice versa. These findings are argued to be consistent with discriminative models of associative learning, where cues are used to predict subsequent outcomes, and the learner only acquires associations from predictive cues to outcomes. Predictive cues allow the learner to discriminate between cue sets followed by a particular outcome and those followed by other outcomes, and the downweighting of

20

unpredictive cues is central to learning to appropriately discriminate. Further evidence for

the importance of the temporal relationship for learning is provided by Arnon & Ramscar

(2012), who show that the temporal relationship is crucial for the ability of the learner to

learn to use gendered articles to predict upcoming nouns. Ramscar (2013) shows that

prefixes make the following nominal items more predictable, whereas suffixes make the

preceding nominal items more similar to each other. These results suggest that the

temporal relationship between form and meaning determines what is learned about the

form-meaning relationship. In the present work, we argue that the temporal relationship

between paradigmatically related forms is a crucial determinant of what is learned about

the paradigm.

1.4.3. The neurological underpinnings of learning

Much prior research has focused on the influence of error-driven predictive learning

on language acquisition (Baayen et al., 2011; Ellis, 2006; Lim et al., 2014; Plunkett &

Juola, 1999; Ramscar et al., 2010; Ramscar et al., 2013; Ramscar & Gitcho, 2007;

Ramscar & Yarlett, 2007; Rumelhart & McClelland, 1986; Westermann & Ruh, 2012). In

reality, multiple systems work simultaneously. Neuroscience has shown that the brain

possesses several complementary learning systems that learn in fundamentally different

ways, including at least the posterior neocortex, hippocampus, and striatum (Ashby et al.,

2007; Kumaran et al., 2016). The striatum supports error-driven predictive learning (Lim

et al., 2014; Ramscar et al., 2010; Schultz, 2006; Waelti et al., 2001). The hippocampus

supports a rapid chunking process, which fuses cues together and allows them to acquire

associations that the individual elements lack (O'Reilly & Rudy, 2001; Sutherland &

Rudy, 1989). The neocortex learns in a Hebbian manner, strengthening associations

between co-occurring stimuli regardless of expectations (Ashby et al., 2007; McClelland, 2001). Learning in the striatum is sensitive to prediction error because prediction error inversely correlates with the amount of dopamine in the synapse. The posterior cortical synapses lack dopamine projections, so synaptic connections strengthen whenever the cue co-occurs with the outcome, regardless of whether the outcome's occurrence was expected (Ashby et al., 2007). All of these areas contribute to language learning, so they should all contribute in a learning experiment, as well, though any processes that require sleep (like consolidation in the neocortex) would only surface in an experimental paradigm that spans multiple days. In our experiments, we expect behavioral signatures of error-driven predictive learning to co-exist with chunking, with Hebbian learning contributing minimally if at all.

1.4.4. Temporal contiguity

Cue-outcome contiguity enables predictions of the outcome based on the cue (discrimination of cue configurations, e.g. base forms, leading to different outcomes, e.g. derived forms) and also highlights the differences between the forms (Carvalho & Goldstone 2015; Zaki et al., 2016). Alternating between categories allows learners to identify the discriminative features of category exemplars, the ones that best distinguish between the two categories. This contrasts with a blocked presentation, where learners focus on the features that are most common for that category, regardless of whether they are informative about the differences between categories. This suggests that in order for learners to determine which features distinguish e.g. singulars from plurals in their language, they should encounter corresponding singular and plural forms in close temporal proximity.

Previous studies have not investigated the importance of temporal contiguity between forms, likely because they shared McNeill's (1966) assumption that paradigmatically related words rarely if ever appear next to each other. Work by corpus linguists has shown this assumption does not hold true for antonyms and morphological paradigms; antonyms co-occur more frequently than non-antonyms (Fellbaum, 1996; Jones, 2002; Jones et al., 2007; Justeson & Katz, 1991; Murphy, 2006), and canonical (small/large vs. big/little) co-occur more than non-canonical (Jones et al., 2007). Morphologically related words are more likely to co-occur in a limited text window than other word pairs, and computational models seeking to identify sets of words sharing a stem have been found to benefit from paying attention to co-occurrence (Baroni et al., 2002; Xu & Croft, 1998). If paradigmatically related words do co-occur, then instances of syntagmatic co-occurrence may be crucial for learning paradigmatic mappings like $C_i\#_{Nom.Sg} \sim C_i ov\#_{Gen.Pl}$ and $C_i a\#_{Nom.Sg} \sim C_i\#_{Gen.Pl}$. In the realm of production, there is anecdotal evidence that children spontaneously produce paradigms in monologic word play (Weir, 1962; Nelson, 1989; Saville-Troike, 1988). Whether we learn paradigms from perception (McNeill, 1966; Plunkett & Juola, 1999) or production (Taatgen & Anderson, 2002), or a combination of the two, temporal contiguity may be essential for enabling acquisition of paradigmatic mappings.

If temporal contiguity helps learn implicative relationships, how does it do so? Following Ramscar et al. (2010) and Arnon & Ramscar (2012), we posit that contiguity allows for discrimination of cue configurations that result in distinct outcomes. However, while they argue that forms constitute single undecomposable cues, we believe forms are configurations of somewhat separable cues (Kapatsinski, 2009), and that every

23

phonetic/phonological feature of a wordform can in theory be predictive of semantic and distributional characteristics of the word (see also Arnold et al., 2017; Baayen et al., 2011), including – crucially – what other forms of the word are like, e.g. that a particular type of Russian non-diminutive form predicts a particular type of diminutive suffix. To productively apply implicative relations, a learner needs to know that the 'base' form is predictive of other forms, and the specific paradigmatic mappings between particular phonological features in the base and related/derived forms, which are largely arbitrary and not based on phonetics (e.g. Russian [k]-final nouns take -*ok* in the diminutive) and must be learned through experience.

If individual phonological features are associable, then words are elemental, i.e. composed of a large set of independently associable elements. The evidence for this lies in the fact that although larger chunks like rimes and words are associable, associations of a segment sequence cannot be changed without interfering with the associations of the component parts, since they are recognized in the process of recognizing the whole (Kapatsinski, 2007, 2009). An important part of learning is "chunking," where previously separate cues that are used together fuse together (Bybee, 2002; Ellis, 2017; Goldstone, 2000), though separation of previously fused cues can also occur (Goldstone, 2003).

In this work, we investigate whether contiguity benefits acquisition of two kinds of paradigmatic mappings: unfaithful mappings involving a change to the stem (e.g. $k_{SG} \rightarrow t\int_{PL}$), and faithful mappings that do not (e.g. $k_{SG} \rightarrow k_{PL}$). Theories of grammar differ on whether temporal adjacency should benefit both types of mappings. Network Theory suggests that unfaithful mappings could be carried out under pressure from schemas like "plurals should end in [t∫i]," which do not require noticing paradigmatic relations (Bybee,

2001; Kapatsinski, 2013, 2017b). Perhaps, despite appearances, unfaithful mappings are not learned by observing unfaithful mappings, so temporal contiguity between members of the exemplifying pairs will not benefit acquisition. Faithful mappings have been suggested to be the default (Hayes, 2004; McCarthy, 1998; Pinker & Prince, 1988), and therefore not subject to improvement from adjacency. If this is the case, then temporal contiguity will not benefit acquisition of faithful mappings. However, output-output faithfulness constraints (like Copy, Kapatsinski, 2017b, 2018a, §1.3) can be downweighted with linguistic experience, and English speakers have likely learned that consonants sometimes change (as in *electri*[k]~*electri*[s]*ity*). If that is the case, then making faithful mappings more obvious will strengthen the constraint and extend faithful mappings.

1.4.5. Variation sets

Variation sets are defined as successive utterances containing partial repetitions (Onnis et al., 2008), where the communicative intention is maintained, but any or all of lexical substitution and rephrasing, addition and deletion of specific reference, and reordering are present (Küntay & Slobin, 1996). One advantage of these partial repetitions is that they allow language learners to use local comparison to discover structure, even if their memory is limited, as it is for children. Approximately 20% of child-directed speech appears in variation sets (Küntay & Slobin, 1996), and the same proportion has been shown to be sufficient to assist in learning lexical items and phrasal units in miniature artificial language learning (Onnis et al., 2008).

Like our proposal, variation sets emphasize the importance of temporal contiguity. Variation sets allow comparison and extraction of shared units, as well as anticipation of

the first form encountered in a context when a different form is presented, and discovery

of the variety of contexts that a form can appear in. In Chapters V and VI, we focus on

the effect contiguity has specifically on noticing changes, and the intact singular-plural

pairs can be considered examples of (very small) variation sets.

1.5. Creating novel forms

In order to form new forms of known words, or to re-create a known form that cannot

be accessed quickly enough by lexical retrieval, speakers generate a form to express the

desired meaning by using:

1) meaning→form associations, such as the product-oriented/first-order schemas of

Bybee (1985, 2001), Kapatsinski (2012, 2013), Nesset (2008), and the constructions of

Goldberg (2003),

2) paradigmatic form-form associations, or second-order schemas, which are

necessary for arbitrary paradigmatic mappings like k~s in *electric~electricity* (Booij,

2010; Gouskova & Becker, 2013; Nesset, 2008; Pierrehumbert, 2006),

3) copying from activated wordforms, like the output-output faithfulness constraints

of Benua (1997) and Kenstowicz (1996), and

4) a mechanism for maintaining/re-creating serial order, such as a word-sized

prosodic template that grows with learning to fit the range of experienced wordforms

(Redford, 2015; Vihman & Croft, 2007).

In the present work, we focus on the interaction between (1), (2) and (3), but see

Kapatsinski (2018a) for the full model. Paradigmatic associations (2) are more difficult to

learn when the to-be-associated forms involve different articulators/dissimilar articulatory

gestures, and they compete with the perseveratory tendency to simply output the

activated neuromotor representations (which vary by context and language, based on linguistic experience). When the association is too weak to override the perseveratory tendency, the stem change is leveled and paradigm uniformity arises. Product-oriented schemas may be acquired and used for judgments, even when they are not strong enough to drive production because of a strong opposition from perseveration, so a form that is rarely or never produced may still be accepted in judgment because it contains the correct cues for the meaning (Kapatsinski, 2013).

1.6. Alternative theories

There are a number of competing theories about the origin of paradigm uniformity. We review them in turn below.

1.6.1. Stem and affix

Pinker & Prince (1988) criticized Rumelhart & McClelland (1986) for predicting *membled* as the past tense of *mail*, based on a model similar to that in Figure 1.1 but without Copy outputs, and proposed that words are instead produced from stems and affixes. The tendency to retain too little of the stem is a recurring problem for connectionist models of wug test behavior; without Copy$_{INIT}$, the model in Figure 1.1 predicts that a novel onset will always be replaced with another onset that has already been experienced, with the particulars being determined by co-occurrence with the rest of the segments in the word and what meaning is intended, but Kapatsinski (2018a) shows that Copy resolves that issue.

Morphology as a description of language does not require that novel forms be produced from combining the stem with an affix; it could instead be based on the storage of whole-word representations in a paradigmatic network (Booij, 2010; Bybee, 1985;

Hockett, 1954; Matthews 1965; Robins, 1959; see J. P. Blevins, 2013, for a review).

Complex morphological systems where a speaker needs to know more than one other

complete form of a word in order to derive a novel form cannot be described as a

combination of stem and affix, because the paradigm has multiple principle parts

(Ackerman et al., 2009; Ackerman & Malouf, 2013; J. P. Blevins, 2006). Additionally,

parts of the stem can fuse with the affixes (and the meanings of the affixes) with which

they frequently co-occur, forming a construction/schema with boundaries that do not

correspond to morpheme boundaries: [[]$_N$*holic*]$_A$ has been extracted from *alcoholic* and

generalized to mean 'addicted to', as in *workaholic,* which is puzzling if *alcoholic* is

regularly derived from *alcohol + ic* (Bybee, 1985). Cross-boundary units can form even

very early in acquisition; hearing *blutʃ~blutʃi* strengthens the notion of 'tʃi' as a plural

unit and increases the likelihood of participants believing that *blut* should become *blutʃi*

(Kapatsinski, 2012, 2013). In other words, tʃ~tʃi supports 'tʃi'=PLURAL rather than

PLURAL=stem+i, which is problematic for models that assume morphologically-related

forms are separated into change + stem/context (Albright & Hayes, 2003; Gouskova et

al., 2015). It is likely that the tendency to preserve the stem is not because it is the basic

word form from which others are derived by concatenation, but rather as a consequence

of the fact that affixes are former words that have grammaticalized in situ (Bybee, 1985,

p. 41; Lehmann, 1992).

Additional evidence against PU as a consequence of concatenative morphology lies in

the fact that the bias for preserving the stem is much stronger than the bias for preserving

the affix (Beckman, 1998; Benua, 1997; McCarthy & Prince, 1995). For example, vowel

harmony tends to spread from the stem to affixes, not the other way around: Bakovic

(2003) shows languages preserve the identity of the stem, even when the result is disharmony, and Finley (2015) demonstrates a learning bias in favor of vowel harmony (and against PU) in affixes. These effects are unexpected if stem preservation and affix preservation both originate from a word = stem + affix rule. Following Correspondence Theory (McCarthy & Prince, 1995), we omit the affix from the input in our model, and it is instead activated by the intended meaning and arbitrary paradigmatic associations. Since it is not in the input, it cannot be perseverated on.

Some aspects of the base are more likely to be perseverated on. In the Canadian Raising example in (1), the height of the vowel is preserved, but the duration of the following stop closure and absence of vocal fold vibration are not: Voiced and voiceless stops both become flaps. Any analysis thus needs to be able to target specific submorphological units for perseveration, rather than entire stems and affixes (Prince & Smolensky, 1993/2004). These include phonological units like vowel height, but also subphonemic features of sounds, such as duration of stop/tap closure and phonetic correlates of stress like duration, pitch accent, and vowel quality (Steriade, 2000). The existence of phonological and subphonemic preservation effects motivates our proposal that the root of stem perseveration lies in production perseveration, rather than the allegedly concatenative nature of morphology.

1.6.2. Storage economy

If we assume that words are stored in their phonetic surface form, then related words that share the same phonological structure require less space (Kenstowicz, 1998). However, there is no evidence to suggest that long-term memory storage is limited (Householder, 1966; Johnson, 1997). Humans are capable of storing vast amounts of

information, down to specific episodes. Shepard (1967), Standing et al. (1970), and Standing (1973) show that thousands of novel pictures are automatically memorized based on a single exposure and are retained over several days, so working memory cannot be the sole factor behind PU. Brady et al. (2008) and Konkle et al. (2010) demonstrate that these memories are numerous and fairly detailed. Palmeri et al. (1993) use an old/new recognition task to show that voice-specific memories are stored, with word recognition facilitated when spoken by a familiar speaker over when spoken by a novel speaker, and the same degree of facilitation observed regardless of the number of voices (up to 20). From these results, they argue that long-term voice-specific memories of words are formed, even if 20 distinct voice-specific memories per word would be required. The formation of voice-specific memories is not necessarily automatic; it has been found to occur for old/new speaker identification, but not for lexical decision (Theodore et al., 2015). Nevertheless, the results suggest that an additional form representation per paradigm would not impose a significant load on long-term memory.

Thus far, the accounts discussed do not account for the finding that PU favors small changes over large, in addition to favoring no change over change. Skoruppa et al. (2011) and White (2013, 2014) find that a change in two features is harder to learn than a change in one feature, but it is unclear how storage economy can explain this, since in both cases, a non-uniform paradigm requires more storage space. The preference against large changes motivates the accounts of PU in §1.6.3 and §1.6.4, based on avoidance of perceptual dissimilarity between alternants.

1.6.3. Perceptual similarity

Kenstowicz (1996) argues that PU makes the relationships between related words more obvious, thereby facilitating lexical access (see also Steriade, 2000). The lack of PU may cause a listener to misinterpret a word that would have been parsable, had related words been activated, and speakers therefore avoid large changes because they are sensitive to the potential lack of understanding that could result. However, we do not know that a large change, like [p] to [tʃ], is harder for the listener to undo than a smaller change, like [t] to [tʃ]. Speakers do avoid ambiguity by hyperarticulating cues that distinguish words from their minimal pair neighbors (Baese-Berk & Goldrick, 2009; Wedel et al., 2013), but there are data suggesting that learners are not highly sensitive to homophony resulting from alternation, at least in the lab (Kapatsinski, 2012, 2013).

Steriade (2001/2009) and White (2017) propose that PU arises because speakers avoid producing changes that are easily noticeable. Like other Optimality Theory-based accounts, stem changes are introduced to improve phonotactics/ease of articulation, and speakers are additionally posited to possess a store of perceptual similarities between segments in context (the P-map), which they use to avoid noticeable changes so they do not violate speech norms (Steriade, 2001/2009). In other words, on their account, PU is due to avoidance of perceptual dissimilarity. Speakers may desire to change the language (e.g. to make it more regular or easier to produce), but do not desire that listeners disapprove, so they keep the changes small.

We believe that listener modeling in determining the articulatory details of pronunciation, such as degree of consonant voicing, is unsubstantiated. The online involvement of perceptual representations in production is not a settled matter (Perkell,

31

2012). Modeling the perceiver is involved in the retuning of production targets following target selection in auditory perturbation experiments (Perkell, 2012; Purcell & Munhall, 2006). However, this retuning is widely believed to be an offline process that follows selection of the target (Villacorta et al, 2007; Perkell, 2012; see Norris et al., 2003, and Norris & McQueen, 2008, for arguments against online feedback in perception). The perceptual dissimilarity biases suggested by Kenstowicz (1996) and Steriade (2001/2009) seem to require the influence of perceptual modeling prior to selection of the production target during everyday speech production, but guidance of production by online perception seems too slow to be consistent with how rapidly speech production, and other skilled motor action, proceeds (Elsner & Hommel, 2001; Welsh & Llinas, 1997). Regardless, an online perceptual feedback mechanism seems unnecessary to account for word production, so we would rather not rely on it to explain PU.

1.6.4. Categorization

Moreton & Pater (2012a) propose that the bias against large changes stems from category learning. Arbitrary sound categories requiring multiple features to describe are more difficult to learn than those requiring a single feature (Cristià & Seidl, 2008; Moreton et al., 2017; Pycha et al., 2003; for non-linguisic categories, see Shepard et al., 1961, and Feldman, 2003). The same sounds are perceived as being more similar by speakers of languages where they are allophones of the same phoneme (Boomershine et al., 2008; Johnson & Babel, 2010; Seidl et al., 2009). Moreton & Pater (2012a) suggest that acquiring an alternation involves categorizing the sounds that undergo the alternation separately from those that do not, and that this is easier when the groups can be defined by a single feature; for example, palatalization of labials (p~tʃ) in the absence of

palatalization of alveolars (t~tʃ) and/or velars (k~tʃ) requires categorizing [p] and [tʃ] together and separately from [t] and [k], perhaps as [labial] | ([coronal] and [dorsal]) (because English [tʃ] is [coronal] and [dorsal], Yun, 2006), whereas k~tʃ without t~tʃ and p~tʃ could be captured by [dorsal].

Perceptual category structure can account for the bias against large changes in judgment, but it is not clear that it should apply in production. We are not aware of any studies showing that training learners to categorize sounds together (like with the unimodal distributional training procedure of Maye et al., 2002) improves acquisition of an alternation involving the two sounds. Instead, increased perceptual similarity between alternating sounds may be a side effect of learning to produce the alternation rather than a cause of it.

1.6.5. Perseveration Hypothesis vs. Optimality Theory

We agree with the above accounts that paradigm uniformity exists, and that it influences learnability. However, there are several notable differences between our account and OT. Firstly, OT proposes that PU exists because of highly-ranked universal output-output faithfulness constraints (Hayes, 2004; Kenstowicz, 1997; McCarthy, 1998), whereas we believe it is better explained through perseveration in the production system (§1.2): In the process of generating a novel form of a known word, related forms of the word can become activated and be incorporated into the production plan. Secondly, the order that constraints are ranked determines the learnability of patterns in OT, with constraints being re-ranked after exposure to enough input. We propose that the learnability of an alternation is the result of the difficulty in associating the representations that participate in the alternation (§1.3); representations that are very

dissimilar are more difficult to associate, so it takes longer for them to become strong enough to override the perseveratory tendency. Finally, OT generally considers only the learnability of patterns, not the executability in production, with the notable exception of Do (2018). She shows that Korean children avoid producing the alternating forms of verbs, even after they have learned them; from this, she argues that learning biases influence production preferences as well as learnability. We expect PU to be a factor in production, as perseveration is perpetually present and should therefore continue to affect performance, even after speakers have learned what they should produce. Additionally, judgments of the goodness of forms could be based on first-order schemas, rather than paradigmatic mappings, and therefore speakers may like forms that result from an alternation because they contain the appropriate cues to meaning without being able to produce them themselves.

1.7. Structure of dissertation

Chapter II explores what palatalization is, and why we chose it as the test case. Chapter III discusses the findings from the baseline experiment with no training, to measure the biases subjects bring with them from English. Chapter IV describes the results of the learnability of palatalization of either labials, alveolars, or velars, and differences in production vs. judgment. Chapter V covers the results of experiments varying the adjacency of faithful and unfaithful forms in training for labial or velar palatalization. Chapter VI discusses the model of the findings from Chapter V regarding the influence of different cues on learnability. Chapter VII reviews and concludes. Portions of Chapters II, IV, V, VI, and VII were co-authored with Vsevolod Kapatsinski. Portions of Chapters II, IV, and VII were published as: Smolek, A. & Kapatsinski, V.

(2018). What happens to large changes? Saltation produces well-liked outputs that are

hard to generate. *Laboratory Phonology: Journal of the Association for Laboratory

Phonology*, *9*(1), 10. Portions of Chapters V, VI, and VII are undergoing revision as:

Smolek, A. & Kapatsinski, V. (2019). *Syntagmatic paradigms: Learning correspondence

from contiguity.* Manuscript submitted for publication.

CHAPTER II

INVESTIGATING CHANGE MAGNITUDE EXPERIMENTALLY

Portions of this chapter were taken from:

Smolek, A. & V. Kapatsinski. (2018). What happens to large changes? Saltation produces well-liked outputs that are hard to generate. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *9*(1), 10.

In order to investigate the influence of change magnitude and contiguity on the learnability of phonological patterns, we need a test case. We first discuss previous work on alternation magnitude before turning to the palatalization alternation used in the experiments in this thesis.

2.1. Prior experimental work on change magnitude

2.1.1. Learning in adults

A few experiments have examined the learnability of alternations while manipulating the distance between alternants. Skoruppa et al. (2011) and White (2013, 2014) investigate the learning of alternations involving place, voicing, and manner of obstruents (e.g. p~t vs. p~s vs. p~z; p~v vs. b~v). They find that alternations involving a change to one feature (non-saltatory) are easier to learn/perform than alternations involving more than one feature (saltatory).

Skoruppa et al. (2011) compare the learning rate of saltatory vs. non-saltatory alternations and find that the larger saltatory changes are more slowly acquired, as we would expect under the Perseveration Hypothesis. White (2013, Ch. 3; 2014) trained participants to criterion on alternating sounds and examined overgeneralization to other sounds. In the potentially saltatory condition, participants trained on p~v ("jumping over"

36

[b] and [f]) or t~ð ("jumping over" [d] and [s]), when tested using a two-alternative forced choice task, extend the alternation to [b]/[f] and [d]/[s]. Participants in the control conditions, trained on b~v or d~ð, do not extend the alternation to [p]/[f] or [t]/[θ].

Even when participants are given explicit proof of the alternation being saltatory, they still extend it to the intermediate sounds. Half of the saltatory condition participants in White (2013, 2014) were trained on p~v and half on t~ð, as in the potentially saltatory condition, but they were additionally exposed to either copied intermediate fricatives (f~f or s~s, respectively) or copied intermediate obstruents (b~b or d~d), then tested on the other intermediate sound in a two-alternative forced choice task. Even though only participants who got 80% correct on the trained alternations proceeded to the test portion, they still extend the alternating pattern to the other intermediate sound. Comparable results were obtained through a production task (White, 2013, Ch. 4.5). He concludes that large saltatory changes are taken by participants to imply that the intermediate sounds also change. However, the large changes may be harder to perform: Many participants in the saltatory condition were excluded for failing to reach criterion accuracy on the trained segments and/or failing to select changes on the trained segments during test (17/33 for the saltatory condition vs. 2/22 for the potentially saltatory condition; White, 2013, p. 83). We think that training to criterion is helpful for revealing categorization biases but obscures the bias behind PU, namely that large changes are harder to perform and dissimilar sounds harder to associate.

In the experiments above, the alternations involving multiple features are saltatory, jumping over another sound, and it is unclear if the existence of an intermediary sound is necessary for large changes to be more difficult to learn than small. To our knowledge,

37

no one has investigated the learnability for the same change for learners that possess the intermediate sound in their native language inventory and those that do not (e.g. training Arabic and English speakers on b~f, as Arabic lacks the intermediate [p] and [v] in the native inventory; Watson, 2002). We therefore take these results to indicate that large changes are harder than small changes, in general, which could manifest as a bias against saltatory changes.

2.1.2. Learning in infants

White & Sundara (2014) and White (2013, Ch. 5) investigate a bias against saltatory alternations in 2 year old infants. The infants were placed in one of four conditions, shown in (1)-(4). Groups 1 and 2 were exposed to a saltatory alternation, p~v in (1) and t~z in (2), and were tested on b~v in (1) and d~z in (2). Participants in (1) pay more attention, as evidenced through longer looking times, to d~z trials than b~v, and participants in (2) pay more attention to b~v trials than d~z. The infants in (3) and (4) were trained on b~v and d~z, respectively, and tested on p~v and t~z, but they show no difference in looking times between the alternations at test. In other words, participants in (1) and (2) learn that the alternating forms are allophones in complementary distribution and extend the phoneme to include the intermediate sound, whereas participants in (3) and (4) do not extend the alternating category to [p] or [t], respectively. The authors claim that this provides evidence for a bias against saltatory alternations in infants. We believe it is instead evidence that alternating sounds are grouped together in perception and that the category includes intermediate sounds, unless contradictory evidence is provided (and possibly even then, if the results from White, 2013, 2014, hold true for infants as well), making the bias against saltatory alternations a special case of the bias

against discontinuous categories in perception (Maddox et al., 2005, 2007; Moreton & Pater, 2012a; Moreton et al., 2017). This bias is distinct from the production-internal biases that are the focus of the Perseveration Hypothesis, and we do not believe that it obviates the need for a bias against large changes in production. It is unlikely that infants are responsible for much language change, since they are not in a position to spread their innovations to others (Bybee, 2001). Learning to productively use paradigmatic mappings continues into school years (Berko, 1958), and may not be complete even in adulthood (Dąbrowska, 2012).

1. {*rom;na*}{*t;z*}VCV and *rom p*VCV and *na v*VCV
2. {*rom;na*}{*p;v*}VCV and *rom t*VCV and *na z*VCV
3. {*rom;na*}{*d;z*}VCV and *rom b*VCV and *na v*VCV
4. {*rom;na*}{*b;v*}VCV and *rom d*VCV and *na z*VCV

When producing novel forms of a word, learners of paradigms struggle primarily with changing the sounds that should be changed rather than avoiding changing any intermediate sounds. Children usually make perseveratory errors when producing novel forms of known words, leveling stem changes instead of extending them (Do, 2013; Kerkhoff, 2007; Krajewski et al., 2011), and errors in overgeneralization of a change to intermediate sounds are relatively rare (Bolognesi, 1998; White, 2017). This suggests that saltatory alternations are rare because they are a kind of large change, not because the "jumped over" sounds come to alternate as well. If imperfect learning ever seeds language change, it is more likely to be through persistent paradigmatic perseveration in production.

2.2. Palatalization

In the present work, we train participants on a palatalization alternation, where voiced ([b;d;g]) and voiceless oral stops ([p;t;k]) in singulars become the palato-alveolar affricates [dʒ] or [tʃ], respectively, before *-i* or *-a* in plurals. Palatalization is a common process cross-linguistically (Bateman, 2007; Kochetov, 2011), and has been studied experimentally (Guion, 1998; Kapatsinski, 2013; Stave et al., 2013; Wilson, 2006).

2.2.1. The typology of palatalization

Palatalization of coronals and velars is equally common, whereas labial palatalization is very rare (Bhat, 1978; Chen, 1973). Bateman (2007) argues that there are no cases of productive labial palatalization where the labial articulation is fully suppressed, and Kochetov (2011) proposes an implicational universal: If a language has labial palatalization, then it also has alveolar and/or velar palatalization. Palatalization before *-i* is more common than before any other vowel (Bateman, 2007; Kochetov, 2011). The preference for palatalization before *-i* over other vowels has been argued on articulatory (Anttila, 1989, p. 72-73; Hock, 1991, p. 73-77) and perceptual grounds (Guion, 1998; Ohala, 1989, p. 183-185, 1992, p. 320).

2.2.2. Palatalization and phonetic naturalness

Palatalization allows us to consider two types of phonetic naturalness. The first is contextual (Stave et al., 2013; termed "feature spreading" by Skoruppa et al., 2011 and "contextual relevance" by Peperkamp et al., 2006): Is the result of the alternation phonetically closer to the context than the input is? For example, through coarticulation, the high front vowel [i] fronts preceding velars (Bateman, 2007; Bhat, 1978; Wilson, 2006, Yun, 2006). With sufficient gestural overlap (Bateman, 2007), [k] can move

forward to [tʃ] before -i, making velar palatalization before -i natural in this sense. Palatalization before [a] is less contextually natural, because moving the place of articulation to the front of the mouth does not result in a consonant that is closer (articulatorily or acoustically) to [a] than [k] is.

The second type of phonetic naturalness is how natural the change itself is (Stave et al., 2013), in other words, how similar the alternating forms are to each other (termed "phonetic distance" by Skoruppa et al., 2011 and "phonetic proximity" by Peperkamp et al., 2006). The infrequency of labial palatalization could be attributed to avoidance of either perceptual or articulatory dissimilarity: [tʃ] is articulatorily more similar to [t] and [k] than [p], as [tʃ] and [dʒ] feature coronal and dorsal articulations (Yun, 2006), but not labial. The asymmetry is paralleled in patterns of perceptual similarity (Kochetov, 2011): Wang & Bilger (1973) show that linguals are confused with alveopalatals much more frequently than labials are confused with alveopalatals. Palatalization of [k] is thus more natural than palatalization of [p], whether perceptual or articulatory similarity is considered.

If the *perceptual* similarity between alternants is what determines learnability, as claimed by Hayes & White (2015), Kenstowicz (1996), Steriade (2001/2009), and White (2013, 2014), then palatalization of [k] before -i should be easier to learn than palatalization of [g] before -i: Guion (1998) demonstrates that American English listeners misperceive [ki] as [tʃi] in noise, but mistaking [gi] for [dʒi] is much less common. 40% of velar palatalization involves only [k], whereas there are no cases where only [g] is palatalized (Bhat, 1978). If *articulatory* similarity is what is important, as posited by the Perseveration Hypothesis (§1.1.3), then there should be no difference in the learnability

of palatalization of [k] and [g], since they are equally articulatorily similar to their palatal

counterparts. Despite the greater perceptual similarity between [ki] and [tʃi] than [gi] and

[dʒi] (Guion, 1998), English speakers palatalize [g] more than [k] (Wilson, 2006). This

effect is likely due to first language experience: The English letter <g> is often

pronounced as [dʒ], while <k> and <c> are rarely pronounced as [tʃ] (Gontijo et al.,

2003). A replication of Wilson (2006) would suggest that the perceptual similarity effect

is minor enough to be overcome by orthographic categorization and therefore is likely not

a strong factor in learnability.

2.2.3. Learnability

Typological frequency often maps onto ease of learning (Finley, 2008; Mitrović,

2012; White, 2013, 2014; Wilson, 2006). It has been shown that palatalization before *-i* is

easier to learn than palatalization before any other vowels for artificial and natural

language learning (Mitrović, 2012; Wilson, 2006), in line with a preference for context

naturalness. Contextually unnatural alternations can still be learned, and may even be

more productive than their contextually natural counterpart (e.g. Kapatsinski, 2010), but

they are more difficult to learn and likely to be generalized to the more natural context

(Mitrović, 2012; Wilson, 2006). Based on typological frequency and perceptual accounts

of paradigm uniformity, k~tʃi should be easier to learn than g~dʒi, but there are cases

where typological frequency does not correlate with a difference in learnability (Cristià &

Seidl, 2008; Moreton & Pater, 2012b; Pycha et al., 2003; Skoruppa & Peperkamp, 2011;

Seidl & Buckley, 2005). It seems plausible that the learnability of synchronic alternation

patterns, like palatalization, is the cause of only some of the difference in typological

frequency of those patterns, with the larger part being due to differences in the

frequencies of the diachronic change pathways that result in the alternations (J. Blevins, 2006; Bybee, 2001).

The typological asymmetry could be because [ti] and [ki] are more marked than [pi] (perhaps because they sound more like the palatal before [i]; Guion, 1998; Kochetov, 2011), and palatalization of alveolars and velars improves a bad output. However, Stave et al. (2013) found that [ap]→[atʃa] was palatalized less than [ak]→[atʃa] or [at]→[atʃa], which cannot be ascribed to markedness, as all of the alternations are equally phonetically unmotivated. The rarity of labial palatalization thus suggests the existence of a learning bias against labial palatalization that would cause it to either not be learned well, or to be overgeneralized so that it obeys the implicational universal.

In the present work, we manipulate context naturalness by testing the learnability of palatalization before -*i* (Experiment 2) vs. before -*a* (Experiment 3). We also examine two types of change naturalness, comparing the learnability of p~tʃ vs. t~tʃ vs. k~tʃ, and k~tʃi vs. g~dʒi. In the former, the input consonants differ in how articulatorily and perceptually similar they are to the output, and in the latter, they differ only in the degree of perceptual similarity to the output. Under the Perseveration Hypothesis, we expect to find a difference in palatalization rates by place of articulation (particularly labials vs. linguals) but not by voicing, whereas perceptual similarity-based accounts would expect differences by place of articulation and voicing.

2.2.4. The potential problem with palatalization

While using palatalization allows us to compare our findings to prior work and manipulate change and context naturalness, it does have the disadvantage of being present in English, for velars and alveolars in word pairs like *legal~legislate* and

create~creature and for alveolars in frequent phrases like *did you*. There is therefore a

chance that our findings could be due to first language transfer and not the experimental

manipulations. However, Experiment 1 (Chapter III) tests participants' judgments of

voiced and voiceless labial, alveolar, and velar palatalization before *-i* and *-a*, which we

include as a comparison for Experiment 2. Thus, we can compare the difference in

acceptability of palatalization of labials to palatalization of alveolars before and after

training to determine to what degree the results are due to pre-existing biases vs. learning.

We now turn to the palatalization experiments discussed in this thesis.

2.3. Experiment review

2.3.1. Experiment 1: Baseline

In Experiment 1 (Chapter III), we obtain judgments of palatalization of voiced and

voiceless labial, alveolar, and velar stops before *-i* and *-a* from native English speakers in

the absence of any training, in order to establish the biases our participants bring to the

task. We find that palatalization of alveolars is preferred to palatalization of labials and

velars, but that there is no difference in acceptability of palatalization before *-i* vs. before

-a or between voiced and voiceless velars. For the most part, the differences are minor,

which suggests participants do not come to the experiment with biases that strongly

influence acceptability of different types of palatalization.

2.3.2. Experiment 2: Palatalization before *-i*

Experiment 2 (Chapter IV) trains participants on miniature languages containing

palatalization of voiced and voiceless labials, alveolars, or velars before *-i*. This allows us

to evaluate the learnability of alternations that differ only in change magnitude: All the

conditions have the same target output [tʃi], but alveolars and velars share gestures with

palato-alveolars whereas labials do not, so learning palatalization of labials requires "jumping over" [t] or [k]. We find that participants in all conditions are able to learn to prefer the alternating form over the faithful (e.g. after receiving training on labial palatalization, subjects prefer p~tʃi over p~pi), but that labial palatalization is difficult to produce. Participants in the Alveolar and Velar Palatalization conditions produce palatalization of the target consonants as often as they judge it acceptable, but participants in the Labial Palatalization condition accept it without producing it. Comparison to the baseline shows that training increases acceptability of the trained alternation an equal amount across places; that is to say, the training improves performance an equal amount in judgment. We argue that acceptability judgments may reflect first-order schemas, rather than paradigmatic mappings, which accounts for why Labial Palatalization participants accept labial palatalization but rarely produce it.

The perceptual similarity account proposes that the greater confusability of [ki] and [tʃi] than [gi] and [dʒi] should result in k~tʃi being easier to acquire, but comparison of [k] and [g] reveals that there is no significant difference in palatalization rates in production or acceptability of palatalization in judgment by voicing, and in fact a slight preference for g~dʒi. While the bias against labial palatalization could conceivably be explained through either articulatory or perceptual dissimilarity, the absence of a perceptual similarity effect for [k] vs. [g] suggests that it is likely not the motivator of the dislike of labial palatalization.

Taken together, the results of Experiment 2 show that labial palatalization is more difficult to learn to produce than lingual palatalization, and it is because labials are articulatorily distinct from palato-alveolars, as expected by the Perseveration Hypothesis.

2.3.3. Experiment 3: Palatalization before *-a* with contiguity

Experiment 3 (Chapter V) investigates whether syntagmatic contiguity benefits acquisition of paradigmatic mappings. Participants are trained on either Labial ([p;b]) or Velar ([k;g]) Palatalization, before *-a*. We manipulate whether pairs of related forms are kept intact, with the plural immediately following the corresponding singular, or whether they appear in random order. In two of the trial order conditions (NoChange Obvious and All Obvious), pairs exemplifying faithful mappings (e.g. k~k{i;a}) are kept intact, and in the other two (None Obvious and Change Obvious), the singulars and plurals are randomly sorted, with the same true for pairs exemplifying unfaithful mappings (e.g. p~tʃa is kept intact in Change Obvious and All Obvious, and is not kept intact in None Obvious and NoChange Obvious). Within each language the four conditions are: None Obvious, where none of the pairs are kept intact and therefore none of the mappings are obvious (the same order as in Experiment 2); All Obvious, where all of the pairs are kept intact; Change Obvious, where only unfaithful pairs (e.g. p~tʃa) are kept intact and all other singulars and plurals are randomly ordered; and NoChange Obvious, where only faithful pairs (e.g. k~k{i;a}) are kept intact and all other singulars and plurals are randomly ordered.

Both faithful and unfaithful mappings benefit from temporal contiguity of the singular and the plural: Keeping unfaithful pairs intact in training results in more palatalization in production, and keeping faithful pairs intact in training results in less palatalization in production. Intact faithful and unfaithful mappings both extend beyond their trained contexts, unless the competing pattern is made more obvious: Change Obvious participants palatalize more of the stems that should not be palatalized (e.g.

alveolar- and velar-final stems in Labial Palatalization) than All Obvious participants do, and NoChange Obvious participants retain the stem consonant of the stems they should palatalize more than All Obvious participants do. Having both patterns in temporal contiguity enables speakers to learn what to change and what not to change.

2.3.4. Discriminative model of Experiment 3

In Chapter VI, we describe a discriminative learning model that captures the effects of Experiment 3. We find that keeping singular-plural pairs intact makes the cues of the singular available for predicting the outcome of the plural form. Conditions with intact faithful pairs show higher rates of copying (retaining the input consonant), and conditions with intact unfaithful pairs result in higher rates of palatalization. We also find that patterns extend to the smallest natural class that contains all of the participating segments, which results in copying extending more to velars in the Velar Palatalization condition than it does to labials in the Labial Palatalization condition. Chunking has an effect, as well, with surprise boosting the association between parts: Encountering *blutʃa* after hearing *blup* fuses [tʃa] and makes it easier for one part (e.g. *-a*) to elicit the other (e.g. *tʃ*). Temporal contiguity benefits even salient patterns; participants tend to produce only *-a* unless faithful pairs (half of whose plurals are suffixed with *-i*) are kept intact and *-i* is therefore made obvious. The model is successful at capturing the experimental manipulations and provides support for the importance of domain-general learning mechanisms in language acquisition.

2.3.5. Comparison of Experiments 2 and 3

In Chapter VII, we compare the results of Experiments 2 and 3. Unlike Experiment 2, there is no difference by change magnitude in Experiment 3. Neither labial nor velar

47

palatalization is produced in the None Obvious trial order before *-a*, whereas velar

palatalization is produced much more often than labial palatalization before *-i*. We

propose that this provides evidence for a substantive bias; velar palatalization is learnable

before *-i* because the high front vowel causes the tongue to move towards an alveopalatal,

making k~tʃ motivated before *-i* but not before *-a*.

Like Wilson (2006), we find that participants trained on palatalization before *-a*

generalize the alternation to *-i*, but the reverse is not true, providing evidence that

alternations generalize from less natural to more natural contexts. Unlike Wilson, we do

find a difference in the learnability of palatalization by suffix vowel, with participants

producing more palatalization before *-i* than before *-a*, even when trained on labial

palatalization, which is unmotivated in both contexts. This may reflect a learning bias in

favor of typologically-frequent patterns (White & Sundara, 2014).

In the next chapter, we discuss the results of the baseline judgment experiment.

CHAPTER III

EXPERIMENT 1: PALATALIZATION BASELINE

In order to determine effects of training on learning, a baseline needs to be established. English has examples of alveolar and velar palatalization in word pairs like *create~creature* and *legal~legislative* (though these are likely not productive in the synchronic grammars of speakers), and alveolar palatalization in phrases like *would you* and *bet you* (though these may not be the same process; Zsiga, 1995); because of this, participants may accept alveolar and velar palatalization more than labial palatalization. As discussed in §2.3, Experiments 2 and 3 investigate the effect of change magnitude (Chapter IV) and contiguity (Chapter V) on the learning of alternations. Without a baseline in the absence of training, the patterns of results could be due to pre-existing biases rather than the experimental manipulation. We chose to use judgments of palatalization in the absence of any training, since any production test requires at least some examples to illustrate the pattern in question. Perceptual judgments, however, require no generation on the part of the participants.

Experiments 2 and 3 include extensive training, whereas a baseline is intended to evaluate the influence of native language without training. However, the very act of exposure to training trials may change the interpretation that participants have of the task: Being asked to judge exemplars of an alternation may be less odd, or subject to alternative explanations, after being shown hundreds of nonce-word trials containing the alternation than doing so without training. We tried to minimize any differences by providing similar instructions to both groups, though the degree to which that was a successful strategy is certainly open for debate. The responses during informal post-

experiment interviews were largely the same for participants who were exposed to training and those who were not, though the former group included more frustration and confusion (unsurprising, given the full experiment was much longer and they had to learn a pattern rather than merely judge it). We were concerned that participants without training might rate all of the pairs containing palatalization as bad, but fortunately this was not the case. Given the goal – namely, to obtain data on participants' views of palatalization patterns based solely on native language experience – we believe the chosen experimental paradigm was the best strategy.

3.1. Methods

3.1.1. Participants

12 undergraduates in psychology and linguistics classes at the University of Oregon were recruited through the Human Subject Pool and received partial course credit for their participation. None reported having any speech, visual, auditory, or learning disabilities.

3.1.2. Materials

The test stimuli were 30 unique singular forms, randomly paired with pictures of creatures from the Spore database. The singular forms were all C(C)VC, and the final consonant was an oral stop. The final consonants were evenly divided between place of articulation (labial, alveolar, and velar) and voicing (voiced and voiceless), resulting in 5 tokens of every place * voicing combination.

Each singular had four plurals, crossing whether it was palatalized and what vowel was added (e.g. the singular *smip* had the plurals *smipi, smipa, smitʃi,* and *smitʃa*). Each singular was paired with all four plurals, resulting in 120 pairs. The singular and plurals

forms were recorded by a male native American English speaker from Oregon. The
materials can be found in Appendix A.

3.1.3. Procedure

The singular picture with the singular recording was followed by a 300 ms blank
screen and then the plural picture with one of the four plural recordings (see Figure 3.1).
Every participant received a different random ordering of the pairs. There was a one
second blank screen pause between trials. The experiment was conducted on E-Prime 2.0
Professional (Psychology Software Tools, Pittsburgh, PA) and lasted around 10 minutes.

Figure 3.1. Example display for a stimulus pair with labial palatalization. Participants
saw the creature(s) and heard the associated word (shown in brackets here).

Subjects were told they would hear the names of alien creatures, and that they needed
to indicate using a button box how good a match they thought the plural was for the
singular. The button box had 5 buttons but the results were strongly bimodal: 74.6% of
responses were 1 or 5, 19.7% were 2 or 4, and only 5.7% were 3, so we created a binary
measure where 1 and 2 were coded as 0 and 4 and 5 were coded as 1, with 3s excluded,
but the results are comparable if the full rating range is used. We chose to use the
bimodal responses rather than the full range because there is no difference in the

distribution of 3's across Test Place ($F(2) = 0.86$, $p < 0.43$, *ns*), and the results can be more easily compared to those of Experiments 2.

3.1.4. Measures

We performed generalized logistic linear mixed-effects models with the lme4 package (version 1.1-21, Bates et al., 2015) in R (version 3.6.0, R Development Core Team, 2019). Fixed effects were included for Keep Place (yes [faithful] vs. no [palatalized]), Plural Vowel (*-i* vs. *-a*), Test Place (Labial, Alveolar, and Velar), and Test Voice (Voiced vs. Voiceless), and any significant interactions. Random intercepts were included for Subjects and Singulars/Bases, with the full random effect structure that still allowed the model to converge (with maximally Keep Place, Plural Vowel, Test Place, and Test Vowel within Subjects, and Keep Place and Plural Vowel within Bases). Log likelihood tests on nested models were used to derive significance values. When a contrast that was expected to be significant was not, the evidence for the null hypothesis was evaluated using the BIC approximation to the Bayes Factor (Wagenmakers, 2007), which compares the posterior probabilities of the null and alternative hypotheses assuming their priors are equal and so can provide evidence for the null (unlike frequentist analysis), distinguishing between lack of evidence against the null vs. evidence for the null.

We used Helmert contrast coding for Test Place to compare labial to lingual (alveolar and velar) stems and alveolar to velar stems, because we suspected that the absence of labial palatalization and presence of (limited) velar and alveolar palatalization in English would make participants less likely to accept palatalization of labials than linguals. Visual inspection of the graphs showed that it was actually alveolar-final stems that often

patterned separately from labial and velar stems, so post-hoc tests comparing alveolars to labials were also performed. Tested models and contrast coding are included in footnotes, and the full dataset and code are available at https://app.box.com/s/bd8jhx4g5m7bvlmxb8i4jgtjfjo2x111.

3.1.5. Predictions

We predict that participants will likely judge alveolar and velar palatalization as better than labial, because of the patterns in English. Alveolar and velar palatalization are also much more common cross-linguistically than labial palatalization (Kochetov, 2011; Bateman, 2007).

Following typological patterns, we expect participants to like palatalizing before -*i* better than before -*a* (Bateman, 2007; Chen, 1973; Kochetov, 2011; Wilson, 2006). However, we do not expect the greater typological frequency of voiceless palatalization (Bhat, 1978) to correspond to higher ratings here, because English alveolar palatalization targets voiced and voiceless segments, which could make it likely that palatalization at all places of articulation would be expected to follow the same pattern. In fact, despite the greater perceptual similarity of [ki] and [tʃi] vs. [gi] and [dʒi] (Guion, 1998), English speakers prefer palatalizing [g] (Wilson, 2006), likely because of orthographic overlap between [g] and [dʒ] in <g>, so it could be possible that (at least for velar palatalization) the voiced alternation would be judged better than voiceless. We expect any difference between voiced and voiceless consonants, if present, to be strongest before -*i* and minimal or nonexistent before -*a*, as [tʃa] and [dʒa] are perceptually dissimilar from [ka] and [ga], respectively.

Lastly, we predict that subjects will rate faithful plurals (e.g. *blut bluti* and *blut~bluta*) as better than unfaithful (*blut~blutʃi* and *blut~blutʃa*), because larger changes are liked less (Kenstowicz, 1996; Skoruppa et al., 2011; Steriade, 2001/2009; White, 2014). We predict that there could be a difference in patterning by suffix vowel; due to the acoustic similarity between [ki] and [tʃi] (Guion, 1998), and that palatalizing causes consonants to become articulatorily more similar to a following [i] (Kochetov, 2011), participants might judge palatalization before *-i* as equivalent to, or even better than, lack of palatalization. We do not expect a similar effect before *-a*.

3.2 Results

3.2.1. Judgments of palatalized plurals

There is no significant difference in acceptance rates of labial palatalization vs. lingual palatalization[7] ($b = -0.37$, $se(b) = 0.34$, $z = -1.09$, $p < 0.28$), but alveolar palatalization is liked significantly more than velar palatalization ($b = 1.09$, $se(b) = 0.36$, $z = 3.05$, $p = 0.002$), and including Test Place significantly improved model fit ($\chi^2(2) = 7.01$, $p = 0.03$). Post-hoc tests show that alveolar palatalization is liked significantly more than labial palatalization, as well[8] ($b = 0.72$, $se(b) = 0.33$, $z = 2.16$, $p = 0.03$; $\chi^2(1) = 4.08$, $p = 0.04$). These patterns can be seen in Figure 3.2. and Table 3.1.

While there is no difference in overall acceptance rates of palatalization before *-i* vs. before *-a* ($b = 0.31$, $se(b) = 0.46$, $z = 0.68$, $p < 0.50$, *ns*; according to the BIC approximation to the Bayes Factor, the results provide strong evidence for the null, ΔBIC

---

[7] Rating Bin ~ Test Place * Plural Vowel + Test Voice + (1 + Test Place * Plural Vowel | Subject) + (1 + Plural Vowel | Singular), palatalized plurals only, with Helmert contrasts comparing labial to alveolar/velar stems and alveolar to velar stems.

[8] Rating Bin ~ Test Place + Plural Vowel + Test Voice (1 + Test Place + Plural Vowel | Subject) + (1 + Plural Vowel | Singular), palatalized plurals only, labial vs. alveolar stems.

= 6.14, P_{BIC} = 0.956), the difference between alveolar and velar ratings is smaller before

-*a* than before -*i* (*b* = -1.28, *se(b)* = 0.52, *z* = -2.48, *p* = 0.01), and including the

interaction significantly improves model fit (χ^2(2) = 6.22, *p* = 0.04). The figure clearly

shows where this effect originates: The acceptance rates for palatalized alveolars and

velars are essentially the same before -*a*, whereas palatalized velars are accepted less

often before -*i*.[9] While post-hoc analyses show that palatalized velar-final stems are liked

marginally more before -*a* than -*i*[10] (*b* = 0.94, *se(b)* = 0.55, *z* = 1.71, *p* < 0.09), including

Plural Vowel does not significantly improve the fit of the model and the data provide

positive support for the null, ΔBIC = 2.8, P_{BIC} = 0.802.
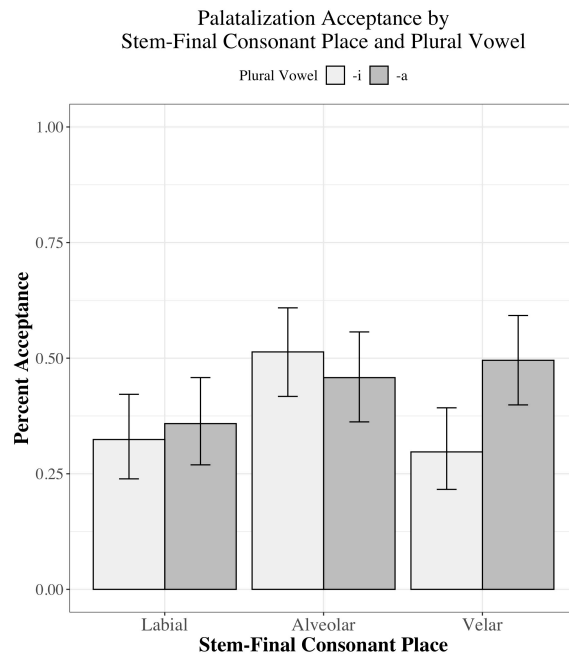
Figure 3.2. Acceptance of palatalized plurals by stem-final consonant place of articulation and plural suffix.

---

[9] Post-hoc tests show no significant interaction of Test Place and Plural Vowel for labial vs. alveolar stems (*b* = -0.50, *se(b)* = 0.44, *z* = -1.12, *p* = 0.26, *ns*) and the results provide positive support for the null, ΔBIC = 4.82, P_{BIC} = 0.918.

[10] Rating Bin ~ Plural Vowel + Test Voice + (1 + Plural Vowel | Subject) + (1 + Plural Vowel | Singular), restricted to palatalized velar plurals.

There is no significant difference between voiced and voiceless palatalization (b =

-0.054, *se(b)* = 0.18, z = -0.30, $p < 0.77$, *ns*; according to the BIC approximation to the

Bayes Factor, the data provide strong evidence for the null, ΔBIC = 6.39, P_{BIC} = 0.961;

Figure 3.3). Even when we consider only velar palatalization before $-i^{11}$ (where

perceptual similarity would suggest favoring [k] over [g]), there is no significant

difference by voicing (b = -0.47, z = -0.81, $p < 0.42$, *ns*; the results provide positive

support for the null according to the BIC approximation to the Bayes Factor, ΔBIC =

3.97, P_{BIC} = 0.879).

Table 3.1. Generalized linear effects model output for acceptance of palatalized plural-singular pairs by stem place of articulation, suffix vowel, voicing, and interactions.

	b	*se(b)*	*z*	*p*	
(Intercept)	-0.60322	0.33065	-1.824	0.0681	.
Labial vs. Alveolar/Velar	-0.36724	0.33662	-1.091	0.27528	
Alveolar vs. Velar	1.08724	0.35648	3.05	0.00229	**
Before -*a*	0.31218	0.4622	0.675	0.4994	
Voiceless	-0.05425	0.18374	-0.295	0.7678	
Labial vs. Alveolar/Velar x Before -*a*	-0.25377	0.47602	-0.533	0.59396	
Alveolar vs. Velar x Before -*a*	-1.28447	0.51763	-2.481	0.01309	*

. Marginally significant
* Significance level of 0.05
** Significance level of 0.01

---
[11] Rating Bin ~ Test Voice + (1 + Test Voice | Subject) + (1 | Singular), restricted to palatalized velars before –*i*.

56
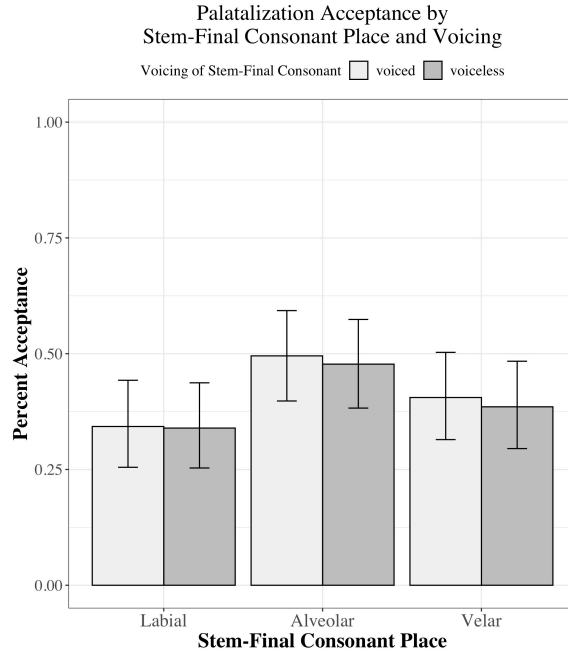
Figure 3.3. Acceptance of palatalized plurals by stem-final consonant place of articulation and voicing.

3.2.2. Judgments of faithful plurals

There are no significant differences between acceptance rates of faithful forms[12], whether for labials vs. the linguals ($b = 0.093$, *se(b)* $= 0.28$, $z = 0.33$, $p = 0.74$, *ns*) or alveolars vs. velars ($b = -0.24$, *se(b)* $= 0.31$, $z = -0.76$, $p = 0.45$, *ns*). The BIC approximation to the Bayes Factor provides very strong evidence for the null, $\Delta BIC =$ 12.42, $P_{BIC} = 0.998$. There is no significant difference of acceptability of faithful forms by suffix vowel ($b = 0.88$, *se(b)* $= 0.82$, $z = 1.07$, $p = 0.29$, *ns*), and according to the BIC approximation to the Bayes Factor, the results provide positive evidence for the null, $\Delta BIC = 5.38$, $P_{BIC} = 0.936$. These results are illustrated in Figure 3.4[13].

---

[12] Rating Bin ~ Test Place + Plural Vowel + Test Voice + (1 + Test Place + Plural Vowel + Test Voice | Subject) + (1 + Plural Vowel | Singular), faithful plurals only, Helmert contrasts comparing labial to alveolar/velar stems and alveolar to velar stems.

[13] Despite appearances, the interaction between Test Place and Plural Vowel is not significant ($b = 0.10$, *se(b)* $= 0.56$, $z = 0.18$, $p = 0.85$, *ns*) and the results provide very strong support for the null, $\Delta BIC = 10.45$, $P_{BIC} = 0.995$.
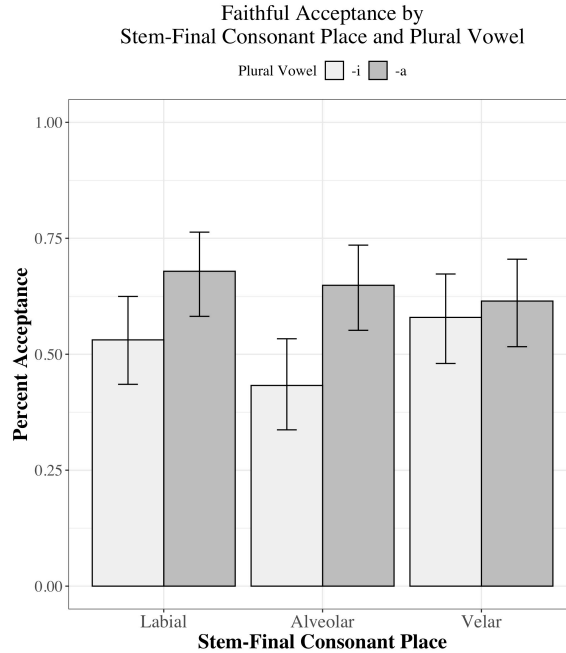
Figure 3.4. Acceptance of faithful plurals by stem-final consonant place of articulation and suffix.

3.2.3. Judgments of plurals before *-i*

Before *-i*[14] *(*see Figure 3.5 and Table 3.2), there is no difference in ratings of labial-final stems vs. lingual-final stems ($b = 0.035$, $se(b) = 0.47$, $z = 0.074$, $p = 0.94$) or between the linguals ($b = -0.79$, $se(b) = 0.58$, $z = -1.37$, $p = 0.17$), and according to the BIC approximation to the Bayes Factor, the results provide strong support for the null ($\Delta BIC = 8.69$, $P_{BIC} = 0.987$). There are also no differences between ratings of palatalized and faithful plurals before *-i* ($b = -0.52$, $se(b) = 0.64$, $z = -0.82$, $p = 0.41$) and the data provide strong support for the null according to the BIC approximation to the Bayes Factor ($\Delta BIC = 6.35$, $P_{BIC} = 0.960$). Post-hoc tests with alveolar as the baseline[15] show that unfaithful labial and velar stems are rated lower than unfaithful alveolar stems ($b =$

---

[14] Rating Bin ~ Keep Place * Test Place + Test Voice + (1 + Keep Place * Test Place | Subject) + (1 + Keep Place | Singular), before *–i*, Helmert contrast coded for labial vs. alveolar/velar stems and alveolar vs. velar stems.

[15] Rating Bin ~ Keep Place * Test Place + Test Voice + (1 + Keep Place * Test Place | Subject) + (1 + Keep Place | Singular), alveolar vs. labial and alveolar vs. velar stems.

-1.38, *se(b)* = 0.69, *z* = -2.00, *p* < 0.05 and *b* = -1.87, *se(b)* = 0.65, *z* = -2.87, *p* = 0.004,

respectively), but the inclusion of the interaction only marginally improves model fit

($\chi^2(2)$ = 5.30, *p* = 0.07) Thus, despite the apparent interaction of faithfulness by stem-

final consonant in Figure 3.5, the results provide strong support for the null according to

the BIC approximation to the Bayes Factor (ΔBIC = 7.67, P_{BIC} = 0.979). Finally, voiced

stems are no different from voiceless stems (*b* = 0.0026, *z* = 0.014, *p* = 0.99) and the BIC

approximation to the Bayes Factor shows that the results provide strong support for the

null (ΔBIC = 6.5, P_{BIC} = 0.963).
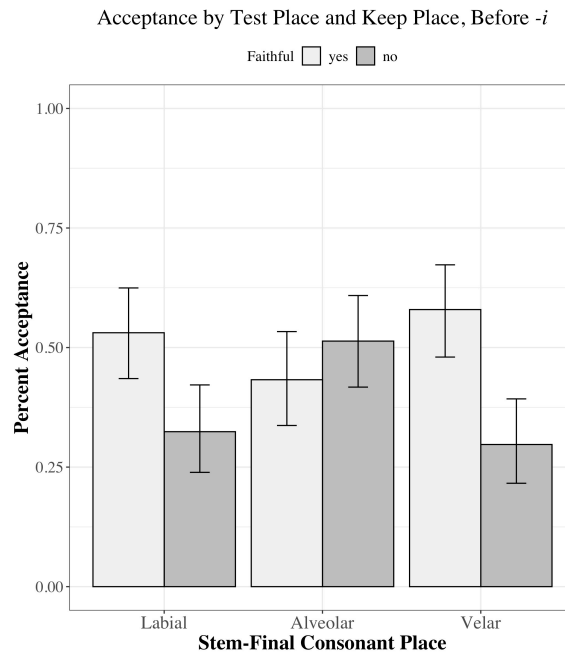
Figure 3.5. Acceptance of plurals before *-i* by stem-final consonant place of articulation
and faithfulness.

Table 3.2. Generalized linear mixed effects model output for acceptance of singular-plural pairs suffixed with -*i* by whether the plural was palatalized, stem-final consonant place of articulation, voicing, and the interaction between palatalization and place of articulation.

	b	*se(b)*	*z*	*p*	
(Intercept)	-0.125993	0.707238	-0.178	0.85861	
Palatalized	-0.521188	0.635601	-0.82	0.41222	
Labial vs. Alveolar/Velar	0.034646	0.467642	0.074	0.94094	
Alveolar vs. Velar	-0.794316	0.579559	-1.371	0.17051	
Voiceless	0.002644	0.192782	0.014	0.98906	
Palatalized x Labial vs. Alveolar/Velar	-0.444719	0.650644	-0.684	0.49429	
Palatalized x Alveolar vs. Velar	1.868223	0.651947	2.866	0.00416	**

** Significance level of 0.01

3.2.5. Judgments of plurals before -*a*

Before -*a*[16] (see Figure 3.6 and Table 3.2), the only significant effect is that palatalized plurals are rated worse than faithful (b = -1.13, $se(b)$ = 0.40, z = -2.80, p = 0.005; the inclusion of Keep Place significantly improves model fit, $\chi^2(1)$ = 6.31, p = 0.01). There is no difference between ratings of labial-final stems and lingual-final stems (b = -0.26, $se(b)$ = 0.27, z = -0.98, p < 0.33) or between alveolar and velar stems (b = 0.10, $se(b)$ = 0.27, z = 0.37, p = 0.71), and according to the BIC approximation to the Bayes Factor, the data provide very strong evidence for the null (ΔBIC = 12, P_{BIC} = 0.998). Voiceless stems are rated no differently than voiced stems (b = 0.0092, $se(b)$ = 0.25, z = 0.037, p = 0.97) and the data provide strong evidence for the null (ΔBIC = 6.48, P_{BIC} = 0.962). None of the interactions are significant.

---

[16] Rating Bin ~ Keep Place + Test Place + Test Voice + (1 + Keep Place + Test Place + Test Voice | Subject) + (1 + Keep Place | Singular), before –*a*, Helmert contrast coded for labial vs. alveolar/velar stems and alveolar vs. velar stems.

Acceptance by Test Place and Keep Place, Before *-a*
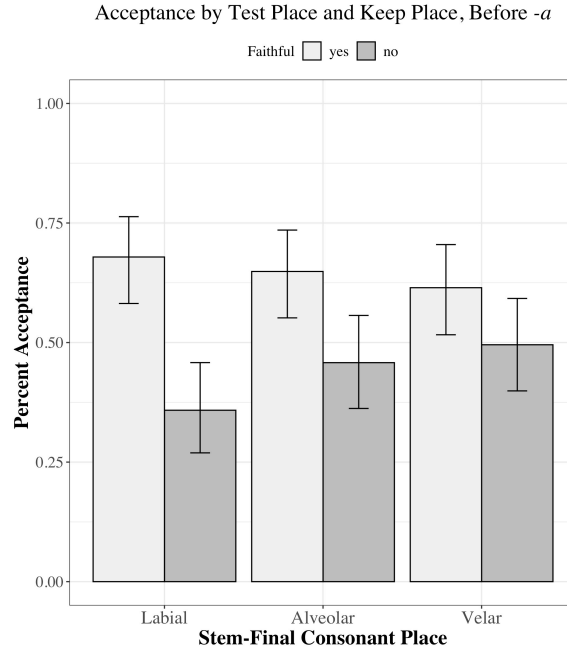
Figure 3.6. Acceptance of plurals before *-a* by stem-final consonant place of articulation and faithfulness.

Table 3.3. Generalized linear mixed effects model output for acceptance of singular-plural pairs suffixed with *-a* by whether the plural was palatalized, stem-final consonant place of articulation, voicing, and the interaction between palatalization and place of articulation.

	b	*se(b)*	*z*	*p*	
(Intercept)	0.83142	0.43778	1.899	0.05754	.
Palatalized	-1.12897	0.40266	-2.804	0.00505	**
Labial vs. Alveolar/Velar	-0.26095	0.26656	-0.979	0.32761	
Alveolar vs. Velar	0.10003	0.27361	0.366	0.71465	
Voiceless	0.00917	0.24826	0.037	0.97054	

. Marginal significance
** Significance level of 0.01

3.2.5. Judgments by faithfulness

Our final model evaluates judgments by Test Place, Plural Vowel, and Palatalization, as well as any significant interactions.

Palatalized forms are rated only marginally worse than faithful forms in the full model[17] (b = -0.70, z = -1.85, p = 0.06), but the inclusion of Keep Place does significantly improve model fit ($\chi^2(1)$ = 5.28, p = 0.02). There is a greater difference between palatalized alveolars and palatalized velars than between faithful alveolars and faithful velars (b = 1.65, z = 3.98, p < 0.001), and while there is no difference between labials and the linguals (b = -0.44, z = -1.24, p = 0.21), the interaction between Keep Place and Test Place does significantly improve model fit ($\chi^2(2)$ = 9.86, p = 0.007). Post-hoc tests[18] reveal that unfaithful alveolars also differ more from unfaithful labials than faithful alveolars do from faithful labials (b = 1.30, $se(b)$ = 0.41, z = 3.15, p < 0.002; $\chi^2(2)$ = 9.28, p = 0.002). Figure 3.7 illustrates the locus of this effect: Participants have no preference for faithful over unfaithful alveolars[19], but dislike palatalized labials and velars.

There is a significant main effect of suffix vowel, with plurals suffixed with -a being preferred to those suffixed with -i (b = 0.62, $se(b)$ = 0.17, z = 3.60, p < 0.001), and the inclusion of Plural Vowel significantly improves model fit ($\chi^2(1)$ = 13.54, p < 0.001). The significant three-way interaction between Keep Place, Test Place, and Plural Vowel indicates that the difference between ratings of palatalized alveolars and palatalized velars is smaller before -a (b = -2.01, $se(b)$ = 0.59, z = -3.44, p < 0.001; including the interaction significantly improves model fit ($\chi^2(1)$ = 12.23, p = 0.002), which can be seen

---

[17] Rating Bin ~ Keep Place * Test Place * Plural Vowel + Test Voice + (1 + Keep Place + Test Place | Subject) + (1 + Keep Place + Plural Vowel | Singular), Helmert contrasts coded for labial vs. alveolar/velar stems and alveolar vs. velar stems.

[18] Rating Bin ~ Keep Place * Test Place * Plural Vowel + Test Voice + (1 + Keep Place + Test Place | Subject) + (1 + Keep Place + Plural Vowel | Singular), restricted to alveolar and labial stems.

[19] Post-hoc test on alveolar stems only, b = 0.35, $se(b)$ = 0.41, z = 0.86, p = 0.39, ns; Rating Bin ~ Keep Place * Plural Vowel + Test Voice + (1 + Keep Place + Plural Vowel | Subject) + (1 + Keep Place | Singular).

by comparing the dark bars in Figures 3.5 and 3.6: Unfaithful alveolars are liked much more than unfaithful velars before *-i*, but before *-a*, they are accepted equally often.

Acceptance by Stem-Final Consonant Place
and Faithfulness

Faithful ☐ yes ■ no
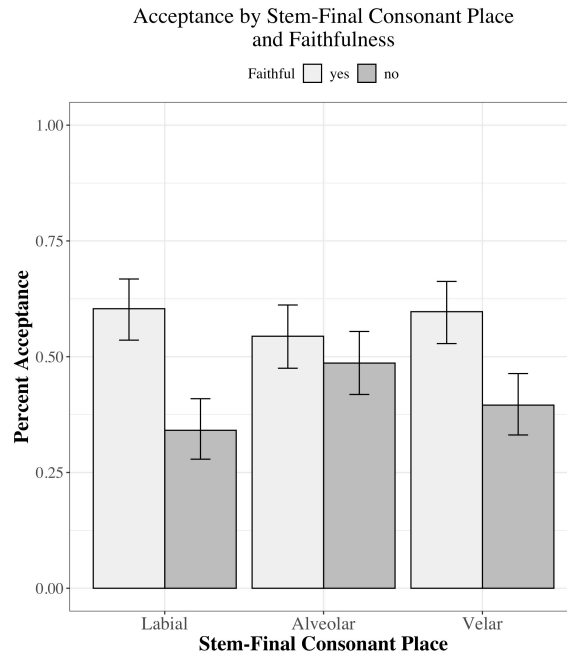
Figure 3.7. Acceptance rates of plurals by stem-final consonant place of articulation and whether the plural was faithful to the singular.

3.3. Discussion

The results suggest that the greater frequency and productivity of alveolar palatalization in English does translate into a (slight) preference in a test setting, with alveolar palatalization rated higher than labial and velar palatalization. Although palatalization before *-i* is much more common cross-linguistically (Bateman, 2007; Chen, 1973; Kochetov, 2011), participants show no preference for the alternation before *-i* vs. before *-a*. This could again be attributed to English patterns: Even though palatalization is triggered in common phrases like *would you* by the glide [j], which is acoustically and articulatorily similar to [i], in its palatalized form the "suffix vowel" is a schwa, so they may have learned to associate [tʃ] with non-high, non-front vowels.

There is no difference is ratings for palatalization of voiced vs. voiceless consonants in general, or velars in particular. This suggests that both the P-map (which suggests a preference for palatalizing voiceless velars; Guion, 1998; Steriade, 2001/2009) and orthographic overlap (which suggests a preference for palatalizing voiced velars, since [g] and [dʒ] can both be written with <g>) either have minimal effect on preference for voicing in palatalization, or the effects cancel each other out. Regardless, the results could indicate that we won't see any differences in learnability of palatalization by voicing in Experiments 2 and 3.

If [ti] is liked less than [ki] or [pi], the preference for alveolar palatalization could be explained as avoidance of a marked structure. However, there are no significant differences between ratings of faithful forms, which suggests that any influence of markedness on the likability and learnability of palatalization is minimal, since all of the faithful forms are considered roughly equivalent (and therefore, none would be especially improved by alternating).

Overall, palatalized forms are rated marginally lower than faithful forms. Given that larger changes tend to be disliked more than smaller (Kenstowicz, 1996; Steriade, 2001/2009), and that adding a vowel and changing a consonant is a larger change than merely adding a vowel, we can hypothesize why we don't see a stronger preference for non-palatalization here. Perhaps their familiarity with English palatalization patterns make participants more likely to accept other palatalization patterns, although first-language transfer may have a limited effect on miniature artificial language learning (Garcia et al., 2017; Mitrović, 2012; Wang & Saffran, 2014). Or perhaps judgment is

lenient (Kempen & Harbusch, 2005), so even if they wouldn't produce a form themselves, participants are still willing to deem it acceptable.

In summary, the baseline experiment shows that whatever differences participants have in the acceptability of palatalization by place of stem-final consonant, vowel, and voicing, they are relatively minor.

In Chapter IV, we compare the baseline to the judgment data after training in Experiment 2 in order to confirm the influence pre-existing biases have on the learning and production of palatalization.

CHAPTER IV

EXPERIMENT 2: PALATALIZATION BEFORE *-i*

Portions of this chapter were taken from:

Smolek, A. & Kapatsinski, V. (2018). What happens to large changes? Saltation produces well-liked outputs that are hard to generate. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *9*(1), 10.

In Experiment 2, we investigate the learnability of three miniature artificial languages containing either labial, alveolar, or velar voiced and voiceless palatalization before *-i*. Like in previous work on the influence of alternation magnitude on the learnability of alternations (Skoruppa & Peperkamp, 2011; White, 2013, 2014; White & Sundara, 2014), the languages differ in the degree of change to the base that they require. Here, the output plural forms always end with the palato-alveolar [tʃi]. Whereas alveolars and velars share gestures with [tʃ] (tongue tip and tongue body, respectively; Yun, 2006), labials do not, so articulatorily, labials require a larger change and palatalization then necessitates saltation over [t] or [k], depending on whether palatalization is reached through [coronal] or [dorsal].

4.1. Predictions and hypotheses

Our principal expectation is that labial palatalization, the saltatory alternation, will be more difficult to learn than alveolar or velar palatalization. In particular, we expect p~tʃ to be harder to produce after training than t~tʃ or k~tʃ, and that it will also be more likely to be overgeneralized to alveolars and velars than alveolar or velar palatalization will be to overgeneralize to labials. We test five explicit hypotheses:

Hypothesis 1: Labial palatalization is hard to learn because of markedness, not faithfulness.

Faithful judgments will not differ across conditions: p~tʃ is liked less than t~tʃ or k~tʃ, but p~pi, t~ti, and k~ki are equivalent. If k~ki and t~ti are worse than p~pi, that alone would make learning palatalization of those stops easier (Pater & Tessier, 2006).

Hypothesis 2: Large alternations, including saltation, are hard to produce (Skoruppa et al., 2011).

Large alternations are those where the alternants are dissimilar (articulatorily, perceptually, and/or featurally). Saltation, where there exists a segment that is more similar to the input than the output is, such as [t] between [p] and [tʃ] in labial palatalization, is one type of large change. This hypothesis proposes that labials will be palatalized less in the Labial Palatalization condition than will alveolars in the Alveolar Palatalization condition or velars in the Velar Palatalization condition. They will also be palatalized less in error in the Alveolar and Velar conditions than alveolars and velars will be (in the Labial and Velar, and Labial and Alveolar conditions, respectively).

Hypothesis 3: Saltatory alternations are likely to be overgeneralized (Hayes & White, 2015; Moreton & Pater, 2012a; White, 2013, 2014, 2017; White & Sundara, 2014).

[t] will be palatalized more after training on Labial Palatalization than Velar Palatalization, because Labial learners will attempt to acquire a simple conjunctively-defined category subsuming [p] and [tʃ] (e.g. [-continuant], which also includes [t] and

[k]), whereas a category subsuming [t] and [tʃ] (e.g. [coronal; -continuant]) would not include [k] and [p], and vice versa.

Hypothesis 4: Large changes are hard to produce, even if they are judged to be preferable.

The difference between the Labial condition and the others will be larger in the production test than in judgment; more specifically, labial palatalization is likely to be accepted in judgment while being rarely produced. The Perseveration Hypothesis proposes that production of an alternation can be difficult even when the product is judged acceptable because production involves overcoming paradigmatic perseveration, the default predisposition to copy activated segments of the source(s) into the production plan, whereas judgment does not. Overcoming paradigmatic perseveration is made easier by the acquisition of paradigmatic associations between segments participating in an alternation. Judgments need not rely on paradigmatic mappings because there is no need to overcome paradigmatic perseveration in judgment. A speaker who has not acquired paradigmatic associations necessary for overcoming paradigmatic perseveration and is therefore unable to produce an alternation could nonetheless judge the alternation to be more acceptable than the faithful mapping s/he would actually produce, by using product-oriented/first-order schemas (Bybee, 1985, 2001; Kapatsinski, 2012, 2013; Nesset, 2008); that is, by judging that the product contains the appropriate cues for the meaning expressed (Kapatsinski, 2013). The avoidance of labial palatalization should therefore be stronger in production than judgment. This mirrors the patterns found for Tagalog by Zuraw (2000), where speakers rarely produce nasal substitution, which requires a change

68

to the base, but still judge it better than nasal assimilation, which does not require a stem change. A dissociation between judgment and production would also address the critique that judgments are merely more tolerant than production (Kempen & Harbusch, 2005), since the less acceptable form is more likely to be produced. In other words, the speaker recognizes the correct form, but is reluctant or unable to produce it.

Hypothesis 5: The bias against labial palatalization is a bias in favor of perceptual *similarity between alternants.*

Perceptual explanations of alternation learnability (Hayes & White, 2015; Kenstowicz, 1996; Moreton & Pater, 2012a; Steriade, 2001/2009) propose that greater perceptual similarity between alternants corresponds with greater ease of acquisition. Previous work by Guion (1998) shows that [ki] and [tʃi] are more confusable than are [gi] and [dʒi], which implies that [k] will be palatalized less than [g] in all conditions.

Under the Perseveration Hypothesis, we expect support for H1, H2, and H4. Categorization based on featural simplicity (Moreton & Pater, 2012a; §1.6.4) would support H2 and H3. Perceptual explanations (Kenstowicz, 1996; Steriade, 2001/2009; White, 2013, 2014; §1.6.3) would support H1, H2, H3, and H5. We further hypothesize that, when comparing effects of alternation magnitude on the segments participants were trained to change as well as those they were trained not to change, we will see over-extension of palatalization in judgment (because judgment is more tolerant, Kempen & Harbusch, 2005; and because listeners may have acquired product-oriented schemas that

favor it, Kapatsinski, 2012, 2013) and underapplication leading to its eventual demise in

production (Skoruppa et al., 2011; Stave et al., 2013).

4.2. Methods

4.2.1. Languages

There were three training languages, consisting of either labial, alveolar, or velar

palatalization. The stems were always C(C)VC and ended in an oral stop [b;p;d;t;g;k].

The plural suffix was *-i* (100% of the time for To-Be-Palatalized consonants, 50% of the

time for Not-To-Be-Palatalized consonants) or *-a* (0% of the time for To-Be-Palatalized

consonants, 50% of the time for Not-To-Be-Palatalized consonants). The To-Be-

Palatalized consonant became [tʃ] if voiceless and [dʒ] if voiced; we included both voiced

and voiceless consonants to test for the influence of perceptual vs. articulatory similarity

on the learnability of alternations (Hypothesis 5). Labial palatalization was saltatory,

because palatals contain both [coronal] and [dorsal] features, which labials lack. Table

4.1 shows the patterns in each of the languages.

Table 4.1. Labial, Alveolar and Velar Palatalization patterns presented to participants in Experiment 2.

	Labial Palatalization	Alveolar Palatalization	Velar Palatalization
Singular	Plural	Plural	Plural
…p	**…tʃi**	…{pi;pa}	…{pi;pa}
…b	**…dʒi**	…{bi;ba}	…{bi;ba}
…t	…{ti;ta}	**…tʃi**	…{ti;ta}
…d	…{di;da}	**…dʒi**	…{di;da}
…k	…{ki;ka}	…{ki;ka}	**…tʃi**
…g	…{gi;ga}	…{gi;ga}	**…dʒi**

70

4.2.2. Participants

107 undergraduates in psychology or linguistics classes at the University of Oregon were recruited through the Human Subject Pool and received partial course credit for participation. 11 were excluded for producing plurals that did not correspond to patterns in the training[20]. After exclusions, there were 32 participants in the Alveolar Palatalization condition, 31 in Labial, and 33 in Velar. All participants were native English speakers with no speech, hearing, language, or learning disabilities.

4.2.3. Materials

4.2.3.1. Training

For each language, there were 28 unique singulars, randomly paired with images from the Spore creature database. Each creature was shown at least once alone, and at least once as part of a group (the same image copied multiple times), with 74 tokens total (frequencies are shown in Appendix B).

Images of the solo creatures were matched with singular form recordings, and the groups of creatures were paired with the corresponding plural form recordings (see Figure 3.1). The words were recorded by an adult male native American English speaker from Oregon. All pairings were in a random order, so corresponding singular-plural pairs were rarely adjacent. This trial order was chosen because random ordering encourages overgeneralization (which is necessary for evaluating Hypotheses 2 and 3), presumably by making it less obvious which singular-final consonants map onto the palatals [tʃ] and [dʒ] (see Chapter V).

---

[20] Including one memorable fellow who, when provided the singular [klip], produced [kliopætra].

12 out of 28 of the singulars ended in the To-Be-Palatalized consonants (half voiced, half voiceless), with the remaining 16 split evenly between the other places * voices. The complete stimulus lists are available in Appendix B.

All participants received an equal amount of training, unlike in White (2013, 2014), where participants were trained to a criterion level of accuracy: Training stopped once a constant level of accuracy was reached and low performers were excluded. We believe that training to criterion obscures the bias against large changes by ensuring that all participants learn the alternation equally well.

4.2.3.2. Production Test

An additional 92 pairs of names and pictures were created. As in the training, the singular picture was paired with a recording of the singular form, but in the test, it was immediately followed by the corresponding plural picture, which had no recording. Subjects were instructed to produce the appropriate plural for the given singular, which was recorded for later coding. 36 out of 92 of the trials ended with the To-Be-Palatalized consonants (half voiced, half voiceless), with the other 56 split evenly between the other places * voices. The complete stimulus lists are available in Appendix B.

4.2.3.3. Judgment Test

The judgment test followed production, since it exposed participants to forms that contradict the training. The stimuli were the same across all conditions, with 30 new singulars, divided equally between places * voices. Each singular had 4 possible plurals, crossing whether it was palatalized and which suffix vowel was added. The singular picture with the singular recording was followed by the plural picture with one of the

plural recordings, and all the pairs were randomly ordered. The complete stimulus list is available in Appendix A.

4.2.4. Procedure

4.2.4.1. Training

At the beginning of the experiment, subjects were informed that each word was either a singular or a plural and that they would be tested on remembering some of them. The recall test occurred after going through the complete list of training trials once; subjects were shown 14 tokens (7 pairs) from the training and were asked to produce the correct form for each picture. The recall test was included to ensure that participants were motivated to pay attention, and the results are not included in any analyses. Following the recall test, the training stimulus list was presented two more times.

For the training trials, each picture was shown for 500 ms, followed by the spoken wordform referring to it. The picture stayed on screen until the offset of the spoken word. There was a 500 ms blank screen pause between trials. The recall trials consisted of a picture on the screen, and participants had 6 seconds to say correct name (or terminate the trial by clicking mouse or hitting the space bar). There was a 1 second blank screen pause between recall trials.

The experiment was conducted on E-Prime 2.0 Professional (Psychology Software Tools, Pittsburgh, PA). Participants wore headphones and could hit the spacebar on a keyboard to advance to the next trial.

4.2.4.2. Production test

The participants were shown a novel singular form, paired with the picture of a novel solo creature, followed 300 ms later by the picture of the corresponding group of

creatures. They were instructed to produce what they thought the correct plural was and had 3 seconds to speak before the trial ended. Trials were separated by a 1 second blank screen pause. The pairs were randomly ordered for every participant. The spoken responses were recorded onto the computer using a Sennheiser HMD 281 headset in an isolated room, and each plural was saved as a separate file for later coding by me or undergraduate RAs. (All of the RA codings were verified for accuracy, and where there was a conflict of opinion, I relied on my judgment or looked at the spectrogram if it was unclear.)

We coded the identity of the stem-final consonant and word-final vowel. If the stem consonant was replaced by either palatal ([tʃ] or [dʒ], regardless of voicing of stem consonant), the plural was coded as palatalized. If the stem consonant was retained, it was coded as not palatalized. If the plural included both the stem consonant *and* a palatal, it was coded as not palatalized because the participant preserved the base (in other words, underapplied the change), but there were relatively few of these and the results are comparable with or without them, or if they are coded as palatalized. Rarely, participants replaced the stem consonant with another non-palatal consonant, which we excluded; any trials that included the English plural -*s* were also excluded, and any participants who produced a majority of such trials were excluded entirely (*n* = 5).

4.2.4.3. Judgment test

Like the production test, the judgment test consisted of novel singular-plural pairs. The singular form recording paired with the singular picture was followed by a 300 ms blank screen pause before the plural form recording played with the plural picture. The pairs were randomly ordered for every subject and were separated by a 1 second blank

74

screen pause. Subjects were instructed to indicate using a button box whether they thought the plural was the right one for the singular. The button box had 5 buttons, but the results were strongly bimodal: 59% of responses were 1 or 5, 29% were 2 or 4, and 12% were 3, so we transformed the scale into a binary dependent measure for easier comparison to the production test: 1s and 2s were coded as 0 and 4s and 5s as 1, with 3s excluded (since they seemed to indicate indecision). See Figure 4.1, below, for the distribution of ratings by training condition and final consonant place of articulation.

We considered using the binarized difference of ratings between the palatalized and non-palatalized plurals of the same singular with the same vowel (e.g. *bup~bupi* minus *bup-butʃi*), excluding trials that were rated equally. This would allow for a straightforward comparison to production (where we assume palatalization is produced if it is more acceptable than non-palatalization), but there were no effects of training on judgments of faithful plurals, so the results remain unchanged from the simpler absolute binary ratings: Since all faithful plurals are accepted equally often, the binarized difference measure reflects only differences in judgments of palatalized plurals, which can be captured just as well by absolute ratings, without the loss of trials where faithful and unfaithful forms were rated equally. We also preferred the absolute ratings because it allows us to examine the effects of condition on judgments of faithful and unfaithful plurals separately, which is necessary for evaluating Hypothesis 1. Prior studies (except for Stave et al., 2013) had participants choose between the faithful and unfaithful forms in production or forced-choice, which conflates preference for one form with dislike of another (e.g., if *blup* is the singular and *blupi* and *blutʃi* are the plurals, participants may choose *blutʃi* because they like it, or because they really dislike *blupi*).

75
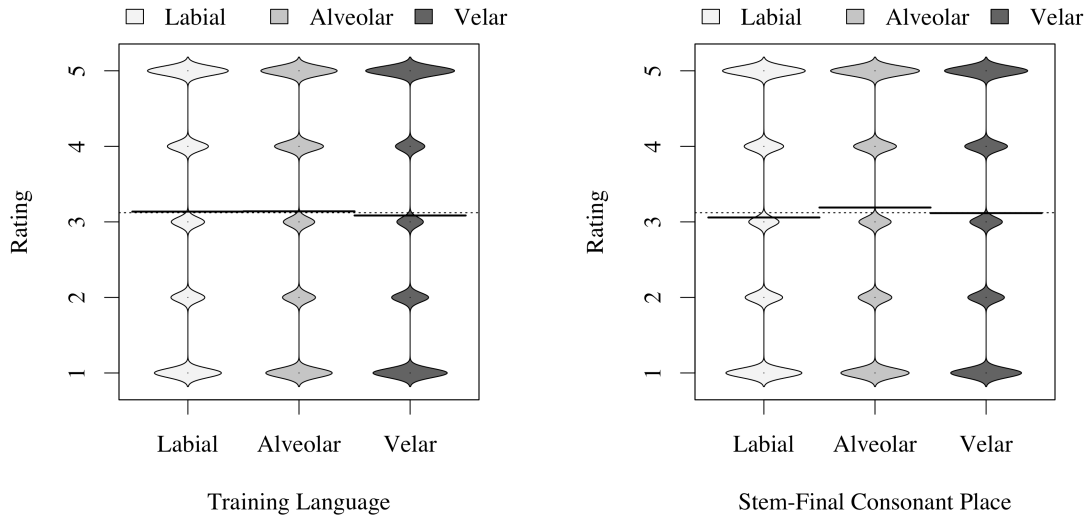
Figure 4.1. Distribution of ratings by training condition (left) and final consonant place of articulation (right). The black bars indicate the means by factor level; the dotted line indicates the overall mean.

4.2.5. Measures

We ran generalized logistic linear mixed-effects models with the lme4 package (version 1.1-21, Bates et al., 2015) in R (version 3.6.0, R Development Core Team, 2019). Fixed effects were included for Training Condition (Labial, Alveolar, and Velar, contrast coded as noted), Plural Vowel (-*i* vs. -*a*), Test Place (labial, alveolar, and velar), TBP (To-Be-Palatalized vs. Not-To-Be-Palatalized, given training condition), Test Voice (voiced vs. voiceless), Test Type (production vs. judgment), and any significant interactions. To evaluate the magnitude of improvement after training, Training (no [baseline data from Experiment 1) vs. yes [data from Experiment 2]) was included as a fixed effect. We included random intercepts for Subjects and Bases, with the full random effect structure that allowed the model to converge (selecting from Plural Vowel, Test Place, TBP, Test Voice, and Test Type within Subjects, and Training Condition, Plural Vowel, and TBP within Bases; random effect structures are included in the footnotes).

76

Log likelihood models on nested models were used to derive significance values. The BIC approximation to the Bayes Factor (Wagenmakers, 2007) was calculated when a contrast that was expected to be significant under a hypothesis was not significant in the model, as it allows us to directly test the degree of evidence supporting the null hypothesis. The tested models are in footnotes, and the full dataset and code are available at https://app.box.com/s/bd8jhx4g5m7bvlmxb8i4jgtjfjo2x111.

4.3. Results

4.3.1. Hypothesis 1: Labial palatalization is hard to learn because of faithfulness, not markedness

It is possible that [t] and [k] are easier to change to [tʃ] before *-i* because [ti] and [ki] are worse (i.e. more marked) than [pi]. Judgments of faithful mappings are informative here; if the bias exists, then p~pi would be rated better than t~ti and k~ki (in the Labial, Alveolar, and Velar Palatalization conditions, respectively). But there is no such pattern[21] (Table 4.2), and in fact there is a slight non-significant trend in the unexpected direction for the Not-To-Be-Palatalized consonants (Table 4.3)[22]. These results are shown in Figure 4.2. The evidence provides very strong evidence for the null (ΔBIC = 14.1, P_{BIC} ($H_0 \mid D$) = 0.999) and are contrary to the markedness explanation: The bias is against *changes*, not against certain output structures.

---

[21] RatingBin ~ Training Condition + Test Voice + Plural Vowel + (1 + Test Voice + Plural Vowel | Subject) + (1 + Plural Vowel | Base), Helmert contrast coded for Labial vs. Alveolar and Velar training and Alveolar vs. Velar training, restricted to ratings of incorrect faithful plurals.

[22] RatingBin ~ Training Condition + Test Voice + (1 + Test Voice | Subject) + (1 + Training Condition | Base), Helmert contrast coded for Labial vs. Alveolar and Velar training and Alveolar vs. Velar training, restricted to ratings of correct faithful plurals.

Table 4.2. Judgments of incorrect faithful mappings for To-Be-Palatalized consonants across training conditions. The inclusion of Training Condition does not significantly improve the fit of the model, $\chi^2(2) = 0.45$, $p = 0.80$, *ns*.

	b	se(b)	z	p	
(Intercept)	-0.48295	0.22952	-2.104	0.0354	*
Labial vs. Alveolar and Velar Training	-0.08744	0.36176	-0.242	0.809	
Alveolar vs. Velar Training	-0.27456	0.42757	-0.642	0.5208	
Voiceless	-0.05531	0.15829	-0.349	0.7268	
-a	0.35555	0.20704	1.717	0.0859	.

. Marginally significant
* Significance level of 0.05

Table 4.3. Judgments of correct faithful mappings for Not-To-Be-Palatalized consonants across training conditions; Training Condition does not significantly improve the fit of the model, $\chi^2(2) = 0.098$, $p = 0.95$, *ns*.

	b	se(b)	z	p	
(Intercept)	0.29886	0.13278	2.251	0.0244	*
Labial vs. Alveolar and Velar Training	-0.07223	0.24345	-0.297	0.7667	
Alveolar vs. Velar Training	-0.03532	0.28153	-0.126	0.9002	
Voiceless	0.1244	0.10086	1.233	0.2174	

* Significance level of 0.05
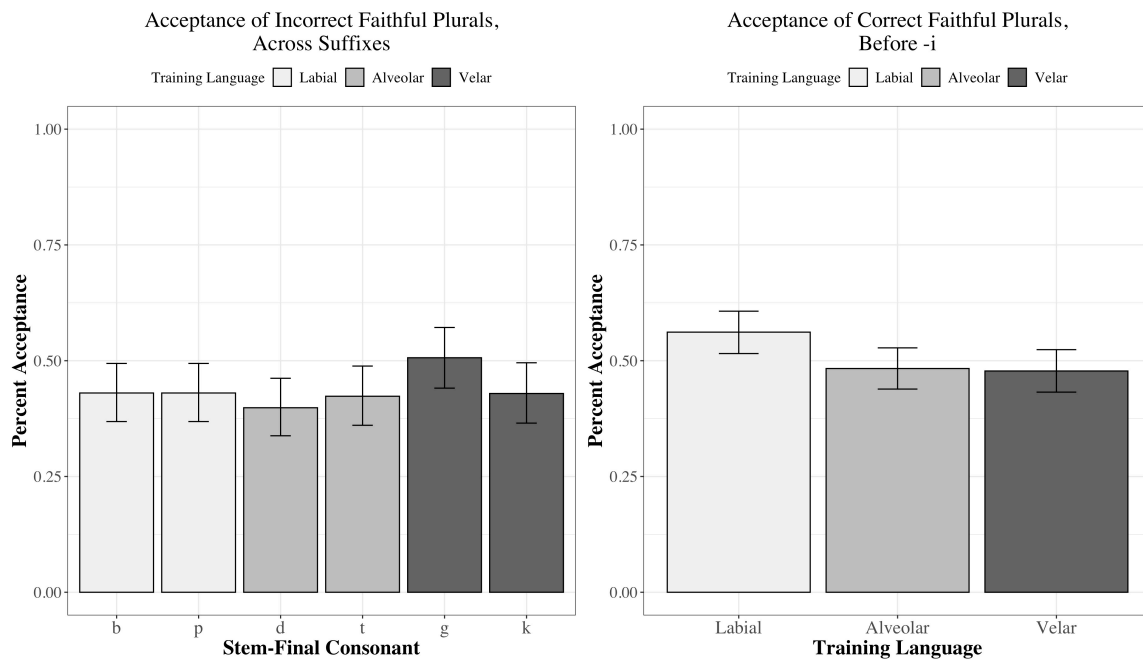
Figure 4.2. Judgments of faithful plurals across conditions. Left: judgments of incorrect faithful To-Be-Palatalized stems. Right: judgments of correct faithful Not-To-Be-Palatalized stems.

4.3.2. Hypothesis 2: Large alternations, including saltation, are hard to produce

Following Hypothesis 2, To-Be-Palatalized consonants should be palatalized less often when they require a larger change, which here is in the Labial Palatalization condition. There are large differences between the Labial and the lingual (Alveolar/Velar) conditions, as expected: To-Be-Palatalized consonants are palatalized significantly less often after Labial Palatalization training than after Alveolar and Velar Palatalization training[23] (Table 4.4, Figure 4.3). There is no effect of or interaction with voicing, and the inclusion of Training Condition significantly improves model fit, $\chi^2(2) = 29.47$, $p < 0.001$. Participants in the Velar Palatalization condition learn to palatalize velars, and those in the Alveolar Palatalization condition learn to palatalize alveolars, but there is no difference in palatalization rates of labials vs. the linguals for participants trained on Labial Palatalization (and in fact, they actually palatalize the To-Be-Palatalized consonants slightly *less* than the Not-To-Be-Palatalized consonants after Labial Palatalization training). That the larger labial palatalization change is applied to the to-be-changed segments less often compared to smaller changes mirrors the results of Skoruppa et al. (2011) and is contra White (2013).

---

[23] Keep Place ~ Training Condition + Test Voice + (1 + Test Voice | Subject) + (1 | Base); Helmert coded for Labial vs. Alveolar and Velar and Alveolar vs. Velar Palatalization training conditions, restricted to productions of To-Be-Palatalized consonants before -*i*.

Rates of Correct vs. Incorrect Palatalization,
Before -i

Palatalization Expected ☐ yes ☐ no
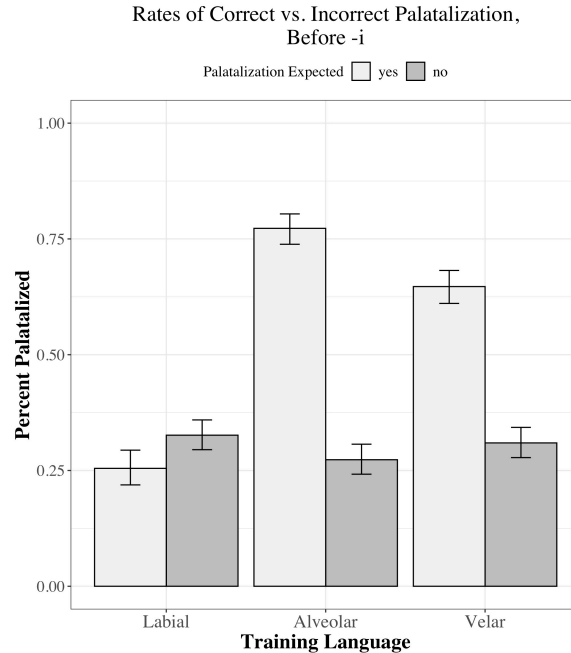
Figure 4.3. Palatalization rates before *-i* in production, by training condition. The bars represent the rate of palatalization in production of To-Be-Palatalized (light) and Not-To-Be-Palatalized (dark) consonants, grouped by training language, which determines (Not-)To-Be-Palatalized status of consonants.

Labials should also be *erroneously* palatalized less often than alveolars and velars, which is true and can be seen in Figure 4.4, where overgeneralization of palatalization to labials is shown by the dark bars and overgeneralization to linguals by the light bars. Table 4.5 shows that training on Labial Palatalization overgeneralizes more to alveolar-final stems than Alveolar Palatalization training overgeneralizes to labial-final stems[24]. Test Place significantly improves the fit of the model, $\chi^2(1) = 13.13$, $p < 0.001$. Table 4.6 shows that velars are palatalized more after Labial Palatalization training than are labials after Velar Palatalization training[25], with Test Place significantly improving the model fit, $\chi^2(1) = 7.13$, $p = 0.008$. In other words, labial palatalization is less likely to be produced

---

[24] Keep Place ~ Test Voice + Test Place + (1 + Test Voice | Subject) + (1 | Base), restricted to Alveolar training palatalization of labials and Labial training palatalization of alveolars.

[25] Keep Place ~ Test Voice + Test Place + (1 + Test Voice | Subject) + (1 | Base), restricted to Velar training palatalization of labials and Labial training palatalization of velars.

80

than alveolar or velar palatalization, whether or not participants are exposed to it in

training.

Table 4.4. The effect of Training Language on (erroneous) retention rates of To-Be-Palatalized consonants in production, before -*i*. Negative regression coefficients indicate higher rates of palatalization (less retention of the base consonant), which in this case means higher accuracy.

	b	se(b)	z	p	
(Intercept)	0.01121	0.36582	0.031	.976	
Labial vs. Alveolar and Velar Training	4.17856	0.80836	5.169	<.00001	***
Alveolar vs. Velar Training	-1.20023	0.86301	-1.391	.164[a]	
Voiceless	0.12622	0.26445	0.477	.633	

*** Significance level of 0.001
[a] According to the BIC approximation to the Bayes Factor, the results provide positive evidence for the null (ΔBIC = 3.8, P_{BIC} (H_0 | D) = 0.87).
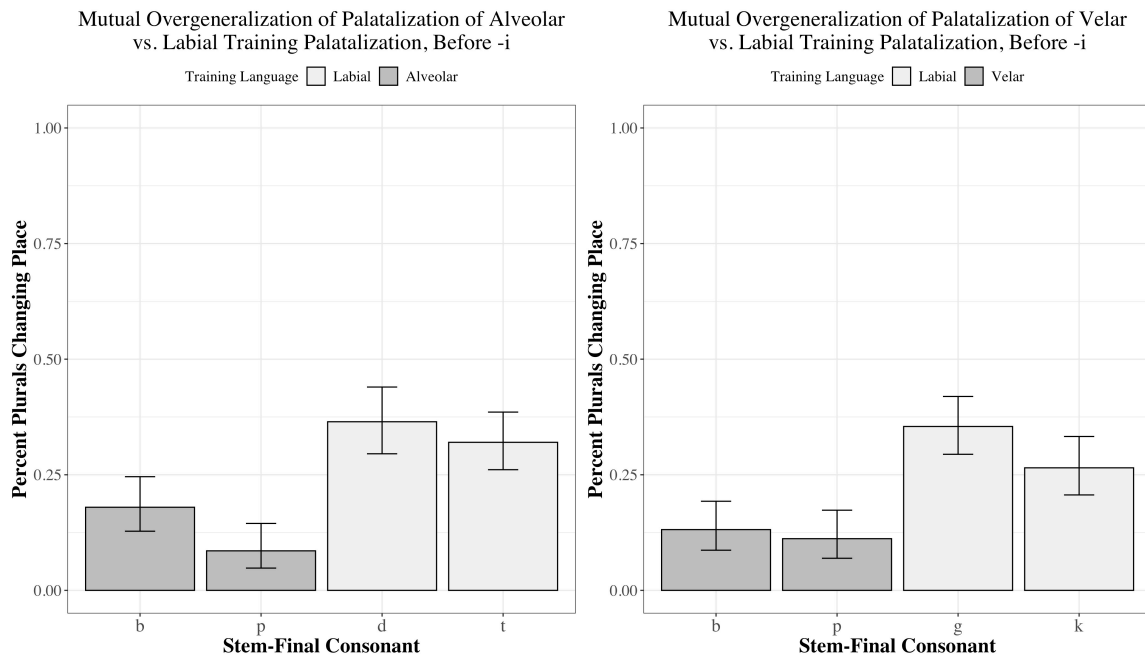
Figure 4.4. Palatalization rates of Not-To-Be-Palatalized consonants by stem-final consonant place of articulation and training condition. Left panel: Overgeneralization of alveolar palatalization to labials (light) vs. overgeneralization of labial palatalization to alveolars (dark). Right panel: Overgeneralization of velar palatalization to labials (light) vs. overgeneralization of labial palatalization to velars (dark).

Table 4.5. Overgeneralization of palatalization from alveolars to labials and labials to alveolars.

	b	se(b)	z	p	
(Intercept)	4.2555	0.9343	4.555	<.00001	***
Voiceless	0.7503	0.7457	1.006	.31432	
Alveolar Stem	-3.3044	1.0073	-3.28	.00104	**

** Significance level of 0.01
*** Significance level of 0.001

Table 4.6. Overgeneralization of palatalization from velars to labials and labials to velars.

	b	se(b)	z	p	
(Intercept)	3.8124	0.784	4.863	<.00001	***
Voiceless	0.9275	0.7749	1.197	.23131	
Velar Stem	-2.3129	0.8975	-2.577	.00996	**

** Significance level of 0.01
*** Significance level of 0.001

4.3.3. Hypothesis 3: Saltatory alternations are likely to be overgeneralized

According to Hypothesis 3, alveolars and velars should be palatalized more by participants in the Labial Palatalization condition than by participants in the Velar and Alveolar Palatalization conditions, respectively. Since palatals are [coronal] and [dorsal] (Yun, 2006), labials ([labial]) need to jump over [coronal] and or [dorsal], whereas alveolars ([coronal]) and velars ([dorsal]) have a direct route without intermediate segments. In other words, alveolar palatalization doesn't need to imply velar palatalization, and vice versa.

In production, participants in the Labial and Velar Palatalization conditions palatalize alveolar-final stems equally frequently[26] (Figure 4.5, left panel; $b = -0.46$, $se(b) = 0.68$, $z = -0.67$, $p = 0.50$), as do Labial and Alveolar Palatalization participants for velar-final

---

[26] Keep Place ~ Training Condition + Test Voice + Plural Vowel + (1 + Test Voice + Plural Vowel | Subject) + (1 + Training Condition + Plural Vowel | Base), restricted to Velar and Labial training palatalization of alveolars.

stem palatalization[27] (Figure 4.5, right panel; $b = -0.46$, *se(b)* $= 0.67$, $z = -0.69$, $p = 0.49$). The BIC approximation to the Bayes Factor provides strong support for the null hypothesis in both cases ($\Delta BIC = 6.9$, P_{BIC} ($H_0 \mid D$) $= 0.97$ and $\Delta BIC = 7$, P_{BIC} ($H_0 \mid D$) $= 0.97$, respectively). The study is sufficiently powerful to provide positive evidence in favor of the hypothesis that a saltatory change is no more likely to overgeneralize than a non-saltatory change (contra White, 2013, 2014).
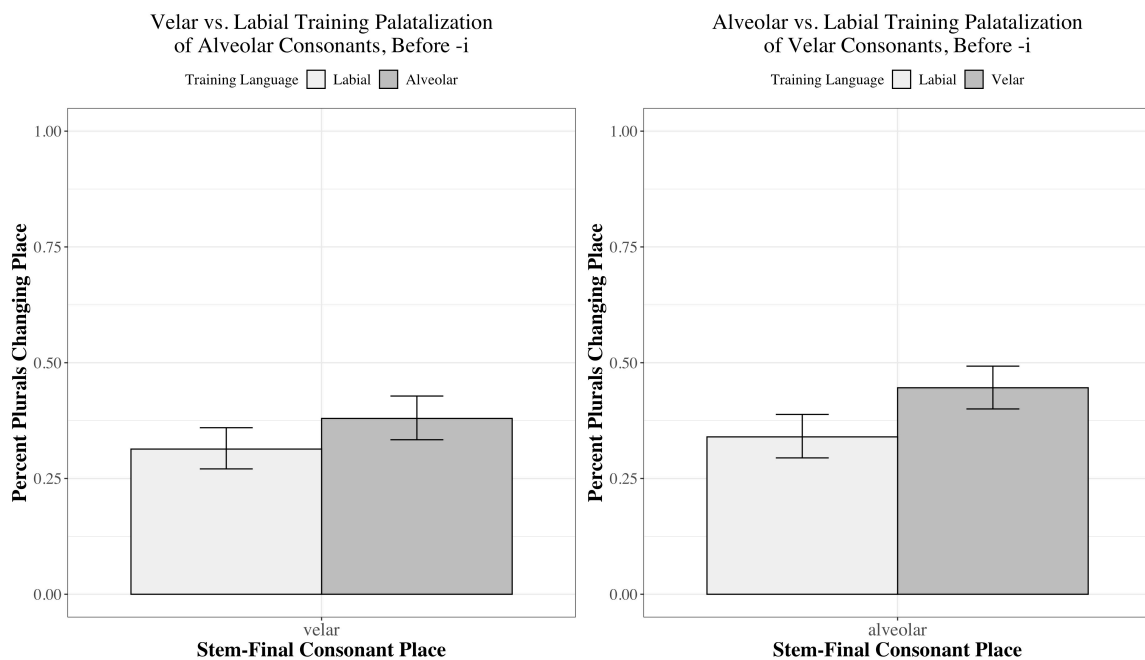
Figure 4.5. Overgeneralization of palatalization depending on magnitude. Overgeneralization of labial palatalization is shown by the light bars, and overgeneralization of lingual palatalization by the dark bars.

In judgment, there is no difference in judgments of palatalized velars for participants in the Labial and Alveolar Palatalization conditions (Figure 4.6), both across vowel

---

[27] Keep Place ~ Training Condition + Test Voice + Plural Vowel + (1 + Test Voice + Plural Vowel | Subject) + (1 + Training Condition + Plural Vowel | Base), restricted to Alveolar and Labial training palatalization of velars.

83

contexts[28] (b = -0.02, $se(b)$ = 0.31, z = -0.70, p = 0.94, ns) and before -*i*, the palatalizing

suffix[29] (b = -0.40, $se(b)$ = 0.44, z = -0.91, p = 0.36, ns). The data provide strong evidence

for the null across suffixes (ΔBIC = 6.8, P_{BIC} (H_0 | D) = 0.97) and positive evidence

before -*i* (ΔBIC = 5.4, P_{BIC} (H_0 | D) = 0.94).
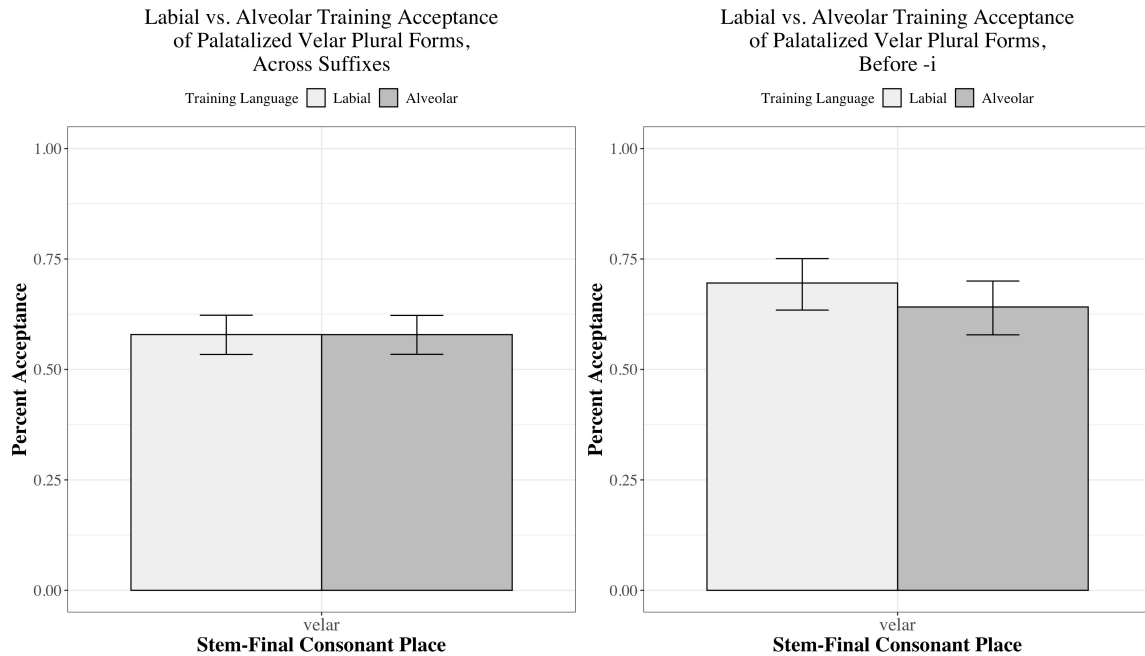
Figure 4.6. Acceptance of overgeneralization of palatalization to velars.

There is also no difference in the rate of acceptance for palatalized alveolars after

Velar and Labial Palatalization training (Figure 4.7) across vowel suffixes[30] (b = -0.63,

$se(b)$ = 0.39, z = -1.59, p = 0.11, ns), and according to the BIC approximation to the

Bayes Factor, the results provide positive evidence for the null hypothesis (ΔBIC = 4.4,

---

[28] RatingBin ~ Training Condition + Test Voice + Plural Vowel + (1 + Test Voice + Plural Vowel |
Subject) + (1 + Training Condition | Base), restricted to Labial and Alveolar Palatalization training ratings
of palatalized velars.

[29] RatingBin ~ Training Condition + Test Voice + (1 + Test Voice | Subject) + (1 + Training Condition |
Base), restricted to Labial and Alveolar Palatalization training ratings of palatalized velars before –*i*.

[30] RatingBin ~ Training Condition + Test Voice + Plural Vowel + (1 + Test Voice | Subject) + (1 +
Training Condition + Plural Vowel | Base), restricted to Labial and Velar Palatalization training ratings of
palatalized alveolars.

P_{BIC} ($H_0 \mid D$) = 0.90). However, before -*i*, Velar Palatalization participants give marginally lower ratings of palatalized alveolars than Labial Palatalization participants[31] (*b* = -0.89, *se(b)* = 0.47, *z* = -1.86, *p* = 0.06; Trained Place marginally improves model fit, $\chi^2(1)$ = 3.40, *p* = 0.06, though according to the BIC approximation to the Bayes factor, the results still provide positive evidence for the null hypothesis, ΔBIC = 2.78, P_{BIC} ($H_0 \mid$ D) = 0.80).
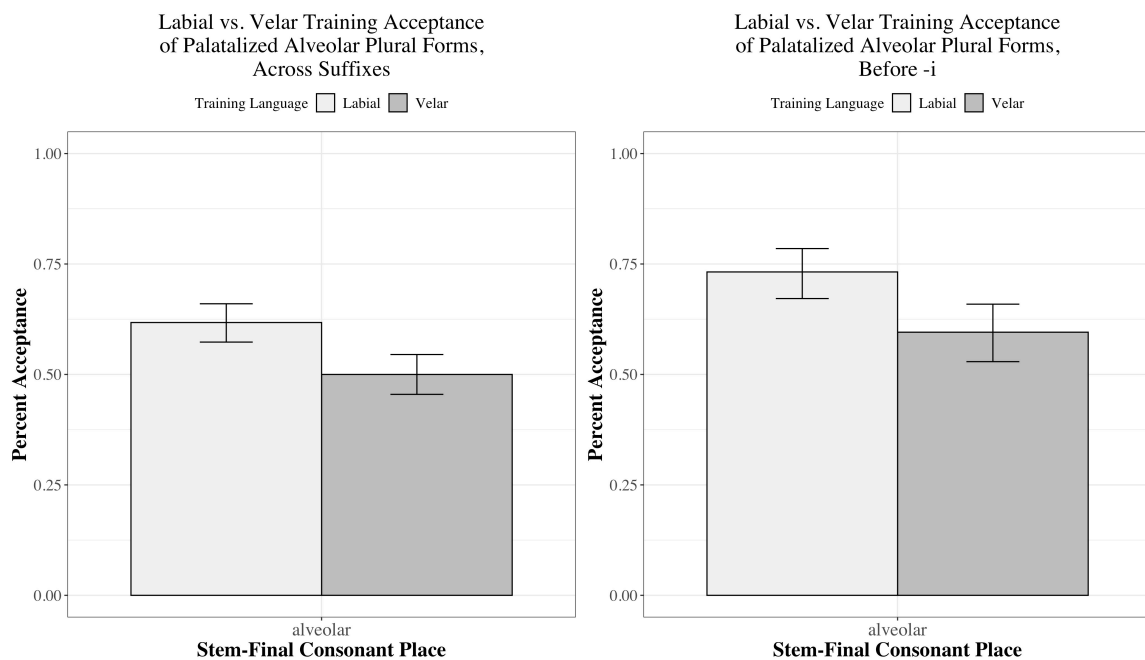
Figure 4.7. Acceptance of overgeneralization of palatalization to alveolars, across suffixes (left panel) and before -*i* (right panel).

There is little evidence for saltation overgeneralizing, even in judgment, contrary to Hypothesis 3, and what differences there are are quite small, compared to the very large differences in production (Figure 4.3, page 80).

4.3.4. Hypothesis 4: Large changes are hard to produce, even if they are judged to be preferable

---

[31] RatingBin ~ Training Condition + Test Voice + (1 + Test Voice | Subject) + (1 + Training Condition | Base), restricted to Labial and Velar Palatalization training ratings of palatalized alveolars before –*i*.

The bias against labial palatalization is expected to be stronger in production than judgment. Even if judgment is based on a model of production, participants are provided with a product form in the judgment task, minimizing differences in activation between forms that are associated with the source form to different degrees (see Harmon & Kapatsinski, 2017; Luce & Pisoni, 1998, for the same difference arising in open-set vs. closed-set tasks). Product-oriented schemas may also be sufficient to judge the product of palatalization to be more acceptable than an unpalatalized form even when they are not strong enough to overcome paradigmatic perseveration in production (Kapatsinski, 2013); if participants learn that plurals often end in [tʃi], they may consider all plurals ending in [tʃi] well-formed, regardless of the input segment or whether they would produce it themselves. Judgments of faithful plurals are the same across conditions (Figure 4.2), so any differences in judgment between conditions must be driven by ratings of palatalized forms (Figure 4.8,).

In Table 4.7, the dependent variable was coded as 0 if the trial was non-palatalized in production and the non-palatalized form had a rating under 3 in judgment, and 1 if the trial was palatalized in production and the palatalized form was rated over 3 in judgment. We can see the dissociation between production and judgment after Labial Palatalization training (left bars of Figure 4.8 and 4.2; see also Figure 4.10): Labial palatalization is accepted, but rarely produced. The interaction between Labial Training and Test is significant in Table 4.7[32], and it significantly improves model fit ($\chi^2(1) = 21.94$, $p < 0.001$). This interaction is not present for Velar or Alveolar Palatalization training, where subjects produce palatalization as often as they accept it.

---

[32] Dependent Variable ~ Test Voice + Labial Training * Test + (1 + Test Voice + Test | Subject) + (1 | Base), restricted to To-Be-Palatalized trials in production and judgments of palatalized forms, before –i.
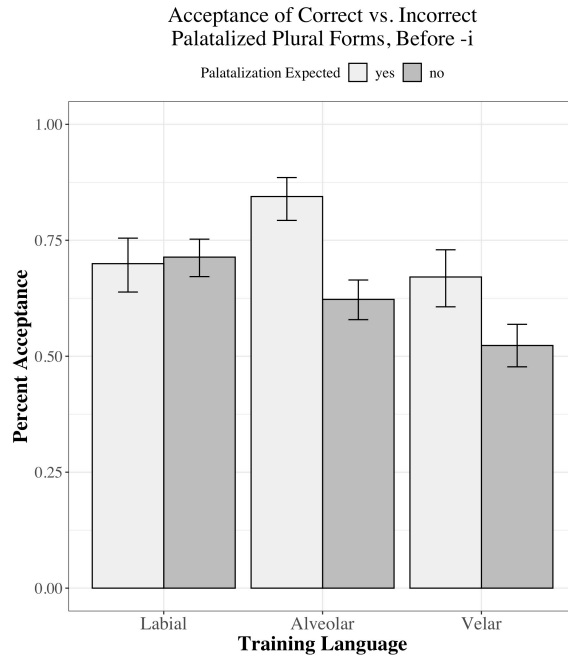
Figure 4.8. Acceptance of palatalization in judgment; take special notice of the high rate of acceptance of labial palatalization.

Table 4.7. The effects of training on Labial vs. Alveolar and Velar Palatalization on judgment vs. production of palatalized forms before -*i*.

	b	se(b)	z	p	
(Intercept)	-1.3751	0.4113	-3.344	0.000827	***
Voiceless	0.218	0.2077	1.049	0.294075	
Labial Training	4.0657	0.7809	5.207	1.92E-07	***
Judgment Test	-0.3777	0.4064	-0.929	0.352711	
Labial Training x Judgment Test	-3.6816	0.786	-4.684	2.81E-06	***

*** Significance level of 0.001

It is not just that judgments are more lenient than production (Kempen & Harbusch, 2005). Figure 4.9 shows that judgments of faithful p~pi (left bars) are lower for Labial Palatalization participants than palatalized p~tʃi (right bars), and Keep Place significantly improves model fit, both before -*i*[33] ($\chi^2(1) = 11.86$, $p < 0.001$) and across suffix vowels[34]

---

[33] RatingBin ~ Test Voice + Keep Place + (1 + Keep Place + Test Voice | Subject) + (1 + Keep Place | Base), restricted to Labial Palatalization training ratings of To-Be-Palatalized labial stems before –*i*.

[34] RatingBin ~ Test Voice + Keep Place + Plural Vowel + (1 + Keep Place + Test Voice + Plural Vowel | Subject) + (1 + Keep Place | Singular), restricted to Labial Palatalization training ratings of To-Be-Palatalized labial stems.

$(\chi^2(1) = 4.53, p = 0.03)$. Palatalization is preferred to non-palatalization in the judgment task after training, but it nonetheless fails to be produced. In other words, the mapping that is preferred in production is dispreferred in judgment.
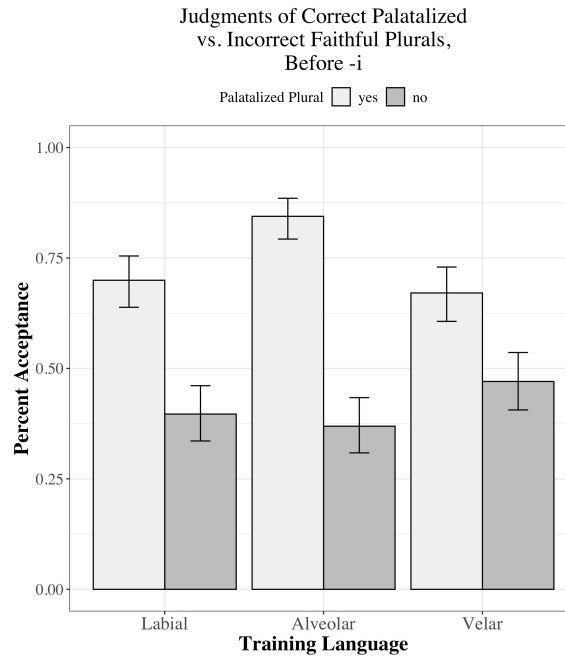
Figure 4.9. Judgments of To-Be-Palatalized plurals by faithfulness, before *-i*.

These effects are reflected in the 3-way interaction in Table 4.8, shown in Figure 4.10[35]. After Alveolar and Velar Palatalization training, palatalization of To-Be-Palatalized consonants is produced about as often as it is accepted (left panel), but palatalization of Not-To-Be-Palatalized consonants is still accepted more than half the time, where it is seldom produced (right panel). Additionally, To-Be-Palatalized velars are palatalized more than To-Be-Palatalized labials, but correct palatalization of velars is accepted as much as correct palatalization of labials.

---

[35] Dependent Variable ~ Test Voice + Labial Training * Test * To-Be-Palatalized Place + Plural Vowel + (1 + Test * To-Be-Palatalized Place | Subject) + (1 + Labial Training + To-Be-Palatalized Place | Base).
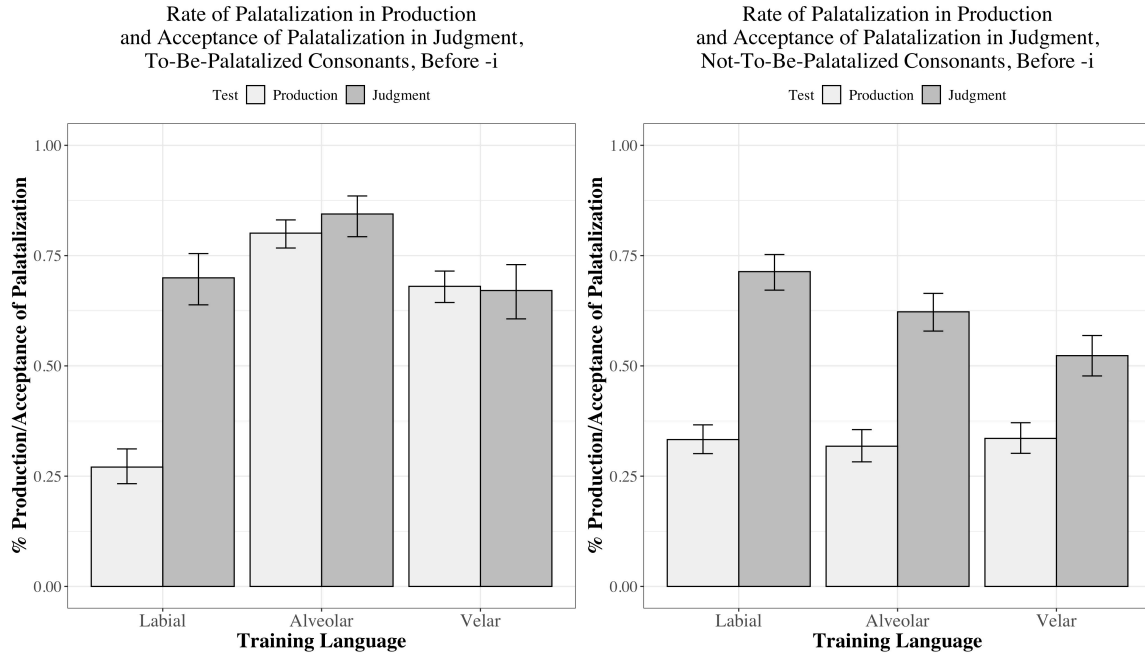
Figure 4.10. Comparison of the rate of palatalization in production (light bars) to the rate of acceptance of palatalized plurals in judgment (dark bars), by training language. Left panel: To-Be-Palatalized stems, before -*i*. Right panel: Not-To-Be-Palatalized stems, before -*i*.

Table 4.8. The effects of training on Labial vs. Lingual palatalization on correct vs. erroneous palatalization and judgments of correct vs. erroneous palatalization. Bolded rows show effects of interest. Palatalization is accepted more often than produced but especially so after labial training. Labial training reduces or reverses the difference in palatalization rates and judgments of palatalization between To-Be-Palatalized and Not-To-Be-Palatalized consonants. However, this reduction is smaller in judgments than in production.

	b	*se(b)*	*z*	*p*	
(Intercept)	0.16494	0.24954	0.661	0.509	
Voiceless	0.06059	0.0944	0.642	0.521	
Labial Training	3.03852	0.58809	5.167	2.38E-07	***
Rating Test	-1.74	0.23899	-7.281	3.33E-13	***
Not-To-Be-Palatalized Place	1.85559	0.25219	7.358	1.87E-13	***
-a	2.23907	0.12595	17.778	2.00E-16	***
Labial Training x Rating Test	**-2.81071**	**0.59131**	**-4.753**	**2.00E-06**	***
LabialTraining x Not-To-Be-Palatalized Place	**-3.38016**	**0.42421**	**-7.968**	**1.61E-15**	***
Not-To-Be-Palatalized Place x Rating Test	-1.44839	0.2256	-6.42	1.36E-10	***
Labial Training x Rating Test x Not-To-Be-Palatalized Place	**3.4364**	**0.4113**	**8.355**	**2.00E-16**	***

*** Significance level of 0.001

89

The production data suggests that Labial Palatalization participants fail to learn the paradigmatic association between labials and alveopalatals, and are therefore unable to produce labial palatalization, but the judgment data shows that they still learn that palatalization is better than lack thereof. Prior work has shown that participants in similar experiments acquire first-order schemas (like "plurals end in [tʃi]"; Kapatsinski, 2012, 2013), which can explain the preference for palatalized over faithful forms, even after Labial training. In fact, the first-order schema might even be stronger after Labial training. Because they have failed to acquire an association between labials and alveopalatals, every time [tʃi] occurs it is surprising, which may help notice that plurals often end in [tʃi], strengthening the first-order schema. By definition, a first-order schema applies to all inputs equally, and this is all that participants in the Labial condition can rely on because they have failed to learn what should be changing (i.e., to acquire the paradigmatic mapping p→tʃi). Alveolar and Velar Palatalization participants do acquire the relevant paradigmatic mappings, as evidenced in their higher rates of production of palatalization for the To-Be-Palatalized consonants (see Figure 4.3), and they can use those to drive their judgments, which accounts for the higher ratings of correct palatalization over incorrect in those conditions (see Figure 4.8). However, there is still evidence that they acquire first-order schemas, as well: Faithful plurals of Not-To-Be-Palatalized stems are rated lower than palatalized plurals before $-i$[36] ($b = -0.59$, $se(b) = 0.21$, $z = -2.85$, $p = 0.004$; Keep Place significantly improves model fit, $\chi^2(1) = 7.61$, $p < 0.006$), even after Alveolar and Velar Palatalization training, as shown in Figure 4.11. Labial Palatalization training results in higher ratings for Not-To-Be-Palatalized stems,

---

[36] Rating Bin ~ Training Language + Keep Place + Test Voice + (1 + Keep Place + Test Voice | Subject) + (1 | Base), restricted to judgments of Not-To-Be-Palatalized plurals before $-i$.

overall ($b = 0.68$, $se(b) = 0.29$, $z = 2.37$, $p < 0.02$), and the inclusion of Training

Language significantly improves model fit ($\chi^2(2) = 6.89$, $p = 0.03$), but all training

languages result in preference for the unfaithful plural over the faithful. The lingual

conditions participants' preference for palatalizing the correct consonants over the

incorrect consonants can be attributed to paradigmatic mappings, but their preference for

palatalizing the incorrect consonants over not palatalizing them must be due to product-

oriented schemas favoring plurals ending in [tʃi] and [dʒi]: Changes that are rarely

produced are still accepted, presumably because the resulting structure has become

associated with the intended meaning (Kapatsinski, 2013).

Rate of Acceptance of Not-To-Be-Palatalized
Plurals by Faithfulness, Before -i
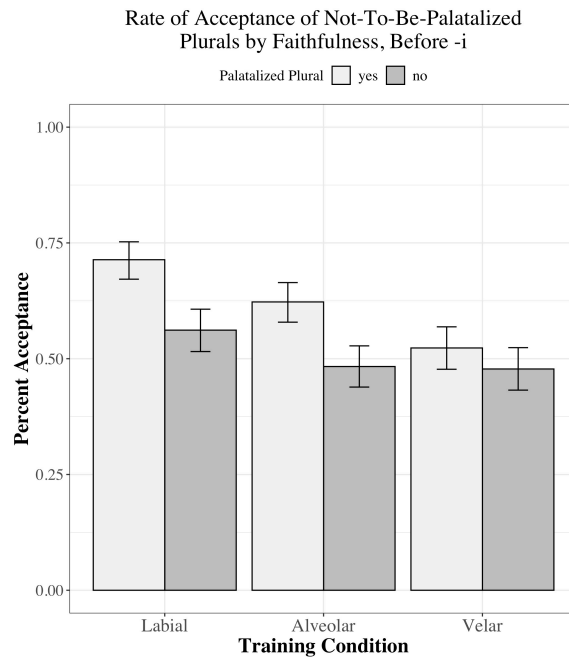
Figure 4.11. Judgments of Not-To-Be-Palatalized plurals by faithfulness, before -*i*.

4.3.5. Hypothesis 5: The bias against labial palatalization is due to perceptual dissimilarity

Hypothesis 5 claims that the bias against labial palatalization is perceptual, because

[pi] and [tʃi] are acoustically very dissimilar. By extension, [k] should be palatalized

more than [g], since [ki] is more confusable with [tʃi] than [gi] is with [dʒi] (Guion, 1998).

Comparing palatalization rates of voiced and voiceless velars by training condition (Figure 4.11, left panel) reveals that [k] is significantly less likely to be palatalized than [g][37] ($b = 0.77$, $se(b) = 0.28$, $z = 2.75$, $p = 0.006$; Test Voice significantly improves model fit, $\chi^2(1) = 7.43$, $p = 0.006$). In judgment (Figure 4.12, right panel), palatalized plurals of [k]-final stems are liked marginally less than palatalized plurals of [g]-final stems[38] ($b = -0.59$, $se(b) = 0.33$, $z = -1.80$, $p = 0.07$; Test Voice marginally improves model fit, $\chi^2(1) = 3.31$, $p < 0.07$).
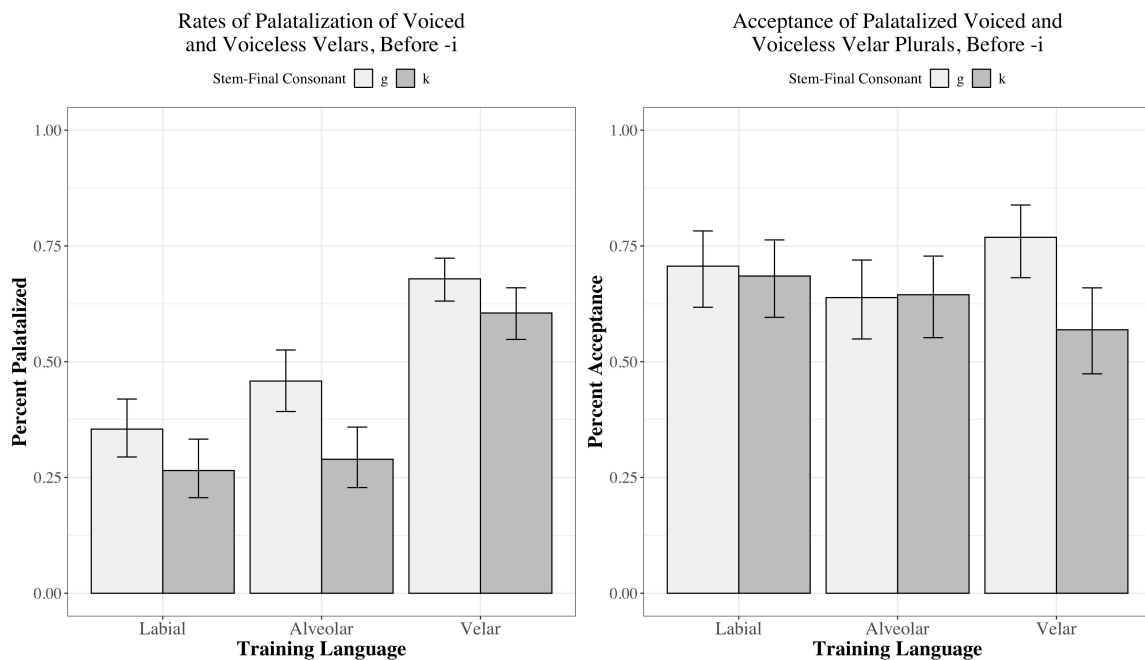
Figure 4.12. Rates of palatalization in production (left panel) and acceptance of palatalization in judgment (right panel) of velars before *-i* by voicing and training condition.

---

[37] Keep Place ~ Training Condition + Test Voice + (1 + Test Voice | Subject) + (1 | Base), restricted to productions of velar-final stems before –i.

[38] RatingBin ~ Training Condition + Test Voice + (1 + Test Voice | Subject) + (1 + Training Condition | Singular), restricted to ratings of palatalized velar-final stems before –i.

Combined, the results show that, contrary to the perceptual similarity hypothesis, the less similar alternants g~dʒ are preferable to the more similar k~tʃ. This is likely due to categorization, since in English orthography, [g] and [dʒ] can both be written with <g> (Gontijo et al., 2013), whereas [k] and [tʃ] have minimal orthographic overlap. This result mirrors Wilson's (2006) findings. We would expect that languages with different phonology-orthography mappings would result in different learning patterns. For example, in Italian, [g] and [dʒ] can both be written as <g>, and [k] and [tʃ] can both be written as <ch> and <c> (Proudfoot & Cardo, 2005). To the extent that orthographic overlap determines alternation learnability (aside from other factors like frequency), Italian orthography favors equivalent learning of k~tʃ and g~dʒ. Turkish orthography writes [k], [g], [tʃ], and [dʒ] with distinct characters (<k>,<g>,<ç>, and <c>, respectively; Underhill, 1976). We would still expect k~tʃ and g~dʒ to be learned equally well, but because Turkish orthography does not group the stop and corresponding alveopalatal together, Turkish speaker-writers should show overall lower rates of alternation acquisition compared to Italian. Future work is needed to evaluate this hypothesis. For the time being, our results suggest that whatever effect perceptual similarity may have on the learnability of alternations, it is insufficient to overcome other categorization effects. By extension, the perceptual dissimilarity between [p] and [tʃi] should have minimal effect on the learnability of the p~tʃi alternation, for which other explanations must be explored – such as the proposed articulatory effects.

4.3.6. Effect of training

In order to ensure that the patterns found are due to learning and not pre-existing biases, we compare the judgments from Experiment 1 (Chapter III) to the judgment data

from Experiment 2. In particular, we compare the acceptance of palatalization (in the judgment task) of To-Be-Palatalized plurals before -*i* after training to the acceptance of palatalization in the baseline experiment, with Training as the new factor. The dependent variable was coded as 0 if the palatalized form had a rating under 3 and 1 if the palatalized form was rated over 3 in judgment.

For the judgment test, subjects who received training judge palatalized plurals as being better than subjects without training[39] ($b = -1.97$, *se(b)* $= 0.32$, $z = -6.14$, $p < 0.001$), and the inclusion of Training significantly improves model fit ($\chi^2(1) = 29.33$, $p < 0.001$). This can be seen in the comparison of the light bars (no training) and dark bars (judgment test) in Figure 4.13, where the dark bars are all higher than the corresponding light bars. However, there is no interaction between Test Place and Training[40] (labials vs. linguals, $b = -0.11$, *se(b)* $= 0.58$, $z = -0.19$, $p = 0.85$, *ns*; alveolars vs. velars, $b = -0.12$, *se(b)* $= 0.50$, $z = -0.25$, $p = 0.80$, *ns*), indicating that all conditions learn an equal amount about the acceptability of palatalizing what should be palatalized; the difference between the grey bars and white bars are roughly uniform across all places. According to the BIC approximation to the Bayes Factor, the data provide very strong support for the null ($\Delta BIC = 13.9$, $P_{BIC}(H_0 \mid D) = 0.999$).

---

[39] Dependent Variable ~ Training + Test Place + Test Voice + (1 + Test Voice + Test Place | Subject) + (1 + Training | Singular), restricted to ratings before –*i* of all plurals from No Training condition and only To-Be-Palatalized plurals from the Experiment 2 training data.

[40] Dependent Variable ~ Training * Test Place + Test Voice + (1 + Test Voice + Test Place | Subject) + (1 + Training | Base), restricted to ratings before –*i* of all plurals from No Training condition and only To-Be-Palatalized plurals from the Experiment 2 training data.
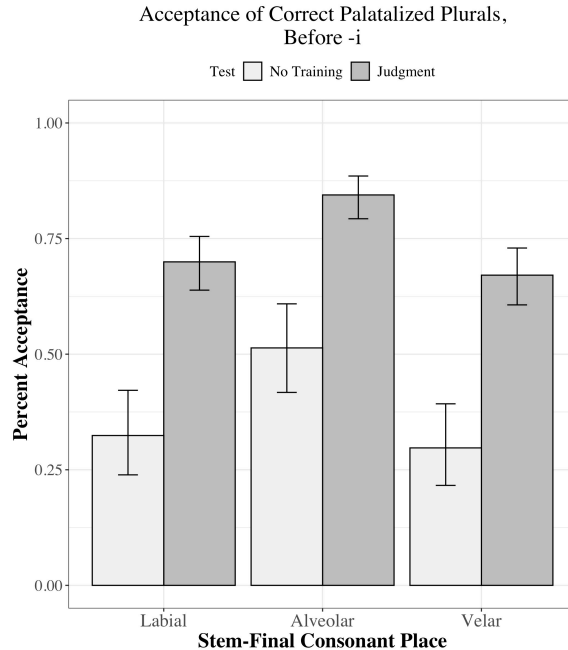
Figure 4.13. Acceptance of correct palatalized plurals, before -*i*. The light bars represent the acceptance rate of palatalized plurals before -*i* at each of the places of articulation in the baseline judgment condition, without training. The dark bars indicate the acceptance rate of *correct* (To-Be-Palatalized) palatalized plurals before -*i* at each of the places of articulation (which here are also the training conditions, i.e. Labial participants' judgments of correct palatalized labial-final stems) in the judgment test after training.

If the strength of associations between production representations at least partly

determines whether an alternation will be produced, as proposed by the Perseveration

Hypothesis, then the lack of production of Labial Palatalization would suggest that the

paradigmatic associations between labials and alveopalatals have not been formed. What,

then, explains the increase in ratings of labial palatalization after training? We propose

that this increase in judgments is based on the acquisition of product-oriented schemas

(Kapatsinski, 2012, 2013), so the improvement in judgments after training reflects that

participants have learned that [tʃi] and [dʒi] are good indicators of plurality. The fact that

judgments change as a result of training indicates that even Labial Palatalization

95

participants learn first-order schemas, but these schemas are insufficient to overcome the production-internal bias against changing labials into alveopalatals[41].

4.4. Discussion

Judgments of faithful plurals, for both To-Be-Palatalized and Not-To-Be-Palatalized stems, are equal for all training languages. The results support Hypothesis 1: Alveolars and velars are not palatalized more than labials because [ti] and [ki] are poorly formed and palatalizing results in a less marked output, but because the change from [p] to [tʃ] is undesirable (see also §4.4.1.1.1). Labials are correctly palatalized less than alveolars and velars are, and are also palatalized in error less than alveolars and velars are, supporting Hypothesis 2. Labial palatalization is not more likely to be overgeneralized in general, however; while Labial Palatalization generalizes to alveolars and velars more than Alveolar and Velar Palatalization generalizes to labials, it does not generalize more to the linguals than the linguals do to each other, so Hypothesis 3 is not supported. While Labial Palatalization participants rarely palatalize in production, they accept palatalized labials more than faithful labials in judgment. Even though they know what they should produce, they are unable to do so, just as Hypothesis 4 predicts. Contrary to the perceptual similarity hypothesis in Hypothesis 5, palatalization of [k] is not produced or accepted more than palatalization of [g], even though [ki] is perceptually and acoustically more similar to [tʃi] than [gi] is to [dʒi] (Guion, 1998). In fact, palatalization of [g] is slightly preferred to palatalization of [k], as was the case in Wilson (2006), likely due to English orthography.

---

[41] *Map constraints (Zuraw, 2007) could capture the bias against labial palatalization by ranking *Map(p,tʃ) / *Map(b,dʒ) higher than *Map(t,tʃ) / *Map(d,dʒ) and *Map(k,tʃ) / *Map(g,dʒ). White (2013) uses *Map constraints based on the P-map (Steriade, 2001/2009) to capture saltation, whereas we propose that articulatory similarity is the relevant metric.

Despite the low rate of production palatalization after Labial Palatalization training, judgments of labial palatalization increase (compared to the baseline) an equal amount in all conditions. If paradigmatic associations between production representations drive production of an alternation, as proposed by the Perseveration Hypothesis, the results suggest that Labial Palatalization participants fail to learn the mapping between labials and alveopalatals, but the improvement in judgments of labial palatalization indicates that they do learn something. We propose that they learn product-oriented schemas (Bybee, 1985, 2001; Kapatsinski, 2012, 2013; Nesset, 2008), which describe the characteristics certain types of forms are likely to have (like "plurals end in [tʃi]"), and can be used in making judgments. Even though Labial Palatalization participants do not manage to associate labials with alveopalatals, they still notice the high rate of [tʃi] in plurals, which allows them to acquire the first-order schema. They then apply the schema to all input consonants equally, having no paradigmatic mappings to tell them otherwise. Alveolar and Velar Palatalization participants also acquire the first-order schema, as shown by the preference for palatalized Not-To-Be-Palatalized stems over faithful. However, they also learn to associate alveolars and velars, respectively, with alveopalatals, and these mappings compete with the first-order schema, driving down the rates of acceptance of palatalized Not-To-Be-Palatalized consonants compared to palatalized To-Be-Palatalized consonants.

4.4.1. Implications for other theories

4.4.1.1. Perceptual similarity

Steriade (2001/2009) proposed the P-map, a store of perceptual similarities between segments in context, which speakers rely on in their quest to avoid noticeable changes

that violate speech norms. In our study, as in Zuraw's (2000) study of Tagalog, the speech norms (as reflected in judgment scores) encourage changing the stem, but speakers still fail to produce the change. Perhaps first-language experience makes English speakers prefer some alternations, but the equivalent judgments of faithful mappings show that the preference cannot be reduced to phonotactics, and the same pattern of results was found in Stave et al. (2013) (though with overall less likelihood of palatalization, since the context was phonetically unnatural; Mitrović, 2012; Wilson, 2006).

The bias against labial palatalization is against certain *changes*, not for/against certain *structures*. The data suggest that English speakers know that labials are less changeable than velars, which are less changeable than alveolars. However, White (2013) found no evidence for alternations targeting labials (p~v) to be any harder to learn than those targeting alveolars (t~θ). Together, the results indicate that learners assign prior probabilities to alternations/paradigmatic mappings, which can be captured by *Map constraints in OT (Zuraw, 2007) and operations in rule-based phonology.

One difference between our results and White's (2013, 2014) is that in our experiment no-change errors on the trained segments (i.e. failing to change what should be) are the most common error in production, and in his experiment large changes overgeneralize more than small in production and a two-alternative forced choice task.[42] Our results are similar to Skoruppa et al. (2011), and we think the difference is due to the exclusion of subjects who made too many no-change errors in White (2013, 2014), which

---

[42] In our experiment we do find greater overgeneralization of the large change in judgment, with Labial Palatalization extending to alveolars and velars more than Alveolar and Velar Palatalization extend to labials, which we attribute to the product-oriented schema having no competition from paradigmatic mappings in the Labial condition, which allows it to apply everywhere.

affected the large change condition more than the small (White, 2013, p. 72). Had those participants been included, there would have been a higher proportion of no-change errors after exposure to the large change.

Another possibility is that perhaps our and Skoruppa et al.'s (2011) results can be explained by first-language experience, whereas White's (2013, 2014) results cannot. Maybe familiarity affects the learnability of a change (unfamiliar changes are harder to notice in training and perform in production), and magnitude affects the likelihood the change will be overgeneralized (large changes are more likely to be overgeneralized than small changes). We doubt this proposal, however, because the changes presented in White (2013, 2014) were not necessarily novel, either: Turning a voiceless stop into a voiced fricative between vowels was compared to intervocalic lenition of a voiced stop, and while English does not have categorical intervocalic stop lenition, it does have fairly common variable lenition (Davidson, 2011; Honeybone, 2001; Sangster, 2001; Warner & Tucker, 2011). Subjects at UCLA likely have exposure to Spanish and Spanish-accented English, which have voiced stop lenition that tends to preserve voicing (Zampini, 1996). Therefore the small changes in White (2013, 2014) may have been more familiar than the large. Additionally, diachronic patterns suggest that large changes are more likely to lose productivity than they are to be generalized (Bybee, 2008). We should still replicate our results with speakers of other languages, especially languages with labial palatalization, like the Southern Bantu languages Xhosa and Sotho (Bennett & Braver, 2015; Ohala, 1978); their exposure to labial palatalization may result in different biases, if the experience of producing and hearing labials palatalizing is able to overcome the greater articulatory distance between labials and palatals, but we expect results to be comparable.

99

Serbian has productive velar palatalization before [e] but not [i], but Serbian speakers still learned velar palatalization before [i] better than before [e] in an artificial language experiment (Mitrović, 2012). It would also be interesting to test speakers of languages where none of the palatalization patterns are productive (e.g. Catalan, Finnish, Kannada, Tagalog, Tamil; Bateman, 2007).

4.4.1.1.1. Influence of markedness

Prior researchers have proposed that markedness has an effect on the learnability of alternations. Wilson (2006) found that palatalizing [k] before *-i* is more likely than palatalizing [k] before *-e*, because [ki] is more marked than [ke]. White (2017) argued that learning to produce large changes requires a very high weight for the competing markedness constraint. However, our subjects show no difference in the acceptability of [pi] vs. [ti] vs. [ki], even though they are more willing to change [t] and [k] than [p]. The prior studies did not have participants judge faithful forms, so it is unclear if Wilson (2006) and White (2013, 2014) would show similar patterns, had that comparison been included. Contrary to the standard assumptions of OT and the models of Wilson (2006) and White (2017), we think there is no particular connection between the markedness of a structure (like [ki]) and the productivity of a change that avoids that structure (like [k]→[tʃi]). We view production of alternations as being driven by attraction to particularly good outputs (Bybee, 2001; Kapatsinski, 2013) following the paths of paradigmatic associations when these are available, rather than avoidance of marked outputs.

4.4.1.2 Storage economy

Kenstowicz's (1996) proposal that Paradigm Uniformity improves lexical access
fares better given our results, assuming that the bias against large changes is within the
production system. Perhaps speakers assume that p~tʃ is harder for listeners to undo than
k~tʃ, and so avoid the larger change because of the potential for misunderstanding. There
is evidence that speakers avoid ambiguity by hyperarticulating cues that distinguish a
word from its minimal pair neighbors (Baese-Berk & Goldrick, 2009; Wedel et al.,
2013). We consider that a desire to maintain recoverability of the base (or underlying)
form is an unlikely explanation for the avoidance of large alternations because
neutralization of underlying contrasts does not lead participants to avoid alternations that
result in it in comparable experiments. For example, adding examples of tʃ~tʃi makes
participants more likely to change [k] into [tʃi], even though these examples make it
unclear whether a [tʃi]-final plural originated from a [k]- or [tʃ]-final singular
(Kapatsinski, 2009, 2012, 2013). The only circumstance in which neutralization has been
shown to result in avoidance of an alternation is when it both resulted in complete
homonymy between corresponding forms, and these forms were adjacent to each other
(Kapatsinski, 2017b). In the present experiment, neither large nor small alternations result
in homonymy, and the corresponding forms are randomly ordered.

4.4.1.3. Categorization

According to the categorization account (Moreton & Pater, 2012a), labial
palatalization should overgeneralize to alveolar and velar palatalization, since any
category including labials and alveopalatals (such as [-continuant]) also includes
alveolars and velars. Our results (Hypothesis 3) show no more overgeneralization from

labials to alveolars than from velars to alveolars, or from labials to velars than alveolars to velars. When subjects are not trained to criterion (cf., White, 2013, 2014), they are not able to overcome the bias against the large changes, making large changes no more likely to be overgeneralized, merely less likely to be produced.

There is previous research that has shown that the featural complexity of categories corresponds to learnability. In Skoruppa et al. (2011) and White (2013, 2014), alternations that differed by one feature were easier to learn and perform than those that differed by more than one feature. However, the alternations with more than one feature different were also saltatory. Prior work has suggested these alternations are harder to learn because of category learning (Moreton & Pater, 2012a; White, 2013, 2017), but we suggest they are more difficult because they are a type of large change, which are harder to learn than small changes *generally* (§2.2.1). It is unclear whether the existence of an intermediary sound is necessary for the larger change to be harder to learn (e.g. in an alternation between [b] and [f], [p] and [v] are intermediary), so future work should investigate the learnability of the same change for speakers who have the intermediate sound in their inventory vs. those who do not.

4.4.1.4. Learnability

If the learnability of an alternation is based on perceptual similarity (Wilson, 2006), then k~tʃi should be easier to learn than g~dʒi, since [ki] and [tʃi] sound more similar than [gi] and [dʒi] (Guion, 1998). Our results contradict this, with [g] being palatalized significantly more often than [k] (and palatalized [g]-final stems being rated marginally better than palatalized [k]-final stems, Hypothesis 5), which suggests that the link

102

between perceptual similarity and learnability is not a causal one.[43] Alternatively, perhaps it is precisely because it is more noticeable that g~dʒ is easier to acquire than k~tʃ. However, were this the case, we would expect labial palatalization to be more productive, or at least learned faster, than velar or alveolar palatalization, which is distinctly not what we find. The effect is therefore probably due to orthography: [k] and [tʃ] spellings are largely distinct (<k> always maps onto [k], <c> maps onto [tʃ] 3% of the time, and <ch> maps onto [tʃ] 87% of the time), whereas [g] and [dʒ] often overlap (<g> maps onto [dʒ] about 30% of the time; Gontijo et al., 2003). Prior research has shown that it is easier to learn alternations between sounds that can be grouped into one category (Moreton & Pater, 2012a), and that participants tend to convert sounds into their orthographic representation (White, 2013). The shared spellings of [g] and [dʒ] in English appear to make them easier to associate than the distinct spellings of [k] and [tʃ]. Future research should investigate whether the asymmetry between palatalization of [g] and [k], unexpected on the basis of perception, holds for other languages with different orthographic correspondences (see §4.3.5), and for pre-literate children.

Another possibility is that the confusability values from Guion (1998) are not the best measure to evaluate perceptual similarity. Wang & Bilger (1973) showed that [g] and [dʒ] are more confusable than [k] and [tʃ] if confusions before [u], [i], and [ɑ] are combined, raising the question of whether context-sensitive or context-independent values should be used. Steriade (2000, 2001/2009) proposed that perceptual similarity must be context-specific to account for typological asymmetries in assimilation patterns. White (2017) uses context-independent confusion probabilities from Wang & Bilger

---

[43] Additionally, all training languages increase judgments of palatalized forms an equal amount over the no-training baseline; in other words, labial palatalization (perceptually dissimilar) is learned in judgment as well as alveolar and velar palatalization (perceptually similar).

(1973) as an estimate of perceptual similarity between alternating segments, which enables his model to capture the slight preference for [g] found here and in Wilson (2006). Palatalization may be harder to learn in a phonetically-unmotivated context (Mitrović, 2012; Wilson, 2006; see also Chapter VII), but is the influence of the context segments independent from the identity of the input segments? In other words, do speakers assign probabilities to rules (p→tʃ/_a), or do they assign probabilities to changes (p→tʃ) and their outputs (tʃa) and use the combined probability to evaluate the likelihood of a particular change resulting in a particular output (Labov, 1969)? The prior probability of a change is at least partially context-independent, since p→tʃ is harder to learn than t→tʃ or k→tʃ, whether before -*i* (as in the present experiment) or -*a* (Stave et al., 2013). However, Experiment 2 and the experiment in Stave et al. (2013) differ in other respects than just the suffix vowel. Comparison of Experiment 2 to Experiment 3, which differ only in the magnitude of change (labial vs. velar) and the context of the change (-*i* vs. -*a*), shows that velar palatalization is produced more than labial palatalization only before -*i*, with no difference found before -*a*. The degree to which probability of change is context-dependent vs. context-independent is therefore still an open question. One final possibility is that perceptual similarity is more abstract than confusability (i.e. originating from a higher level of representation) and could be affected by, for example, how often two sounds share a spelling. Future work should test how well an alternation between two articulatorily dissimilar and acoustically or perceptually similar segments is learned; articulatory similarity predicts that an alternation like f~θ would be difficult to learn, where perceptual similarity predicts the opposite.

4.4.2. Limitations

The primary limitation of this study is that all participants were native American English speakers, who may have generalized from their knowledge of English or imposed English patterns on the artificial language (Finn & Hudson Kam, 2009). This subject pool allows us to make comparisons to the perceptual data from Guion (1998) and previous palatalization learning experiments (Wilson, 2006; Kapatsinski, 2012, 2013; Stave et al., 2013), but it also means the results could be due to first-language phonological experience rather than the difference in change magnitude. In particular, English has alveolar palatalization before glides in frequent phrases like *would you* and *bet you* and in words like *creature* (cf. *create*) or *torture* (cf. *extort*). While the former do not involve a complete change in place of articulation (Zsiga, 1995) and the latter are of doubtful productivity, the existence of the patterns may have made alveolar palatalization easier to learn to produce. First-language transfer may not be an insurmountable factor in miniature artificial language learning (Garcia et al, 2017; Mitrović, 2012; Wang & Saffran, 2014), but it would still be worthwhile to see the pattern of results for speakers of languages without alveolar palatalization.

The other potential limitation is that which plagues all artificial language learning experiments, namely that learning in the lab may not be the proxy for learning in more natural contexts that we take it to be. Learners in the lab are exposed to a much more impoverished input, for a much shorter period of time, than learners of natural languages. While this allows a degree of control over the variables of interest that is not possible with natural language, it may also lead to different learning strategies than would normally be applied. However, Friederici et al. (2002) showed that patterns of brain

activation when processing a miniature artificial language are comparable to those activated when processing natural (first) language, and Ettlinger et al. (2016) found that performance in artificial language tasks correlates with performance of L2 Spanish learners, suggesting that the same abilities are recruited in both cases. While the simplified structure and limited exposure to linguistic input in the lab is a definite limitation, it seems that similar resources are relied on as are in natural language. There is always the possibility that the findings from laboratory experiments do not all generalize to natural language, but they at least provide a starting point for further investigation.

4.4.3. Summary

The Perseveration Hypothesis as an explanation for Paradigm Uniformity proposes that stem changes are leveled by paradigmatic perseveration within the production system. When creating a novel form, the speaker activates other related forms and the meaning to be expressed, and articulatory gestures from the base are incorporated into the form being produced through a blending process. When too much of the base is copied, the stem change is leveled. Paradigmatic associations between related forms prevent leveling by specifying that X gesture in the base form activates Y gesture in the target form, and it is harder to learn associations between dissimilar representations because they require greater synaptic modification. Paradigmatic associations can drive judgments, and we argue they are a factor in palatalizing the To-Be-Palatalized consonants in the small change conditions (Alveolar and Velar Palatalization). As a result, learners in these conditions judge palatalization of To-Be-Palatalized consonants as more acceptable than palatalization of Not-To-Be-Palatalized consonants. However, product-oriented schemas (like "plurals end in [tʃi]") also influence judgments. In the

absence of paradigmatic associations, only product-oriented schemas are present. Because participants in the large change condition (Labial Palatalization) are unable to form the appropriate paradigmatic associations, [tʃi] is always a surprise when it occurs, which makes it more salient and therefore available and activated for all inputs, resulting in overgeneralization of palatalization from labials to velars and alveolars in judgment. Palatalization of Not-To-Be-Palatalized consonants benefits from the salience of the [tʃi] schema in the labial condition, whereas palatalization of To-Be-Palatalized consonants is hurt by the absence of the second-order schema. As a result, the difference in judgments between the two is much smaller in the labial condition compared to the others.

The ultimate goal of any theory of Paradigm Uniformity is to explain natural language, not just metalinguistic judgments or elicited production. If we are interested in modeling the process responsible for the avoidance of large changes in language, the production task is the more informative, and it tells us that large changes are rare cross-linguistically because they are hard to perform, so they lose productivity. This loss of productivity is extremely common, and may in fact be a diachronic universal (Bybee, 2008); for an example in English, consider the k~s alternation in *electric~electricity*, and how it is not extended to the intermediate [t] (Pierrehumbert, 2006). Based on this, we expect that labial palatalization will be lost in languages that have it (like Southern Bantu; Ohala, 1978), rather than be generalized to all stops. In the first study on the productivity of labial palatalization, Bennett & Braver (2015) found that it is indeed only partially productive in Xhosa. Based on universal diachronic patterns, we believe that paradigmatic perseveration is partially responsible for the rarity of large stem changes

typologically: They are hard to perform in production, and difficult stem changes are especially likely to be leveled by performance pressure.

In Chapter V we present the results of Experiment 3, which was designed to investigate whether temporal contiguity of related pairs can make the paradigmatic associations exemplified by the pairs easier to acquire. Of particular interest was whether making large changes obvious would provide enough evidence to allow for paradigmatic associations to form between labials and alveopalatals.

CHAPTER V

EXPERIMENT III: EFFECTS OF ADJACENCY ON LEARNABILITY OF

PALATALIZATION BEFORE -*a*

Portions of this chapter were taken from:

Smolek, A. & Kapatsinski, V. (2019). *Syntagmatic paradigms: Learning correspondence from contiguity.* Manuscript submitted for publication.

Experiment 2 shows that it is difficult to learn to produce alternations involving dissimilar sounds, and that those alternations are likely to be extended beyond their original context in judgments. This raises the question of how alternations are acquired: What allows speakers to learn an alternation, and to restrict it to the appropriate context? We propose that temporal contiguity of related forms is an integral part of the acquisition of morphophonology. Erwin (1961) pointed out that morphologically related words tend to occur in similar contexts, allowing the listener to "bring them into contiguity" by using the context to anticipate one of the words as the other is perceived (see also McNeill, 1966; Onnis, Waterfall & Edelman, 2008; Slobin & Küntay, 1996). Furthermore, contrary to McNeill (1966), paradigmatically related words also co-occur in natural language (Baroni et al., 2002; Fellbaum, 1996; Jones et al., 2007; Murphy, 2006; Xu & Croft, 1998), allowing for the listener to anticipate one form of a word as another is perceived. In both cases, the corresponding forms are in contiguity because the listener activates one form of a word (whether perceived or anticipated on the basis of context) as s/he is about to perceive another form.

Work on category learning has shown that contiguity between members of contrasting categories helps learners pick up on discriminative features, the ones that best

distinguish between the categories (Arnon & Ramscar, 2012; Ramscar et al., 2010; Carvalho & Goldstone, 2015); by extension, learners may use corresponding singulars and plurals occurring in contiguity in order to learn what distinguishes singulars from plurals. For example, in a language with labial palatalization before the plural suffix -*a*, a final …*p* indentifies the form as singular, and a final …*a* identifies it as plural. Here, we argue that temporal contiguity also allows for learners to predict one of the forms of a word in the context of the other form, and therefore using the other form as a set of predictive cues. We show that discriminative learning under contiguity allows the learner to identify cues that discriminate among singulars corresponding to different plural patterns (e.g. learning that a final …*p* in a singular indicates that the plural will end in …*tʃa*).

Theories of grammar differ in whether they predict contiguity to benefit faithful and unfaithful mappings. For example, in Optimality Theory, faithful mappings are due to output-output (OO) faithfulness constraints, which are thought to start out at the top of the constraint ranking (Hayes, 2004; McCarthy, 1998). Similarly, Taatgen and Anderson (2002) propose that faithful mappings are produced by a default "do nothing" rule that remains the default unless an alternative rule is learned. If OO faithfulness constraints or "do nothing" are still at the top for our participants, then contiguity of faithful forms should have no effect on the acquisition of the faithful mapping, since it is already at ceiling. However, exposure to English should downweight OO-faith[alveolar] and OO-faith[velar], but not OO-faith[labial], which would mean that intact pairs including faithful alveolars and velars in training would increase retention of stem-final alveolar

and velar consonants at test, but adjacent faithful labials would have no effect on retention of stem-final labial consonants.

In Usage-based Phonology, unfaithful mappings have been attributed to product-oriented schemas that are acquired by generalizing over forms belonging to a particular cell in the morphological paradigm (such as "plural"), without reference to any base form (Bybee, 1985, 2001; Kapatsinski, 2012, 2013). If unfaithful mappings are always attributable to product-oriented schemas, then unfaithful mappings (like p~tʃa) should not benefit from contiguity as much as faithful mappings (like t~ta and k~ka), because a product-oriented schema about plurals is acquired without reference to the singular (e.g., "plurals end in [tʃa]" can be learned without knowing what the corresponding singulars are like; Bybee, 2001; Kapatsinski, 2013, 2017b).

However, it is also possible that unfaithful paradigmatic mappings are at least helped by generalizations over corresponding pairs of words. This is true in rule-based phonological models that map surface forms onto each other by context-specific transformations (Albright & Hayes, 2003). It is also true in usage-based models that include a role for second-order schemas or paradigmatic mappings (Booij, 2010; Kapatsinski, 2018; Nesset, 2008). Finally, generalization over both singulars and plurals is required in discriminative learning models (Baayen et al., 2011), where learners would learn what plurals are like by determining what phonological features discriminate singulars from plurals. That is, …tʃa is associated with plurality not because it is common in plurals but because it is more common in plurals than in non-plurals.

Unfaithful paradigmatic mappings are notoriously difficult to acquire in artificial and natural language contexts (Braine et al., 1990; Brooks et al., 1993; Dąbrowska &

111

Sczerbinski, 2006; Krajewski et al., 2011). Unfaithful mappings may benefit *more* from contiguity if they are created by application of a second-order schema or rule, because they have more room for improvement.

The following experiment explores these effects through training on voiced and voiceless palatalization before -*a*, manipulating the order of the training trials. For the purposes of this experiment, only pairs where the singular is immediately followed by the corresponding plural are considered temporally contiguous. While temporal contiguity in natural language includes cases where the related forms are separated by some number of other words but are connected by occuring in a similar context, and sharing the context has been shown to work in the lab (Onnis et al., 2008), here we implement contiguity by actual adjacency. That is, temporal contiguity is implemented by keeping pairs of corresponding words "intact," by presenting the corresponding singulars and plurals adjacent to each other. This manipulation is, however, intended to control, in the most direct way possible, whether the singular form is available to compare to or predict the plural form when the plural is presented.

5.1. Methods

5.1.1. Participants

152 native American English speakers with no reported history of speech, language, hearing, or learning disabilities were recruited from the Psychology/Linguistics Human Subject Pool at the University of Oregon. There were 38 participants per trial order condition (split between training languages), and they received partial course credit for participation.

112

5.1.2. Languages

There were two training languages, containing either labial or velar palatalization. Just as in Experiment 2 (Chapter IV), the stems were always C(C)VC and ended in an oral stop [b;p;d;t;g;k]. The plural suffix was -*a* (100% of the time for To-Be-Palatalized consonants, 50% of the time for Not-To-Be-Palatalized consonants) or -*i* (0% of the time for To-Be-Palatalized consonants, 50% of the time for Not-To-Be-Palatalized consonants). The To-Be-Palatalized consonant became [tʃ] if voiceless and [dʒ] if voiced, before -*a*. Velar palatalization is perceptually (Guion, 1998; Ohala, 1989, p. 183-185, 1992, p. 320) and articulatorily (Anttila, 1989, p. 72-73; Hock, 1991, p. 73-77) motivated before -*i*, whereas labial palatalization is not; neither are phonetically motivated before -*a*. In Experiment 2, Velar Palatalization participants produce palatalization of the target consonants much more often than Labial Palatalization participants. [k] is articulatorily closer to [tʃ] than is [p], regardless of the vowel context, so the difference in change magnitude could hold before -*a* as well. However, if learning is affected by substantive bias, as has been proposed and found experimentally (e.g. Do, 2013; Finley, 2008; Hayes & White, 2015; Mitrović, 2012; Stave et al., 2013; White, 2013, 2014; Wilson, 2006), then velar palatalization may only be learned better in Experiment 2 because it is in a phonetically motivated context, so there may be no difference in learnability by change magnitude in Experiment 3. The two experiments are compared in Chapter VII.

See Table 5.1 for the patterns in each language.

Table 5.1. Labial and Velar Palatalization patterns presented to participants in Experiment 3.

	Labial Palatalization	Velar Palatalization
Singular	Plural	Plural
…p	**…tʃa**	…{pi;pa}
…b	**…dʒa**	…{bi;ba}
…t	…{ti;ta}	…{ti;ta}
…d	…{di;da}	…{di;da}
…k	…{ki;ka}	**…tʃa**
…g	…{gi;ga}	**…dʒa**

The identity of the To-Be-Palatalized consonants (labial [p;b] or velar [k;g]) was crossed with trial order, which is the variable of principal interest in this study. Participants were exposed to one of four trial orders, shown here for the Velar Palatalization condition:

–**All Obvious**: All corresponding singular-plural pairs were intact, with the plural immediately following the singular, whether the mapping was faithful or unfaithful (1)

–**None Obvious**: Singulars and plurals were all randomly ordered, so corresponding singulars and plurals were not adjacent at greater than chance frequency (4)

–**Change Obvious**: Unfaithful corresponding singular-plural pairs were intact, but singular-plural pairs exemplifying faithful mappings were split up and randomly ordered (2)

–**NoChange Obvious**: Only singular-plural pairs exemplifying faithful mappings were kept intact; unfaithful pairs were split up and randomly ordered (3)

(1) All Obvious: blup$_{SG}$ blupa$_{PL}$ klut$_{SG}$ kluti$_{PL}$ smak$_{SG}$ smatʃa$_{PL}$…

(2) Change Obvious: kluti$_{PL}$ blup$_{SG}$ klut$_{SG}$ smak$_{SG}$ smatʃa$_{PL}$ blupa$_{PL}$ …

(3) NoChange Obvious: blup$_{SG}$ blupa$_{PL}$ smatʃa$_{PL}$ klut$_{SG}$ kluti$_{PL}$ smak$_{SG}$…

(4) None Obvious: kluti$_{PL}$ blup$_{SG}$ smatʃa$_{PL}$ klut$_{SG}$ smak$_{SG}$ blupa$_{PL}$ …

The trial orders differ on two dimensions: whether unfaithful singular-plural mappings (To-Be-Palatalized, here $smak_{SG}$~$smatʃa_{PL}$) are adjacent and therefore easier to learn, which is the case for the All Obvious and Change Obvious conditions; and whether faithful singular-plural mappings (Not-To-Be-Palatalized, here $blup_{SG}$~$blupa_{PL}$ and $klut_{SG}$~$kluta_{PL}$) are adjacent and therefore easier to learn, which is true for the All Obvious and NoChange Obvious conditions. As discussed earlier, adjacency is a proxy for temporal contiguity of paradigmatically-related words in corpus data. We expect adjacency to influence learnability of the alternation (as reflected by the overall rate of palatalization and differences in this rate across different types of singulars).

We use intactness of pairs and faithfulness as the primary predictors in statistical analyses, and expect both to influence rate of palatalization. We do not have any strong intuitions regarding how they will interact with language, suffix, and final consonant, and we explore those interactions using conditional inference trees.

5.1.2.1. What learners need to weight

To perfectly reproduce the input language, participants need to learn several generalizations:

1) there are two plural suffixes, *-i* and *-a*

2) when to use each suffix, and in particular, that *-a* is the only eligible suffix after consonants that should be changed ([k] and [g] for Velar Palatalization, [p] and [b] for Labial Palatalization); learning (when) to use suffixes is covered in §5.2.1

3) there is palatalization in the language

4) the context of palatalization, in particular which suffix triggers it (§5.2.2)

5) the context of palatalization, in particular which input consonants are afflicted by it (§5.2.3)

The learnability of these generalizations is potentially affected by contiguity, and we explore these effects in §5.2.

5.1.3. Materials

The materials were the same as for Experiment 2 (§4.2.3), except for the trial order manipulations described in §5.1.2 and that all To-Be-Palatalized plurals were suffixed with *-a* instead of *-i*, as well as an additional 24 novel singulars which ended with a palatal (half [tʃ], half [dʒ]) in the prodution test phase. The complete materials are available in Appendix C.

5.1.4. Procedure

5.1.4.1. Training

The training phase procedure was the same as in Experiment 2 (§4.2.4.1), except for the trial order differences. We created two "blocks" of trials within each of the three training blocks; one "block" consisted of word pairs, and the other of single word forms. Each training trial sampled a "block" (with replacement) and then randomly sampled a trial within the "block" (without replacement). In the None Obvious condition, all samples came from the single wordform "block" (this was the same procedure as Experiment 2). In the All Obvious condition, all samples came from the word pair "block". For Change Obvious, the To-Be-Palatalized pairs were selected from the word pair "block" and the Not-To-Be-Palatalized pairs from the single wordform "block," with the reverse true for NoChange Obvious. Table 5.2 shows the block choice by the To-Be-Palatalized status of the stem for each trial order.

Table 5.2. Trial selection "blocks" by trial order and To-Be-Palatalized status of stem-final consonant.

Trial Order	To-Be-Palatalized	Not-To-Be-Palatalized
None Obvious	single wordform block	single wordform block
All Obvious	word pairs block	word pairs block
Change Obvious	word pairs block	single wordform block
NoChange Obvious	single wordform block	word pairs block

5.1.4.2. Test

The procedure for the test was the same as for the production test in Experiment 2 (§4.2.4.2). As for Experiment 2, all the production trials were randomly ordered.

5.1.5. Measures

5.1.5.1. Transcription protocol and exclusions

Each production was transcribed for later analysis. We coded for whether palatalization was present and the identity of the plural suffix. Palatalization was coded as present if the word ended in a vowel preceded by [tʃ] or [dʒ]. Codings were performed by me and a team of undergraduate RAs, with the latter also checked by me; in cases of disagreement, I revisited the recording and looked at the spectrogram if necessary. We excluded productions where the singular-final consonant was replaced with anything other than [tʃ] or [dʒ], or where the plural suffix vowel was anything other than *-i* or *-a*. 85% of observations remained after exclusions ($n = 15,565$). See §5.2.1 for analysis of error patterns.

5.1.5.2. Model structure

The results were analyzed using generalized (logistic) linear mixed-effects models with the lme4 package (version 1.1-21, Bates et al., 2015) in R 3.6.0 (R Core Team, 2019). The maximal random effects structure was used, including random intercepts for Subjects and Singulars and random slopes for Trial Order variables (Faithful Intact,

Unfaithful Intact) within Singulars and To-Be-Palatalized (To-Be-Palatalized vs. Not-To-Be-Palatalized, given training condition) within Subjects. Pairwise interactions between the individual trial order variables and To-Be-Palatalized were included if they were significant. Complete model structures are included in footnotes.

Subsequent exploratory data analysis was conducted using conditional inference trees with the party package (Hothorn et al., 2006). Trees included Language as a predictor of vowel choice and both Language and Plural Vowel as predictors of palatalization. To control for multiple comparisons, the minimum significance level required to split the tree was lowered to 0.001. Theoretically interesting interactions (between, for example, To-Be-Palatalized and Plural Vowel) discovered by inspecting the trees were tested using mixed-effects models to take into account dependencies between observations coming from the same subject or item. We visually inspected the conditional inference trees for any (a priori unexpected) cross-over interactions, and would have included them in mixed-effects regression models had they been found.

5.1.5.3. Predictions

If contiguity helps the acquisition of faithful mappings, then NoChange Obvious and All Obvious will have less palatalization than Change Obvious and None Obvious, because the faithful alternations are intact. If contiguity helps the acquisition of unfaithful mappings, then Change Obvious and All Obvious will have more palatalization than in the conditions where unfaithful alternations are not intact (NoChange Obvious and None Obvious). If contiguity helps both types of mapping, NoChange Obvious should have many faithful mappings, Change Obvious many unfaithful mappings, All Obvious should be somewhere in between, with None Obvious serving as a baseline.

118

5.2. Results

5.2.1. Error patterns

We compared the error rate by trial order and training language, as shown in Figure 5.1. Keeping faithful pairs intact results in lower error rates[44] ($b = -1.47$, $se(b) = 0.27$, $z = -5.40$, $p < 0.001$), and adjacency of faithful pairs significantly improves model fit ($\chi^2(1) = 26.32$, $p < 0.001$): The proportion of acceptable plurals is higher for participants in the NoChange Obvious and All Obvious conditions than those in the None Obvious and Change Obvious conditions. Keeping unfaithful pairs intact has no significant effect on error rates ($b = 0.046$, $se(b) = 0.27$, $z = 0.17$, $p = 0.87$, ns), and according to the BIC approximation to the Bayes Factor, the results provide very strong support for the null, $\Delta BIC = 10$, $P_{BIC} (H_0 \mid D) = 0.993$. Participants in the Velar Palatalization language produce fewer errors than those in Labial Palatalization ($b = -0.70$, $se(b) = 0.27$, $z = -2.56$, $p = 0.01$), and Training Language significantly improves model fit ($\chi^2(1) = 6.49$, $p = 0.01$): The darker bars, representing Velar Palatalization, are all higher than the lighter bars, representing Labial Palatalization. None of the interactions are significant.

---

[44] Errors ~ Unfaithful Intact + Faithful Intact + Training Language + (1 | Subject) + (1 | Singular)
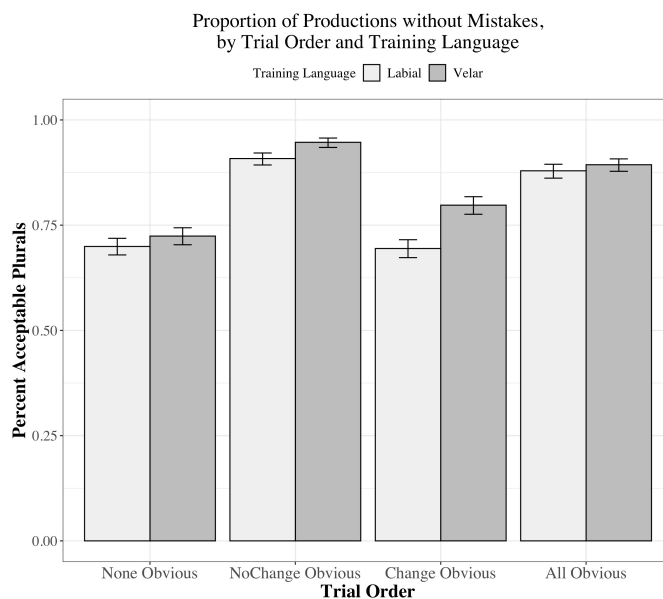
Figure 5.1. Percentage of plural productions without mistakes by trial order and training condition.

5.2.1.1. Consonant and vowel error types

We looked closer at the types of errors that participants in each language and training condition produce. We first separated consonant errors and vowel errors, for ease of visualization[45]. There are three types of consonant errors, as shown in Figure 5.2: "Stop+Palatal," where the plural form contains both the stop and the palatal (e.g. *streik~streiktʃa*); "Bizarre," where the stem-final consonant is replaced with a non-palatal consonant (e.g. *trab~traga*) or the stem consonant is retained and followed by a non-palatal consonant (e.g. *drag~dragda*); and "Absent," where no consonant is produced (e.g. *ʃlud~ʃlu*, but more commonly *traɪk~???*). There are four types of vowel errors, as shown in Figure 5.3: "New Vowel," where the suffix is a vowel other than *-i* or *-a* (e.g. *smuk~smuku*); "-s," where an *-s* is attached directly to the stem (e.g. *stɛd~stɛds*) or to a suffix vowel (e.g. *klub~klubas*); "Bizarre," where the suffix is a syllabic consonant (e.g.

---

[45] Fewer than 5% of productions contained both errors in the consonant and the vowel, and 88% of those are trials where participants failed to produce any plural form.

dig~dign) or a sequence of segments (e.g. *roʊp~roʊpakaɪ*); and "Absent," when no plural

suffix is produced, either because the singular is repeated (e.g. *kwug~kwug*) or, more

commonly, nothing is produced (e.g. *flad~???*). For the analyses, we ran separate logistic

regressions for each error category within error type (e.g. comparing "-s" to all other

plural consonant types, including faithful and unfaithful); none of the interactions were

significant. All inferential statistics were performed on the entire data set, but for ease of

visual comparison, the figures only show the numbers of error-containing productions.

5.2.1.2. Consonant errors

Figure 5.2 shows the distribution of consonant errors by training language (right

panel) and trial order (left panel). Keeping unfaithful pairs intact results in more

consonant errors, overall[46] ($b = 0.66$, *se(b)* $= 0.27$, $z = 2.47$, $p = 0.01$), and inclusion of

Unfaithful Intact significantly improves model fit ($\chi^2(1) = 5.83$, $p < 0.02$): Change

Obvious and All Obvious produce more consonant errors than do None Obvious and

NoChange Obvious (which from the figure seems driven entirely by the high rate of

"Stop+Palatal" productions in Change Obvious; see below for further analysis). Keeping

faithful pairs intact reduces the number of consonant errors ($b = -1.09$, *se(b)* $= 0.27$, $z =$

-4.05, $p < 0.001$), and including Faithful Intact significantly improves model fit ($\chi^2(1) =$

15.32, $p < 0.001$), as can be seen by comparing the bars in NoChange Obvious and All

Obvious to those in None Obvious and Change Obvious. Lastly, participants produce

fewer consonant errors after training on Velar Palatalization compared to Labial

Palatalization ($b = -0.95$, *se(b)* $= 0.27$, $z = -3.53$, $p < 0.001$), and Training Language

significantly improves model fit ($\chi^2(1) = 12.19$, $p < 0.001$).

---

[46] Consonant Mistakes ~ Unfaithful Intact + Faithful Intact + Training Language + (1 | Subject) + (1 | Singular)
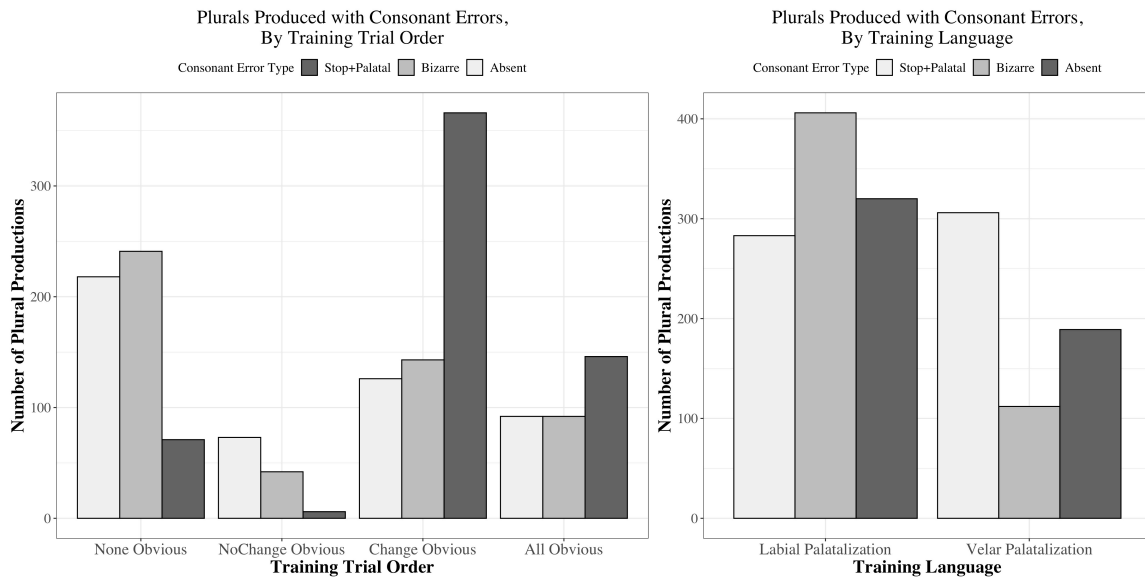
Figure 5.2. Plural productions containing consonant errors by error type. Left panel: Error rates by training trial order. Right: Error rates by training language.

Breaking down the consonant errors by type, we find that for "Stop+Palatal" errors (light bars), keeping unfaithful pairs intact (in Change Obvious and All Obvious) results in higher rates of producing both the stem-final consonant and a palatal[47] ($b = 2.86$, *se(b)* = 0.69, $z = 4.15$, $p < 0.001$), and including Unfaithful Intact significantly improves model fit ($\chi^2(1) = 16.37$, $p < 0.001$). Keeping faithful pairs intact (in NoChange Obvious and All Obvious) results in lower rates of "Stop+Palatal" productions ($b = -2.03$, *se(b)* = 0.70, $z = -2.91$, $p < 0.004$), and including Faithful Intact significantly improves model fit ($\chi^2(1) = 8.59$, $p = 0.003$). Change Obvious, where unfaithful pairs are intact and faithful pairs are not, has a much higher rate of productions retaining the stem-final consonant and also producing a palatal, which seems like a compromise between implementing the obvious change and the effort of overriding perseveration. The majority of "Stop+Palatal" productions are suffixed with -*a* (537/585, or 92%), suggesting that Change Obvious

---

[47] Stop+Palatal Errors ~ Unfaithful Intact + Faithful Intact + Training Language + (1 | Subject) + (1 | Singular)

participants have learned that [tʃa] indicates plurality. (See §6.3.2.2.1 and §6.4.2.2 for further evidence and discussion of [tʃa] as a "chunk.") There is no significant difference in "Stop+Palatal" productions by training language (b = -0.21, $se(b)$ = 0.68, z = -0.31, p = 0.76, ns) and according to the BIC approximation to the Bayes Factor the data provide very strong support for the null, ΔBIC = 9.5, P_{BIC} (H_0 | D) = 0.991.

Intact faithful pairs results in fewer errors of the "Bizarre" variety[48] (b = -0.79, $se(b)$ = 0.29, z = -2.77, p < 0.007; including Faithful Intact significantly improves model fit, $\chi^2(1)$ = 7.55, p < 0.006), as shown in the comparison of the grey bars in None Obvious and Change Obvious vs. NoChange Obvious and All Obvious. There is no significant effect of Unfaithful Intact (b = 0.21, $se(b)$ = 0.28, z = 0.74, p < 0.46), and the results provide strong support for the null according to the BIC approximation to the Bayes Factor, ΔBIC = 9, P_{BIC} (H_0 | D) = 0.989. Finally, after Velar Palatalization training, participants produce fewer "Bizarre" consonants (b = -1.18, $se(b)$ = 0.29, z = -4.09, p < 0.001), and including Training Language significantly improves model fit ($\chi^2(1)$ = 16.62, p < 0.001), as seen in the medium grey bars in the right panel of Figure 5.2.

There is no significant effect of Unfaithful Intact on the rate of failing to produce a plural consonant[49] ("Absent," dark grey bars in Figure 5.2; b = -0.37, $se(b)$ = 0.33, z = -1.11, p < 0.27), and according to the BIC approximation to the Bayes Factor, the results provide strong support for the null, ΔBIC = 8.3, P_{BIC} (H_0 | D) = 0.984. Keeping faithful pairs intact results in marginally fewer "Absent" productions (b = -0.60, $se(b)$ = 0.33, z = -1.81, p = 0.07), but according to the BIC approximation to the Bayes Factor, the data

---

[48] Bizarre Errors ~ Unfaithful Intact + Faithful Intact + Training Language + (1 | Subject) + (1 | Singular)

[49] Absent Consonant Errors ~ Unfaithful Intact + Faithful Intact + Training Language + (1 | Subject) + (1 | Singular)

provide strong support for the null, ΔBIC = 6.4, P_{BIC} ($H_0 \mid D$) = 0.961. Training on Velar

Palatalization results in lower rates of "Absent" consonants (b = -1.00, $se(b)$ = 0.33, z =

-2.99, p < 0.003), and including Training Language significantly improves model fit

($\chi^2(1)$ = 9.14, p < 0.003).

5.2.1.3. Vowel errors

Figure 5.3 shows comparisons of vowel errors. Keeping unfaithful pairs intact results

in fewer mistaken vowel productions[50] (b = -0.91, $se(b)$ = 0.32, z = -2.85, p = 0.004), and

Unfaithful Intact significantly improves model fit ($\chi^2(1)$ = 7.77, p = 0.005). Intact faithful

pairs also results in fewer vowel errors (b = -1.27, $se(b)$ = 0.32, z = -4.00, p < 0.001), and

including Faithful Intact significantly improves model fit ($\chi^2(1)$ = 14.62, p < 0.001). None

Obvious, where neither faithful nor unfaithful pairs are kept intact, has a much higher

error rate compared to the other conditions, as shown in the left panel of Figure 5.3. Velar

Palatalization participants produce fewer errors in plural vowels than Labial

Palatalization participants do (b = -0.64, $se(b)$ = 0.32, z = -2.02, p = 0.04), and including

Training Language significantly improves model fit ($\chi^2(1)$ = 4.05, p = 0.04). The right

panel of Figure 5.3 suggests that difference is driven by higher rates of "Other Vowel"

and "Absent" productions after Labial Palatalization training, but we evaluate differences

by error type, below, to confirm.

---

[50] Vowel Errors ~ Unfaithful Intact + Faithful Intact + Training Language + (1 | Subject) + (1 | Singular)
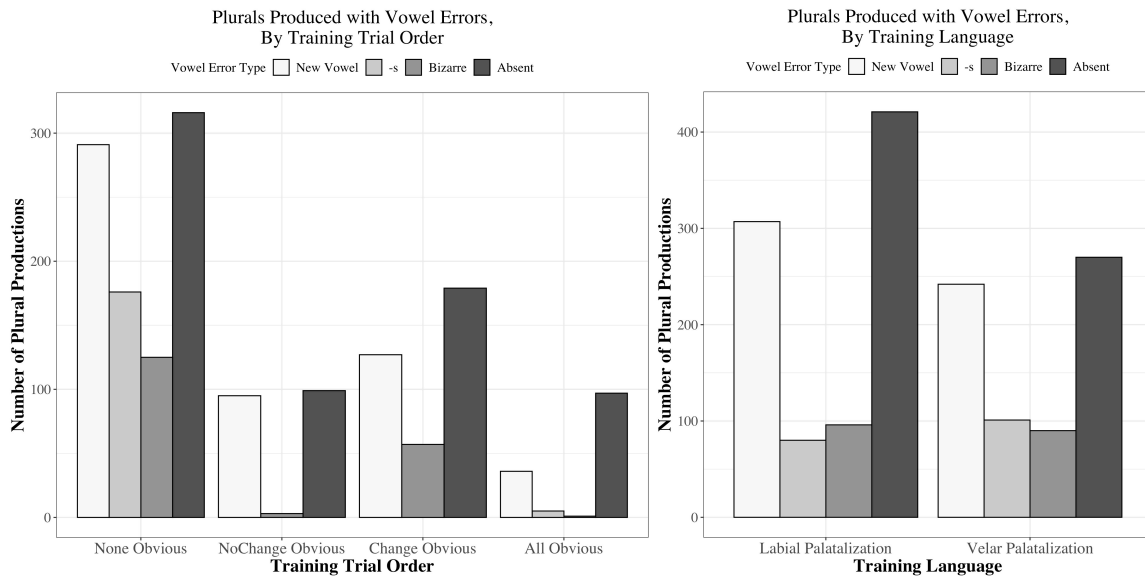
Figure 5.3. Plural productions containing vowel errors by error type and training trial order. Left panel: Error rates by training trial order. Right: Error rates by training language.

Keeping unfaithful pairs intact results in fewer "New Vowel" productions[51] (white bars, $b = -1.49$, $se(b) = 0.68$, $z = -2.20$, $p < 0.03$), and Unfaithful Intact significantly improves model fit ($\chi^2(1) = 4.57$, $p = 0.03$). Keeping faithful plurals intact has no effect on "New Vowel" productions ($b = -1.07$, $se(b) = 0.68$, $z = -1.58$, $p = 0.11$), and according to the BIC approximation to the Bayes Factor, the results provide strong support for the null, $\Delta BIC = 7.2$, $P_{BIC}(H_0 \mid D) = 0.973$. There is no effect of training language ($b = -0.24$, $se(b) = 0.68$, $z = -0.35$, $p < 0.73$), and according to the BIC approximation to the Bayes Factor, the data provide very strong support for the null, $\Delta BIC = 9.5$, $P_{BIC}(H_0 \mid D) = 0.991$.

None of the effects are significant in the "-s" vowel error model[52] (light grey bars, all $z < 1$) or the "Bizarre" vowel error model[53] (dark grey bars, all $z < 1.3$). The vast majority

---

[51] New Vowel Errors ~ Unfaithful Intact + Faithful Intact + Training Language + (1 | Subject) + (1 | Singular)

[52] -s Vowel Errors ~ Unfaithful Intact + Faithful Intact + Training Language + (1 | Subject) + (1 | Singular)

of "-s" vowel errors occur in the None Obvious condition (176/181), and none at all occur in the NoChange Obvious and Change Obvious conditions. All but 4 of the "Bizarre" plural vowels are produced in None Obvious (125/186) or Change Obvious (57/186), the conditions where faithful pairs are not intact. Only when the faithful pairs are intact are both vowels made obvious (since both *-a* and *-i* attach to Not-To-Be-Palatalized stems, whereas only *-a* attaches to To-Be-Palatalized stems); otherwise, it seems that some participants notice that there is *something* going on with the vowel but cannot determine what the particulars are.

For "Absent" errors (black bars), there is no effect of keeping unfaithful pairs intact[54] (b = -0.41, *se(b)* = 0.29, z = -1.43, p = 0.15), and the data provide strong support for the null according to the BIC approximation to the Bayes Factor, ΔBIC = 7.5, P_{BIC} (H_0 | D) = 0.977. Keeping faithful pairs intact reduces "Absent" productions (b = -0.89, *se(b)* = 0.29, z = -3.06, p < 0.003), and Faithful Intact significantly improves model fit (χ^2(1) = 9.17, p = 0.002). Finally, training on Velar Palatalization results in fewer "Absent" errors (b = -0.81, *se(b)* = 0.29, z = -2.76, p < 0.006), and including Training Language significantly improves model fit (χ^2(1) = 7.82, p = 0.005).

In summary, keeping unfaithful pairs intact results in more consonant errors (driven by higher rates of "Stop+Palatal" and "Bizarre" errors), but fewer vowel errors (driven by fewer "New Vowel" errors). Keeping faithful pairs intact results in fewer consonant errors (driven by lower rates of "Stop+Palatal" errors) and fewer vowel errors (driven by

---

[53] Bizarre Vowel Errors ~ Unfaithful Intact + Faithful Intact + Training Language + (1 | Subject) + (1 | Singular)

[54] Absent Vowel Errors ~ Unfaithful Intact + Faithful Intact + Training Language + (1 | Subject) + (1 | Singular)

lower rates of "Absent" vowel errors). Velar Palatalization results in fewer consonant

errors (driven by lower rates of "Bizarre" and "Absent" errors) and fewer vowel errors

(driven by lower rates of "Absent" errors). Bizarre, unattested patterns seem to reflect

general confusion ("I guess it's just random"), and silence seems to indicate profound

indecision ("I have no idea what I am supposed to do"), which is evidently a larger

problem when the change is obvious – where participants can more easily notice that

something is changing but may struggle to figure out exactly what – and for participants

receiving training on the larger change. There are many more vowel errors in None

Obvious than the other trial orders, suggesting that exposure to randomly-ordered

phonetically-unnatural patterns confuses participants, leading them to either assume any

vowel can be used (white bars), or giving up on even trying (black bars).

Overall, the error distributions suggest that labial palatalization is more confusing and

challenging for participants to learn, and that the lack of any training structure in the

None Obvious condition is an obstacle to learning.

5.2.2. Suffix choice

5.2.2.1. Suffix frequency

Participants generally learn that there are two suffix vowels; 86% use each suffix at

least once, but there is a dramatic effect of trial order. Figure 5.4 shows the distribution of

suffix choice probabilities for each trial order condition. It illustrates a large effect of

contiguity on the system learned: When faithful (Not-To-Be-Palatalized) pairs are kept

intact $(0\{i;a\}/C_{\text{not-TBP}}\_$ such as *hɛt~hɛta* and *naɪd~naɪdi*, right column), participants use *-i*

and *-a* at similar rates to the input probabilities and few subjects regularize (i.e. use one

suffix 100% of the time). When faithful pairs are not intact (left column), participants

strongly favor -*a* and often regularize. There is a weaker effect of unfaithful pair adjacency (To-Be-Palatalized, C_{TBP}tʃa such as *bluk~blutʃa* and *fɛp~fɛtʃa*), but when they are kept intact, use of -*a* increases (bottom row).
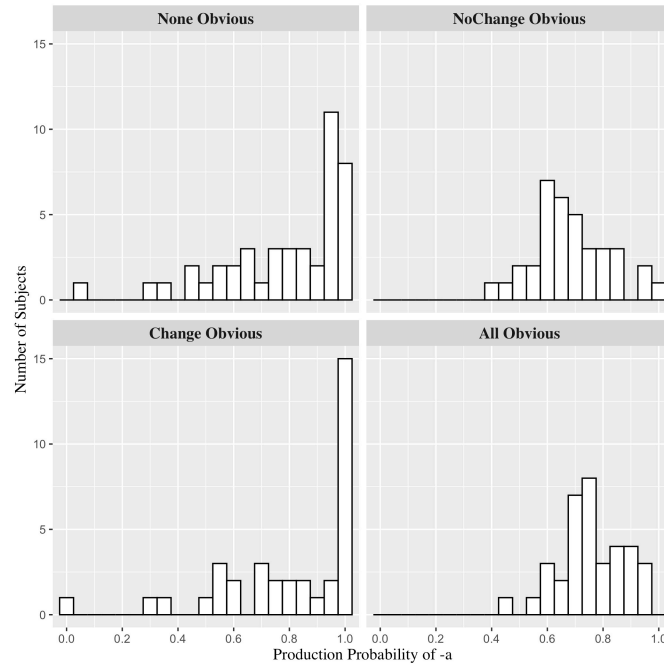
Figure 5.4. Suffix choice probabilities across trial order conditions. In the 'NoChange Obvious' and 'All Obvious' conditions, word pairs exemplifying 0{i;a}/$C_{not\text{-}TBP}$_ were kept intact. In the 'Change Obvious' and 'All Obvious' conditions, word pairs exemplifying C_{TBP}tʃa were kept intact.

Table 5.3[55] shows that keeping Not-To-Be-Palatalized pairs (faithful forms, 0{a;i}/$C_{not\text{-}TBP}$_) intact significantly increases the use of -*i* (the effect is significant in model comparisons, $\chi^2(1) = 17.20$, $p < 0.001$). There is also a significant effect of To-Be-Palatalized, which favors -*a*, especially when To-Be-Palatalized pairs are kept intact in training. Including the interaction significantly improves model fit ($\chi^2(1) = 7.45$, $p = 0.006$), as does To-Be-Palatalized alone ($\chi^2(1) = 23.45$, $p < 0.001$). In sum, participants

[55] Plural Vowel ~ No Change Order + Change Order * To-Be-Palatalized + (1 + To-Be-Palatalized | Subject) + (0 + No Change Order | Base) + (0 + Change Order | Base), restricted to non-palatal-final singular stems

learn that *-a* is favored by the consonants it palatalizes, and they learn it better when pairs

exemplifying *-a* combined with palatalization are intact.

Table 5.3. Generalized linear mixed-effects model output for suffix choice by trial
adjacency and To-Be-Palatalized.

	b	se(b)	z	p	
(Intercept)	-1.997	0.2382	-8.385	<.0001	***
Intact Not-To-Be-Palatalized Pairs	1.1722	0.2793	4.196	<.0001	***
Intact To-Be-Palatalized Pairs	-0.2887	0.2827	-1.021	0.30712	
To-Be-Palatalized = yes	-0.3191	0.1354	-2.357	0.01842	*
Adjacent To-Be-Palatalized Pairs x To-Be-Palatalized = yes	-0.5391	0.1927	-2.798	0.00514	**

* Significance level of 0.05
** Significance level of 0.01
*** Significance level of 0.001

5.2.2.2. Consonant choice effect

Figure 5.5 shows the factors that influence the choice of suffix vowel using a

conditional inference tree. The effect of consonant choice (whether the stem is To-Be-

Palatalized) on suffix choice is weaker than in the input language (and completely absent

for subjects who regularized). The 103 subjects who use each suffix more than 10% of

the time show sensitivity to the To-Be-Palatalized status of the consonant, with To-Be-

Palatalized disfavoring *-i* but still using it 26% of the time (vs. 0% in the input) and Not-

To-Be-Palatalized stems using *-i* 36% of the time (vs. 50% in the input). The effect of

To-Be-Palatalized[56] is significant in the mixed-effects regression model (b = -0.28, *se(b)*

= 0.14, z = -2.06, p < 0.04; including To-Be-Palatalized improves model fit, $\chi^2(1)$ =

21.62, p < 0.001), as is the interaction between To-Be-Palatalized and whether the

---

[56] Plural Vowel ~ Change Order * To-Be-Palatalized + No Change Order + (1 + To-Be-Palatalized |
Subject) + (1 + Change Order + No Change Order | Base), data restricted to subjects who used each suffix
>10% of the time and stems ending in a non-palatal consonant.

unfaithful pairs are kept intact (b = -0.55, $se(b)$ = 0.21, z = -2.66, p = 0.008; including the interaction significantly improves model fit, $\chi^2(1)$ = 6.82, p = 0.009).
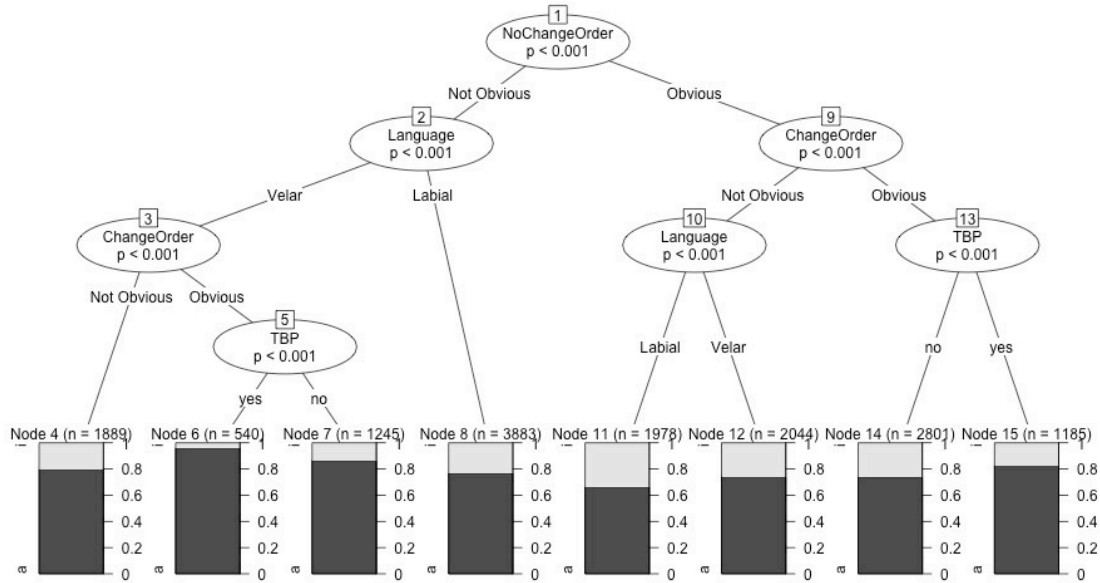
Figure 5.5. Conditional inference tree of the factors that influence suffix vowel choice.

5.2.2.3. Other effects

The strongest effect is whether faithful pairs are kept intact (NoChangeOrder, Node 1): doing so increases the frequency of *-i*, though it never becomes the majority variant. Keeping C_{TBP}~tʃa pairs intact is also a significant effect (Nodes 3 and 9, ChangeOrder), with consonants showing increased use of *-a* in the All Obvious (Node 13) and Change Obvious (Node 5) conditions, where To-Be-Palatalized pairs are kept intact. The effect is especially pronounced for To-Be-Palatalized consonants (Nodes 6 and 15). Finally, training on velar palatalization decreases the use of *-i* compared to training on labial palatalization (Nodes 2 and 10, TrainPlace).

5.2.3. Suffix content

5.2.3.1. Trial order effects

Figure 5.6 shows the production probabilities of palatalization across subjects, and as in Figure 5.4, there is a clear effect of trial order. Keeping pairs exemplifying palatalization intact leads to more palatalization[57] (Change Obvious and All Obvious, bottom row) ($b = 4.50$, $se(b) = 0.69$, $z = 6.52$, $p < 0.001$; including Unfaithful Intact significantly improves model fit, $\chi^2(1) = 47.38$, $p < 0.001$). There is also a significant effect of whether faithful pairs are kept intact (NoChange Obvious and All Obvious, right column), with less palatalization when they are ($b = -1.30$, $se(b) = 0.64$, $z = -2.03$, $p = 0.04$; Faithful Intact significantly improves model fit, $\chi^2(1) = 4.17$, $p = 0.04$). The interaction is not significant ($z < 1$, $p = 0.73$, ns), and according to the BIC approximation to the Bayes Factor, the data provide strong support for the null ($\Delta BIC = 8.1$, $P_{BIC}(H_0 \mid D) = 0.983$).

---

[57] Palatalization ~ Change Order + No Change Order + (1 | Subject) + (1 | Base), restricted to To-Be-Palatalized consonants before -*a*
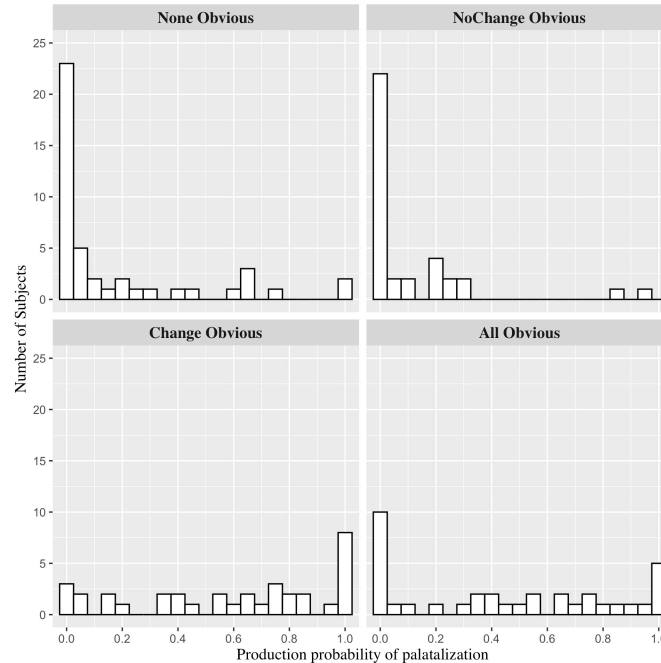
Figure 5.6. Palatalization rates across conditions of the appropriate consonants in the appropriate context (a To-Be-Palatalized consonant before -*a*). In the "Change Obvious" and "All Obvious" conditions, word pairs exemplifying palatalization were kept intact in training. In the "NoChange Obvious" and "All Obvious" conditions, word pairs exemplifying faithful mappings (lack of palatalization) were kept intact.

5.2.3.2. Suffix vowel effects

Participants that use each suffix more than 10% of the time show a significant effect of plural vowel on the probability of palatalization. In line with the training data, palatalization is favored by -*a* and disfavored by -*i*[58] ($b = -1.35$, *se(b)* $= 0.27$, $z = -4.94$, $p < 0.001$; Plural Vowel significantly improves model fit, $\chi^2(1) = 29.83$, $p < 0.001$). There is no interaction between final vowel and trial order, which is to be expected: The suffix vowel occurs in the same form as the palatalization, so the dependency between them should be equally apparent regardless of trial ordering. (However, the interaction between final vowel and adjacency of unfaithful (To-Be-Palatalized) pairs approaches

---

[58] Palatalization ~ Change Order + No Change Order + Plural Vowel + (1 + Plural Vowel | Subject) + (1 + Change Order + No Change Order + Plural Vowel | Base), restricted to subjects who palatalized.

significance, $z = -1.87$, $p = 0.06$.) These effects are shown in the conditional inference tree in Figure 5.7.
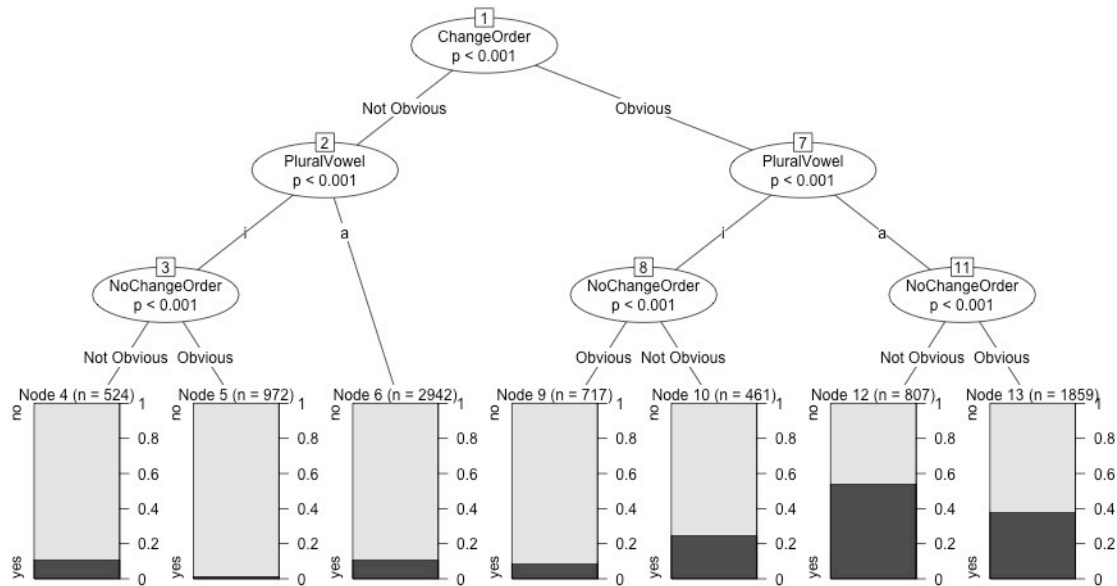
Figure 5.7. Conditional inference tree of the effects of vowel suffix and trial order on the probability of palatalization (dark) for subjects who used each suffix vowel >10% of the time.

5.2.4. Input consonant

Finally, participants need to learn not only that palatalization exists, and that it is triggered by -*a*, but also that it only affects certain consonants (namely the labial [p] and [b] in the Labial Palatalization condition and the velar [k] and [g] in the Velar Palatalization condition). In previous work using the None Obvious trial order, this proved difficult to learn, with subjects tending to overgeneralize velar palatalization to the alveolar stops and labial palatalization to both alveolar and velar stops, when they were able to learn to palatalize at all (Smolek & Kapatsinski, 2018; Stave et al., 2013; see also Chapter IV). We were curious here about whether participants could learn to restrict palatalization to the appropriate context if singular-plural mappings were made more

133

obvious by changing the trial order. The results below are for the -*a* suffix, which triggers palatalization in the training and strongly favors palatalization at test, but there are no crossover interactions between vowel and any of the predictors we discuss, and the results hold if both vowel contexts are included.

5.2.4.1. Trial order effects on overgeneralization of palatalization

Figure 5.8 shows that faithful and unfaithful mappings both benefit from keeping the pairs exemplifying them intact, and keeping both types of mappings intact is necessary for participants to learn when they should be unfaithful to the input. To test for the effects of trial order on palatalization rates and the probability of overgeneralization to Not-To-Be-Palatalized consonants across languages, we restricted the analysis to To-Be-Palatalized consonants and the Not-To-Be-Palatalized consonants that palatalization is overgeneralized to in the None Obvious order. In other words, we excluded labial-final stems from the set of Not-To-Be-Palatalized consonants for Velar Palatalization, because subjects do not overgeneralize to it in the None Obvious condition so there is no way for adjacency to help (but the results are the same if they are included).
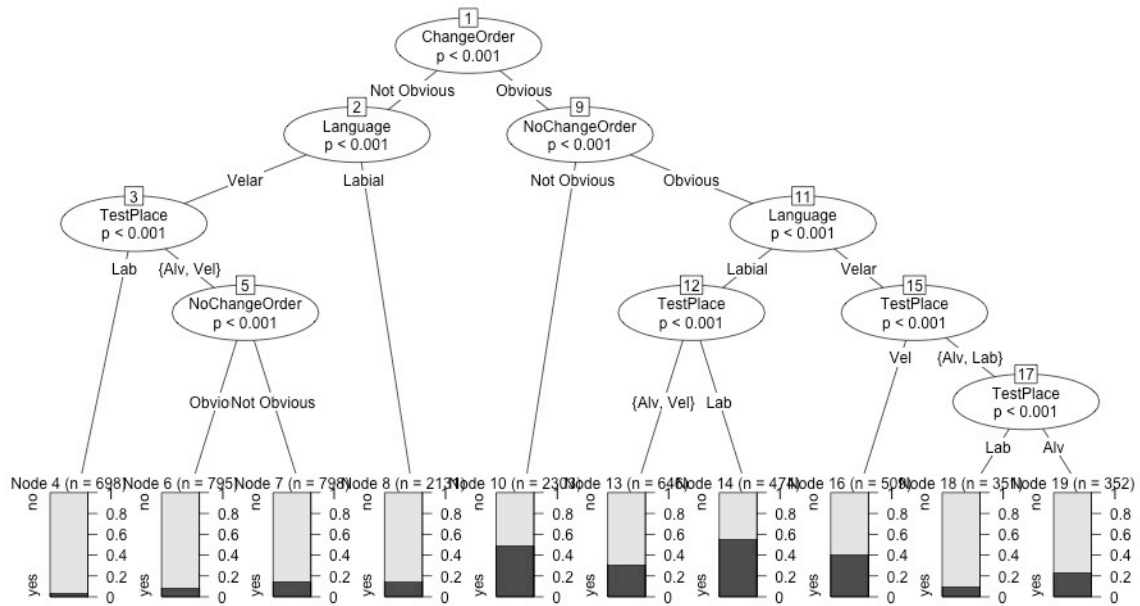
Figure 5.8. An overview of the results on the probability of palatalization (before *-a*).

ChangeOrder (Intact Unfaithful) has the strongest effect on the probability of palatalization, increasing its incidence: Palatalization is more productive when clearly visible in training. When word pairs exemplifying palatalization are not kept intact (the subtree below Node 2), participants trained on velar palatalization palatalize alveolars and velars equally. Participants trained on labial palatalization palatalize everything equally. These results replicate Chapter IV: Participants eliminate saltatory alternation patterns (see also White, 2013, 2014). When only the word pairs exemplifying palatalization are kept intact (Nodes 9-10), participants palatalize all consonants equally, even when trained on velar palatalization. That is, with intact To-Be-Palatalized and non-intact Not-To-Be-Palatalized pairs (Change Obvious condition), participants palatalize both types of consonants at a high rate. However, when all word pairs are kept intact (All Obvious, the subtree below Node 11), participants are able to learn to palatalize only the

135

consonants they are trained to palatalize. That is, keeping Not-To-Be-Palatalized pairs

intact reduces their palatalization rates.

The results are shown in Table 5.4[59]. In the baseline None Obvious condition,

participants palatalize Not-To-Be-Palatalized stems as much as To-Be-Palatalized,

reflecting a bias in favor of palatalizing alveolars and against palatalizing labials ($z < 1$, p

$= 0.33$, ns), which replicates prior work (Smolek & Kapatsinski, 2018; Stave et al., 2013;

Chapter IV). Keeping pairs exemplifying unfaithful mappings intact increases

palatalization overall ($\chi^2(1) = 44.20$, $p < 0.001$), and the effect is stronger for To-Be-

Palatalized than Not-To-Be-Palatalized consonants ($\chi^2(1) = 24.09$, $p < 0.001$). Keeping

pairs exemplifying faithful mappings intact reduces palatalization rates ($\chi^2(1) = 6.62$, $p =$

0.01), especially for Not-To-Be-Palatalized consonants ($\chi^2(1) = 18.03$, $p < 0.001$).

Because of the interactions, keeping both types of mappings intact reduces

overgeneralization: Participants learn to palatalize only the To-Be-Palatalized

consonants.

Table 5.4. The influence of trial order on palatalization rates.

	b	se(b)	z	p	
(Intercept)	-3.0428	0.3954	-7.695	<.00001	***
Intact To-Be-Palatalized Pairs	3.4397	0.462	7.445	<.00001	***
To-Be-Palatalized = no	0.134	0.1377	0.973	.33	
Intact Not-To-Be-Palatalized Pairs	-0.8616	0.4515	-1.908	.056	
Intact To-Be-Palatalized Pairs x To-Be-Palatalized = no	-0.7212	0.147	-4.907	<.00001	***
Intact Not-To-Be-Palatalized Pairs x To-Be-Palatalized = no	-0.5945	0.1401	-4.242	.00002	***

*** Significance level of 0.001

---

[59] Palatalization ~ Change Order * To-Be-Palatalized + No Change Order * To Be Palatalized + (1 | Subject) + (1 | Singular), restricted to non-palatal-final singulars before -*a* for Labial Palatalization subjects and non-palatal, non-labial-final singulars before -*a* for Velar Palatalization subjects.

5.3. Summary

Retaining adjacency of word pairs exemplifying a given paradigmatic mapping helps faithful *and* unfaithful mappings, stem changes triggered by particular affixes, and the affixes themselves. For example, many participants in the None Obvious and Change Obvious conditions never use *-i*, even though it occurs 25% of the time in training and adults tend to probability match in tasks like this (Harmon & Kapatsinski, 2017; Schwab et al., 2018), whereas participants in the All Obvious and NoChange Obvious conditions use *-i* at roughly the same rate as in training. The difference is that the latter groups encountered intact faithful pairs, where the addition of *-i* is clearly seen. Similarly, many participants in the None Obvious and NoChange Obvious never palatalize, even though half of the input plurals contain palatalization/end in *-tʃa*, vs. participants in the All Obvious and Change Obvious conditions, who have seen pairs like *blup~blutʃa*. Contiguity makes the mapping more obvious and helps the learner notice it.

However, making the mapping obvious does not prevent overgeneralization to contexts not seen in training. Rather, overgeneralization of a mapping to a particular consonant is prevented by making the competing mapping obvious: Making p→tʃ obvious increases t→tʃ and k→tʃ, and it is only when t→t and k→k are also made obvious that palatalization is properly restricted to [p].

In sum, temporal adjacency helps acquisition of both faithful and unfaithful mappings, and both are necessary to learn when each should apply. These results therefore require participants to generalize over pairs of corresponding forms to learn both faithful and unfaithful mappings. Unfaithful mappings cannot be attributed entirely to first-order / product-oriented schemas acquired by generalizing over forms within a

137

paradigm cell. Faithful mappings cannot be attributed entirely to a default "do nothing" rule or top-ranked output-output faithfulness constraints.

In Chapter VI, we discuss a discriminative model that captures the results of Experiment 3, suggesting that temporal contiguity of related forms allows learners to use phonological characteristics of a singular form to predict the characteristics of the corresponding plural form. This allows participants to form singular-to-plural paradigmatic mappings that encode what kinds of singulars are mapped onto …*tʃa* in the plural. Implications for linguistic theory and learnability of morphology are discussed further in §6.4, and comparison of the results of Experiments and 3 is presented in Chapter VII.

CHAPTER VI

COMPUTATIONAL MODEL: EFFECTS OF ADJACENCY ON LEARNABILITY OF

PALATALIZATION BEFORE -*a*

Portions of this chapter were taken from:

Smolek, A. & Kapatsinski, V. (2019). *Syntagmatic paradigms: Learning correspondence from contiguity.* Manuscript submitted for publication.

In Chapter V, we discussed the results of an experiment on the learnability of labial and velar palatalization before -*a*, varying whether pairs exemplifying faithful and unfaithful mappings were kept intact in training. We found that obvious examples of -*i* suffixation (present when faithful pairs, like *blut~bluti*, were kept intact) increase the usage of -*i* in production; that adjacency helps both faithful and unfaithful mappings and that participants are more likely to produce palatalization before -*a* than -*i* after training. In Chapter VI, we explore these effects through computational modeling in order to discover the potential learning mechanisms behind them. We use a custom function in R (R Core Team, 2019) based on the original code implementing the Rescorla-Wagner (1972) discriminative learning model, obtained from R.H. Baayen. All code and data are available at https://app.box.com/s/bd8jhx4g5m7bvlmxb8i4jgtjfjo2x111.

6.1. Discriminative learning

Research over the last 30 years has explored the possibility that domain-general learning mechanisms could explain much of language acquisition. Morphology is a particularly fruitful area for exploration due to the amount of intricate and language-specific structure; it is also the only domain where paradigmatic mappings are clearly necessary (Kapatsinski, 2018a, 2018b). It began with Rumelhart & McClelland (1986),

and subsequent work includes Albright & Hayes (2003), Baayen et al. (2011), Ramscar et al. (2013), and Westermann & Ruh (2012). We focus here on discriminative learning.

Discriminative learning was originally applied to classical conditioning in animals. Here (as elsewhere, cf. Baayen et al., 2011; Caballero & Kapatsinski, 2019; Kapatsinski, 2017b, 2018b; Ramscar et al., 2013), we apply it to morphological learning. The fundamental structure of discriminative learning is that it is *predictive*, i.e. the learner aims to make predictions, based on environmental *cues* (e.g. Pavlov's (1927) dogs, who used lights and sounds to predict whether food would be available). In the Rescorla-Wagner model, the strength of association from a particular cue to an outcome tracks the statistic Δp, which is the probability of the outcome when the cue occurs minus the probability of the outcome when the cue does not occur (Ellis, 2006; Harmon & Kapatsinski, 2017). The strongest cues are those that greatly increase the outcome's likelihood of occurrence, in other words, the ones that best *discriminate* between the contexts where an outcome occurs and where it does not.

The model increases on a trial-by-trial basis the weights of cues present when (or before) the outcome unexpectedly occurs, and the more surprising the outcome, the greater the increase in cue weights. Cue weights are decreased when the outcome is unexpectedly absent in the presence of the cues. Only weights of present cues are updated. Equation 1 applies for present outcomes and equation 2 for absent; w is the cue weight (the weight of the connection from cue to outcome), and a is the activation of the outcome (the sum of the present cue weights, equivalent to the model's expectation whether the outcome will (not) occur at time t). When the outcome does (not) occur, the model updates its expectations for the future. The weight w of the cue-outcome

connection at time t is incremented by an amount proportional to learning rate Λ (set to 1), multiplied by the difference between the correct expectation (1 for present, 0 for absent) and the model's actual expectation given current weights. If the model's prediction is accurate, there is no need to update the weights. Whenever the prediction is inaccurate, the weights must be updated, and the more confident it is in its erroneous prediction, the greater the adjustment to the weights. This makes the model *error-driven*.

(1) $w_{t+1}^{C \to O} = w_t^{C \to O} + (1 - a_t^O) \times \Lambda$

(2) $w_{t+1}^{C \to O} = w_t^{C \to O} + (0 - a_t^O) \times \Lambda$

6.2. Model design

6.2.1. Relevant cues and outcomes

To capture the effects of plural production Experiment 3, the relevant cues are the meaning of the form (PL) and the stem-final consonant of the base, singular form and the relevant outcomes are the stem-final consonant of the plural form and the suffix vowel (e.g. $blut_{SG}$~$bluta_{PL}$ yields cues PL and 't' and outcomes 't' and 'a').

6.2.2. Capturing implicational hierarchies

To capture the effects of Change Order (whether unfaithful pairs were kept intact) on overgeneralization, we include the assumption that encountering k→tʃ supports t→tʃ but not p→tʃ, and encountering p→tʃ supports t→tʃ and k→tʃ. This is a common assumption for capturing implicational hierarchies in patterns of overgeneralization (Hayes & White, 2015; Steriade, 2008). To capture this representationally, we assume that an example like $bluk_{SG}$~$blutʃa_{PL}$ contains the cues 't' and 'k' and PL, and an example like $blup_{SG}$~$blutʃa_{PL}$ contains 'k' and 't' and 'p' and PL. See §6.3.2.1 and §6.3.2.2 for further discussion.

6.2.3. Trial order effects on cue availability

The trial order manipulations of Experiment 3 influence the availability of cues for predicting the phonological shape of the corresponding plural form when it is encountered. When related singulars and plurals are adjacent, the stem-final consonant is available to predict the plural form. When they are not adjacent, only the plural meaning is available. For example, when *blutʃa*$_{PL}$ is preceded by anything other than the singular, only PL is available to predict 'tʃ' and 'a', so only PL gets updated. This means that in the None Obvious condition, singular-final weights are never updated and reflect the prior beliefs that subjects bring to the experiment.

6.2.4. Prior beliefs

Participants' prior beliefs are represented by the initial connection weights in the model. They include that subjects are least reluctant to palatalize alveolars and most reluctant to palatalize labials, with velars falling in between. This may be due in part to English palatalization of alveolars and velars in word pairs like *create~creature* and *legal~legislative*, and alveolar palatalization in frequent phrases like *would you*. Although the word-internal alternations only occur in a few words and are of doubtful productivity, and the cross-boundary alternations may be a different process (Zsiga, 1995), the presence of these processes in English may make participants more willing to palatalize alveolars and velars. The bias could instead (or additionally) be due to cross-linguistic biases against large changes, whether the measure of magnitude is perceptual (Hayes & White, 2015; Steriade, 2008) or articulatory (Smolek & Kapatsinski, 2018; Chapter IV). We remain agnostic about the source of the bias for the present work and encode it by mapping alveolar, velar, and labial stops onto themselves, pre-training the model with 2

142

examples of t➔t (weight of 0.19), 4 examples of k➔k (weight of 0.34), and 6 examples of p➔p (weight of 0.47). This is not intended to be a representation of English, merely a way of 'telling' the model to have a strong expectation of [p] remaining [p] and a weak expectation of [t] remaining [t].

We employ a traditional two-stage architecture, with morphology preceding phonology, which is widely assumed by generative linguistics (see Scheer, 2011, for a review). Recent theoretical and experimental work has questioned this assumption and instead proposed that a suffix and the stem change thought to be triggered by it are in fact chosen in parallel[60] (Bybee, 2001; Kapatsinski, 2010; Prince & Smolensky, 1993/2004).

The modeling results below are consistent with suffix choice initiating first, or suffix and stem change initiating simultaneously. However, in order for the suffix and the stem change to not be independent (which they should not be, given the relationship in the training data between suffix vowel and presence of palatalization), one process needs to complete first so that it can condition the other. Any choice is the outcome of a process that takes time, with harder decisions requiring more time (Usher & McClelland, 2001). Our results are consistent with suffix choice requiring less time than palatalization.

If the stem allomorph were chosen first, then the stem-final consonant would be available to select the suffix vowel, regardless of the proximity of the singular form. However, suffix vowel choice is predictive of palatalization regardless of trial order, but trial order is predictive of suffix choice; when the stem-final consonant is selected first, the model predicts no influence of trial order on suffix choice, beyond its effect on the

---

[60] OT and Harmonic Grammar consider competing output forms, which obviates the need to order processes, but here we only generate one form, so that is not an option.

143

choice of stem-final consonant. We suspect the suffix vowel is (usually) chosen first, perhaps because it is more salient in the final position.

A third possibility we considered was to separate the suffix and stem-final consonant and treat them as bigrams (e.g. [tʃa], [ki]), but the dependency between vowel and consonant is stronger in the input than in subject productions. If participants were choosing from a set of bigrams, they would never produce unattested combinations like [ka] in the Velar Palatalization condition, but they do (quite often, in fact).

Perhaps all three processing routes are used, and speakers choose based on how easy it is to select the suffix vs. the stem-final consonant (or change) on any given occasion (Kapatsinski, 2013; Westermann & Ruh, 2012). Selection of stem-final consonant may be faster than suffix choice when the plural and the singular have the same stem allomorph (i.e. the mapping is faithful), but here, choosing the suffix before the stem 100% of the time captures all of the effects we see in the experimental data.

Other criteria for model selection could lead to differing conclusions: The best stem-before-vowel model fits the palatalization probability across the cells of the experimental design better (pseudo-R^2 = 86% vs. 96%), but this ordering fails to capture the dependency between trial order and vowel choice (§5.2.1). It is possible that alternating between routes, dependent on context, could capture the significant patterns without sacrificing fit. Since we were only interested in explaining the mechanisms behind trial order effects rather than the structure of the grammar, the latter must be explored in future work.

6.2.5. Linking hypothesis

The activation levels must be connected to the observable dependent variable of production probability, which necessitates a theory of the task participants perform in the experiment. We follow the Luce choice rule in (3) (Luce, 1959), where the probability of choosing response *i* tracks the ratio of the activation of *i* and the activation of all response choices. This captures the probability matching seen in open-set tasks like production, and the equalizing/irregularization behavior seen in closed-set forced choice tasks (Harmon & Kapatsinski, 2017; Luce & Pisoni, 1998).

(3) $p(i) = \frac{a_i}{\sum_{i=1}^{n} a}$

6.3. Modeling results

As discussed above (§6.2.1), the baseline model's cues include the stem-final consonant in the singular and the plural meaning, and the outcomes include the stem-final consonant in the plural and the suffix vowel. When the baseline model's expectations differ systematically from participants' behavior, we modify the model to better reflect the observed results. The modifications are additional cues and outcomes (§6.3.2.2.2) or adjustments to connection weights (§6.3.1.2, §6.3.2.2.2), discussed in detail below.

6.3.1. Suffix vowel choice

6.3.1.1. Baseline model results

As in Chapter V, we start with an exploration of the effects on suffix vowel choice. We compare the observed values (based on the participant data from Experiment 3) to the expected values (based on the model). As shown in Figure 6.1, the base model captures 49% of the variance. In particular, it captures that *-a* is disfavored by Not-To-Be-

Palatalized consonants, and that consonants have a weaker effect on subject productions than is present in the input data.
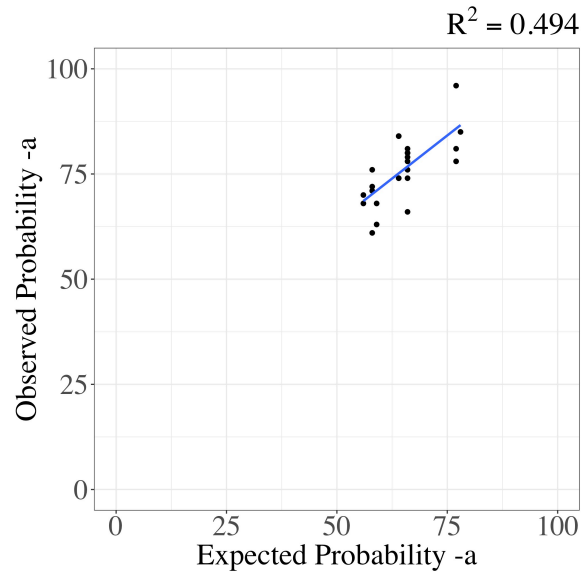
Figure 6.1. Expected (x-axis) vs. observed (y-axis) vowel choice probabilities, using unaltered probabilities from the model. Each dot corresponds to one cell in Table 6.1.

Table 6.1. Expected (model) / observed (experiment) production probabilities for the palatalizing suffix -*a* (vs. -*i*) across training language (Velar vs. Labial), place of articulation of the consonant at the end of the singular form, and trial order.

Velar Language	TBP	None Obvious	NoChange Obvious	Change Obvious	All Obvious
Alveolar	-	66% / 80%	59% / 68%	64% / 84%	58% / 71%
Velar	+	66% / 81%	66% / 76%	77% / 96%	77% / 81%
Labial	-	66% / 79%	58% / 72%	64% / 84%	56% / 70%
Labial Language					
Alveolar	-	66% / 74%	59% / 63%	64% / 74%	58% / 76%
Velar	-	66% / 78%	58% / 61%	64% / 74%	56% / 68%
Labial	+	66% / 80%	66% / 66%	77% / 78%	78% / 85%

As shown in Table 6.1, the model captures the significant effects of trial order on suffix choice probabilities. Making -*a* addition to To-Be-Palatalized stems more obvious increases the use of -*a* with To-Be-Palatalized consonants: For velars in the Velar Palatalization condition and labials in the Labial Palatalization condition, the Change Obvious and All Obvious conditions pattern alike and separate from None Obvious and

146

NoChange Obvious. It is also able to capture that making *-i* addition to Not-To-Be-Palatalized stems obvious increases its usage with Not-To-Be-Palatalized consonants: For labials and alveolars in the Velar Palatalization condition and alveolars and velars in the Labial Palatalization condition, NoChange Obvious and All Obvious pattern alike and separately from None Obvious and Change Obvious.

6.3.1.2. Shortcomings and modifications

The model does fall short in some important ways. Firstly, it underpredicts the use of *-a* overall; the Luce choice rule predicts probability matching, but subject responses fall between probability matching and maximizing (Harmon & Kapatsinski, 2017). Secondly, it fails to capture the difference in *-a* usage between training languages. Labial Palatalization subjects use *-a* less than Velar Palatalization subjects do, perhaps because it is associated with a difficult-to-produce alternation that they tend to avoid (a strategy employed by child learners as well; Schwartz & Leonard, 1982). Labial palatalization is very difficult to execute (Smolek & Kapatsinski, 2018; Stave et al., 2013; see also Chapter IV), so it is unsurprising that subjects avoid the suffix that triggers it. Fit improves if the production probability of *-a* is incremented by 4% in each language (to account for maximizing), and by an additional 6% in Velar Palatalization (to account for avoidance of a difficult change in Labial Palatalization). However, the effect of language is weak in the mixed-effects analysis of the human data and would not survive a Bonferroni correction ($b = 0.69$, *se(b)* = 0.29, $z = 2.37$, $p = 0.018$), and the confidence intervals for the observed-expected correlation coefficients for the original vs. adjusted models overlap substantially ($0.42 \leq r \leq 0.86$ vs. $0.54 \leq r \leq 0.90$).

6.3.2. Palatalization

6.3.2.1. Baseline model results

Tables 6.2 and 6.3 and Figure 6.2 show the probabilities of palatalization for the unmodified model across training language, training trial orders, place of articulation of stem-final consonant, and suffix. One of the most robust findings from this line of research is that labial palatalization overgeneralizes to alveolars and velars, and velar palatalization overgeneralizes to alveolars but not labials. Encoding the implicational hierarchy in the training data improves the model: Table 6.2 shows that in the None Obvious condition, Labial Palatalization participants palatalize alveolars and velars more than labials, and Velar Palatalization participants palatalize alveolars as often as velars, but do not palatalize labials (see also Smolek & Kapatsinski, 2018; Stave et al., 2013; Chapter IV). Only in the All Obvious condition do participants learn to palatalize only the To-Be-Palatalized consonants (at least, more than the Not-To-Be-Palatalized consonants that are articulatorily or perceptually closer to [tʃ]). We must assume that palatalization of [k] provides evidence for t→tʃ, and palatalization of [p] provides evidence for t→tʃ and k→tʃ, because if they only provide evidence for palatalizing themselves, the predicted palatalization rates of alveolars and velars in the Labial Palatalization condition are far too low (and the R^2 is only 46%).

Table 6.2. Expected / observed palatalization rates before -*a*, the palatalizing suffix in the input. Expectations from the unmodified Rescorla-Wagner model. Large deviations from human behavior are bold.

Velar Language	TBP	None Obvious	NoChange Obvious	Change Obvious	All Obvious
Alveolar	-	22% / 13%	20% / 6%	**28% / 47%**	22% / 23%
Velar	+	22% / 17%	25% / 10%	*42% / 56%*	*47% / 40%*
Labial	-	15% / 5%	14% / 1%	**14% / 32%**	11% / 10%
Labial Language					
Alveolar	-	22% / 23%	19% / 15%	**31% / 49%**	22% / 34%
Velar	-	19% / 16%	16% / 9%	**29% / 47%**	19% / 28%
Labial	+	17% / 13%	19% / 16%	*38% / 57%*	*38% / 56%*

Table 6.3. Expected / observed palatalization rates before -*i*, the suffix that does not trigger palatalization in the input. Expectations from the unmodified Rescorla-Wagner model. Large deviations from human behavior are bold.

Velar Language	TBP	None Obvious	NoChange Obvious	Change Obvious	All Obvious
Alveolar	-	8% / 9%	5% / 2%	7% / 1%	13% / 5%
Velar	(+)	8% / 13%	10% / 2%	*26% / 30%*	*26% / 9%*
Labial	-	6% / 1%	3% / 0%	5% / 1%	4% / 3%
Labial Language					
Alveolar	-	8% / 12%	6% / 2%	**14% / 31%**	7% / 13%
Velar	-	7% / 14%	4% / 1%	**14% / 34%**	6% / 11%
Labial	(+)	6% / 10%	8% / 2%	*22% / 24%*	*22% / 19%*

$R^2 = 0.657$
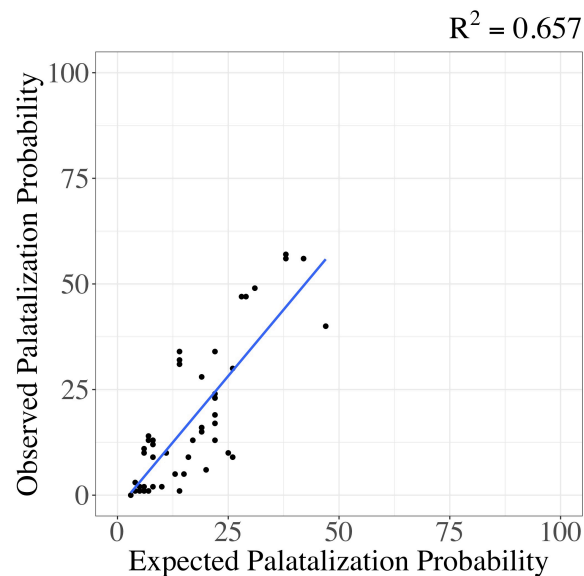
Figure 6.2. Expected (x-axis) vs. observed (y-axis) palatalization probabilities. Each dot corresponds to one cell in Tables 6.2-6.3.

As shown in Figure 6.2, the model succeeds at fitting the proportions in Tables 6.2 and 6.3, in particular capturing the difference between To-Be-Palatalized and Not-To-Be-Palatalized consonants, the difference in palatalization rates between *-i* and *-a* suffixes, and the increased palatalization rates of Not-To-Be-Palatalized consonants in the Labial Palatalization condition. It is also quite successful at capturing the ordinal effects of trial order, with regards to the relationship between conditions and interactions between trial order and the (Not-)To-Be-Palatalized status of consonants.

6.3.2.2. Shortcomings and modifications

The model also differs systematically from human learners. It sometimes underestimates increases in palatalization of Not-To-Be-Palatalized stems in the Change Obvious conditions vs. the All Obvious and None Obvious conditions (bold non-italic in Tables 6.2 and 6.3). In particular, note the bottom rows of None Obvious and Change Obvious in the Velar Palatalization language in Table 6.2: Making velar palatalization obvious dramatically increases the rate of labial palatalization (from 5% to 32%), but it has no effect on model expectations (15% vs. 14%). Note also that this increase in palatalization rates is specific to the palatalizing suffix. The palatalization rates of Not-To-Be-Palatalized consonants increase in Change Obvious before *-a* (Rows 1 and 3 in Table 6.2) but not *-i* (Rows 1 and 3 in Table 6.3). The model is correct about the pre-*i* context but not the pre-*a*.

6.3.2.2.1. Perceptual contrast and chunking

To capture the effect of obvious unfaithful alternations on palatalization rates of Not-To-Be-Palatalized consonants before *-a*, we propose that encountering bluk→blutʃa or blup→blutʃa boosts the association between [tʃ] and [a]. In other words, noticing that

something changes into [tʃa] makes it easier for [a] to trigger a preceding [tʃ] in general: -a fuses with [tʃ], increasing the automaticity of palatalization before -a, which we implement as a boost with constant magnitude (λ). The perceptual contrast of [k] or [p] becoming [tʃa] makes the listener notice the [tʃa] and increment the association between [tʃ] and [a], resulting in [tʃa] being treated as a single chunk. Alternatively, the boost could be dependent on the surprisal value of [tʃa], such that the first encounter increases the association between [tʃ] and [a] substantially with subsequent encounters increasing it less. Regardless, the boost is *in addition* to the boost driven by the surprise at an unexpected [tʃ] from the Rescorla-Wagner model. This can also explain the cross-linguistic process of suffixes fusing with the stem-final consonants that frequently precede them (and which are themselves usually the result of a stem change triggered by the suffix; Haspelmath, 1995).

6.3.2.2.2. Overgeneralization asymmetries in To-Be-Palatalized consonants

Finally, we need to explain why To-Be-Palatalized consonants are palatalized less in the All Obvious condition vs. the Change Obvious condition for Velar Palatalization, but there is no difference between conditions for Labial Palatalization (bold italic in Tables 6.2 and 6.3). We propose this difference arises because of overgeneralization of Copy$_{FIN}$, the relevant outcome for faithful mappings, which is taken to apply to the smallest natural class that includes all copied segments; in other words, faithful mappings are produced via a constraint like "Copy$_{FIN}$ applies when the coda consonant is [X]," where [X] is a the natural class including all segments subject to being copied. For Labial Palatalization, stem-final alveolars and velars are associated with Copy$_{FIN}$, and they are both made using parts of the tongue that are not independently controlled until quite late in development

151

(Gibbon, 1999), frequently pattern together (Christdas, 1988; Clements & Hume, 1995; Mielke, 2004), and have been argued to form the natural class [lingual] (Browman & Goldstein, 1989; Christdas, 1988; Clements & Hume, 1995). The overgeneralization of copying to velars in Velar Palatalization suggests that alveolars and labials do not form a natural class (Clements & Hume, 1995; though cf. Chomsky & Halle, 1968, and [anterior]), so any class that includes both must also include velars, and therefore $Copy_{FIN}$ spreads to the To-Be-Palatalized in Velar Palatalization All Obvious. We can augment trials containing [k] and [t] with an additional [lingual] cue (i.e. t+i+lingual → t+copy) to capture that they are alike and separate from labials, which accounts for the lack of overgeneralization of copying in Labial Palatalization. However, the model does not make global inferences, even with the inclusion of [lingual], because it aims to distinguish how inputs behave differently (i.e., to *discriminate* among them), not what features they share. As such, we added a mechanism to increment the weight of $Copy_{FIN}$ by λ whenever the natural class of copied segments comprises all consonants in the language (i.e. in the Velar Palatalization language). This is a post-hoc hack intended to capture the Hebbian learning mechanism that wires co-occurring cues together, regardless of surprise, which slightly reduces the rate of palatalization of To-Be-Palatalized consonants in the Velar All Obvious condition. The evidence is somewhat questionable, as the effect on the model fit is quite small, and eliminating the mechanism still achieves an R^2 of 80% so long as [lingual] is accessible. But without it, the model cannot reproduce the significant differences between palatalization rates of To-Be-Palatalized consonants in the Velar All Obvious and Change Obvious conditions. See

152

Tables 6.4 and 6.5 for the expected and observed values for the modified model and

Figure 6.3 for the scatterplot after adjusting the model.

Table 6.4. Expected / observed palatalization rates before -*a*, the palatalizing suffix in the input. Expectations from the modified Rescorla-Wagner model.

Velar Language	TBP	None Obvious	NoChange Obvious	Change Obvious	All Obvious
Alveolar	-	21% / 13%	15% / 6%	36% / 47%	20% / 23%
Velar	+	21% / 17%	18% / 10%	49% / 56%	31% / 40%
Labial	-	15% / 5%	10% / 1%	22% / 32%	14% / 10%
Labial Language					
Alveolar	-	21% / 23%	12% / 15%	50% / 49%	32% / 34%
Velar	-	21% / 16%	14% / 9%	44% / 47%	29% / 28%
Labial	+	15% / 13%	9% / 16%	46% / 57%	41% / 56%

Table 6.5. Expected / observed palatalization rates before -*i*, the suffix that does not trigger palatalization in the input. Expectations from the modified Rescorla-Wagner model.

Velar Language	TBP	None Obvious	NoChange Obvious	Change Obvious	All Obvious
Alveolar	-	8% / 9%	4% / 2%	6% / 1%	1% / 5%
Velar	(+)	8% / 13%	8% / 2%	20% / 30%	12% / 9%
Labial	-	6% / 1%	2% / 0%	1% / 1%	0% / 3%
Labial Language					
Alveolar	-	8% / 12%	6% / 2%	27% / 31%	8% / 13%
Velar	-	7% / 14%	4% / 1%	22% / 34%	6% / 11%
Labial	(+)	6% / 10%	8% / 2%	26% / 24%	20% / 19%
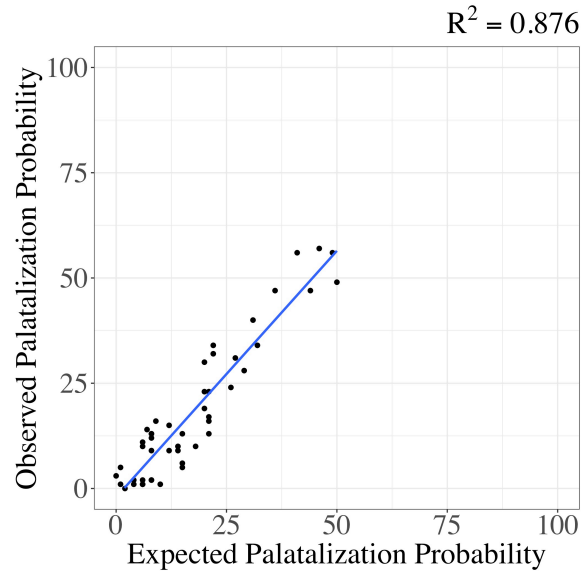
Figure 6.3. Expected (x-axis) vs. observed (y-axis) palatalization probabilities after model adjustments. Each dot corresponds to one cell in Tables 6.4-6.5.

6.4. General discussion

6.4.1. Implications for learning

6.4.1.1. Discriminative framework

To review, a simple discriminative learning model (Rescorla & Wagner, 1972) can

capture trial order effects and the interaction of trial order with training language. These

results are consistent with extensive prior work showing that language learning can be

captured using domain-general associative learning models originally developed for

conditioning experiments (Arnold et al., 2017; Arnon & Ramscar, 2012; Baayen et al.,

2011, 2016; Kapatsinski & Harmon, 2017; Kruschke, 1992; Lim et al., 2014; McMurray

et al., 2012; Mirman et al., 2006; Olejarczuk et al., 2018; Ramscar et al., 2010, 2013;

Ramscar & Yarlett, 2007; Rumelhart & McClelland, 1986; Yu & Smith, 2012; inter alia).

In a discriminative learning framework, outcomes are predicted from cues, so the

availability of cues should be critical to whether learners use the cue as a predictor of the

outcome (MacWhinney et al., 1985). In the present work, the outcomes are the varying

154

plural forms, and the trial order manipulations are intended to influence the availability of the cues contained in the corresponding singular form. The model captures the effects of the manipulation, showing that the manipulation was successful: The cues in the singular are used to predict the plural when the singular is adjacent to the plural. When the singular and plural are not adjacent, then the cues contained in the singular are largely or entirely unavailable for predicting outcomes, and only the semantics can be used. This proves true for faithful *and* unfaithful singular-plural mappings, concatenative suffixes, and stem changes.

6.4.1.2. Saliency and adjacency

A surprising illustration of the importance of contiguity is visible in Figure 5.1 (§5.2.1; see also Figure 6.1). Many None Obvious and Change Obvious participants always choose *-a*, which might be optimal behavior because it better matches the unconditioned suffix probabilities, and in the absence of obvious examples of *-i*, participants do not acquire the probabilistic conditioning of suffix choice. The participants in the NoChange Obvious and All Obvious conditions, however, tend to probability match. This is what we would normally expect them to do, since participants believe their task is to reproduce the training data as accurately as possible (Harmon & Kapatsinski, In prep; Perfors, 2016), and probability matching is observed in a variety of language production tasks, both artificial and natural (Harmon & Kapatsinski, 2017; Hayes et al., 2009; Hudson Kam & Newport, 2005; Kapatsinski, 2010; Ramscar & Yarlett, 2007). These results indicate that the addition of the suffix needs to be noticed (just as changes to the stem need to be noticed) in order for probability matching to occur. When the plural forms ending in a suffix are not adjacent to the corresponding

155

singular forms, a significant proportion of subjects fail to notice the different suffix, as if it were not there at all. We were very surprised by this finding, since we thought the suffixes were salient enough to be noticed without explicit form-form comparisons, but the result suggests that even very salient concatenative morphemes benefit from the perceptual contrast between contiguous paradigmatically-related forms.

6.4.2. Implications for phonological theory

The most fundamental implication of the present results is that both faithful and unfaithful mappings rely, at least in part, on generalizing over the relationships between source forms and the corresponding products. In particular, the present results are accounted for by discriminating among source forms that lead to distinct product patterns, resulting in paradigmatic mappings that serve the role of second-order schemas or rules. They differ from second-order schemas in being directional: Being able to anticipate the plural based on the singular does not imply also being able to anticipate the singular based on the plural (e.g., Krajewski et al., 2011). They differ from rules in that they are not context-specific. Rather than mappings being enacted by changes occurring in certain contexts, the present model allows the output to be gradiently conditioned by many different features of the context, where the corresponding source segment is only one such feature with no special status.

Discriminative learning needed to be combined with Hebbian learning that boosted outcomes that frequently occur within products across source contexts, and strengthening of syntagmatic bonds between the parts of unexpected chunks. The former correspond to first-order / product-oriented schemas and indicate that the goal of learning does not reduce to discriminating among source forms. Rather, participants learn what plurals are

156

like in general (Bybee, 1985, 2001; Kapatsinski, 2012, 2013). The latter correspond to phonotactic or morphotactic dependencies resulting in the frequent phenomenon of affixes fusing with the parts of the stem they change (Haspelmath, 1995).

Overgeneralization patterns we observe are also informative regarding the structure of the phonological similarity space. In experimental lab work, Cristià & Seidl (2008) used patterns of (over)generalization to argue that English nasals are [-continuant] like stops and not [+continuant] like fricatives. The pattern of overgeneralization in the present experiment is informative of the structure of the phonological similarity space: Participants overgeneralize non-palatalization (faithful copying) from alveolars and labials to velars (in the Velar Palatalization language) but not from alveolars and velars to labials (in the Labial Palatalization language). Because a pattern is assumed to apply to all members of the smallest natural class that contains the segments sharing a behavior (Moreton & Pater, 2012a; White, 2013, 2014), the results suggest that velars and alveolars belong to the [lingual] natural class (Browman & Goldstein, 1989; Clements & Hume, 1995), but that labials and alveolars do not belong to [anterior] (Chomsky & Halle, 1968) and thus anything that applies to both must also apply to velars.

The patterns of overgeneralization of faithful mappings therefore support a featural organization of the segment inventory. The fact that faithful mappings can be overgeneralized at all in turn supports the notion of conditioned copying (Kapatsinski, 2017; §1.2.2, §6.3.2.2.2). Copying here means incorporating memory representations into the production plan (and is, for example, why [p] mapping onto [p] is the same as [k] mapping onto [k]), and we propose it can be conditional on various cues. That copying can be conditioned predicts that speakers need to learn when and what to copy, and that

copying can be over/undergeneralized, depending on the conditions observed where copying does/does not occur. This proposal explains why keeping faithful pairs (like *blut~bluta*) intact facilitates faithful mappings, and why this extends to the unattested k~ka mapping Velar Palatalization but not p~pa in Labial Palatalization: In Labial Palatalization, copying is conditioned on the [lingual] feature, whereas in Velar Palatalization it is not conditioned on a place feature and so can be extended to velars.

6.4.2.1. Retreating from overgeneralization

6.4.2.1.1. Entrenchment and pre-emption

Preventing overgeneralization has been of concern in work on construction learning. The two most prominent mechanisms for explaining overgeneralization patterns are:

1) Entrenchment (Ambridge et al., 2008, 2012; Blything et al., 2012; Braine & Brooks, 1995; Brooks et al., 1999; Harmon & Kapatsinski, 2017; Regier & Gahl, 2004; Stefanowitsch, 2008; Xu & Tenenbaum, 2007): If a form occurs often in one context, learners infer it does not occur in other contexts, and

2) Statistical pre-emption (Boyd & Goldberg, 2011; Goldberg, 2011): Forms are pushed out of contexts by other forms, which act as pre-emptors; without pre-emptors, forms can extend to any context.

Both palatalization and faithful copying into the output form are subject to overgeneralization, and making the paradigmatic context where the form occurs more obvious through temporal contiguity does not restrict it to that context/diminish overgeneralization, contra the entrenchment explanation. For example, making p→tʃ obvious in Labial Palatalization Change Obvious makes palatalization rates for *all* consonants higher than in Labial None Obvious (§5.2.3, §6.3.2.1), and making p→p and

158

t→t obvious in Velar All Obvious increases the rate of velar copying (in other words, reduces the palatalization rate of [k], §6.3.2.2.2). The increase in availability of [tʃ] or Copy_FIN overrides the evidence of it being restricted to a particular context.

This finding parallels Harmon & Kapatsinski (2017) for form-meaning mappings. They showed that the increase in frequency makes forms more available to extend to new contexts, but it also makes it less likely that subjects will map it onto new contexts. In other words, frequency only leads to entrenchment when accessibility differences between frequent and infrequent forms are neutralized. The results hold as long as the occurrence of the form-meaning pairing strengthens the relationship between the form and all features comprising the meaning. If any feature is presented as part of a novel related meaning, then the form paired with the original meaning usually receives the greatest activation. We see the same effect in the present experiment. When the availability of paradigmatic mappings is boosted in Change Obvious and NoChange Obvious, the outputs of those mappings become more available for use with related inputs that share some features with the original. One prediction for future work is that if participants are given a plural form and asked to choose the singular, they will be more likely to pick [tʃa] plurals correctly in Change Obvious than in NoChange Obvious, and more likely to pick faithful plurals correctly in NoChange Obvious than Change Obvious.

The data present a strong case for statistical pre-emption: Participants in All Obvious, and only All Obvious, learn which consonants should be mapped onto [tʃ]. When both faithful and unfaithful mappings are strong, they can pre-empt each other; obvious faithful mappings pre-empt unfaithful mappings from affecting Not-To-Be-Palatalized

159

consonants, and obvious unfaithful mappings pre-empt faithful mappings from affecting To-Be-Palatalized consonants.

6.3.2.1.2. Other accounts of overgeneralization

Rule-based grammars assume that "do nothing" is the "elsewhere" condition, and therefore cannot be conditioned by context (Chomsky & Halle, 1965, 1968; Pinker & Prince, 1988). If we disregard this assumption, it is straightforward to implement context-sensitive copying, which provides a parsimonious account of the present data (for computational implementations, see Albright & Hayes, 2003; Allen & Becker, 2015; Taatgen & Anderson, 2002; for an application to morphologically-conditioned palatalization, see Kapatsinski, 2010).

Optimality Theory captures avoidance of unfaithful mappings through output-output constraints (Benua, 1997; Kenstowicz, 1996), but it is not clear how that could capture the overgeneralization of non-palatalization to To-Be-Palatalized velars but not To-Be-Palatalized labials. Labial palatalization may violate Ident-[lingual] and Ident-[delayed-release], velar palatalization may violate Ident-[dorsal], and alveolar palatalization may violate Ident-[coronal], but evidence against alveolar and labial palatalization does not result in evidence against velar palatalization; Ident-[anterior] and Ident-[lingual] do not help Ident-[dorsal]. A mechanism to prevent a specified class of segments from changing (to [tʃ]) is necessary. Zuraw (2007) proposes that the grammar contains a set of *Map constraints, which militate against specific segment-segment mappings, so *Map(C→tʃ) could be upweighted whenever an input segment fails to change into [tʃ]. Hayes & White (2015) use these kinds of constraints to account for overgeneralization of changes to intermediate segments (e.g. overgeneralization of p→v to b→v in White, 2014).

160

However, it is not clear how that can account for overgeneralization of faithful mappings. If we assume that t→ta and p→pa provide evidence for *Map(C→tʃ), then it would seem that they should also provide evidence against the faithful tʃ→tʃ mapping. While this seems implausible, it needs to be directly tested.

6.4.2.2. Schemas

The present results are challenging for the proposal that unfaithful mappings are all due to a product-oriented/first-order schema like "plurals end in [tʃa]," which is learned by generalizing over forms in a paradigm cell, in this case plurals (Bybee, 2001; Kapatsinski, 2013). If this were true, then contiguity of unfaithful forms should have no effect on acquisition of unfaithful mappings; bluk→blutʃa should not help the plural [tʃa] schema any more than would [blutʃa] in isolation. Perhaps these examples help not by facilitating discovery of the paradigmatic k→tʃ mapping, but by strengthening the association between elements of the schema via perceptual chunking, such that when one is selected, the other must follow. However, Table 5.4 shows that such examples are particularly helpful for To-Be-Palatalized stems, meaning that they do help the particular paradigmatic mappings they exemplify. These results provide evidence that both faithful and unfaithful mappings rely on conditioned paradigmatic mappings, and the conditioning of the mapping is acquired most easily when forms are adjacent.

6.4.2.3. Implicational hierarchy

Labial palatalization overgeneralizes to alveolars and velars, but velar palatalization only overgeneralizes to alveolars, not labials (§5.2.3, §6.3.2). This pattern mirrors prior work and provides further support for the implicational hierarchy that velar palatalization implies alveolar palatalization and labial palatalization implies alveolar and velar

161

palatalization (Smolek & Kapatsinski, 2018; Stave et al., 2013). It is also reminiscent of other findings where a change from A to C generalizes to a sound intermediate between A and C (B) (Skoruppa et al., 2011; White, 2014). The model can capture the extension to intermediate sounds by the existence of a [lingual] feature, but the implicational hierarchy can only be captured if the model assumes that perceiving a change like A→C involves perceiving A→B→C, where the intermediate sound B is actually perceived to change in the output, on at least some occasions. Future work should investigate if B must be phonetically intermediate between A and C, or if it just needs to be a priori more likely to change into C than A is. We are not confident that we can argue that [t] is between [tʃ] and [k], though it is perceptually closer to [tʃ] than [k] (much less [p]) and is considered more likely to turn into [tʃ] by English speakers. If the results are due to a priori likelihood of change and not intermediacy, then they may be better captured by stronger surprisal-dependent boosts to [tʃ] or post-hoc inference of the type "if [p] changes to [tʃ], then surely everything must" rather than an online perceptual mechanism.

6.4.2.4. Morphology feeds phonology

Lastly, the results provide some support for the traditional rule-based view of the architecture of the grammar, where morphology feeds phonology (Chomsky & Halle, 1965, 1968; cf. Bybee, 1985, 2001; Kapatsinski, 2010). The crucial finding is that the choice of suffix vowel is influenced by the identity of the singular consonant, even if the plural consonant is also taken into account: [tʃ]'s that are made from To-Be-Palatalized consonants favor -*a* over [tʃ]'s that are made from Not-To-Be-Palatalized consonants. The model fails to learn this dependency if the suffix is chosen after the choice of whether to palatalize, or if the choices are made in parallel and independently. If the

model chooses CV sequences like [tʃa] vs. [ka] vs. [ki] vs. [tʃi], consonant-vowel dependencies are learned *too* well, so that unattested (in training) sequences like [ka] in the Velar Palatalization language are never produced. The suffix does not need to always be chosen first, or always before the associated stem change; we rather propose that whichever change is easier to perform is applied before harder changes, and in this experiment, vowel addition is apparently easier to notice and enact than changing the consonant (see also Smolek & Kapatsinski, 2018; Stave et al., 2013; and Chapter IV). The vowel is therefore likely to condition the consonant change, as we would otherwise expect all [tʃ]'s to be suffixed with *-a* at equivalent rates.

6.4.3. Limitations

One important caveat is that this experiment only looked at the very early stages of language learning. Research has shown the importance of sleep for consolidation of knowledge (Cai et al., 2009; Davis et al., 2009), which is outside the purview of the Rescorla-Wagner (1972) associative learning model we use here. Such models instead provide good descriptions of rapid error-driven learning subserved by subcortical structures including the basal ganglia (for procedural memory, Ashby et al., 2007; Lim et al., 2014) and hippocampus (for declarative memory, Davis et al., 2009; McClelland et al., 1995).

In the Complementary Learning Systems framework, the consolidation process involves rapidly learning subcortical structures by 'replaying' the events of the day to the slowly learning neocortex (Kumaran et al., 2016; McClelland et al., 1995). There are reasons to believe that paradigmatic mappings could emerge from this process without the requirement of temporal contiguity: The hippocampus has been shown by lesioning

studies in non-human animals to be particularly important for learning associations in the absence of contiguity ("trace conditioning," Bangasser et al., 2006). It is possible that the hippocampus acquires associations between non-contiguous words but is unable to motivate behavior at test without "enlisting" the neocortex by sharing what it has learned. The neocortex learns more slowly and represents knowledge in a more distributed code than do the rapidly-learning subcortical systems (McClelland et al., 1995; Davis et al., 2009), and the more distributed coding brings out similarities in the coded events, leading to new generalities (Cai et al., 2009; Lewis & Durrant, 2011). Newly learned words are integrated over time with the previously learned words that form the native lexicon in word learning experiments (Davis & Gaskell, 2009; Davis et al., 2009; Dumay & Gaskell, 2007), so it is possible that paradigmatic relations between words could emerge here, when corresponding forms are integrated. Davis et al. (2009) propose that consolidation is necessary for learning to expand beyond the trained sensory regions; in other words, exposure to an alternation may not result in production of that alternation until time and/or sleep has passed. English speakers, due to their experience with (limited) alveolar and velar palatalization, may already have built the connections necessary to employ the alternation after brief exposure, whereas labial palatalization is novel enough that the articulatory architecture is not yet in place, but that given time for incubation/consolidation could be created. Future work should examine possible restructuring of lexical knowledge and bring models of consolidation (Ashby et al., 2007; McClelland et al., 1995) to bear on changes to paradigmatic structure after sleep.

6.5. Summary

A simple two-layer discriminative model consisting of cues and outcomes successfully captures the significant effects of Experiment 3. Regarding suffix vowel choice, the model captures that keeping unfaithful and faithful pairs intact results in increased use of the associated suffix vowel (*-a* and *-i*, respectively), especially for the associated consonants (To-Be-Palatalized and Not-To-Be-Palatalized, respectively). The baseline model underpredicts the use of *-a*, the more frequent suffix, so we increment *-a* by 4% in each language and an additional 6% in Velar Palatalization, since Labial Palatalization subjects seem to avoid the suffix that triggers the difficult-to-produce labial palatalization alternation.

For palatalization, the model captures that intactness of pairs allows access to the phonological form cues in the singular, which affects faithful and unfaithful mappings, stem changes, and suffix vowel choice. We encode the implicational hierarchy of more difficult changes implying easier changes, in particular that labial palatalization implies alveolar and velar palatalization, and that velar palatalization implies alveolar but not labial palatalization. This allows the model to capture that there is more palatalization of the Not-To-Be-Palatalized consonants in the Labial language than Velar. One shortcoming is that the model underestimates the increase in palatalization of Not-To-Be-Palatalized consonants in Change Obvious compared to All Obvious and None Obvious. We propose that the perceptual contrast of encountering singulars adjacent to the corresponding unfaithful plurals leads participants to "chunk" the stem change and the triggering vowel together, boosting the association between them and allowing them to select for one another. This boost may be due to surprise at an unexpected outcome (like

encountering [tʃa] when expecting [pa]), which is separate from the standard RW effects of surprise.

The other major shortcoming in the model is its failure to capture overgeneralization of faithful mappings to the To-Be-Palatalized consonants in Velar Palatalization All Obvious, but not in Labial. We propose that faithful mappings are the result of the outcome Copy$_{FIN}$, which mandates retention of the singular-final consonant, and is cued by various phonological features of the singular form as well as the semantics to be expressed in the product form.

It appears that Copy$_{FIN}$ applies to all the consonants in the smallest natural class that includes the copied segments attested in training. This can be captured by the baseline discriminative learning model when the Copy$_{FIN}$ outcome is triggered by a subset of singular-final consonants. Thus, for the Labial Palatalization language, Copy$_{FIN}$ applies to alveolars and velars, which are included in [lingual]; since labials are excluded from the [lingual] natural class, copying does not extend to the To-Be-Palatalized consonants in Labial Palatalization. The model is able to discriminate between linguals, which trigger Copy$_{FIN}$, and non-linguals, which do not. In Velar Palatalization, however, Copy$_{FIN}$ applies to alveolars and labials, which do not form a natural class that excludes velars, so Copy$_{FIN}$ is taken to apply to all consonants. The model, however, cannot generalize to all inputs, discriminating among alveolars and labials on the one hand and velars on the other. The results are captured only if we increment the activation Copy$_{FIN}$ for labials when the natural class of segments triggering Copy$_{FIN}$ includes all consonants in the language, as it does in Velar Palatalization.

We suspect that this increment comes from an additional Hebbian learning mechanism that complements the discriminative learning mechanism implemented in the baseline model. This mechanism appears to increment associations between cues and co-occurring outcomes even when those cues are not discriminative. Thus, even though all singulars end in [+cons] segments, rendering [+cons] powerless to discriminate among singulars, it can become associated with copying of the final segment, allowing for faithful copying to be overgeneralized form alveolars and labials to velars.

Overall, these results are therefore consistent with a "maximalist," all-of-the-above view of grammar in which forms are produced using a variety of partially-redundant generalizations (e.g., Bybee, 1985; Langacker, 1987) that include both product-oriented and source-oriented schemas, phonotactic dependencies within product forms, and knowledge about what parts of activated representations ought to be copied into the production plan being constructed.

CHAPTER VII

REVIEW, GENERAL DISCUSSION, AND CONCLUSIONS

Portions of this chapter were taken from:

Smolek, A. & Kapatsinski, V. (2019). *Syntagmatic paradigms: Learning*

correspondence from contiguity. Manuscript submitted for publication.

Smolek, A. & Kapatsinski, V. (2018). What happens to large changes? Saltation

produces well-liked outputs that are hard to generate. *Laboratory Phonology: Journal of*

the Association for Laboratory Phonology, *9*(1), 10.

Paradigm Uniformity (PU) is the preference for consistency in forms sharing a stem[61]

(Benua, 1997; Kenstowicz, 1996; Steriade, 2000). Phonologically dissimilar allomorphs

are particularly dispreferred, and learning alternations is harder with dissimilar alternants

(Hayes & White, 2015; Moreton & Pater, 2012a; Peperkamp et al., 2006; Skoruppa et al.,

2011; Steriade, 2001/2009; White, 2014).

The Perseveration Hypothesis proposes that motor perseveration in the production of

a novel form of a known word causes the observed avoidance of stem changes.

Paradigmatic perseveration conflicts with paradigmatic associations, which require

particular relationships between related forms of a word (e.g. in Russian the nominative

trop 'trope' corresponds to the genitive plural *tropov* 'tropes.GEN' but the nominative

tropa 'path' corresponds to the genitive plural *trop* 'paths.GEN'). For the Perseveration

Hypothesis, the relevant associations are between production representations (such as the

gestures of Articulatory Phonology, Browman & Goldstein, 1989) of related forms. The

reason that large changes are especially difficult to learn is that learning associations

between alternants is more difficult when the to-be-associated parts of the alternants are

---

[61] Kenstowicz's (1998) Uniform Exponence allows for consistency in affixes as well.

dissimilar (Moreton, 2008, 2012; Warker & Dell, 2006). If a paradigmatic association requires a change to the base, then obeying paradigmatic perseveration is a PU error. A poorly-acquired association is a lesser obstacle to paradigmatic perseveration, making PU more likely to arise for large changes.

One consequence of the production-locus of the Perseveration Hypothesis is that we expect PU to be stronger in production than in perception, contrary to proposals that privilege perceptual similarity (Kenstowicz, 1996; Steriade, 2001/2009; White, 2017). We experimentally investigate this prediction in Experiment 2, where we find that labial palatalization is accepted in judgment while being rarely produced, confirming our proposal. We suggest that participants in the Labial Palatalization condition are unable to acquire associations between labials and alveopalatals because of articulatory dissimilarity, which prevents them from producing the alternation. However, they do learn the product-oriented schema "plurals should end in [tʃi]," and because they do not have any competing paradigmatic mappings, the schema freely applies to all inputs, resulting in high ratings of palatalization of all consonants.

In Experiment 3, we investigate how associations between dissimilar alternants could be learned by varying the presentation order of training trials. We hypothesize that temporal contiguity between related forms allows participants to notice the relationship between the singular and plural forms of a word and home in on the relevant cues that distinguish the forms. We vary adjacency of faithful and unfaithful pairs separately to determine whether contiguity affects the acquisition of mappings that require a change to the base and those that do not. If unfaithful mappings are learned from product-oriented schemas like "plurals should end in [tʃi]" (Bybee, 2001; Kapatsinski, 2013, 2017b), then

169

adjacent unfaithful pairs should not help acquisition of the unfaithful mapping, whereas if they are learned through paradigmatic relations, they should improve when unfaithful pairs are adjacent. If faithful mappings are the default (Pinker & Prince, 1988), then adjacent faithful pairs should not help acquisition of a faithful mapping. While output-output faithfulness constraints (Hayes, 2004; McCarthy, 1998) are thought to be initially high-ranked and therefore "default," they are subject to demotion with linguistic experience; since our participants are adults, they have presumably learned that related forms do not always share their form (e.g. the past tense of [kip] is [kɛpt], not [kipt]), which would indicate that (at least some of) their faithfulness constraints are not dominant and could benefit from adjacency. Experiment 1 shows that participants like palatalization of alveolars more than labials or velars, suggesting that the corresponding faithfulness constraints have different weights – and even labial and velar palatalization is accepted sometimes, so it seems unlikely that any of the faithfulness constraints remain at the top of the hierarchy.

Experiment 3 shows that both faithful and unfaithful mappings benefit from contiguous pairs of corresponding wordforms in training, and it is only when both types are exemplified by contiguous word pairs that participants learn to produce the correct forms in the correct contexts. We show that a simple domain-general discriminative model based on Rescorla-Wagner (1972) can capture the effects of Experiment 3, supporting the idea that alternations are produced via learned paradigmatic associations acquired using domain-general mechanisms of associative learning.

In this final chapter, we review the results from Experiments 1, 2, and 3 and explore the theoretical implications of our findings before concluding.

7.1. Review of results

All three experiments shared similar stimuli and structure: Participants were tasked with producing (Experiments 2 and 3) and judging (Experiments 1 and 2) singular-plural pairs of alien creatures. The singulars all ended with oral stops ([b;p;d;t;g;k]) and the plurals were suffixed with *-i* or *-a*. Depending on the training language, labial, alveolar, or velar stem-final consonants in the singular alternated with palatal affricates ([tʃ] for voiceless stops and [dʒ] for voiced) in the plural. In Experiment 1, participants provided ratings for singular-plural mappings without any training, in order to establish pre-existing biases. In Experiments 2 and 3, participants were exposed to training exemplifying faithful and unfaithful mappings; these results are informative with regards to the learnability and executability of alternations.

7.1.1. Experiment 1

The baseline experiment (Chapter III) evaluates participants' judgments of palatalization by stem-final consonant place of articulation (labial, alveolar, and velar), voicing of stem-final consonant (voiced or voiceless), and suffix vowel (*-i* or *-a*).

The primary preference found is for palatalization of alveolars: Palatalized alveolars are accepted more often than palatalized labials or velars, and are also accepted as often as faithful alveolars, whereas faithful labials and velars are preferred to alveopalatals. The general preference for perseveration on the base (Do, 2013; Hayes, 2004; Kerkhoff, 2007; McCarthy, 1998) is absent for alveolar-final stems for native English speakers, perhaps because of their experience with palatalization in frequent phrases like *let you*.

There is no preference for palatalization before *-i* vs. before *-a* without training, despite the former being the typologically more frequent pattern (and the context where

171

articulatory and perceptual factors favor palatalization; Anttila, 1989; Guion, 1998). Wilson (2006) trained participants on velar palatalization before either [i] or [e] and found that the alternation was learned equally well in both contexts, but generalized from [e] to [i] (and [ɑ]), but not from [i] to [e] (and [ɑ]) (see §7.1.4.2 for further discussion of generalization by vowel). In the baseline experiment, palatalization of velars is preferred marginally more before -*a* than before -*i*. We propose that the lack of preference for the more motivated context is due to the frequency with which palatalization is followed by a low vowel in English. Even though the high glide [j] triggers the palatalization in e.g. *bet you*, it is not present in the production of *betcha*, so listeners (and speakers) may have become accustomed to palatalization before non-high vowels.

There is also no preference for voiceless palatalization over voiced, even for velars before -*i*, where perceptual similarity patterns suggest that [ki]~[tʃi] should be preferred over [gi]~[dʒi], as the former are more confusable than the latter (Guion, 1998). The lack of bias for the more perceptually similar alternants suggests that perceptual similarity does not have a strong influence on the acquisition of alternations and that differences in learnability or willingness to produce an alternation are better viewed as arising from articulatory (dis)similarity.

Taken together, the results of the baseline experiment suggest that alveolar palatalization may be preferred to (and more frequently produced than) labial or velar palatalization, and that suffix vowel and stem-final consonant voicing may play a limited role.

7.1.2. Experiment 2

Experiment 2 trained participants on palatalization before *-i* of voiced and voiceless labials, alveolars, or velars, and all training trials occured in a random order.

Figure 7.1 shows the rate of palatalization in production by training language and stem-final consonant. The productivity of palatalization for the To-Be-Palatalized consonants is highest for alveolars, lower for velars, and lowest for labials, providing support for our claim that large changes are harder to (learn to) produce than small changes are. In other words, paradigm uniformity exerts a stronger force on the large-change labial palatalization than small-change alveolar and velar palatalization. There is no difference in the rates of overgeneralization to linguals across conditions, so it is not that labial palatalization is more likely to extend to intermediate sounds (as proposed by White, 2014), but rather that it is less likely to be produced in the first place.
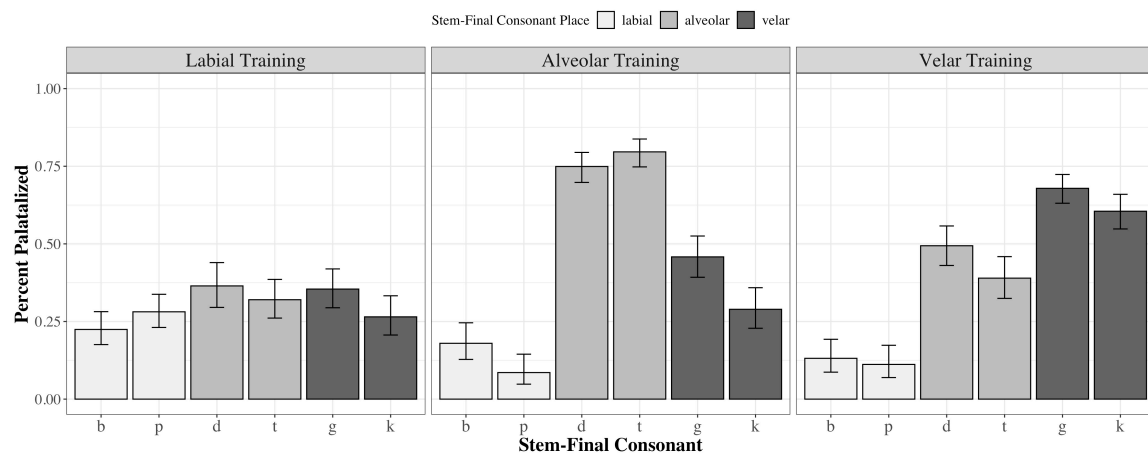
Figure 7.1. Differences in palatalization rates in production before *-i* across individual stops and training languages. Shading indicates place of articulation from labial (lightest) through alveolar (medium) to velar (darkest). Voiced consonants are on the left within shading, while voiceless are on the right. Left panel: Labial Palatalization. Center panel: Alveolar Palatalization. Right panel: Velar Palatalization.

Nonetheless, participants in the Labial Palatalization condition still learn that they *should* palatalize labials. Figures 7.2 and 7.3 show the rate of acceptance in judgment of faithful and palatalized plurals, respectively, by training language and stem-final consonant. There are no differences in acceptability of faithful plurals, and participants in all conditions learn to prefer the unfaithful plural over the faithful for To-Be-Palatalized consonants. But whereas participants in the Velar and Alveolar Palatalization languages produce palatalization of the target consonants as often as they accept it, participants in the Labial Palatalization condition accept it much more frequently than they produce it. The judgments of unfaithful mappings suggest that Labial Palatalization participants like to palatalize the best of everyone, and yet they produce palatalization the least often. Zuraw (2000) found similar results for nasal substitution vs. assimilation in Tagalog, where the more difficult substitution was judged as being better, but nasal assimilation was produced more often, although the small sample size ($n = 9$) was insufficient for inferential statistics on production or the difference between production and judgment. Our results provide stronger evidence for a dissociation between production and judgment.
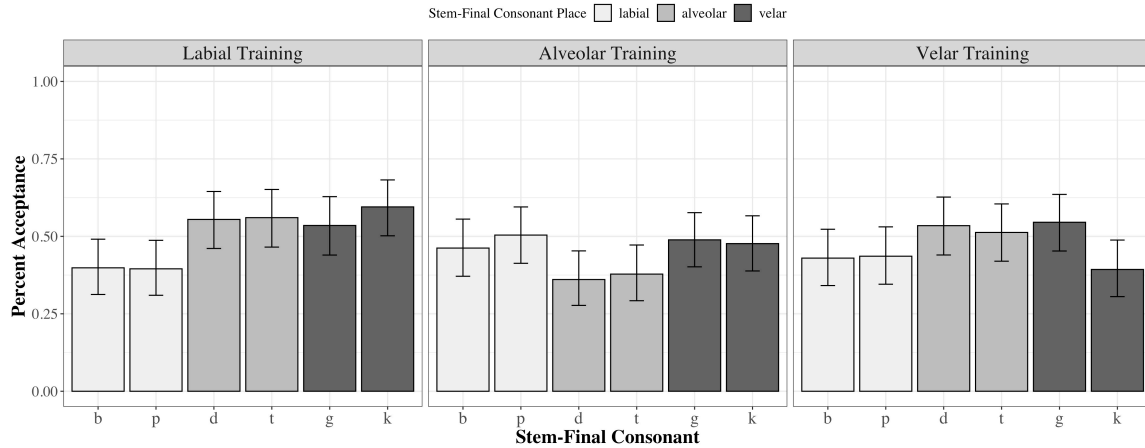
Figure 7.2. Differences in judgments of faithful mappings before *-i* across individual stops and training languages. Shading indicates place of articulation from labial (lightest) through alveolar (medium) to velar (darkest). Voiced consonants are on the left within shading, while voiceless are on the right. Left panel: Labial Palatalization. Center panel: Alveolar Palatalization. Right panel: Velar Palatalization.
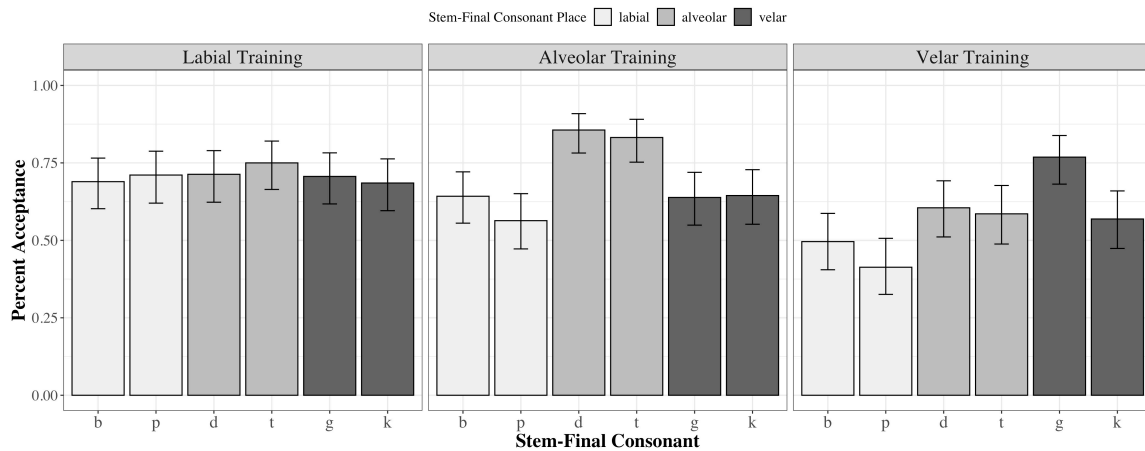
Figure 7.3. Differences in judgments of palatalization before *-i* across individual stops and training languages. Shading indicates place of articulation from labial (lightest) through alveolar (medium) to velar (darkest). Voiced consonants are on the left within shading, while voiceless are on the right. Left panel: Labial Palatalization. Center panel: Alveolar Palatalization. Right panel: Velar Palatalization.

Prior work on alternation-learning largely examined judgment only (Moreton, 2008; Skoruppa et al., 2011; White & Sundara, 2014; but see White, 2013, Ch. 4.5, 2014; Wilson, 2008). Our work, and the results of Stave et al. (2013), shows that the bias against changing labials manifests differently in production and judgment. In both cases,

training on labial palatalization makes participants not prefer palatalizing labials over other consonants, but in production, that is because they tend to palatalize nothing, and in perception, because they tend to accept palatalization of everything. After training on alveolar or velar palatalization, participants learn to palatalize, and to prefer palatalization of, the To-Be-Palatalized consonants more than Not-To-Be-Palatalized (although they still accept incorrect palatalization more often than they produce it). Overgeneralizing to more likely targets is typical in the judgment task (White, 2013; Wilson, 2006), and substantive biases are designed to capture this effect (Moreton & Pater, 2012b; White, 2017; Wilson, 2006). Our data show that in production, however, the problem lies not in overgeneralizing the change, but in failing to change what should be changed.

While [ki] is perceptually more similar to [tʃi] than [gi] is to [dʒi] (Guion, 1998), [g] is palatalized more than [k], and preferred marginally more in judgment. [g] and [k] are equally articulatorily similar to [dʒi] and [tʃi], respectively, and this combined with the orthographic overlap of [g] and [dʒ] under <g> seems sufficient to overcome any effect perceptual similarity plays on learnability in this instance.

Comparison to the baseline experiment with no training reveals that all training languages result in an equal increase in judgments of palatalization of the target consonants; in other words, acceptance of labial palatalization improves after training as much as acceptance of alveolar or velar palatalization do. It thus seems that the difference in production rates is not because subjects are unable to learn the pattern, but rather that they are unable to apply that knowledge in production. We propose that the articulatory dissimilarity between labials and alveopalatals prevents participants from acquiring paradigmatic mappings in the Labial Palatalization condition, and they are thus unable to

produce palatalization. Because they do not know what inputs correspond to [t∫i] in the output, every occurrence of [t∫i] is surprising, which strengthens the first-order schema, as reflected in their high ratings of palatalized forms. The rarity of labial palatalization is due to the articulatory difference between [p] and [t∫] being so large. The fact that the effect of articulatory dissimilarity manifests itself only in production, leaving judgments and the learning rate as indicated by pre- and posttest judgment task comparisons unaffected, indicates that it is a channel bias, not an inductive bias.

The Perseveration Hypothesis provides the best account for the results: Labial palatalization is avoided because of articulatory dissimilarity, not perceptual dissimilarity, and the bias against labial palatalization is much stronger in production than perception.

7.1.3. Experiment 3 and a discriminative model

In Experiment 3 (Chapter V), we trained participants on palatalization of labials or velars before -*a*, with non-alternating alveolars, labials (for Velar Palatalization), and velars (for Labial Palatalization) suffixed with -*i* or -*a* (50% of stems for each), varying the order in which training trials appeared. Adjacency of faithful and unfaithful mappings was manipulated such that faithful mappings were adjacent in half of the conditions and randomly ordered in the other half, with the same true for unfaithful mappings. The four conditions were None Obvious (neither faithful nor unfaithful mappings presented in contiguity), All Obvious (both faithful and unfaithful mappings presented in contiguity), Change Obvious (only unfaithful mappings presented in contiguity), and NoChange Obvious (only faithful mappings presented in contiguity).

We model the results of Experiment 3 with code implementing the Rescorla-Wagner (1972) discriminative learning model. The model includes two layers of nodes, one for cues and one for outcomes, with association weights connecting them. The model uses the cues to predict the outcomes, and when the prediction is wrong (an unexpectedly absent or present outcome, given the set of active cues), weights are adjusted, with more unexpected outcomes resulting in greater adjustments. The cues in the baseline model include the stem-final consonant of the singular and the semantic meaning (PL), and the outcomes are the stem-final consonant of the plural and the plural vowel. The initial connection weights reflect English speakers' biases, namely that alveolars are most acceptable to change, followed by velars, with labial alternations disliked. The model captures the results below when the suffix vowel is chosen before the plural stem-final consonant allomorph (see §6.2.2 for discussion). To encode the implicational hierarchy of larger changes implying smaller (Skoruppa et al., 2011; White, 2013, 2014), we specify that k→tʃ provides support for t→tʃ as well, and that p→tʃ provides support for both t→tʃ and k→tʃ. The trial order manipulation determines the availability of cues: When the singular and plural forms are adjacent, the phonological form of the singular is available to predict the phonological form of the plural, but when they are not adjacent, only the semantics (PL) can be used. Since only the weights of present cues can be updated, mappings that are not adjacent can not be updated and thus in the None Obvious condition, they reflect the initial biases. In the sections below, we review the results from the experiment and discuss how successful the model is at capturing them, as well as any modifications (additional cues/outcomes or adjustments to association weights) we apply.

7.1.3.1. Suffix vowel

Keeping faithful pairs intact (NoChange Obvious and All Obvious) results in greater use of the suffix *-i*, with most participants approximately probability matching. Not keeping faithful pairs intact (None Obvious, Change Obvious) leads many participants to use only *-a*, with many participants regularizing. This can be seen in Figure 7.4, with lower rates of *-a* for the NoChange Obvious and All Obvious conditions compared to None Obvious and Change Obvious. There is a weaker effect for keeping unfaithful pairs intact (Change Obvious, All Obvious), which results in more use of *-a*, especially for the To-Be-Palatalized consonants, seen by comparing the light and dark bars in Change Obvious and All Obvious vs. None Obvious and NoChange Obvious.
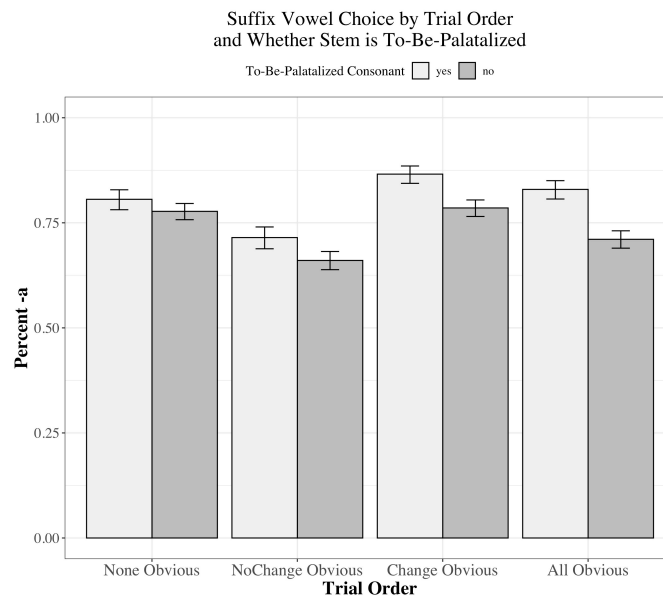
Figure 7.4. Percent of plurals suffixed with *-a* by training trial order and To-Be-Palatalized status of stem.

The baseline model captures that To-Be-Palatalized stems are more likely to be suffixed with *-a* than Not-To-Be-Palatalized stems, that exposure to intact pairs exemplifying faithful mappings results in more use of *-i*, and that exposure to intact pairs

exemplifying unfaithful mappings results in more use of -*a*, particularly for the To-Be-Palatalized consonants. However, it underpredicts the frequency of -*a* overall. We increment -*a* by 4% in both languages to account for how participants tend to fall between probability matching and maximizing in their vowel choice.

The other way the model falls short is that it fails to capture that participants use -*a* more when trained on the Velar Palatalization language than on Labial Palatalization, as shown in Figure 7.5. Palatalization is triggered by -*a* in this experiment, and labial palatalization (like other large changes) is difficult to produce (Chapter IV), so we propose that Labial Palatalization participants choose to avoid -*a* in order to avoid the challenging palatalization alternation. If it is difficulty that makes them avoid -*a*, we would expect that were labial palatalization less difficult to perform, Labial Palatalization participants would trend more towards maximizing/regularization, as the Velar Palatalization participants do. To represent this in the model, we increment -*a* by an additional 6% in Velar Palatalization.
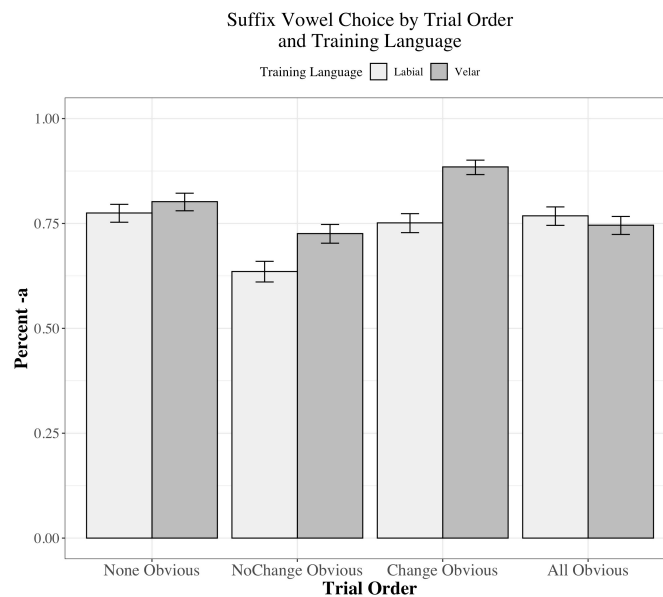
Figure 7.5. Percent of plurals suffixed with -*a* by training trial order and training language.

180

7.1.3.2. Palatalization

Palatalization occurs before *-a* in training, and the participants who produced each suffix on at least 10% of trials learn to palatalize before *-a* more than before *-i*, as shown in Figure 7.6. The baseline model captures this finding.

Palatalization by Plural Vowel and Trial Order

Plural Vowel ☐ -a ◼ -i
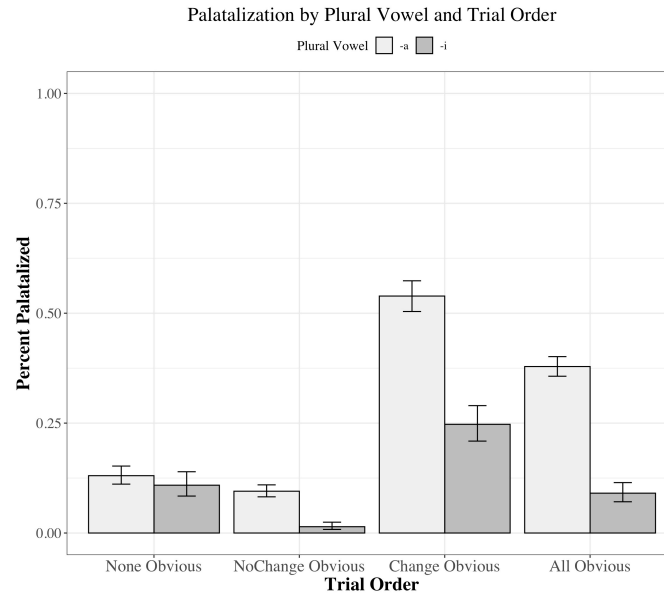
Figure 7.6. Palatalization probability by training trial order and suffix vowel.

Looking only at the palatalizing suffix *-a*, keeping unfaithful pairs intact leads to more palatalization, especially of To-Be-Palatalized consonants, and keeping faithful pairs intact leads to less palatalization, with no significant interaction. Figure 7.7 shows this effect: There is substantially more palatalization in the Change Obvious and All Obvious conditions, especially for To-Be-Palatalized consonants, than there is in the None Obvious and NoChange Obvious conditions. Keeping both types of mappings intact results in production of palatalization, but restricted to the appropriate consonants.
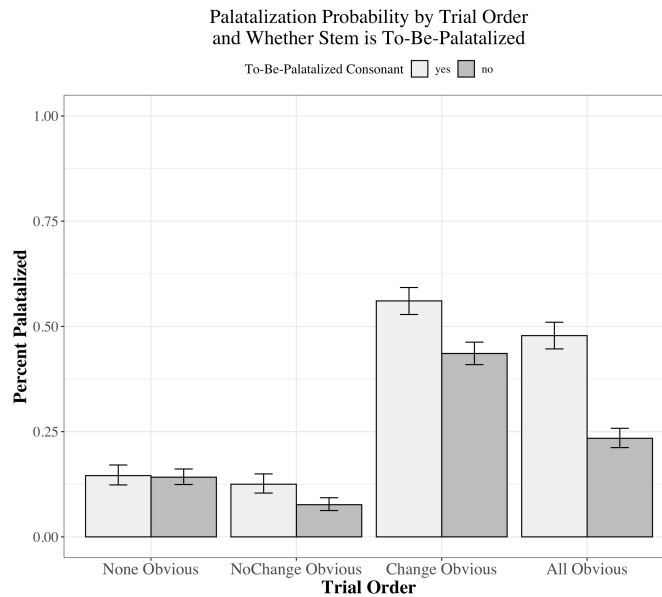
Figure 7.7. Production of palatalization before -*a* by training trial order and To-Be-Palatalized status of stem.

The model underestimates the increase in palatalization of Not-To-Be-Palatalized consonants in the Change Obvious vs. All Obvious/None Obvious conditions before -*a*. We suggest that adjacent unfaithful pairs allows participants to notice, through perceptual contrast and/or surprise, the association between [tʃ] and [a], fusing them together into a single "chunk" (see §7.2.3, below, for further discussion). This chunking allows [tʃ] and [a] to elicit one another, making it more likely that [a] will trigger [tʃ]. In the absence of adjacent faithful pairs to draw notice to [a] without [tʃ], the chunk can apply in all contexts. We implement the association of [tʃ] and [a] in the model as a boost with constant magnitude, after which the model makes the correct prediction of greater overgeneralization in Change Obvious.

The baseline model also fails to predict that To-Be-Palatalized consonants are palatalized less in All Obvious than Change Obvious for Velar Palatalization, but not Labial Palatalization. This is shown in Figure 7.8: The rate of palatalization of To-Be-

182

Palatalized consonants is the same for Change Obvious and All Obvious for Labial Palatalization (light bars), but lower for All Obvious than Change Obvious for Velar Palatalization (dark bars). We propose this is because faithful mappings (produced using Copy$_{FIN}$) are taken to apply to the smallest natural class including the copied segments. In Labial Palatalization, alveolars and velars are both [lingual], but the copied segments in Velar Palatalization are alveolars and labials, which do not form a natural class excluding velars, so copying is free to apply to all consonants. We modify the model by adding [lingual] as a cue for singulars ending with [t] or [k] and incrementing Copy$_{FIN}$ whenever the natural class of copied segments includes all consonants in the language (i.e. in Velar Palatalization).
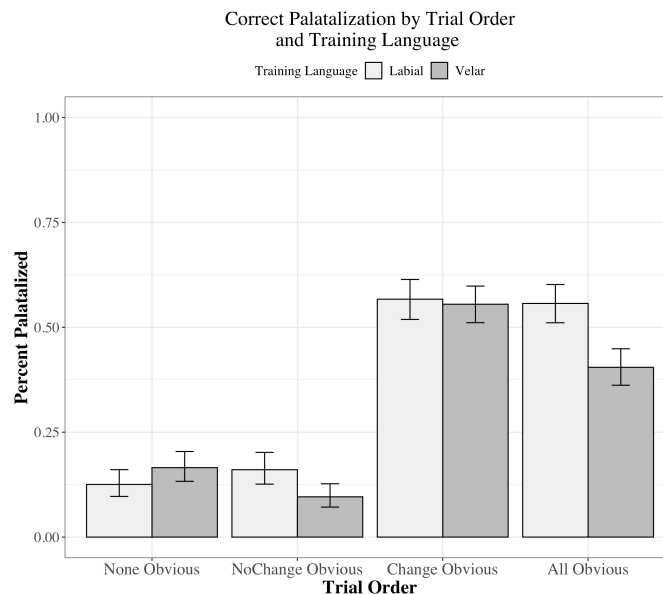
Figure 7.8. Rates of palatalization in production of To-Be-Palatalized consonants by training trial order and training language.

7.1.4. Context naturalness and alternation learnability

Experiment 2 shows that the naturalness of the alternation affects acquisition of paradigmatic mappings, with the larger, less natural labial palatalization being produced

183

less often than the smaller, more natural velar and alveolar palatalization. Labial Palatalization participants still learn the product-oriented schema "plurals should end in [tʃi]," leading to high acceptance rates of palatalization in judgment. To evaluate the effect that context naturalness has on learnability, we compare the palatalization rates in production of To-Be-Palatalized consonants for Experiment 2 to Experiment 3. Palatalization is triggered by -*i* in Experiment 2, which is the typologically more common and phonetically more natural context (Bateman, 2007; Kochetov, 2011), whereas it is triggered by -*a* in Experiment 3. Wilson (2006) shows that palatalization is learned equally well before [i] and before [e], but that palatalization before [e] generalizes to [i] more than the reverse. Our previous work on palatalization before -*a* shows lower rates of palatalization compared to -*i* (Stave et al., 2013, vs. Smolek & Kapatsinski, 2018), but the same bias against labial palatalization. However, the experiments are not strictly equivalent in training design.

Experiments 2 and 3 can be directly compared, since the training has the same number of trials distributed over the consonants in the same way, with certain restrictions. We included only Labial and Velar Palatalization participants from Experiment 2 (since Alveolar Palatalization is not included in the languages of Experiment 3), and only None Obvious participants from Experiment 3 (because all participants in Experiment 2 were trained with trials in a random order). In Experiment 2, we excluded participants who produced particularly egregious plurals, whereas we included every participant from Experiment 3. To ensure that the Experiment 3 participants were not overall worse-performing than those in Experiment 2, we excluded any participants from the former who produced more errors (see pp. 188) than the worst-

184

performing participant in the latter, which removed one subject (from Velar Palatalization None Obvious). We ran generalized logistic linear mixed-effects models with the lme4 package (version 1.1-21, Bates et al., 2015) in R (version 3.6.0, R Development Core Team, 2019). Fixed effects were included for Training Language (Labial and Velar), Vowel Context (Correct vs. Incorrect, based on training; i.e. *-i* is coded as Correct for Experiment 2 and Incorrect for Experiment 3, and *-a* is coded as Incorrect for Experiment 2 and Correct for Experiment 3), To-Be-Palatalized (yes and no), and Experiment (2 and 3), and any significant interactions. Random intercepts were included for Subjects and Singulars, with no random slopes. Log likelihood models on nested models were used to derive significance values, and for contrasts that were not significant and of theoretical interest, evidence for the null hypothesis was evaluated using the BIC approximation to the Bayes Factor (Wagenmakers, 2007).

7.1.4.1. Palatalization of To-Be-Palatalized consonants in the triggering context

First, we compare the rate of palatalization in production of the To-Be-Palatalized consonants in the triggering context (before *-i* for Experiment 2 and before *-a* for Experiment 3)[62] to determine whether the effect of change magnitude holds for both natural and unnatural contexts. Velars are palatalized more than labials ($b = -3.63$, $se(b) = 1.00$, $z = -3.64$, $p < 0.001$, and Training Language significantly improves model fit ($\chi^2(1) = 7.86$, $p = 0.005$). The difference is smaller for Experiment 3 ($b = 3.68$, $se(b) = 1.55$, $z = 2.37$, $p < 0.02$) and the interaction of Experiment by Training Language significantly improves model fit ($\chi^2(1) = 5.64$, $p < 0.02$). Figure 7.9 shows why: Palatalization is learned better before *-i* than before *-a*, overall ($b = 0.84$, $se(b) = 1.10$, $z = 0.76$, $p = 0.45$

---

[62] Keep Place ~ Experiment * Training Language + (1 | Subject) + (1 | Singular), data restricted to To-Be-Palatalized consonants before *-i* for Experiment 2 and before *-a* for Experiment 3

but Experiment significantly improves model fit, $\chi^2(1) = 12.12$, $p < 0.001$), but before *-i*, correct velar palatalization is learned much better than correct labial palatalization, whereas there is no difference by language before *-a*.

Palatalization of To-Be-Palatalized Consonants
in Palatalizing Context, by Experiment
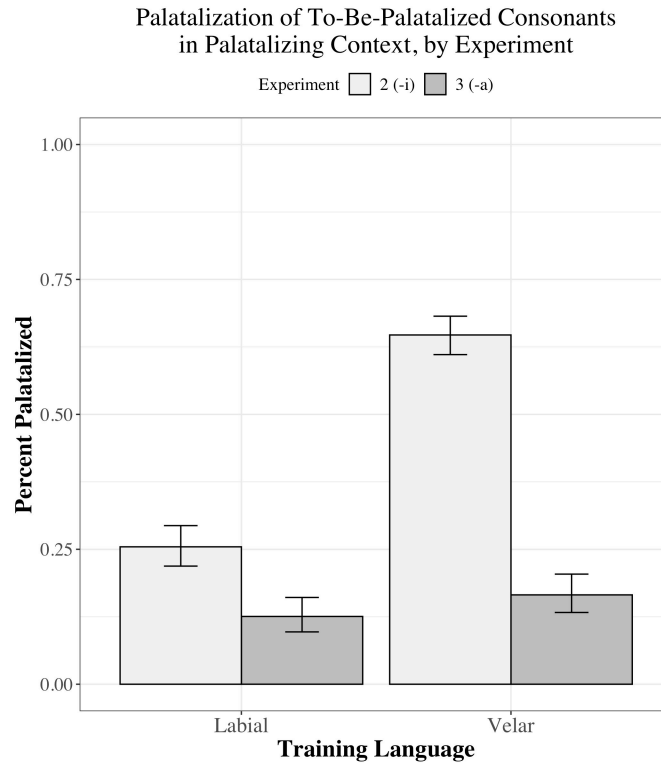
Figure 7.9. Rates of palatalization of To-Be-Palatalized consonant in palatalization-triggering context by training language and experiment.

The lack of difference before *-a* could be due to participants in Experiment 3 failing to learn either alternation, perhaps because palatalization before *-a* is phonetically unmotivated. Stave et al. (2013) find a difference between languages before *-a*, but the stem vowel was always [a] in palatalized forms, so perhaps the contextual consistency made the pattern easier to notice and apply.

We first compare the proportion of palatalized plurals produced by participants in each condition. We would expect that as difficulty increases, so does the number of participants who produce minimal palatalization, with more participants at the higher end

of the scale for easier patterns. We calculated the proportion of palatalized plurals for To-Be-Palatalized stems by dividing the number of plurals (before *-i* or *-a*; for comparison of learning by suffix vowel context, see §7.1.4.2) by 36.

Figure 7.10 shows that many more participants learn to palatalize the correct consonants in Experiment 2 than Experiment 3. 93% of Velar Palatalization participants palatalize velars at least once, and 43% palatalize more than half the time; 50% of Labial Palatalization participants palatalize labials at least once, and 13% palatalize more than half the time. In Experiment 3, however, only around half of participants in either language palatalize at all (52% for Labial Palatalization and 57% for Velar Palatalization), and none palatalize half the time (22% palatalized was the highest proportion for a participant in Labial Palatalization and 42% for Velar Palatalization). Experiment 3 shows lower rates of palatalization because very few participants manage to palatalize at all; palatalizing before *-a* was evidently much more difficult to learn, even for velars.

Another proxy for experimental difficulty is the rate of "acceptable" plurals produced, where by "acceptable" we mean plurals whose plural consonant is either palatalized or retained from the singular, and whose plural vowel is either *-i* or *-a* (i.e. the plurals that were not excluded from analysis). If participants produce more "unacceptable" plurals (e.g. *roʊp~roʊpakaɪ*, *gwæp~gwæpeɪd*) or fail to produce a plural at all, we can assume that this reflects confusion about the pattern. For these purposes, faithful plurals of To-Be-Palatalized consonants and unfaithful plurals of Not-To-Be-Palatalized consonants are still "acceptable," in that the form of the plural obeys the rules regarding how plurals can look.

187
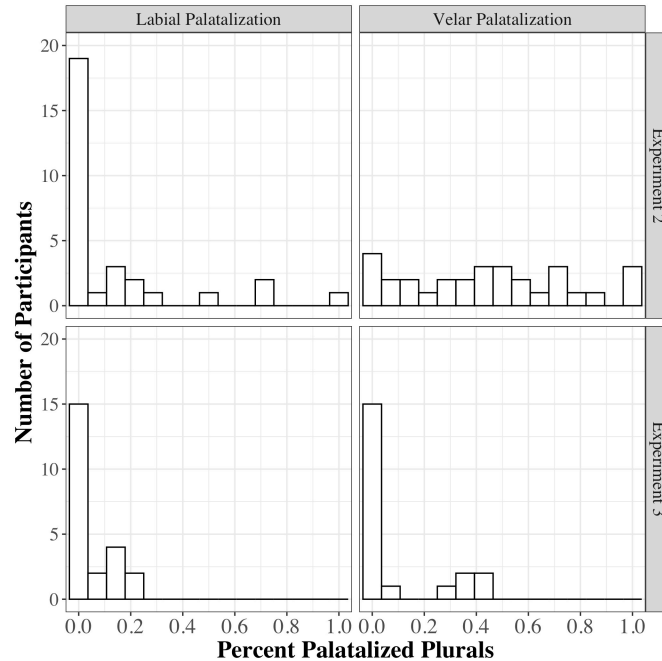
Figure 7.10. Proportion of plurals of To-Be-Palatalized consonants that were palatalized. Upper row: Experiment 2. Lower row: Experiment 3, None Obvious training order. Left column: Labial Palatalization condition. Right column: Labial Palatalization condition.

We created histograms of the proportion of "acceptable" plurals for every subject, by training language and experiment, as shown in Figure 7.11. (Note that the sample size was not the same for every experiment; we are interested not in the raw counts themselves but rather in the shape of the distribution.) The proportion was calculated by dividing the number of "acceptable" plurals by 92, the total number of plurals. Comparison of the top row (Experiment 2) to the bottom row (Experiment 3) is informative: In Experiment 2, most participants produce a majority, even a large majority, of "acceptable" plurals, whereas in Experiment 3, many participants fall below 50% "acceptable" plurals and the distributions are more uniform. For both experiments, Labial Palatalization (left column) has more participants at the lower end of acceptability.

Statistical analyses[63] show that participants in Experiment 3 produce significantly more "unacceptable" forms than participants in Experiment 2 ($b = 0.99$, *se(b)* = 0.36, $z = 2.72$, $p = 0.006$; Experiment is significant in model comparisons, $\chi^2(1) = 7.14$, $p < 0.008$), and Velar Palatalization training results in fewer errors ($b = -0.82$, *se(b)* = 0.36, $z = -2.29$, $p = 0.02$; including Training Language significantly improves model fit, $\chi^2(1) = 5.11$, $p = 0.02$), with no significant interaction. From this, we can conclude that the participants in Experiment 3 struggle to extract the appropriate patterns, often producing inexplicable plurals or failing to produce a plural at all, and participants trained on the large change alternation have more difficulty learning the patterns.
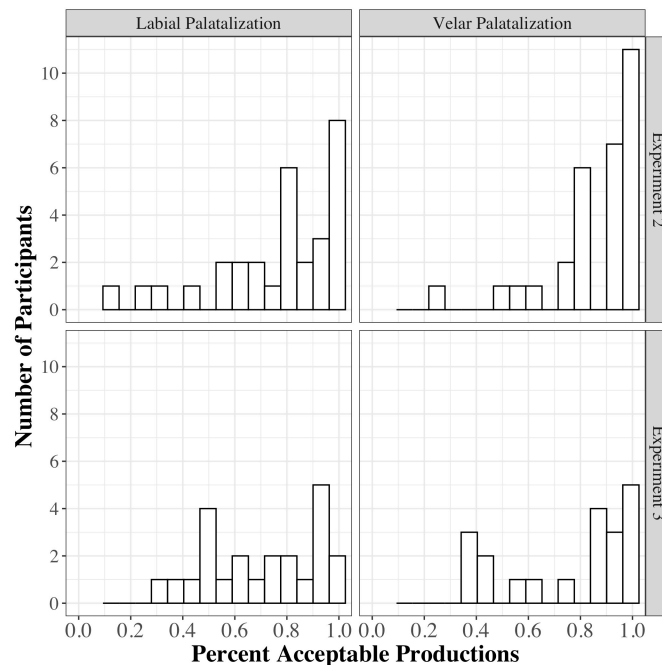
Figure 7.11. Proportion of plurals that followed patterns included in training. Upper row: Experiment 2. Lower row: Experiment 3, None Obvious trial order. Left column: Labial Palatalization condition. Right column: Velar Palatalization condition.

---

[63] Errors ~ Experiment + Training Language + (1 | Subject) + (1 | Singular), including participants from trained on Labial and Velar Palatalization in Experiment 2 and participants in the None Obvious trial order in Experiment 3.

The lack of a difference between Labial and Velar Palatalization in Experiment 3 suggests that substantive biases affect learning, as proposed by Wilson (2006) and White (2013, 2014), among others (cf. J. P. Blevins, 2006; Bybee, 2001; Hale & Reiss, 2000; Moreton & Pater, 2012b, for skepticism). Velar palatalization before -*i* is more natural than before -*a*, in that [k] is articulatorily and perceptually closer to [tʃ] before -*i*, but not before -*a*; velar palatalization is also more natural than labial palatalization, because [k] shares articulators with [tʃ] but [p] does not. The results suggest that it is only when both context and change are natural that the alternation is learned, at least when related forms are not presented in contiguity. The results of Experiment 3 show that unnatural alternations can be acquired when related forms are temporally contiguous, with both labial and velar palatalization before -*a* being produced at high rates in the Change Obvious and All Obvious conditions.

7.1.4.2. Generalization to the "wrong" suffix

We evaluate the extent to which palatalization generalizes from the unnatural context to the natural context, as shown by Wilson (2006) and Mitrović (2012), by comparing rates of palatalization of To-Be-Palatalized consonants by Vowel Context and Experiment[64]. Figure 7.12 is informative: When trained to palatalize before -*i* (left panel), neither Labial nor Velar Palatalization participants palatalize before -*a* at an appreciable rate. When trained to palatalize before -*a* (right panel), both Labial and Velar Palatalization participants palatalize before -*i* at a rate comparable to before -*a*.

In other words, participants trained on the change in an unnatural context generalize to producing the change in the natural context, but not the reverse ($b = -2.07$, *se(b)* =

---

[64] Keep Place ~ Experiment * Training Language * Vowel Context + (1 | Subject) + (1 | Singular), restricted to To-Be-Palatalized consonants

0.70, $z = -2.96$, $p = 0.003$; the interaction of Experiment by Vowel Context significantly improves model fit, $\chi^2(1) = 59.78$, $p < 0.001$). The generalization difference is especially stark for Velar Palatalization, as shown in the difference in heights of the light (correct vowel context) and dark (incorrect vowel context) bars for Velar Palatalization compared to the same for Labial Palatalization in Figure 7.12.
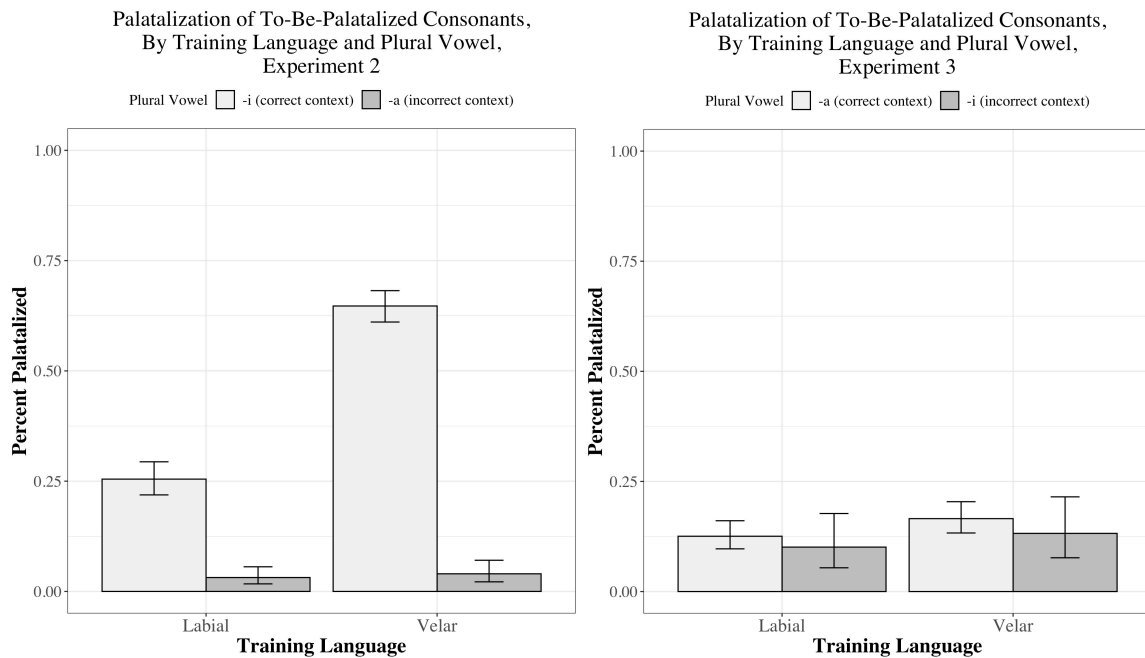
Figure 7.12. Rates of palatalization of To-Be-Palatalized consonants by training language and plural vowel. Left panel: Experiment 2, where participants were trained to palatalize before -*i* and not -*a*. Right panel: Experiment 3, where participants were trained to palatalize before -*a* and not before -*i*.

To confirm that it is overgeneralization and not just that some participants learn the wrong pattern, we evaluated individual differences in overgeneralization of palatalization by vowel context. We restricted the analysis to participants who palatalized at least once in the correct context (excluding one participant from Experiment 3 Velar Palatalization, who palatalized 3 times before -*i* but never before -*a*, suggesting they did not learn the

191

target pattern) and at least twice overall (any fewer than that, and it would not be possible for them to show any differences by suffix vowel). For Experiment 2, 8 out of the 22 Labial Palatalization participants who qualified palatalize before both -*i* and -*a* (36.4%), as do 9 out of the 28 Velar Palatalization participants (32.1%). For Experiment 3, 8 of the 14 Labial Palatalization participants palatalize before both vowels (57.1%), as do 5 of the 8 Velar Palatalization participants (62.5%). There are fewer people who learn to palatalize in Experiment 3, but more than half of those who do extend the alternation to the more natural context, whereas only around a third of participants trained on the natural context extend it to the unnatural. Most of the participants in Experiment 2 who palatalize in both contexts palatalize in the incorrect context very rarely, whereas the participants in Experiment 3 do not show such a skewed distribution (though the range is roughly the same for both), as shown in Figure 7.13. However, Fisher's exact tests show no significant difference in the proportion of overgeneralizers by experiment for Labial ($p = 0.31$) or Velar ($p = 0.22$) Palatalization, so these conclusions are tentative. The results suggest that large, unnatural changes – when they are learned – may be taken to imply smaller, natural changes, as proposed by Wilson (2006) and Mitrović (2012).
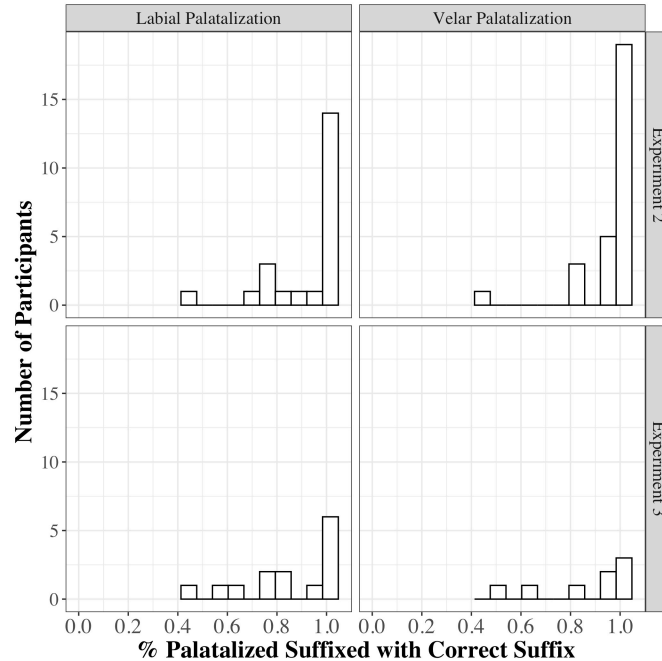
Figure 7.13. Proportion of palatalized plurals that were suffixed with the correct vowel (*-i* for Experiment 2, *-a* for Experiment 3). Only participants who palatalized at least once in the correct context and at least twice overall were included.

7.2. Theoretical implications

7.2.1. The fate of large changes

The dissociation between judgment and production in Experiment 2 introduces some uncertainty regarding the fate of large changes. They are likely leveled by the speaker but judged unacceptable by the listener; if the speaker obeys the listener, they may avoid the faithful form in the future. Speakers do adjust production in response to listener feedback (Buz et al., 2016; Goldstein et al., 2003; Maniwa et al., 2009; Schertz, 2013; Warlaumont et al., 2014), which provides evidence that listeners' beliefs about speakers' productions, if made apparent and heeded by the speaker, can influence their future productions. However, listeners' beliefs are based on the productions they hear, so often in language change, "use leads, and belief follows" (Harmon & Kapatsinski, 2017). Sociolinguistics is full of dissociations between judgment and production, where speakers produce an

193

innovative form but judge it unacceptable due to stigma (Labov, 1975, 1996), but it is unclear whether these judgments result in avoidance of the unacceptable forms or limit their spread. More research on the interaction between belief and use is needed by, for example, performing observational studies on the impact of social acceptability on use and implementation of more interactive tasks on how judgment and production interact (Buz et al., 2016), varying the order of production and judgment tasks (Harmon & Kapatsinski, 2017), and examination of the time course of development of judgment and production in the acquisition of alternations (Kerkhoff, 2007).

7.2.2. The importance of syntagmatic co-occurrence

The results of Experiment 3 show that paradigmatic mappings may actually be learned syntagmatically, strengthening when related forms occur next to each other in time. McNeill (1966) was skeptical that associative models could capture acquisition of paradigmatic mappings because paradigmatic associates do not appear in contiguity, whether in speech or through erroneous anticipation, but this has proven to be unfounded. Corpus studies show that members of morphological paradigms occur near each other more often than other word pairs (Baroni et al., 2002; Xu & Croft, 1998), so learning paradigmatic mappings does not require any superhuman capabilities like perfect recall or anticipation, but rather merely noticing related forms when they occur together. While paradigmatically related words "have a relation to one another different from co-occurrence" (McNeill, 1966, p. 543), that relation is nonetheless learned in the presence of co-occurrence. Associative models are able to acquire paradigmatic mappings precisely because temporal contiguity matters. Mechanisms that acquire paradigmatic mappings in the absence of contiguity (Albright & Hayes, 2003; Ervin, 1961; McNeill,

1966; Plunkett & Juola, 1999) may be overly powerful and exceed the ability of actual learners

7.2.3. Chunking and common fate

Incrementing the association between [tʃ] and [a] in the model successfully captures the greater overgeneralization of palatalization in Change Obvious over All Obvious, which suggests that adjacency of corresponding forms is not just helpful for making the cues comprising the singular form more available for predicting the plural form, but also brings out differences between the corresponding forms. Recent work on category learning has shown that temporal adjacency between exemplars from multiple categories make participants focus on the discriminative features (Carvalho & Goldstone, 2015; Zaki & Salmi, 2019), whereas adjacent exemplars of the same category make participants notice the shared features, even if they are also shared by exemplars of other categories. In the present work, the "chunking" of [tʃa] in Change Obvious motivates our claim that placing singulars and plurals next to each other also makes participants notice the parts they do not have in common.

The results also suggest that elements that "move together" when an exemplar from one category is placed next to an exemplar from another "fuse together," and each becomes able to evoke the other in production and perception. Noticing that the [k] or [p] of the singular has been replaced by [tʃa] in the plural when the forms are adjacent helps the -*a* suffix, once chosen for production, to evoke [tʃ] and thus palatalize the preceding consonant. This can be considered an instance of the "principle of common fate," the basic mechanism of perceptual grouping (Köhler, 1929; Wertheimer, 1923/1938; Uttal et al., 2000). In speech processing, the principle of common fate manifests as the grouping

of auditory elements that change in amplitude together or are frequency-modulated into a single stream (Bregman & Pinker, 1978; but see Böhm et al., 2003, for evidence against common fate in perceptual grouping). Goodsitt et al. (1993) and Kuhl (2000) reference the principle of common fate in their suggestion that infants group together sounds that occur together in words (i.e. that aren't separated by a word boundary). However, Baayen et al. (2016) demonstrate that this clustering can be described using a baseline discriminative learning model, as long as upcoming elements are predicted from preceding ones, without needing to include common fate/chunking. The present results, on the other hand, suggest that discriminative learning needs to be supplemented with a chunking mechanism. Without a chunking mechanism, the baseline model is unable to account for the overgeneralization differences between Change Obvious and All Obvious, suggesting that common fate may aid in strengthening associations between elements that replace another element in a shared context. For example, saying *I went to pet the pengui-, no, to pet my cat on the sofa* may strengthen the association between *my* and *cat* by making the shared context obvious, creating a variation set.

7.2.4. Variation sets

In variation sets, successive utterances present different morphemes in a constant communicative context, which facilitates acquisition of words and morphemes (Küntay & Slobin, 1996; Onnis et al., 2008; Schwab & Lew-Williams, 2016; Tal & Arnon, 2018; Waterfall, 2006). Following Ervin (1961), variation sets are usually thought to indicate to the learner that certain morphemes are interchangeable in a paradigm. A variation set may additionally teach the learner that all segments of a morpheme belong together: Following the principle of common fate, all segments comprising a morpheme "change

together" and therefore can fuse and evoke one another. In other words, in addition to teaching that two morphemes belong together in a paradigm, adjacent utterances may also teach that the segments comprising a morpheme (or word, or phrase) belong together.

7.2.5. Surprise!

The other major tweak to the model was to provide an additional role for surprise, which goes beyond the influence on learning rate that is proposed by all error-driven models (Baayen et al., 2016; Olejarczuk et al., 2018; Rescorla, 1988). The goal of the Rescorla-Wagner model is to make correct predictions when presented with certain cues. Surprise determines learning rate in that whenever an outcome is expected given the preceding cues, the model's beliefs are correct and need not be updated. The extent to which the model's beliefs are updated is proportional to how surprising the event is (see (1-a) and (0-a) terms in equations (1) and (2) in §6.1). The present results suggests that the Rescorla-Wagner model underestimates the importance of surprise, as the occurence of [tʃa] when the singular ends in [p] or [k], which leads the learner to expect [p] or [k] in the plural, boosts [tʃa] more strongly than predicted. This could be driven by a disconfirmation or novelty bias, where surprising events change beliefs more than would normatively be warranted (see Olejarczuk et al., 2018, for evidence of disconfirmation bias in phonological learning). Disconfirmation bias contrasts with confirmation bias, which is discounting evidence inconsistent with prior beliefs (in other words, underutilizing information from surprising events; Bacon, 1620/1932; see Klayman, 1995, Nickerson, 1988 for reviews). Confirmation bias is well-documented in other domains but there is limited evidence of it in language learning outside of experiments

197

where learners are asked to discover rules by explicitly testing the grammaticality of different sentences (Robinson, 1996). Studies have shown that evidence for a phonetically unmotivated pattern is taken by participants to imply the existence of the phonetically motivated counterpart: Exposure to palatalization before [e] leads to production of palatalization before [i] as well (Wilson, 2006), and training on a saltatory alternation (e.g. p~v) generalizes to the intermediate segments (e.g. b~v and f~v) (White, 2013, 2014). Even when explicitly trained to *not* alternate the intermediate segments, participants still often prefer to do so (White, 2013, 2014), disregarding the disconfirming evidence. While these results suggest that surprising linguistic events do not necessarily result in corresponding adjustment of beliefs, much more research is needed to determine under what circumstances confirmation vs. disconfirmation bias takes precedence.

7.3. Conclusion

In this work, we proposed the Perseveration Hypothesis, a novel explanation for Paradigm Uniformity, the avoidance of stem changes (especially large ones). The Perseveration Hypothesis claims that stem changes are leveled by paradigmatic perseveration within the production system. When trying to produce a novel form of a known word, other forms of the word are activated along with production schemas linked to the meaning to be expressed (e.g. PLURAL). The articulatory gestures comprising these forms are incorporated into the novel form through a process of blending the activated production representations (Kapatsinski, 2013; Taylor, 2012). When too much is copied, the stem change is leveled. To prevent wanton paradigm leveling, speakers learn paradigmatic associations between related forms, which specify that activation of a particular gesture in the base form should activate a different gesture in the to-be-

198

produced form, which is copied into the production plan under construction. These paradigmatic associations are harder to learn when gestures are dissimilar, because linking dissimilar representations requires modifying more synapses in the brain (Kapatsinski, 2011; Warker & Dell, 2006).

Paradigmatic perseveration and the bias against associating dissimilar forms comprises the Perseveration Hypothesis: Perseveration conflicts with changes mandated by paradigmatic associations, and associations between dissimilar representations are more difficult to acquire.

The Perseveration Hypothesis makes the unique claim that the bias against large changes is strongest in production, because *performing* the change is what is difficult. We do see a dissociation between production and judgment: In Experiment 2, participants trained on labial palatalization (a large change) judge palatalization to be better than non-palatalization, but are unlikely to produce it. Overgeneralization of large changes to small changes is present in the judgment task, but overgeneralization is not itself the cause of Paradigm Uniformity. Participants may accept alternating forms in judgment because they contain the appropriate cues to meaning (e.g. by following a first-order schema like "plurals end in [tʃi]") without being able to produce the forms themselves. Difficulties in performing the stem change, especially when it is large, erode its productivity: A change that is hard to produce is more likely to be leveled, preventing future speakers from encountering the alternation and falling out of use over time, and this loss of productivity is widespread enough to arguably be universal (Bybee, 2008).

Acquiring morphology is often considered the hardest part of learning language (Slabakova, 2008). Morphological cues are often low in salience and therefore are often

missed or underutilized (MacWhinney et al., 1985). Morphology is also rife with paradigmatic mappings, which play a relatively minor role elswhere in the grammar (Kapatsinski, 2018a). In Experiment 3, we show that both issues are less severe when corresponding forms are in temporal contiguity. Through perceptual grouping, or the principle of common fate (Wertheimer, 1923/1938), parts of a form that jointly cue a difference in meaning between forms "pop out," which makes them easier to cue each other. Contiguity may be essential for learning novel unfaithful mappings, in that paradigmatic mappings become more apparent as contiguity makes the phonological properties of one form available for predicting properties of the other. It may also be important for learning *faithful* mappings, which require copying elements of the base form into the output: Both faithful and unfaithful mappings can be overgeneralized, in a way that is sensitive to the similarity structure of phonological space. Contiguity between forms exemplifying a mapping does not, on its own, prevent overgeneralization. For example, noticing k→tʃ (through contiguous singulars and plurals) favors mapping both [k] and [t] onto [tʃ]; noticing p→tʃ favors mapping all consonants onto [tʃ]; and noticing t→t and p→p favors k→k. In the condition where only unfaithful mappings are intact, changes are overgeneralized, and when only faithful mappings are intact, changes avoid notice and are rarely produced. Only when both are in contiguity are changes constrained to the appropriate context, as discriminative learning partitions the space of inputs between faithful and unfaithful mappings.

Early work assumed that morphologically related words are rarely if ever in contiguity (McNeill, 1966), but recent work in corpus and computational linguistics shows that contiguity between related words is a ubiquitous feature of natural language

and can be helpful for detecting morphological paradigms like that *went* is the past tense of *go* (Baroni et al., 2002; Xu & Croft, 1998). We suspect, based on the results of Experiment 3, that contiguity is not only helpful for computational models but is also essential for human learners of morphology and morphophonology. Paradigmatic relations may only be learnable because paradigmatically-related words are subject to syntagmatic co-occurrence, which allows the patterns between related forms to be noticed and acquired.

APPENDIX A

EXPERIMENT 1 AND EXPERIMENT 2 JUDGMENT STIMULUS LISTS

In the lists below, the plurals are formed by suffixing the vowel to the singular stem(e.g. blæb(-i,-a)→blæbi, blæba) or by removing the stem-final consonant and replacing it with the palatal-vowel combination(e.g. blæb(-dʒi,-dʒa)→blædʒi, blædʒa).

Labial stems:
blæb(-i,-a,-dʒi,-dʒa), fraɪb(-i,-a,-dʒi,-dʒa), kwoʊb(-i,-a,-dʒi,-dʒa), prub(-i,-a,-dʒi,-dʒa), smæb(-i,-a,-dʒi,-dʒa), blaɪp(-i,-a,-tʃi,-tʃa), frɛp(-i,-a,-tʃi,-tʃa), præp(-i,-a,-tʃi,-tʃa), skip(-i,-a,-tʃi,-tʃa), smip(-i,-a,-tʃi,-tʃa)

Alveolar stems:
bloʊd(-i,-a,-dʒi,-dʒa), frɑd(-i,-a,-dʒi,-dʒa), kwæd(-i,-a,-dʒi,-dʒa), preɪd(-i,-a,-dʒi,-dʒa), smɑd(-i,-a,-dʒi,-dʒa), blit(-i,-a,-tʃi,-tʃa), frɛt(-i,-a,-tʃi,-tʃa), kweɪt(-i,-a,-tʃi,-tʃa), prut(-i,-a,-tʃi,-tʃa), smɑt(-i,-a,-tʃi,-tʃa)

Velar stems:
blɪg(-i,-a,-dʒi,-dʒa), fraɪg(-i,-a,-dʒi,-dʒa), kwɪg(-i,-a,-dʒi,-dʒa), prɪg(-i,-a,-dʒi,-dʒa), smɪg(-i,-a,-dʒi,-dʒa), bleɪk(-i,-a,-tʃi,-tʃa), frik(-i,-a,-tʃi,-tʃa), kwɑk(-i,-a,-tʃi,-tʃa), praɪk(-i,-a,-tʃi,-tʃa), smɛk(-i,-a,-tʃi,-tʃa)

APPENDIX B

EXPERIMENT 2 STIMULUS LISTS

Training
Labial Palatalization training language
brib(-dʒi, 1), tʃaɪb(-dʒi, 1), gib(-dʒi, 1), hɛb(-dʒi, 4), paɪb(-dʒi, 4), vab(-dʒi, 1), broup(-tʃi, 4), glɑp(-tʃi, 1), heɪp(-tʃi, 1), naɪp(-tʃi, 1), slæp(-tʃi, 4), snip(-tʃi, 1), tʃaɪd(-a, 1), dæt(-a, 2), drid(-a, 2), feɪd(-i, 2), flaɪt(-a, 1), hɛd(-i, 1), lɑt(-i, 2), preɪt(-i, 1), tʃik(-a, 1), faɪk(-a, 2), glug(-i, 2), gwig(-a, 2), heɪg(-a, 1), noʊk(-i, 2), roʊg(-i, 1), spaɪk(-i, 1)

Alveolar Palatalization training language
brid(-dʒi, 1), tʃaɪd(-dʒi, 1), gid(-dʒi, 1), hɛd(-dʒi, 4), paɪd(-dʒi, 4), vad(-dʒi, 1), brout(-tʃi, 4), glɑt(-tʃi, 1), heɪt(-tʃi, 1), naɪt(-tʃi, 1), slæt(-tʃi, 4), snit(-tʃi, 1), tʃaɪp(-a, 1), dæp(-a, 2), drip(-a, 2), feɪp(-i, 2), flaɪp(-a, 1), hɛp(-i, 1), lɑp(-i, 2), preɪp(-i, 1), tʃik(-a, 1), faɪk(-a, 2), glug(-i, 2), gwig(-a, 2), heɪg(-a, 1), noʊk(-i, 2), roʊg(-i, 1), spaɪk(-i, 1)

Velar Palatalization training language
brig(-dʒi, 1), tʃaɪg(-dʒi, 1), gig(-dʒi, 1), hɛg(-dʒi, 4), paɪg(-dʒi, 4), vag(-dʒi, 1), brouk(-tʃi, 4), glɑk(-tʃi, 1), heɪk(-tʃi, 1), naɪk(-tʃi, 1), slæk(-tʃi, 4), snik(-tʃi, 1), tʃip(-a, 1), faɪp(-a, 2), glub(-i, 2), gwib(-a, 2), heɪb(-a, 1), noʊp(-i, 2), roʊb(-i, 1), spaɪp(-i, 1), tʃaɪd(-a, 1), dæt(-a, 2), drid(-a, 2), feɪd(-i, 2), flaɪt(-a, 1), hɛd(-i, 1), lɑt(-i, 2), preɪt(-i, 1)

Production
Labial Palatalization training language
vɛg, θug, strig, sprug, smoʊg, slɛg, skwæg, sɛg, kwug, krig, klɪg, frig, fleɪg, drɑg, vɪk, traɪk, θoʊk, streɪk, smuk, slak, skwaɪk, sɪk, plæk, kweɪk, klak, fræk, foʊk, faɪk, wɛd, trid, ðoʊd, θaɪd, stɛd, snud, ʃroʊd, ʃlud, proʊd, præd, gwɪd, flad, faɪd, blud, waɪt, θut, ðat, stɪt, sprɛt, snɛt, ʃræt, ʃaɪt, plat, kraɪt, gweɪt, fɛt, draɪt, blɪt, smeɪb, skab, sɪb, sib, ʃrib, ʃlab, proʊb, nab, kwæb, klub, gwab, froʊb, fleɪb, blub, slɛp, slæp, skæp, ʃroʊp, ʃlɪp, roʊp, plɪp, kwap, krip, klip, gwæp, frɪp, dræp, bloʊp, θip, strup, spreɪp, smoʊp, trab, θaɪb, streɪb, snɛb

Alveolar Palatalization training language
vɛg, θug, strig, sprug, smoʊg, slɛg, skwæg, sɛg, kwug, krig, klɪg, frig, fleɪg, drɑg, vɪk, traɪk, θoʊk, streɪk, smuk, slak, skwaɪk, sɪk, plæk, kweɪk, klak, fræk, foʊk, faɪk, smeɪb, skab, sɪb, sib, ʃrib, ʃlab, proʊb, nab, kwæb, klub, gwab, froʊb, fleɪb, blub, slɛp, slæp, skæp, ʃroʊp, ʃlɪp, roʊp, plɪp, kwap, krip, klip, gwæp, frɪp, dræp, bloʊp, trid, ðoʊd, θaɪd, stɛd, snud, ʃroʊd, ʃlud, proʊd, præd, gwɪd, flad, faɪd, blud, waɪt, θut, ðat, stɪt, sprɛt, snɛt, ʃræt, ʃaɪt, plat, kraɪt, gweɪt, fɛt, draɪt, blɪt, moʊt, dʒeɪt, dit, chut, vid, slɛd, hoʊd, krɪd

Velar Palatalization training language
smeɪb, skab, sɪb, sib, ʃrib, ʃlab, proʊb, nab, kwæb, klub, gwab, froʊb, fleɪb, blub, slɛp, slæp, skæp, ʃroʊp, ʃlɪp, roʊp, plɪp, kwap, krip, klip, gwæp, frɪp, dræp, bloʊp, wɛd, trid, ðoʊd, θaɪd, stɛd, snud, ʃroʊd, ʃlud, proʊd, præd, gwɪd, flad, faɪd, blud, waɪt, θut, ðat, stɪt,

sprɛt, snɛt, ʃræt, ʃaɪt, plɑt, kraɪt, gweɪt, fɛt, draɪt, blɪt, vɛg, θug, strig, sprug, smoʊg, slɛg, skwæg, sɛg, kwug, krig, klɪg, frig, fleɪg, drɑg, vɪk, traɪk, θoʊk, streɪk, smuk, slɑk, skwaɪk, sɪk, plæk, kweɪk, klɑk, fræk, foʊk, faɪk, nik, mɛk, tʃuk, boʊk, pɑg, heɪg, dig, blɪg

Judgment Test
See Appendix A.

EXPERIMENT 3 STIMULUS LISTS

Training
Labial Palatalization training language
hɛt(-a, 2), paɪt(-a, 1), vɑt(-i, 2), git(-i, 1), broʊd(-a, 2), slæd(-a, 1), heɪd(-i, 2), naɪd(-i, 1), brik(-a, 2), tʃaɪk(-a, 1), feɪk(-i, 2), hɛk(-i, 1), dæg(-a, 2), flaɪg(-a, 1), lɑg(-i, 2), preɪg(-i, 1), gwip(-tʃa, 4), heɪp(-tʃa, 4), glup(-tʃa, 1), roʊp(-tʃa, 1), θap(-tʃa, 1), gɛp(-tʃa, 1), snaɪb(-dʒa, 4), brub(-dʒa, 4), nib(-dʒa, 1), spab(-dʒa, 1), pæb(-dʒa, 1), toʊb(-dʒa, 1)

Velar Palatalization training language
hɛt(-a, 2), paɪt(-a, 1), vɑt(-i, 2), git(-i, 1), broʊd(-a, 2), slæd(-a, 1), heɪd(-i, 2), naɪd(1 -i), brip(-a, 2), tʃaɪp(-a, 1), feɪp(-i, 2), hɛp(-i, 1), dæb(-a, 2), flaɪb(-a, 1), lɑb(-i, 2), preɪb(-i, 1), gwik(-tʃa, 4), heɪk(-tʃa, 4), gluk(-tʃa, 1), roʊk(-tʃa, 1), θak(-tʃa, 1), gɛk(-tʃa, 1), snaɪg(-dʒa, 4), brug(-dʒa, 4), nig(-dʒa, 1), spag(-dʒa, 1), pæg(-dʒa, 1), toʊg(-dʒa, 1)

Production
Labial Palatalization training language
blɪt, blud, draɪt, faɪd, fɛt, flad, gweɪt, gwɪd, kraɪt, plat, præd, proʊd, ʃlaɪt, ʃlud, ʃræt, ʃroʊd, snɛt, snud, sprɛt, stɛd, stɪt, θaɪd, ðat, ðoʊd, θut, trid, waɪt, wɛd, bloʊp, blub, dræp, fleɪb, frɪp, froʊb, gwab, gwæp, klip, klub, krip, kwæb, kwap, nab, plɪp, proʊb, roʊp, ʃlab, ʃlɪp, ʃrib, ʃroʊp, sib, sɪb, skab, skæp, slæp, slɛp, smeɪb, drag, faɪk, fleɪg, foʊk, fræk, frig, klak, klɪg, krig, kweɪk, kwug, plæk, sɛg, sɪk, skwæg, skwaɪk, slak, slɛg, smoʊg, smuk, sprug, streɪk, strig, θoʊk, θug, traɪk, vɛg, vɪk, smoʊp, snɛb, spreɪp, streɪb, strup, θaɪb, θip, trab, troʊdʒ, trædʒ, θidʒ, ʃrɛdʒ, ʃlædʒ, saɪdʒ, sædʒ, prɪdʒ, kwɛdʒ, kradʒ, kludʒ, klɪdʒ, fridʒ, drɪdʒ, θɛtʃ, θætʃ, sutʃ, staɪtʃ, slaɪtʃ, ʃratʃ, ʃlɪtʃ, prɛtʃ, pratʃ, plutʃ, kwatʃ, kloʊtʃ, gweɪtʃ, gwætʃ

Velar Palatalization training language
blɪt, blud, draɪt, faɪd, fɛt, flad, gweɪt, gwɪd, kraɪt, plat, præd, proʊd, ʃlaɪt, ʃlud, ʃræt, ʃroʊd, snɛt, snud, sprɛt, stɛd, stɪt, θaɪd, ðat, ðoʊd, θut, trid, waɪt, wɛd, bloʊp, blub, dræp, fleɪb, frɪp, froʊb, gwab, gwæp, klip, klub, krip, kwæb, kwap, nab, plɪp, proʊb, roʊp, ʃlab, ʃlɪp, ʃrib, ʃroʊp, sib, sɪb, skab, skæp, slæp, slɛp, smeɪb, drag, faɪk, fleɪg, foʊk, fræk, frig, klak, klɪg, krig, kweɪk, kwug, plæk, sɛg, sɪk, skwæg, skwaɪk, slak, slɛg, smoʊg, smuk, sprug, streɪk, strig, θoʊk, θug, traɪk, vɛg, vɪk, boʊk, chuk, dig, glug, heɪg, mɛk, nik, pag, troʊdʒ, trædʒ, θidʒ, ʃrɛdʒ, ʃlædʒ, saɪdʒ, sædʒ, prɪdʒ, kwɛdʒ, kradʒ, kludʒ, klɪdʒ, fridʒ, drɪdʒ, θɛtʃ, θætʃ, sutʃ, staɪtʃ, slaɪtʃ, ʃratʃ, ʃlɪtʃ, prɛtʃ, pratʃ, plutʃ, kwatʃ, kloʊtʃ, gweɪtʃ, gwætʃ

REFERENCES CITED

Ackerman, F., Blevins, J. P., & Malouf, R. (2009). Parts and wholes: Implicative patterns in inflectional paradigms. In J. P. Blevins & J. Blevins (Eds.), *Analogy in grammar* (pp. 54-82). Oxford University Press. doi:10.1093/acprof:oso/9780199547548.003.0003.

Ackerman, F., & Malouf, R. (2013). Morphological organization: The low conditional entropy conjecture. *Language*, *89*(3), 429-464. doi:10.1353/lan.2013.0054.

Albright, A. (2008). Explaining universal tendencies and language particulars in analogical change. In J. Good (Ed.), *Linguistic universals and language change*, 144-181. Oxford University Press. doi:10.1093/acprof:oso/9780199298495.003.0007.

Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition, 90*(2), 119-161. doi:10.1016/s0010-0277(03)00146-x.

Alderete, J. D. (2001). Dominance effects as transderivational anti-faithfulness. *Phonology*, *18*(2), 201-253. doi:10.1017/s0952675701004067.

Allen, B., & Becker, M. (2015). *Learning alternations from surface forms with sublexical phonology.* Unpublished manuscript, University of British Columbia, Vancouver, Canada, and Stony Brook University, Stony Brook, NY. Retrieved from https://ling.auf.net/lingbuzz/002503.

Ambridge, B., Pine, J. M., Rowland, C. F., & Chang, F. (2012). The roles of verb semantics, entrenchment and morphophonology in the retreat from dative argument structure overgeneralization errors. *Language, 88*(1), 45-81. doi:10.1353/lan.2012.0000.

Ambridge, B., Pine, J. M., Rowland, C. F. & Young, C. R. (2008). The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgments of argument structure overgeneralisation errors. *Cognition, 106*(1), 87-129. doi:10.1016/j.cognition.2006.12.015.

Andersen, H. (1973). Abductive and deductive change. *Language*, *49*(4), 765-793. doi:10.2307/412063.

Anttila, R. (1989). *Historical and comparative linguistics.* John Benjamins.

Arnold, D., Tomaschek, F., Sering, K., Lopez, F., & Baayen, R. H. (2017). Words from spontaneous conversational speech can be recognized with human-like accuracy by an error-driven learning algorithm that discriminates between meanings straight from smart acoustic features, bypassing the phoneme as recognition unit. *PLoS One, 12*(4), e0174623. doi:10.1371/journal.pone.0174623.

Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition, 122*(3), 292-305. doi:10.1016/j.cognition.2011.10.009.

Ashby, F. G., Ennis, J. M., & Spiering, B. J. (2007). A neurobiological theory of automaticity in perceptual categorization. *Psychological Review, 114*, 632-656. doi:10.1037/0033-295x.114.3.632.

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language, 37*(1), 94-117. doi:10.1006/jmla.1997.2509.

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review, 118*(3), 438-481. doi:10.1037/a0023851.

Baayen, R. H., Shaoul, C., Willits, J., & Ramscar, M. (2016). Comprehension without segmentation: A proof of concept with naive discriminative learning. *Language, Cognition and Neuroscience*, *31*(1), 106-128. doi:10.1080/23273798.2015.1065336.

Bacon, F. (1939). Novum organum. In E. A. Burtt (Ed.), *The English philosophers from Bacon to Mill* (pp. 24-123). Random House. (Original work published in 1620)

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes, 24*(4), 527-554. doi:10.1080/01690960802299378.

Bakovic, E. (2003). Vowel harmony and stem identity. *San Diego Linguistic Papers, 1*, 1-42. Retrieved from https://cloudfront.escholarship.org/dist/prd/content/qt7zw206pt/qt7zw206pt.pdf.

Bangasser, D. A., Waxler, D. E., Santollo, J., & Shors, T. J. (2006). Trace conditioning and the hippocampus: The importance of contiguity. *Journal of Neuroscience, 26*(34), 8702-8706. doi:10.1523/jneurosci.1742-06.2006.

Baroni, M., Matiasek, J., & Trost, H. (2002). Unsupervised discovery of morphologically related words based on orthographic and semantic similarity. In *Proceedings of the ACL-02 workshop on morphological and phonological learning, vol. 6* (pp. 48-57). Association for Computational Linguistics. doi:10.3115/1118647.1118653.

Bateman, N. (2007). *A crosslinguistic investigation of palatalization* (Doctoral dissertation, University of California San Diego). Available from ProQuest Dissertations and Theses database. (3262182)

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67(1)*, 1-48. doi:10.18637/jss.v067.i01.

Becker, M., & Gouskova, M. (2016) Source-oriented generalizations as grammar inference in Russian vowel deletion. *Linguistic Inquiry, 47*(3), 391-425. doi:10.1162/ling_a_00217.

Beckman, J. N. (1998). *Positional faithfulness* (Doctoral dissertation, University of Massachusetts). Available from ProQuest Dissertation and Theses database. (9823717)

Bennett, W. G., & Braver, A. (2015). The productivity of 'unnatural' labial palatalization in Xhosa. *Nordlyd*, *42*, 33-44. doi:10.7557/12.3738.

Benua, L. (1997). *Transderivational identity: Phonological relations between words* (Doctoral dissertation, University of Massachusetts). Available from ProQuest Dissertation and Theses database. (9809307)

Berko, J. (1958). The child's learning of English morphology. *Word, 14*, 150–177. doi:10.1080/00437956.1958.11659661.

Bhat, DN. (1978). A general study of palatalization. *Universals of Human Language*, *2*, 47-92.

Bickel, B., Banjade, G., Gaenszle, M., Lieven, E., Paudyal, N. P., Rai, I. P., ... & Stoll, S. (2007). Free prefix ordering in Chintang. *Language*, *83*(1), 43-73. doi:10.1353/lan.2007.0002.

Biedermann, B., Beyersmann, E., Mason, C., & Nickels, L. (2013). Does plural dominance play a role in spoken picture naming? A comparison of unimpaired and impaired speakers. *Journal of Neurolinguistics*, *26*(6), 712-736. doi:10.1016/j.jneuroling.2013.05.001.

Blevins, J. (2006). A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics*, *32*(2), 117-166. doi:10.1515/tl.2006.009.

Blevins, J. P. (2006). Word-based morphology. *Journal of Linguistics*, *42*(3), 531-573. doi:10.1017/s0022226706004191.

Blevins, J. P. (2013). Word-based morphology from Aristotle to modern WP (word and paradigm models). In K. Allan (Ed.), *The Oxford handbook of the history of linguistics* (pp. 41-85). Oxford University Press. doi:10.1093/oxfordhb/9780199585847.013.0017.

Blevins, J. P., Milin, P., Ramscar, M. (2017). The Zipfian paradigm cell filling problem. In F. Kiefer, J. P. Blevins, & H. Bartos (Eds.), *Morphological paradigms and functions* (pp. 139-158). Brill. doi:10.1163/9789004342934_008.

Blything, R. P., Ambridge, B., & Lieven, E. V. (2014). Children use statistics and semantics in the retreat from overgeneralization. *PLoS One, 9*(10), e110009. doi:10.1371/journal.pone.0110009.

Böhm, T. M., Shestopalova, L., Bendixen, A., Andreou, A. G., Georgiou, J., Garreau, G., ... & Winkler, I. (2013). The role of perceived source location in auditory stream segregation: Separation affects sound organization, common fate does not. *Learning & Perception, 5*(Supplement 2), 55-72. doi:10.1556/lp.5.2013.suppl2.5.

Bolognesi, R. (1998). *The phonology of Campidanian Sardinian: A unitary account of a self-organizing structure*. Holland Institute for Generative Linguistics.

Bonami, O., & Beniamine, S. (2016). Joint predictiveness in inflectional paradigms. *Word Structure, 9*(2), 156-182. doi:10.3366/word.2016.0092.

Bonami, O., & Strnadová, J. (2019). Paradigm structure and predictability in derivational morphology. *Morphology*, *29*(2), 167-197. doi:10.1007/s11525-018-9322-6.

Booij, G. (2010). Construction morphology. *Language and Linguistics Compass*, *4*(7), 543-555. doi:10.1111/j.1749-818x.2010.00213.x.

Boomershine, A., Hall, K. C., Hume, E., & Johnson, K. (2008). The impact of allophony versus contrast on speech perception. In P. Avery, B. E. Dresher & K. Rice (Eds.), *Contrast in phonology: Theory, perception, acquisition* (pp. 145-171). Mouton de Gruyter. doi:10.1515/9783110208603.2.145.

Boyd, J. K., & Goldberg, A. E. (2011). Learning what not to say: The role of statistical preemption and categorization in a-adjective production. *Language, 87*(1), 55-83. doi:10.1353/lan.2011.0012.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325-14329. doi:10.1073/pnas.0803390105.

Braine, M. D., Brody, R. E., Brooks, P. J., Sudhalter, V., Ross, J. A., Catalano, L., & Fisch, S. M. (1990). Exploring language acquisition in children with a miniature artificial language: Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of Memory and Language, 29*(5), 591-610. doi:10.1016/0749-596x(90)90054-4.

Braine, M. D., & Brooks, P. J. (1995). Verb argument structure and the problem of avoiding an overgeneral grammar. In M. Tomasello & W. E. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 353-376). Lawrence Erlbaum Associates. doi:10.4324/9781315806860.

Braver, A., & Bennett, W. G. (2015, January). *Phonology or morphology: Inter-speaker differences in Xhosa labial palatalization*. Paper presented at the 89th Annual Meeting of the Linguistic Society of America, Portland, OR.

Bregman, A. S., & Pinker, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology, 32*(1), 19-31. doi:10.1037/h0081664.

Brooks, P. J., Braine, M. D. S., Catalano, L., Brody, R. E., & Sudhalter, V. (1993). Acquisition of gender-like noun subclasses in an artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language, 32*, 79–95. doi:10.1006/jmla.1993.1005.

Brooks, P. J., Tomasello, M., Dodson, K., & Lewis, L. B. (1999). Young children's overgeneralizations with fixed transitivity verbs. *Child Development, 70*(6), 1325–1337. doi:10.1111/1467-8624.00097.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, *6*(2), 201-251. doi:10.1017/s0952675700001019.

Brown, R., & Berko, J. (1960). Word association and the acquisition of grammar. *Child Development, 31*(1), 1-14. doi:10.1111/j.1467-8624.1960.tb05779.x.

Burzio, L. (1996). Surface constraints versus underlying representations. In J. Durand & B. Laks (Eds.), *Current trends in phonology: Models and methods, vol. 1* (pp. 123-142). University of Salford. Retrieved from https://pdfs.semanticscholar.org/1cb4/4e69e127fe04be25fe744409f87e310d8e86.pdf.

Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language*, *89*, 68-86. doi:10.1016/j.jml.2015.12.009.

Bybee, J. (1985). *Morphology: A study of the relation between meaning and form*. John Benjamins. doi:10.1075/tsl.9.

Bybee, J. (2001). *Phonology and language use*. Cambridge University Press. doi:10.1017/cbo9780511612886.

Bybee, J. (2002). Sequentiality as the basis of constituent structure. In T. Givon & B. F. Malle (Eds.), *The evolution of language out of pre-language* (pp. 109–132). John Benjamins. doi:10.1075/tsl.53.07byb.

Bybee, J. (2008). Formal universals as emergent phenomena: The origins of structure preservation. In J. Good (Ed.), *Linguistic universals and language change* (pp. 108-121). Oxford University Press. doi:10.1093/acprof:oso/9780199298495.003.0005.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press. doi:10.1017/cbo9780511750526.

Bybee, J., & Slobin, D. I. (1982). Why small children cannot change language on their own: Suggestions from the English past tense. In A. Alqvist (Ed.), *Papers from the 5th International Conference on Historical Linguistics* (pp. 29-37). John Benjamins. doi:10.1075/cilt.21.07byb.

Caballero, G. (2010). Scope, phonology and morphology in an agglutinating language: Choguita Rarámuri (Tarahumara) variable suffix ordering. *Morphology*, *20*(1), 165-204. doi:10.1007/s11525-010-9147-4.

Caballero, G. & Kapatsinski, V. (2019). How agglutinative? Searching for cues to meaning in Choguita Rarámuri (Tarahumara) using discriminative learning. In A. Sims, A. Ussishkin, J. Parker, & S. Wray (Eds.), *Morphological typology and linguistic cognition*. Cambridge University Press.

Cai, D. J., Mednick, S. A., Harrison, E. M., Kanady, J. C., & Mednick, S. C. (2009). REM, not incubation, improves creativity by priming associative networks. *Proceedings of the National Academy of Sciences, 106*(25), 10130-10134. doi:10.1073/pnas.0900271106.

Carvalho, P. F., & Goldstone, R. L. (2015). The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. *Psychonomic Bulletin & Review, 22*(1), 281-288. doi:10.3758/s13423-014-0676-4.

Chen, M. (1973). On the formal expression of natural rules in phonology. *Journal of Linguistics, 9*(02), 223-249. doi:10.1017/s0022226700003765.

Chomsky, N., & Halle, M. (1965). Some controversial questions in phonological theory. *Journal of Linguistics, 1*(2), 97-138. doi:10.1017/s0022226700001134.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper & Row. Retrieved from ERIC database. (ED020511)

Christdas, Prathima. (1988). *The phonology and morphology of Tamil* (Doctoral dissertation, Cornell University). Avaliable from ProQuest Dissertation and Theses database. (8900809)

Clark, R. (1974). Performing without competence. *Journal of Child Language, 1*, 1-10. doi:10.1017/s0305000900000040.

Clark, R. (1977). What's the use of imitation? *Journal of Child Language, 4*, 341-58. doi:10.1017/s0305000900001732.

Clements, G. N., and Hume, E. (1995). The internal organization of speech sounds. In J. Goldsmith (Ed.), *Handbook of phonological theory* (pp. 245–306). Blackwell.

Cristià, A., & Seidl, A. (2008). Is infants' learning of sound patterns constrained by phonological features? *Language Learning and Development*, *4*(3), 203-227. doi:10.1080/15475440802143109.

Dąbrowska, E. (2012). Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism*, *2*(3), 219-253. doi:10.1075/lab.2.3.01dab.

Dąbrowska, E., & Szczerbiński, M. (2006). Polish children's productivity with case marking: the role of regularity, type frequency, and phonological diversity. *Journal of Child Language*, *33*(3), 559-597. doi:10.1017/s0305000906007471.

Davidson, L. (2011). Characteristics of stop releases in American English spontaneous speech. *Speech Communication*, *53*(8), 1042-1058. doi:10.1016/j.specom.2011.05.010.

Davis, M. H., Di Betta, A. M., Macdonald, M. J., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience, 21*(4), 803-820. doi:10.1162/jocn.2009.21059.

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 364*(1536), 3773-3800. doi:10.1098/rstb.2009.0111.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283-321. doi:10.1037/0033-295x.93.3.283.

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, *104*(1), 123-147. doi:10.1037//0033-295x.104.1.123.

Do, Y. A. (2013). *Biased learning of phonological alternations*. (Doctoral Dissertation, MIT). Retrieved from https://dspace.mit.edu/bitstream/handle/1721.1/84416/868024936-MIT.pdf?sequence=2&isAllowed=y.

Do, Y. (2018). Paradigm uniformity bias in the learning of Korean verbal inflections. *Phonology*, *35*(4), 547-575. doi:10.1017/s0952675718000209.

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18*(1), 35-39. doi:10.1111/j.1467-9280.2007.01845.x.

Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied linguistics*, *27*(1), 1-24. doi:10.1093/applin/ami038.

Ellis, N. C. (2017). Chunking in language usage, learning and change: I don't know. In M. Hundt, S. Mollin, & S. E. Pfenninger (Eds.), *The changing English language: Psycholinguistic perspectives* (pp. 113-147). Cambridge University Press. doi:10.1017/9781316091746.006.

Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 229-240. doi:10.1037//0096-1523.27.1.229.

Ervin, S. M. (1961). Changes with age in the verbal determinants of word-association. *American Journal of Psychology*, *74*, 361-372. doi:10.2307/1419742.

Ettlinger, M., Morgan-Short, K., Faretta-Stutenberg, M., & Wong, P. C. (2016). The relationship between artificial and second language learning. *Cognitive Science*, *40*(4), 822-847. doi:10.1111/cogs.12257.

Farrar, M. J. (1992). Negative evidence and grammatical morpheme acquisition. *Developmental Psychology*, *28*(1), 90-98. doi:10.1037//0012-1649.28.1.90.

Feldman, J. (2003). The simplicity principle in human concept learning. *Current Directions in Psychological Science*, *12*(6), 227-232. doi:10.1046/j.0963-7214.2003.01267.x.

Fellbaum, C. (1996). Co-occurrence and antonymy. *International Journal of Lexicography, 8*(4), 281-303. doi:10.1093/ijl/8.4.281.

Finkel, R., & Stump, G. (2007). Principal parts and morphological typology. *Morphology, 17*(1), 39-75. doi:10.1007/s11525-007-9115-9.

Finley, S. (2008). *Formal and cognitive restrictions on vowel harmony*. (Doctoral dissertation, Johns Hopkins University). Available from ProQuest Dissertations and Theses database. (3339713)

Finley, S. (2015). Learning exceptions in phonological alternations. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 698-703). Cognitive Science Society. Retrieved from https://pdfs.semanticscholar.org/9e95/b5c10202ff40e74c3406c2c6f398016b0377.pdf.

Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition*, *108*(2), 477-499. doi:10.1016/j.cognition.2008.04.002.

Frigo, L., & McDonald, J. L. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language, 39*(2), 218-245. doi:10.1006/jmla.1998.2569.

Garcia, C., van Horne, K. D., & Hartshorne, J. (2017). *Replication of Finn & Hudson Kam (2008) The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation, Exp. 1*. Retrieved from PsyArXiv. doi:10.17605/OSF.IO/2XCWK.

Gibbon, F. E. (1999). Undifferentiated lingual gestures in children with articulation/phonological disorders. *Journal of Speech, Language, and Hearing Research, 42*(2), 382-397. doi:10.1044/jslhr.4202.382

Goldberg, A. E. (1995). *Constructions: A Construction Grammar approach to argument structure.* Chicago University Press.

Goldberg, A. E. (2002). Surface generalizations: An alternative to alternations. *Cognitive Linguistics*, *13*(4), 327-356. doi:10.1515/cogl.2002.022.

Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences*, *7*(5), 219-224. doi:10.1016/s1364-6613(03)00080-9.

Goldberg, A. E. (2011). Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics, 22*(1), 131-153. doi:10.1515/cogl.2011.006.

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, *100*(13), 8030-8035. doi:10.1073/pnas.1332441100.

Goldstone, R. L. (2000). Unitization during category learning. *Journal of Experimental Psychology: Human Perception and Performance, 26*(1), 86-112. doi:10.1037//0096-1523.26.1.86.

Goldstone, R. L. (2003). Learning to perceive while perceiving to learn. In R. Kimchi, M. Behrmann, & C. R. Olson (Eds.), *Perceptual organization in vision* (pp. 245-290). Psychology Press.

Gontijo, P. F., Gontijo, I., & Shillcock, R. (2003). Grapheme–phoneme probabilities in British English. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 136-157. doi:10.3758/bf03195506.

Goodsitt, J. V., Morgan, J. L., & Kuhl, P. K. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language, 20*(2), 229-252. doi:10.1017/s0305000900008266.

Gouskova, M., & Becker, M. (2013). Nonce words show that Russian yer alternations are governed by the grammar. *Natural Language & Linguistic Theory*, *31*(3), 735-765. doi:10.1007/s11049-013-9197-5.

Gouskova, M., Newlin-Łukowicz, L., & Kasyanenko, S. (2015). Selectional restrictions as phonotactics over sublexicons. *Lingua*, *167*, 41-81. doi:10.1016/j.lingua.2015.08.014.

Guion, S. G. (1998). The role of perception in the sound change of velar palatalization. *Phonetica*, *55*(1-2), 18-52. doi:10.1159/000028423.

Hale, M. & Reiss, C. "Substance abuse" and "dysfunctionalism": Current trends in phonology. *Linguistic Inquiry, 31*(1), 157-169. https://doi.org/10.1162/002438900554334.

Harmon, Z., & Kapatsinski, V. (2017). Putting old tools to novel uses: The role of form accessibility in semantic extension. *Cognitive Psychology*, *98*, 22-44. doi:10.1016/j.cogpsych.2017.08.002.

Harmon, Z., & Kapatsinski, V. (2019). *The target grammar is variable: Speakers' beliefs about the optimality of probability matching*. Manuscript in preparation.

Haspelmath, M. (1995). The growth of affixes in morphological reanalysis. In G. Booij (Ed.), *Yearbook of Morphology 1994* (pp. 1-29). Springer. doi:10.1007/978-94-017-3714-2_1.

Hayes, B. (2004). Phonological acquisition in Optimality Theory: The early stages. In R. Kager, J. Pater & P. Zonneveld (Eds.), *Constraints in phonological acquisition* (pp. 158-203). Cambridge University Press. doi:10.1017/cbo9780511486418.006.

Hayes, B., Siptár, P., Zuraw, K. & Londe, Z. (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Language, 85*(4), 822-863. doi:10.1353/lan.0.0169.

Hayes, B. & White, J. (2015). Saltation and the P-map. *Phonology*, *32*(2), 1–36. doi:10.1017/s0952675715000159.

Hluštík, P., Solodkin, A., Noll, D. C., & Small, S. L. (2004). Cortical plasticity during three-week motor skill learning. *Journal of Clinical Neurophysiology*, *21*(3), 180-191. doi:10.1097/00004691-200405000-00006.

Hock, H. H. (1991). *Principles of historical linguistics*. Mouton de Gruyter. doi:10.1515/9783110219135.

Hockett, C. F. (1954). Two models of grammatical description. *Word*, *10*, 210-234. doi:10.1080/00437956.1954.11659524.

Hockett, C. F. (1967). The Yawelmani basic verb. *Language, 43,* 208-222. doi:10.2307/411395.

Honeybone, P. (2001). Lenition inhibition in Liverpool English. *English Language & Linguistics*, *5*(2), 213-249. doi:10.1017/s1360674301000223.

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical statistics, 15*(3), 651-674. doi:10.1198/106186006x133933.

Householder, F. W. (1966). Phonological theory: A brief comment. *Journal of Linguistics*, *2*(1), 99-100. doi:10.1017/s0022226700001353.

Hudson Kam, C.L., & Newport, E.L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development, 1*(2), 151-195. doi:10.1207/s15473341lld0102_3.

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, *59*(1), 30-66. doi:10.1016/j.cogpsych.2009.01.001.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp.145-165). Morgan Kaufmann. doi:10.1002/9780470757024.ch15.

Johnson, K., & Babel, M. (2010). On the perceptual basis of distinctive features: Evidence from the perception of fricatives by Dutch and English speakers. *Journal of Phonetics*, *38*(1), 127-136. doi:10.1016/j.wocn.2009.11.001.

Jones, S. (2002). *Antonymy: A corpus-based approach*. Routledge. doi:10.4324/9780203166253.

Jones, S., Paradis, C., Murphy, M. L., & Willners, C. (2007). Googling for 'opposites': A web-based study of antonym canonicity. *Corpora, 2*(2), 129-154. doi:10.3366/cor.2007.2.2.129.

Joos, M. (1942). A phonological dilemma in Canadian English. *Language, 18*, 141-4. doi:10.2307/408979

Justeson, J. S., & Katz, S. M. (1991). Co-occurrences of antonymous adjectives and their contexts. *Computational Linguistics, 17*, 1-19. Retrieved from https://www.aclweb.org/anthology/J91-1001.

Kager, R. (1999). *Optimality Theory*. Cambridge University Press. doi:10.1017/cbo9780511812408.

Kapatsinski, V. (2007). Implementing and testing theories of linguistic constituency I: English syllable structure. *Research on Spoken Language Processing Progress Report, 28*, 241-76. Retrieved from https://www.researchgate.net/profile/Vsevolod_Kapatsinski/publication/237119532_Implementing_and_Testing_Theories_of_Linguistic_Constituency_I_English_Syllable_Structure_1/links/00b7d526751ad24005000000.pdf.

Kapatsinski, V. (2009). Testing theories of linguistic constituency with configural learning: The case of the English syllable. *Language, 85*(2), 248-277. doi:10.1353/lan.0.0118.

Kapatsinski, V. (2010). Velar palatalization in Russian and artificial grammar: Constraints on models of morphophonology. *Laboratory Phonology*, *1*(2), 361-393. doi:10.1515/labphon.2010.019.

Kapatsinski, V. (2011). Modularity in the channel: The link between separability of features and learnability of dependencies between them. *Proceedings of the XVIIth International Congress of Phonetic Sciences*, 1022-1025. Retrieved from https://www.researchgate.net/profile/Vsevolod_Kapatsinski/publication/258205598_Modularity_in_the_channel_The_link_between_separability_of_features_and_learnability_of_dependencies_between_them/links/0c960527358263d2f6000000/Modularity-in-the-channel-The-link-between-separability-of-features-and-learnability-of-dependencies-between-them.pdf.

Kapatsinski, V. (2012). What statistics do learners track? Rules, constraints and schemas in (artificial) grammar learning. In S. Th. Gries & D. Divjak (Eds.), *Frequency effects in language learning and processing* (pp. 53-73). Mouton de Gruyter. doi:10.1515/9783110274059.53.

Kapatsinski, V. (2013). Conspiring to mean: Experimental and computational evidence for a usage-based harmonic approach to morphophonology. *Language*, *89*(1), 110-148. doi:10.1353/lan.2013.0003.

Kapatsinski, V. (2017a). Copying, the source of creativity. In A. Makarova, S. M. Dickey & D. Divjak (Eds.), *Each venture a new beginning: Studies in honor of Laura A. Janda* (pp. 57-70). Slavica. Retrieved from https://blogs.uoregon.edu/ublab/files/2017/10/JandaCopying-1j97p9i.pdf.

Kapatsinski, V. (2017b). Learning a subtractive morphological system: Statistics and representations. *Proceedings of the 41ˢᵗ Annual Boston University Conference on Language Development*, 357-372. Retrieved from https://www.researchgate.net/profile/Vsevolod_Kapatsinski/publication/332353115_Learning_a_Subtractive_Morphological_System_Statistics_and_Representations/links/5caf7361299bf120975f695f/Learning-a-Subtractive-Morphological-System-Statistics-and-Representations.pdf.

Kapatsinski, V. (2018a). *Changing minds changing tools: From learning theory to language acquisition to language change*. MIT Press. doi:10.7551/mitpress/11400.001.0001.

Kapatsinski, V. (2018b). Learning morphological constructions. In G. Booij (Ed.), *The construction of words: Advances in construction morphology, Vol. 4* (pp. 547-581). Springer. doi:10.1007/978-3-319-74394-3_19.

Kapatsinski, V., & Harmon, Z. (2017). A Hebbian account of entrenchment and (over)-extension in language learning. *Proceedings of the Annual Meeting of the Cognitive Science Society, 39,* 2366-2371. Retrieved from https://pdfs.semanticscholar.org/a13f/16376b3bedb073d6ecacc9abfcf47c6fa1b2.pdf.

Kempen, G., & Harbusch, K. (2005). The relationship between grammaticality ratings and corpus frequencies: A case study into word order variability in the midfield of German clauses. In T. Pechmann & C. Habel (Eds.), *Linguistic evidence: Empirical, theoretical, and computational perspectives* (pp.329-349). Mouton de Gruyter. doi:10.1515/9783110197549.329.

Kenstowicz, M. (1996). Base identity and uniform exponence: Alternatives to cyclicity. In J. Durand & B. Laks (Eds.), *Current trends in phonology: Models and methods, Vol. 1* (pp. 363-393). University of Salford. Retrieved from https://rucore.libraries.rutgers.edu/rutgers-lib/39725/PDF/1/.

Kenstowicz, M. (1998). *Uniform exponence: Exemplification and extension*. Unpublished manuscript, Massachusetts Institute of Technology, Cambridge, MA. Retrieved from https://rucore.libraries.rutgers.edu/rutgers-lib/39727/PDF/1/.

Kerkhoff, A. O. (2007). *Acquisition of morpho-phonology: The Dutch voicing alternation* (Doctoral dissertation, University of Nijmegen, Nijmegen, Netherlands). Retrieved from https://dspace.library.uu.nl/bitstream/handle/1874/22598/full.pdf%3Bjsessionid%3DD62E8DEC366B4193947FD8ECF10CC8DC?sequence%3D1.

Klayman, J. (1995). Varieties of confirmation bias. *Psychology of Learning and Motivation, 32*, 385-418. doi:10.1016/s0079-7421(08)60315-1.

Kochetov, A. (2011). Palatalization. In C. Ewen, B. Hume, M. van Oostendorp, & K. Rice (Eds.), *Blackwell companion to phonology* (pp. 1666-1690). Wiley-Blackwell. doi:10.1002/9781444335262.wbctp0071.

Köhler, W. (1929). *Gestalt Psychology.* Liveright.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science*, *21*(11), 1551-1556. doi:10.1177/0956797610385359.

Krajewski, G., Theakston, A. L., Lieven, E. V., & Tomasello, M. (2011). How Polish children switch from one case to another when using novel nouns: Challenges for models of inflectional morphology. *Language and Cognitive Processes*, *26*(4-6), 830-861. doi:10.1080/01690965.2010.506062.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*(1), 22-44. doi:10.1037//0033-295x.99.1.22.

Kuczaj, S. A. (1977). The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior, 16*(5), 589-600. doi:10.1016/s0022-5371(77)80021-2.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences, 97*(22), 11850-11857. doi:10.1073/pnas.97.22.11850.

Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What learning systems do intelligent agents need? Complementary learning systems theory updated. *Trends in Cognitive Sciences, 20*(7), 512-534. doi:10.1016/j.tics.2016.05.004.

Küntay, A. and Slobin, D. (1996). Listening to a Turkish mother: Some puzzles for acquisition. In D. Slobin & J. Gerhardt (Eds.), *Social interaction, social context, and language: Essays in honor of Susan Ervin-Tripp* (pp. 265-286). Erlbaum.

Kurisu, K. (2001). *The phonology of morpheme realization* (Doctoral dissertation, UC Santa Cruz). Available from ProQuest Dissertation and Theses database. (3029802)

Labov, W. (1969). Contraction, deletion, and inherent variability of the English copula. *Language*, *45*(4), 715-762. doi:10.2307/412333.

Labov, W., & Austerlitz, R. (1975). Empirical foundations of linguistic theory. In R. Austerlitz (Ed.), *The scope of American linguistics* (pp. 77-133). Peter de Ridder. doi:10.1515/9783110857610-006.

Labov, W. (1996). When intuitions fail. *Chicago Linguistic Society*, *32*, 76-106.

Lehmann, C. (1992). Word order change by grammaticalization. In M. Gerritsen & D. Stein (Eds.), *Internal and external factors in syntactic change* (pp. 395-416). Mouton de Gruyter. doi:10.1515/9783110886047.395.

Lewis, P. A., & Durrant, S. J. (2011). Overlapping memory replay during sleep builds cognitive schemata. *Trends in Cognitive Sciences, 15*(8), 343-351. doi:10.1016/j.tics.2011.06.004.

Lim, S. J., Fiez, J. A., & Holt, L. L. (2014). How may the basal ganglia contribute to auditory categorization and speech perception? *Frontiers in Neuroscience, 8*, 230. doi:10.3389/fnins.2014.00230.

Lobben, M. (1991). *Pluralization of Hausa nouns, viewed from psycholinguistic experiments and child language data* (Master's thesis, University of Oslo, Oslo, Norway). Retrieved from https://www.academia.edu/170251/Pluralization_of_Hausa_nouns_-_viewed_from_psycholinguistic_experiments_and_child_language_data.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*(1), 1-36. doi:10.1097/00003446-199802000-00001.

Luce, R. D. (1959). *Individual choice behavior.* Wiley. doi:10.1037/14396-000.

MacWhinney, B., Pleh, C., & Bates, E. (1985). The development of sentence interpretation in Hungarian. *Cognitive Psychology, 17*(2), 178-209. doi:10.1016/0010-0285(85)90007-6.

Maddox, W. T., Filoteo, J. V., Lauritzen, J. S., Connally, E., & Hejl, K. D. (2005). Discontinuous categories affect information-integration but not rule-based category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(4), 654-669. doi:10.1037/0278-7393.31.4.654.

Maddox W. T., Filoteo J. V., Lauritzen J. S. (2007). Within-category discontinuity interacts with verbal rule complexity in perceptual category learning. *Journal of Experimental Psychology: Learning, Memory & Cognition, 33*, 197–218. doi:10.1037/0278-7393.33.1.197.

Magomedova, V. (2017). Pseudo-allomorphs in Modern Russian. *University of Pennsylvania Working Papers in Linguistics, 23*(1), 16. Retrieved from https://repository.upenn.edu/cgi/viewcontent.cgi?article=1956&context=pwpl.

Malouf, R. (2017). Abstractive morphological learning with a recurrent neural network. *Morphology, 27*(4), 431-458. doi:10.1007/s11525-017-9307-x.

Maniwa, K., Jongman, A., & Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. *The Journal of the Acoustical Society of America*, *125*(6), 3962-3973. doi:10.1121/1.2990715.

Martin, A. T. (2007). *The evolving lexicon*. (Doctoral dissertation, University of California Los Angeles). Available from ProQuest Dissertation and Theses database. (3302537)

Matthews, P. H. (1965). The inflectional component of a word-and-paradigm grammar. *Journal of Linguistics*, *1*(2), 139-171. doi:10.1017/s0022226700001146.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101-B111. doi:10.1016/s0010-0277(01)00157-3.

McCarthy, J. J. (1998). *Morpheme structure constraints and paradigm occultation*. Unpublished manuscript, University of Massachusetts, Amherst. Retrieved from https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1045&context=linguist_faculty_pubs.

McCarthy, J. J. & Prince, A. (1995). Faithfulness and reduplicative identity. *Linguistics Department Faculty Publication Series, 10*. Retrieved from https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1009&context=linguist_faculty_pubs.

McClelland, J. L. (2001). Failures to learn and their remediation: A Hebbian account. In J. L. McClelland & R. Siegler (Eds.), *Mechanisms of cognitive development: Behavioral and neural perspectives* (pp. 109-134). Psychology Press. doi:10.4324/9781410600646.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*(3), 419-457. doi:10.1037//0033-295x.102.3.419.

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review, 119*(4), 831-877. doi:10.1037/a0029872.

McNeill, D. (1963). The origin of associations within the same grammatical class. *Journal of Verbal Learning & Verbal Behavior, 3*, 250-262. doi:10.1016/s0022-5371(63)80091-2.

McNeill, D. (1966). A study of word association. *Journal of Verbal Learning & Verbal Behavior, 5*, 548-557. doi:10.1016/s0022-5371(66)80090-7.

Mielke, J. (2004). The emergence of distinctive features (Doctoral dissertation, The Ohio State University). Retrieved from https://etd.ohiolink.edu/!etd.send_file?accession=osu1092833440&disposition=inline

Mirman, D., McClelland, J. L., & Holt, L. L. (2006). An interactive Hebbian account of lexically guided tuning of speech perception. *Psychonomic Bulletin & Review, 13*(6), 958-965. doi:10.3758/bf03213909.

Mitroff, S. R., Simons, D. J., & Levin, D. T. (2004). Nothing compares 2 views: Change blindness can occur despite preserved access to the changed information. *Perception & Psychophysics, 66*(8), 1268-1281. doi:10.3758/bf03194997.

Mitrović, I. (2012). A phonetically natural vs. native language pattern: An experimental study of velar palatalization in Serbian. *Journal of Slavic Linguistics*, *20*(2), 229-268. doi:10.1353/jsl.2012.0011.

Moreton, E. (2008). Analytic bias and phonological typology. *Phonology*, *25*(1), 83-127. doi:10.1017/s0952675708001413.

Moreton, E. (2012). Inter-and intra-dimensional dependencies in implicit phonotactic learning. *Journal of Memory and Language*, *67*(1), 165-183. doi:10.1016/j.jml.2011.12.003.

Moreton, E., & Pater, J. (2012a). Structure and Substance in Artificial-phonology Learning, Part I: Structure. *Language and Linguistics Compass*, *6*(11), 686-701. doi:10.1002/lnc3.363.

Moreton, E., & Pater, J. (2012b). Structure and substance in artificial-phonology learning, Part II: Substance. *Language and Linguistics Compass*, *6*(11), 702-718. doi:10.1002/lnc3.366.

Moreton, E., Pater, J., & Pertsova, K. (2017). Phonological concept learning. *Cognitive Science, 41*(1), 4-69. doi:10.1111/cogs.12319.

Murphy, M. L. (2006). Antonyms as lexical constructions: or, why paradigmatic construction is not an oxymoron. *Constructions, SV1*(8), 1-37. Retrieved from https://journals.linguisticsociety.org/elanguage/constructions/article/download/23/23-81-1-PB.pdf.

Nelson, K. (1989). *Narratives from the crib.* Harvard University Press.

Nesset, T. (2008). *Abstract phonology in a concrete model: Cognitive linguistics and the morphology-phonology interface*. Mouton de Gruyter. doi:10.1515/9783110208368.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175-220. doi:10.1037//1089-2680.2.2.175.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204-238. doi:10.1016/s0010-0285(03)00006-9.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357. doi:10.1037/0033-295x.115.2.357.

O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: principles of cortical and hippocampal function. *Psychological Review*, *108*(2), 311-345. doi:10.1037/0033-295x.108.2.311.

Ohala, J. J. (1978). Southern Bantu vs. the world: The case of palatalization of labials. *Berkeley Linguistics Society*, *4*, 370-386. doi:10.3765/bls.v4i0.2218.

Ohala, J. J. (1989). Sound change is drawn from a pool of synchronic variation. In L. E. Breivik & E. H. Jahr (Eds.), *Language change: Contributions to the study of its causes* (pp. 173-198). Mouton de Gruyter. doi:10.1515/9783110853063.173.

Olejarczuk, P., Kapatsinski, V., & Baayen, R. H. (2018). Distributional learning is error-driven: The role of surprise in the acquisition of phonetic categories. *Linguistics Vanguard, 4*(s2). doi:10.1515/lingvan-2017-0020.

Onnis, L., Waterfall, H. R., & Edelman, S. (2008). Learn locally, act globally: Learning language from variation set cues. *Cognition*, *109*(3), 423-430. doi:10.1016/j.cognition.2008.10.004.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 309-28. doi:10.1037/0278-7393.19.2.309.

Pater, J., & Tessier, A. M. (2003). Phonotactic knowledge and the acquisition of alternations. *Proceedings of the 15th International Congress on Phonetic Sciences*, 1177-1180. https://pdfs.semanticscholar.org/3d47/386f875afd41b2b019f3d40413dfecb18f47.pdf.

Peperkamp, S., Le Calvez, R., Nadal, J.P., & Dupoux, E. (2006). The acquisition of allophonic rules: Statistical learning with linguistic constraints. *Cognition, 101*(3), B31-B41. doi:10.1016/j.cognition.2005.10.006.

Perfors, A. (2016). Adult regularization of inconsistent input depends on pragmatic factors. *Language Learning and Development, 12*(2), 138-155. doi:10.1080/15475441.2015.1052449.

Perkell, J. S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics*, *25*(5), 382-407. doi:10.1016/j.jneuroling.2010.02.011.

Pierrehumbert, J. B. (2006). The statistical basis of an unnatural alternation. In L. Goldstein, D. H. Whalen & C. Best (Eds.), *Laboratory phonology 8* (pp. 81-107). Mouton de Gruyter. doi:10.1515/9783110197211.1.81.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, *28*(1), 73-193. doi:10.1016/0010-0277(88)90032-7.

Plunkett, K., & Juola, P. (1999). A connectionist model of English past tense and plural morphology. *Cognitive Science, 23*, 463-490. doi:10.1207/s15516709cog2304_4.

Prince, A., & Smolensky, P. (1993/2004). *Optimality Theory: Constraint interaction in generative grammar*. Wiley. doi:10.1002/9780470759400.

Proudfoot, A., & Cardo, F. (2005). *Modern Italian grammar: A practical guide*. Routledge. doi:10.4324/9780203085035.

Psychology Software Tools, Inc. E-Prime 2.0 [Computer software]. Retrieved from https://www.pstnet.com.

Purcell, D. W., & Munhall, K. G. (2006). Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America*, *119*(4), 2288-2297. doi:10.1121/1.2173514.

Pycha, A., Nowak, P., Shin, E., & Shosted, R. (2003). Phonological rule-learning and its implications for a theory of vowel harmony. *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, *22*, 101-114. Retrieved from https://www.researchgate.net/profile/Anne_Pycha/publication/247174599_Phonologi cal_Rule-Learning_and_Its_Implications_for_a_Theory_of_Vowel_Harmony/links/55185a4d0 cf2f7d80a3df7fa.pdf.

R Core Team (2019). R: A language and environment for statistical computing (Version 3.1.1) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing.

Raffelsiefen, R. (2005). Paradigm uniformity effects vs. boundary effects. In L. J. Downing, T. A. Hall & R. Raffelsiefen (Eds.), *Paradigms in phonological theory* (pp. 211-262). Oxford University Press. doi:10.1093/acprof:oso/9780199267712.003.0009.

Ramscar, M. (2002). The role of meaning in inflection: Why the past tense does not require a rule. *Cognitive Psychology, 45*(1), 45-94. doi:10.1016/s0010-0285(02)00001-4.

Ramscar, M. (2013). Suffixing, prefixing, and the functional order of regularities in meaningful strings. *Psihologija, 46*(4), 377-396. doi:10.2298/psi1304377r.

Ramscar, M., Dye, M., & McCauley, S. M. (2013). Error and expectation in language learning: The curious absence of mouses in adult speech. *Language, 89*(4), 760-793. doi:10.1353/lan.2013.0068.

Ramscar, M., & Gitcho, N. (2007). Developmental change and the nature of learning in childhood. *Trends in Cognitive Sciences, 11*(7), 274-279. doi:10.1016/j.tics.2007.05.007.

Ramscar, M., & Yarlett, D. (2007). Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. *Cognitive Science, 31*(6), 927-960. doi:10.1080/03640210701703576.

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science, 34*(6), 909-957. doi:10.1111/j.1551-6709.2009.01092.x.

Redford, M. A. (2015). Unifying speech and language in a developmentally sensitive model of production. *Journal of Phonetics*, *53*, 141-152. doi:10.1016/j.wocn.2015.06.006.

Regier, T., & Gahl, S. (2004). Learning the unlearnable: The role of missing evidence. *Cognition, 93*(2), 147-155. doi:10.1016/j.cognition.2003.12.003.

Rescorla, R. A. (1986). Two perceptual variables in within-event learning. *Animal Learning & Behavior*, *14*(4), 387-392. doi:10.3758/bf03200083.

Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist, 43*(3), 151-160. doi:10.1037//0003-066x.43.3.151.

Rescorla, R. A., & Furrow, D. R. (1977). Stimulus similarity as a determinant of Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *3*(3), 203-215. doi:10.1037//0097-7403.3.3.203.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). Appleton-Century-Crofts. Retrieved from https://pdfs.semanticscholar.org/afaf/65883ff75cc19926f61f181a687927789ad1.pdf.

Robins, R. H. (1959). In defence of WP. *Transactions of the Philological Society*, *58*(1), 116-144. doi:10.1111/j.1467-968x.1959.tb00301.x.

Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search, and instructed conditions. *Studies in Second Language Acquisition, 18*(1), 27-67. doi:10.1017/s0272263100014674.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1), 107-142. doi:10.1016/0010-0277(92)90041-f.

Rubino, R. B., & Pine, J. M. (1998). Subject–verb agreement in Brazilian Portuguese: what low error rates hide. *Journal of Child Language*, *25*(01), 35-59. doi:10.1017/s0305000997003310.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In D. E. Rumelhart, J. L. McClelland & The PDP Research Group (Eds.), *Parallel distributed processing*, *Vol. 2*. MIT Press. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/a164233.pdf.

Sangster, C. M. (2001). Lenition of alveolar stops in Liverpool English. *Journal of Sociolinguistics*, *5*(3), 401-412. doi:10.1111/1467-9481.00156.

Saville-Troike, M. (1988). Private speech: Evidence for second language learning strategies during the 'silent'period. *Journal of Child Language, 15*(3), 567-590. doi:10.1017/s0305000900012575.

Scheer, T. (2011). *A guide to morphosyntax-phonology interface theories: How extra-phonological information is treated in phonology since Trubetzkoy's Grenzsignale*. Walter de Gruyter. doi:10.1515/9783110238631.

Schertz, J. (2013). Exaggeration of featural contrasts in clarifications of misheard speech in English. *Journal of Phonetics*, *41*(3), 249-263. doi:10.1016/j.wocn.2013.03.007.

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology*, *57*, 87−115. doi:10.1146/annurev.psych.56.091103.070229.

Schwab, J. F., & Lew-Williams, C. (2016). Repetition across successive sentences facilitates young children's word learning. *Developmental Psychology*, *52*(6), 879–86. doi:10.1037/dev0000125.

Schwartz, R. G., & Leonard, L. B. (1982). Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language, 9*(2), 319-336. doi:10.1017/s0305000900004748.

Seidl, A., & Buckley, E. (2005). On the learning of arbitrary phonological rules. *Language Learning and Development, 1*(3-4), 289-316. doi:10.1207/s15473341lld0103&4_4.

Seidl, A., Cristià, A., Bernard, A., & Onishi, K. H. (2009). Allophonic and phonemic contrasts in infants' learning of sound patterns. *Language Learning and Development, 5*(3), 191-202. doi:10.1080/15475440902754326.

Seyfarth, S., Ackerman, F., & Malouf, R. (2014). Implicative organization facilitates morphological learning. *Annual Meeting of the Berkeley Linguistics Society, 40*, 480-494. doi:10.3765/bls.v40i0.3154.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6*(1), 156-163. doi:10.1016/s0022-5371(67)80067-7.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied, 75*(13), 1-42. doi:10.1037/h0093825.

Sims, A. D., & Parker, J. (2016). How inflection class systems work: On the informativity of implicative structure. *Word Structure, 9*(2), 215-239. doi:10.3366/word.2016.0094.

Skoruppa, K., Lambrechts, A., & Peperkamp, S. (2011). The role of phonetic distance in the acquisition of phonological alternations. *Proceedings of the 39th Annual Meeting of the North East Linguistic Society*, 464-475. Retrieved from http://repository.essex.ac.uk/4251/1/SkoruppaetalNELS.pdf.

Skoruppa, K., & Peperkamp, S. Adaptation to novel accents: Feature-based learning of context-sensitive phonological regularities. *Cognitive Science, 35*(2), 348-366. doi:10.1111/j.1551-6709.2010.01152.x.

Slabakova, R. 2008. *Meaning in the second language*. Mouton de Gruyter. doi:10.1515/9783110211511.

Smith, L. B., Thelen, E., Titzer, R., & McLin, D. (1999). Knowing in the context of acting: the task dynamics of the A-not-B error. *Psychological Review, 106*(2), 235-260. doi:10.1037//0033-295x.106.2.235.

Smolek, A. & Kapatsinski, V. (2018). What happens to large changes? Saltation produces well-liked outputs that are hard to generate. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 9*(1), 10. doi:10.5334/labphon.93.

Smolek, A. & Kapatsinski, V. (2019). *Syntagmatic paradigms: Learning correspondence from contiguity.* Manuscript submitted for publication.

Standing, L. (1973). Learning 10000 pictures. *The Quarterly Journal of Experimental Psychology*, *25*(2), 207-222. doi:10.1080/14640747308400340.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, *19*(2), 73-74. doi:10.3758/bf03337426.

Stave, M., Smolek, A., & Kapatsinski, V. (2013). Inductive bias against stem changes as perseveration: Experimental evidence for an articulatory approach to output-output faithfulness. *Proceedings of the 35ᵗʰ Annual Meeting of the Cognitive Science Society*, 3454-3459. Retrieved from https://cloudfront.escholarship.org/dist/prd/content/qt293733rg/qt293733rg.pdf.

Stefanowitsch, A. (2008). Negative entrenchment: A usage-based approach to negative evidence. *Cognitive Linguistics, 19*(3), 513-531. doi:10.1515/cogl.2008.020.

Steriade, D. (2000). Paradigm uniformity and the phonetics-phonology boundary. In M. B. Broe & J. B. Pierrehumbert (Eds.), *Papers in laboratory phonology V: Acquisition and the lexicon* (pp. 313-334). Cambridge University Press. Retrieved from https://www.researchgate.net/profile/Donca_Steriade/publication/2495311_Paradigm _Uniformity_and_the_Phonetics-Phonology_Boundary/links/549047eb0cf214269f2664c6.pdf.

Steriade, D. (2001/2009). The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In K. Hanson & S. Inkelas (Eds.), *The nature of the word: Studies in honor of Paul Kiparsky* (pp. 151-170). MIT Press. doi:10.7551/mitpress/9780262083799.003.0007.

Stump, G., & Finkel, R. A. (2013). *Morphological typology: From word to paradigm*. Cambridge University Press. doi:10.1017/cbo9781139248860.

Sutherland, R. J., & Rudy, J. W. (1989). Configural association theory: The role of the hippocampal formation in learning, memory, and amnesia. *Psychobiology*, *17*(2), 129-144. Retrieved from https://link.springer.com/content/pdf/10.3758/BF03337828.pdf.

Taatgen, N. A., & Anderson, J. R. (2002). Why do children learn to say "broke"? A model of learning the past tense without feedback. *Cognition, 86*(2), 123-155. doi:10.1016/s0010-0277(02)00176-2.

Tal, S., & Arnon, I. (2018). SES effects on the use of variation sets in child-directed speech. *Journal of Child Language, 45*(6), 1423-1438. doi:10.1017/s0305000918000223.

Taylor, J. R. (2012). *The mental corpus: How language is represented in the mind*. Oxford University Press. doi:10.1093/acprof:oso/9780199290802.001.0001.

Theodore, R. M., Blumstein, S. E., & Luthra, S. (2015). Attention modulates specificity effects in spoken word recognition: Challenges to the time-course hypothesis. *Attention, Perception, & Psychophysics*, *77*(5), 1674-1684. doi:10.3758/s13414-015-0854-0.

Thorndike, E. L. (1898). *Animal intelligence, an experimental study of the associative processes in animals*. Macmillan. doi:10.1037/10780-000.

Thymé, A. (1993). *Connectionist approach to nominal inflection: Paradigm patterning and analogy in Finnish* (Doctoral dissertation, University of California San Diego). Available from ProQuest Dissertations and Theses database. (9317518)

Thymé, A., Ackerman, F., & Elman, J. L. (1994). Finnish nominal inflection: paradigmatic patterns and token analogy. In S. D. Lima, R. Corrigan, & G. K. Iverson (Eds.), *The reality of linguistic rules* (pp. 445-466). John Benjamins. doi:10.1075/slcs.26.25thy.

Tomas, E., van de Vijver, R., Demuth, K., & Petocz, P. (2017). Acquisition of nominal morphophonological alternations in Russian. *First Language*, *37*(5), 453-474. doi:10.1177/0142723717698839.

Umbreit, B. (2011). Motivational networks: An empirically supported cognitive phenomenon. In K.-U. Panther & G. Radden (Eds.), *Motivation in grammar and the lexicon* (pp. 269-286). John Benjamins. doi:10.1075/hcp.27.17umb.

Underhill, R. (1976). *Turkish grammar*. MIT Press. Retrieved from http://www.academia.edu/download/53475783/_Robert_Underhill__Turkish_Grammar__Turk_dili_graBookZZ.org.pdf.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review, 108*(3), 550-592. doi:10.1037//0033-295x.108.3.550.

Uttal, W. R., Spillmann, L., Stürzel, F., & Sekuler, A. B. (2000). Motion and shape in common fate. *Vision Research, 40*(3), 301-310. doi:10.1016/s0042-6989(99)00177-7.

Vihman, M., & Croft, W. (2007). Phonological development: Toward a "radical" templatic phonology. *Linguistics*, *45*(4), 683-725. doi:10.1515/ling.2007.021.

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America*, *122*(4), 2306-2319. doi:10.1121/1.2773966.

Waelti, P., Dickinson, A., & Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, *412*(6842), 43. doi:10.1038/35083500.

Wang, M. D. & Bilger, R. C. (1973). Consonant confusions in noise: A study of perceptual features. *Journal of the Acoustical Society of America, 54*(5), 1248-1266. doi:10.1121/1.1914417.

Wang, T., & Saffran, J. R. (2014). Statistical learning of a tonal language: The influence of bilingualism and previous linguistic experience. *Frontiers in Psychology*, *5*, 953. doi:10.3389/fpsyg.2014.00953.

Warker, J. A., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 387-398. doi:10.1037/0278-7393.32.2.387.

Warker, J. A., Dell, G. S., Whalen, C. A., & Gereg, S. (2008). Limits on learning phonotactic constraints from recent production experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1289-1295. doi:10.1037/a0013033.

Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychological Science*, *25*(7), 1314-1324. doi:10.1177/0956797614531023.

Warner, N., & Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America*, *130*, 1606. doi:10.1121/1.3621306.

Waterfall, H. R. (2006). A little change is a good thing: Feature theory, language acquisition and variation sets (Doctoral dissertation, University of Chicago). Available from ProQuest Dissertations and Theses database. (3219602)

Watson, J. C. (2002). *The phonology and morphology of Arabic*. Oxford University Press.

Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, *128*(2), 179-186. doi:10.1016/j.cognition.2013.03.002.

Weir, R. H. (1962). *Language in the crib*. Mouton.

Welsh, J. P., & Llinás, R. (1997). Some organizing principles for the control of movement based on olivocerebellar physiology. *Progress in Brain Research, 114*, 449-461. doi:10.1016/s0079-6123(08)63380-4.

Wertheimer, M. (1923/1938). Untersuchungen zur Lehre von der Gestalt II. *Psychologische Forschung, 4*, 301-350. doi:10.1007/bf00410640.

Westermann, G., & Ruh, N. (2012). A neuroconstructivist model of past tense development and processing. *Psychological Review, 119*(3), 649-667. doi:10.1037/a0028258.

White, J. (2013). *Bias in phonological learning: Evidence from saltation*. (Doctoral dissertation, University of California Los Angeles). Available from ProQuest Dissertations and Theses database. (3564463)

White, J. (2014). Evidence for a learning bias against saltatory phonological alternations. *Cognition*, *130*(1), 96-115. doi:10.1016/j.cognition.2013.09.008.

White, J. (2017). Accounting for the learnability of saltation in phonological theory: A maximum entropy model with a P-map bias. *Language, 93*(1), 1-36. doi:10.1353/lan.2017.0001.

White, J., & Sundara, M. (2014). Biased generalization of newly learned phonological alternations by 12-month-old infants. *Cognition*, *133*(1), 85-90. doi:10.1016/j.cognition.2014.05.020.

Williams, J. N. (2003). Inducing abstract linguistic representations: Human and connectionist learning of noun classes. In R. van Hout, A. Hulk, F. Kuiken, & R. J. Towell (Eds.), *The lexicon-syntax interface in second language acquisition* (pp. 151-174). John Benjamins. doi:10.1075/lald.30.08wil.

Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*, *30*(5), 945-982. doi:10.1207/s15516709cog0000_89.

Woodrow, H., & Lowell, F. (1916). Children's association frequency tables. *The Psychological Monographs, 22*(5), i-110. doi:10.1037/h0093111.

Xu, J., & Croft, W. B. (1998). Corpus-based stemming using co-occurrence of word variants. *ACM Transactions on Information Systems (TOIS), 16*(1), 61-81. doi:10.1145/267954.267957.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114*(2), 245-272. doi:10.1037/0033-295x.114.2.245.

Yu, C., & Smith, L. B. (2012). Modeling cross-situational word-referent learning: Prior questions. *Psychological Review, 119*(1), 21-39. doi:10.1037/a0026182.

Yun, G. H. (2006). *The interaction between palatalization and coarticulation in Korean and English*. (Doctoral dissertation, University of Arizona). Available from ProQuest Dissertations and Theses database. (3219841)

Zaki, S. R., & Salmi, I. L. (2019). Sequence as context in category learning: An eyetracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. doi:10.1037/xlm0000693.

Zaki, S. R., Rich, A., & Stacy, S. (2016, November). *The sequence of items in category learning: Modeling and eye-tracking data*. Paper presented at the 57th Annual Meeting of the Psychonomic Society, Boston, MA.

Zampini, L. M. (1996). Voiced stop spirantization in the ESL of native speakers of Spanish. *Applied Psycholinguistics*, *17*, 335–354. doi:10.1017/s0142716400007979.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Addison-Wesley.

Zsiga, E. C. (1995). An acoustic and electropalatographic study of lexical and postlexical palatalization in American English. In B. Connell & A. Arvaniti (Eds.), *Phonology and phonetic evidence: Papers in laboratory phonology IV* (pp. 282-302). Cambridge University Press. doi:10.1017/cbo9780511554315.020.

Zuraw, K. (2000). *Patterned exceptions in phonology* (Doctoral dissertation, University of California Los Angeles). Available from ProQuest Dissertations and Theses database. (9979100)

Zuraw, K. (2007). The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from Tagalog. *Language*, *83*, 277–316. doi:10.1353/lan.2007.0105.