

THE ARCHITECTURE OF GRAMMAR IN ARTIFICIAL GRAMMAR LEARNING:  
FORMAL BIASES IN THE ACQUISITION OF MORPHOPHONOLOGY AND THE  
NATURE OF THE LEARNING TASK

Vsevolod Kapatsinski

Submitted to the faculty of the University Graduate School  
in partial fulfillment of the requirements  
for the degree  
Doctor of Philosophy  
in the Departments of Linguistics  
and Cognitive Science  
Indiana University  
May 2009

Accepted by the Graduate Faculty, Indiana University, in partial  
fulfillment of the requirements for the degree of Doctor of Philosophy

---

David B. Pisoni, Ph.D. (Co-chair)

---

Kenneth J. de Jong, Ph.D. (Co-chair)

---

Linda B. Smith, Ph.D.

---

Robert F. Port, Ph.D.

05/14/2009

© 2009

Vsevolod Kapatsinski

ALL RIGHTS RESERVED

## Acknowledgements

I would like to extend my deepest thanks to David Pisoni for his guidance throughout my time here at Indiana and for all the resources and opportunities he provided to me for my studies. The Speech Research Lab has been my home for the last four years, and I can't think of a better place where I could pursue my research. Ken de Jong and Stuart Davis have provided invariably helpful advice on all matters academic and professional. Robert Port, Linda Smith, William Croft, Bruce Hayes, Richard Janda, and Janet Pierrehumbert have given me very helpful and stimulating feedback on the research reported in this thesis. I would also like to specifically thank Joan Bybee for getting me interested in many of the issues addressed in this thesis and for advice and support during my time in New Mexico.

I am very grateful to Luis Hernandez for programming assistance and for always being there to help resolve whatever misunderstandings arose between me and the computers, solve technical glitches, and make sure everything was set up exactly the way I needed it. Many thanks also go to all my colleagues at the Speech Research Lab and fellow linguists and psychologists for all the helpful discussions over the years and for maintaining such a friendly atmosphere to work in.

This work has been supported by the NIH Training Grant No. DC-00012 and Research Grant No. DC-00111 to the Speech Research Laboratory, Indiana University, Bloomington, without which none of this would have been possible.

# THE ARCHITECTURE OF GRAMMAR IN ARTIFICIAL GRAMMAR LEARNING

Vsevolod Kapatsinski

This thesis introduces an experimental paradigm designed to test whether human language learners acquire product-oriented generalizations (e.g., “plurals must end in -i”) and/or source-oriented generalizations (e.g., “add -i to the singular to form the plural”). The paradigm is applied to the morphophonological process of velar palatalization. Ecological validity of the paradigm is confirmed by comparison to corpus data from loanword adaptation in Russian. Characteristics of the training task are shown to influence the extent to which the grammar extracted by a learner is product-oriented or source-oriented. This finding suggests that the shape of the grammar is influenced not only by innate biases of the learner (Universal Grammar) but also characteristics of the learning situation.

Nonetheless, there are regularities that hold across training tasks and languages. First, learners extract both product-oriented and source-oriented generalizations. Thus, learners exposed to a lexicon of singular and plural forms learn at least 1) what typical plurals and singulars are like, 2) which segments of the singular form must be retained in the plural, and 3) which segments of the plural form must be retained in the singular. Second, learners appear to rely on schemas specifying which form classes and paradigmatic mappings are observed frequently (e.g., “plurals should end in -tʃi” or “a [k] in the singular corresponds to a [tʃi] in the plural”), rather than on constraints against underobserved form types (e.g., “plurals must not end in -ki”). Competing generalizations are weighted relative to each other stochastically. Thus, learners obey competing generalizations in proportion to how much statistical support each competitor

receives from the training data, rather than obeying the most strongly supported competitor 100% of the time. Learners do not to obey the Subset Principle, which predicts that the learners should induce the most specific generalizations consistent with the training data. The observed overgeneralization patterns are shown to be expected if we assume a Bayesian approach to speech perception and word recognition, in which the output of perception is not the identity of the most likely structure but rather a probability distribution over possible structures.

---

---

---

---

## TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	1
1.1 Biases and innateness in the acquisition of phonology	2
1.1.1 Formal and substantive biases	4
1.1.2 Formal biases	8
1.2 The need for grammar	13
1.3 Bayesian learning	16
1.4 Source-oriented vs. product-oriented generalizations	20
1.5 Judgments and production data as windows on the grammar	28
1.6 Outline of the thesis	31
 CHAPTER 2: THE LANGUAGES	 33
 CHAPTER 3: A SOURCE-ORIENTED PARADIGM	 41
3.1 The paradigm	42
3.1.1 Tasks	42
3.1.2 Stimuli	45
3.1.3 Procedures	49
3.1.4 Participants	50
3.2 Results I: Errors in perception and memory	50
3.3. Results II: Types of generalizations extracted from the input	53
3.3.1 Elicited production	54
3.3.1.1 Experiment I: Language I vs. Language II	54
3.3.1.2 Experiment II: Languages I-IV	57
3.3.1.3 An apparently product-oriented effect	64
3.3.1.4 A model of source-oriented generalization	70
3.3.2 Plural likelihood rating	77
3.4. Conclusion	86
Appendix 1: Instructions for the source-oriented paradigm	89
Appendix 2: List of stimuli	92
Appendix 3: Features used for training the model	95

CHAPTER 4: A PRODUCT-ORIENTED PARADIGM	96
4.1 The paradigm	97
4.1.1 Tasks	98
4.1.2 Stimuli	99
4.1.3 Procedures	102
4.1.4 Participants	102
4.2 Results I: Errors in recall	103
4.2.1 Overall accuracy of recall across languages	103
4.2.2 Memorizing singulars vs. plurals	105
4.3 Results II: Types of generalizations extracted from the input	111
4.3.1 Elicited production	111
4.3.1.1 Velar-final singulars	111
4.3.1.2 Alveolar-final singulars	116
4.3.1.3 Labial-final singulars	120
4.3.1.4 Alveopalatal-final singulars	123
4.3.1.5 The two suffixes	126
4.3.1.6 Discussion of elicited production	128
4.3.2 The relationship between production and recall	132
4.3.3 Plural likelihood rating	136
4.3.4 Combining results from perception and production	144
4.3.5 The relationship between recall and ratings	147
4.4 Conclusion	148
Appendix 1: Instructions	153
Appendix 2: Stimuli used for training and elicited production	154
CHAPTER 5: A NATURAL LANGUAGE: RUSSIAN	157
5.1 Velar palatalization in Russian	157
5.2 Methods	161
5.2.1 Data collection	161
5.2.2 Modeling	162



5.2.2.1 Training the model	163
5.2.2.2 Testing the model	166
5.3 Results	167
5.4 Explaining successes and failures of the model	172
5.5 The affix and the stem shape are chosen simultaneously	177
5.5.1 The issue of stages in grammatical processing	177
5.5.2 Russian velar palatalization: A single-stage process	180
5.6 Conclusion	182
Appendix 1: Feature sets	185
Appendix 2: Rules extracted by the model	189
 CHAPTER 6: GENERAL DISCUSSION AND CONCLUSION	 193
6.1 Formal and substantive biases in learning morphophonology	193
6.2 Negative product-oriented generalizations and Optimality Theory	197
6.3 Influences on word learning	203
6.4 The influence of the learning task on the architecture of grammar	204
6.5 The number of stages underlying grammatical processing	207
6.6 Individual differences in artificial grammar learning	212
6.7 Implications for language change	214
6.8 Summary	216
6.9 Conclusion	219
Appendix 1: Grammars and training sets for the Gradual Learning Algorithm	222
 References	 225

## CHAPTER 1

### INTRODUCTION

The present thesis develops experimental methods for testing the types of morphophonological generalizations humans extract from a lexicon of a language and use in perception and production, and investigates the extent to which the types of generalizations extracted from a lexicon depend on the way in which the learner is exposed to the lexicon, i.e., the nature of the learning task. In a broader perspective, this thesis addresses the questions of whether the production grammar is the perception grammar, i.e., whether grammatical generalizations receive the same weights relative to each other in perception and production, and whether the shape of the grammar acquired by a human learner depends on the way in which s/he experiences the linguistic data on whose basis the grammar is acquired. Previous work has focused on identifying the biases guiding acquisition of (morpho)phonological grammar. The present thesis extends this work by asking whether biases in favor of certain types of linguistic generalizations are due to the nature of the learners' exposure to language in the course of normal language acquisition (cf. Bybee 2008).

This introductory chapter presents the background for the presented experiments. Section 1.1 introduces the notion of biases constraining the learning phonological and morphophonological generalizations, and reviews the growing experimental literature on this topic, situating the present thesis in relation to the previous studies. Sections 1.2 and 1.3 lay out the theoretical assumptions of the present study. Section 1.2 discusses the evidence for storing generalizations in memory. Section 1.3 presents a Bayesian approach to learning, which clarifies the nature of learning biases and suggests that generalization

is an inevitable part of perception. Section 1.4 discusses theoretical proposals and prior evidence on whether morphophonological generalizations are by default source-oriented or product-oriented. Section 1.5 discusses the relationship between production and perceptual acceptability judgments. The final section outlines the rest of the thesis.

### **1.1. Biases and innateness in the acquisition of phonology**

In the last five years there has been an explosion of interest in experimental exploration of inductive biases, or constraints on learning, using a variety of artificial language learning paradigms. The growing interest is shown by the holding of a special session on ‘An Artificial Grammar paradigm for phonology’ at the 2007 Annual Meeting Linguistic Society of America, a well attended course at the 2008 LSA Summer Institute on ‘Analytic bias in phonology’, and the increase in the number of studies exploring the issue.

The general aim of this line of research has been to determine to what extent linguistic universals can (and should) be attributed to differences in learnability between attested and unattested or frequent and infrequent linguistic structures and to identify the inductive biases that constrain the learning process. Traditionally, generative grammar has assumed that learnability differences are the most likely (and in practice the only) explanation for linguistic universals (Archangeli and Pulleyblank 1994, Chomsky and Halle 1968, Prince and Smolensky 1993/2004, Wexler & Culicover 1980) and that a major source of inductive bias is Universal Grammar, a system of innate linguistic representations. A natural prediction of this hypothesis is that attested linguistic structures should be easier to learn than unattested ones (e.g., Finley 2008, Schane et al. 1974/1975,

Smith et al. 1993). Furthermore, structures that are common in unrelated languages may be assumed to be common because they are easier to learn than the marked structures (see Finley 2008, Goldrick and Larson in press, Hayes 1999, Wilson 2006 for alternative formulations).

This position has been challenged by, among others, Blevins (2004), Bybee (2001), Chang et al. (2001), Hale and Reiss (2000, 2008), Kavitskaya (2002), Mielke (2008), and Ohala (1981, 2005), who have argued that the true universals are universals of sound change and pointed out that a sequence of natural sound changes can result in an unnatural synchronic alternation (also see Newmeyer 2005:174-225 for a review of and a contribution to the same debate in syntax). One example of an unnatural alternation is found in Evenki where /g/, /s/, and /v/ nasalize after nasals, becoming /ŋ/, /n/, and /m/ respectively but other consonants (including /p/, /b/, /t/, /d/, /k/, and /x/) do not (Mielke 2008: 120-121). Not only is the class of changing consonants unnatural but the alternation itself involves relatively non-nasallike segments (like /s/) changing into nasals while more nasallike segments (like /d/) remain unchanged. Some unnatural generalizations have also been shown to be productive and thus to have some degree of psychological reality for the speakers of the language featuring the generalization. One such alternation is velar softening in English where /k/ changes into /s/ before certain Latinate suffixes, e.g., *electri*/k/-*electri*/s/ity, which is unnatural because ‘coarticulation and lenition would yield an aspirated palatal approximant rather than the alveolar fricative /s/’ (Pierrehumbert 2006:84; see also Guion 1998 for perceptual pressures pointing in the same direction) and has been shown to be productive by Pierrehumbert (2006).

Thus unnatural alternations are attested in languages of the world, can be learned by speakers of the languages and, furthermore, according to the only extant large-scale typological study, are quite common (Mielke 2008). While unnatural alternations arise through a sequence of natural sound changes, the naturalness/frequency of a sound change is not necessarily caused by differences in learnability between the original system and the resulting one. Thus, Ohala (1981), Blevins (2004), Hale and Reiss (2000), and Mielke (2008) argue that sound change results from variation in perception of inherently ambiguous information in the signal and can in principle proceed in any direction. Bybee (2001, Hooper 1976a), Browman and Goldstein (1992) and Mowrey and Pagliuca (1995) suggest that the great majority of sound changes are the result of articulatory reduction caused by repeated production of frequent words and sound patterns throughout life. Thus, a pattern's learnability may not go hand in hand with naturalness or cross-linguistic frequency of occurrence (Moreton 2008), which points to a need for experimental studies of learnability, which would complement and could be compared with typological and nonce probe studies of natural language (see Finley 2008, Moreton 2008 for examples of such comparisons yielding interesting theoretical results).

#### 1.1.1. Formal vs. substantive biases

A distinction is often drawn between formal and substantive constraints on learning. For instance, Wilson (2006: 974) suggests that 'the absolute limits on human phonologies' are caused by biases responsible for formal, rather than substantive universals. Formal constraints are 'limitations placed on acquisition by the structural properties of the cognitive system', including 'formal operations that define the space of

possible rules or constraints of the phonological grammar' (Goldrick and Larson, in press). For instance, one formal universal predicted by Optimality Theory (Prince and Smolensky 1993/2004) is that all languages can be described via a strict ranking of markedness and faithfulness constraints (and that language learners naturally come up with a strict ranking of such constraints when exposed to linguistic data, see Guest et al. 2000 for an experimental test of this hypothesis).

Substantive constraints are representations of knowledge about the relative phonetic naturalness of various linguistic patterns (whether acquired prior to the acquisition of the linguistic patterns themselves or genetically preprogrammed). Thus substantive constraints are restrictions on the content of phonological representations rather than their structure (Goldrick and Larson in press, Wilson 2006). For instance, a substantive constraint may be that [i] is assumed to be a more likely trigger of palatalization than [e] because [k] and [tʃ] are more confusable before [i] than before [e] (Wilson 2006). Many researchers have assumed the existence of substantive biases in learning, encoding them in Universal Grammar (in the form of the markedness constraints of Optimality Theory, Prince and Smolensky 1993), considering them as an evaluation metric for alternative grammatical analyses (Archangeli and Pulleyblank 1994, Chomsky and Halle 1968) and incorporating them into the learning algorithm (Finley 2008, Hayes 1999, Wilson 2006). Experimental studies of biases in grammar learning have focused on biases associated with phonetic naturalness and thus potentially responsible for substantive universals. Some examples of studies trying to determine whether phonetically natural rules (or constraints) are easier to learn than phonetically unnatural ones are Finley (2008), Gerken and Bollt (2008), Goldrick and Larson (in

press), Jusczyk et al. (2002, 2003), Koo and Cole (2006), Moreton (2008), Pater and Tessier (2003), Peperkamp et al. (2006), Peperkamp and Dupoux (2007), Pycha et al. (2003), Saffran and Thiessen (2003), Schane et al. (1974/1975), Seidl and Buckley (2005), Wilson (2003, 2006).

Evidence for substantive biases has been provided by Schane et al. (1974/1975), Peperkamp et al. (2006), Gerken and Boltt (2008), Finley (2008), and Goldrick and Larson (in press). Schane et al. (1974/1975) compared the learning of consonant deletion before consonant-initial words (a natural rule) to the learning of consonant deletion before vowel-initial words (an unnatural rule) and found that learning was more rapid when subjects were exposed to the natural rule, although both rules were eventually learned. All of the subsequent studies agree that unnatural rules are learnable, while the difference between natural and unnatural rules was replicated (for different rules) only by Finley (2008), Gerken and Boltt (2008), Peperkamp et al. (2006), and Wilson (2003)<sup>1</sup>. Wilson (2003) exposed adult native English speakers to a phonetically natural rule where the suffix underwent nasal assimilation when following a stem ending in a nasal (-la → -na / [+nasal]\_) or an unnatural rule in which -la became -na if the preceding consonant was velar. Subjects were tested using an old/new word recognition task. Subjects in the natural condition acquired a bias to respond to new words conforming to the rule as old whereas subjects in the unnatural condition were not influenced by whether a new word conformed to the rule they were presented with. Gerken and Boltt (2008) showed that 9-month-old infants were able to generalize a natural rule ‘stress heavy syllables’ to new syllable strings (assessed using the headturn preference procedure) after being exposed to

---

<sup>1</sup> While Pater and Tessier (2003) and Wilson (2006) also claim to demonstrate the result, the presence of confounds make interpretation impossible, as discussed later on.

only three different words exemplifying the rule, whereas they did not generalize an unnatural rule ‘stress syllables beginning with /t/’ with the same amount of exposure. Finley (2008) found that English speakers exposed to roundness harmony exemplified by mid vowels are able to learn the pattern whereas speakers exposed to roundness harmony exemplified by high vowels do not learn it, mirroring the typological observation that mid vowels are better triggers of roundness harmony (due to bearing weaker cues to roundness). However, no preference for high vowel harmony targets was found despite being predicted by the same logic (‘realize the cues that are in danger of being misperceived on the vowels that are best for realizing the cues’).

Wilson (2006) tested native English-speaking adult learners for knowledge that /i/ is a better trigger of velar palatalization than /e/. One group of subjects was exposed to a language in which /i/ triggered velar palatalization while /a/ did not and tested on /e/. The other group was shown that /e/ triggers velar palatalization while /a/ does not and tested on /i/. The subjects in the first group treated /e/ as intermediate between /i/ and /a/, while the subjects in the second group treated /i/ as a trigger of palatalization, like /e/. While Wilson interprets these results as showing a substantive bias, a simpler explanation is available: /e/ actually *is* between /i/ and /a/ in acoustic space, hence subjects do not know whether it should pattern like /i/ or /a/, whereas /i/ is even further from /a/ than /e/, so learners who know that /e/ triggers velar palatalization will infer that /i/ does as well.

Gerken and Boltt (2008) examine generalization of a natural pattern (‘stress heavy syllables’) and an unnatural pattern (‘stress syllables beginning with /t/’). Nine-month-olds acquire the natural pattern but not the unnatural one whereas 7.5-month-olds are able to learn both patterns, suggesting that at least some substantive biases that one might



have ascribed to Universal Grammar are learned. This is likely to be the case for the bias found by Pater and Tessier (2003) who observed that adult native English speakers were able to learn a rule that says that lax vowels must be followed by consonants more easily than they could learn a rule that says that front vowels must be followed by consonants. Since word-final lax vowels are actually banned in English, these results are not evidence of innate bias for natural rules.

The question of whether substantive biases are innate is also addressed by a series of studies that have focused on whether speakers have knowledge about differences in naturalness among structures that are unattested in the language(s) they speak (Albright in prep, Becker et al. 2007, Berent et al. 2007, Davidson 2006, Peperkamp 2007, Pertz and Bever 1975, Zhang and Lai 2006). While such knowledge has been found and is predictable based on the Sonority Sequencing Principle, it is not yet clear that it could not be acquired from the English input (see Albright in prep for a modeling attempt) or from experience with articulation and audition.

#### 1.1.2. Formal biases

The present dissertation is intended to fit into a related body of recent research that tries to determine the shape of the grammar preferentially induced from the data in an artificial grammar learning (AGL) paradigm and to draw inferences about biases responsible for the shapes of natural language grammars (i.e., formal universals, Moreton 2008, Newport and Aslin 2000). The shape of the grammar includes minimally the types of generalizations it contains, a set of functional modules into which the generalizations are divided, a mechanism to weight competing generalizations within a module relative

to each other, and a decision rule to decide between competing generalizations. For example, the generalizations can compete via strict ranking (Prince and Smolensky 1993/2004) or reliability-based weighting (Albright and Hayes 2003). The types of generalizations may be restricted by the types of unit categories they involve, e.g. ‘the grammar cannot make reference to speaker-specific characteristics’, or ‘no syllables’ (Chomsky and Halle 1968), or ‘surface representations only’ (Bybee 2001, Burzio 2002, Hooper 1976b), or ‘classical categories only’ (Chomsky and Halle 1968, Albright and Hayes 2003), and the relationships between representations that can be captured, e.g., ‘no paradigmatic relations in the grammar’ (Hale and Reiss 2008, Marantz and Halle, to appear), or ‘avoid non-local syntagmatic relations’ (Albright and Hayes 2003, Newport and Aslin 2004), or ‘learn categorical relationships only’ (Chomsky and Halle 1968), or ‘learn transitional probability relations’ (Aslin et al. 1998), or ‘learn frequencies of co-occurrence’ (Bybee 1985, 2001, Coleman and Pierrehumbert 1997). One caveat is that experimental evidence is unlikely to show that some grammar type is *impossible*. Rather, as Wilson (2006) and Finley (2008) point out, the working hypothesis is that some generalizations are easier to form than others, so subjects are expected to learn natural generalizations more easily than unnatural ones and, when the data are consistent with multiple generalizations, favor the natural one(s). In Bayesian terms, learners do not start out with a uniform prior (for Bayesian perspectives on possible biases in grammar learning, see Goldrick and Larson in press, Moreton 2008, and Perfors et al. 2006).

The issue of the inventory of grammatical units has been addressed in an AGL paradigm by Cristià and Seidl (2008), Finley (2008), Kapatsinski (in press), Newport and Aslin (2004), and Peperkamp and Dupoux (2007). Newport and Aslin (2004) report that

infants learn dependencies between non-adjacent consonants or vowels but not between non-adjacent syllables, arguing that dependencies between segments are easier to learn.<sup>2</sup> Kapatsinski (in press) shows that rime-affix associations are easier to learn than body-affix associations, arguing for the rime as a grammatical unit. Cristià and Seidl (2008) show that English-learning infants learn that nasals and stops pattern alike and generalize the knowledge to new segments more easily than they learn that nasals and fricatives pattern alike. They suggest that infants classify segments into featural categories, including [-continuant], which subsumes nasals and stops.<sup>3</sup> More evidence for features is provided by Finley (2008) who shows that (in some circumstances) adults generalize vowel harmony to new segments. While the generalization could also proceed by analogy to segments on which the subjects are trained (e.g., Mielke 2008), Finley argues against this possibility using lack of statistically significant differences in accuracy with new and old segments (but see Peperkamp and Dupoux, 2007, for contrasting findings).

Regarding category structure, Saffran and Thiessen (2003) observe that infants find it easier to confine voiced stops or voiceless stops to a single position within the syllable than to confine an arbitrary class like {/p/, /d/, /k/}. This ‘preference for simple rules’ is also shown by Pycha et al. (2003) and Peperkamp et al. (2006) for other rules. It seems likely that this is a domain-general bias for simple, possibly linearly separable category structures (e.g., Shepard et al. 1961), rather than a learning bias specific to grammar and is predicted by most theories of categorization (see Nosofsky et al. 1994 for review), although McKinley and Nosofsky (1995) and Jäkel et al. (2008) point out that

---

<sup>2</sup> Although the results are confounded by the fact that non-adjacent syllables in the experiment are separated by two segments while non-adjacent segments are separated by only one.

<sup>3</sup> It is doubtful that this classification is innate since nasals pattern phonologically with fricatives rather than stops in a great majority of languages (Mielke 2008:65-67).

exemplar models can capture arbitrarily complex category boundaries that humans do not. Another domain-general bias proposed by multiple researchers on the basis of both logical and empirical arguments is the Subset Principle, the tendency to come up with the most specific possible generalization that fits the observed data rather than the most general one (Albright and Hayes 2003, Berwick 1986, Dąbrowska 2008, Dell 1979, Hale and Reiss 2003, 2008:27-57, Langacker 1987, Mitchell 1997:26-28).

The modularity question has been addressed by Finley (2008), Moreton (2008), Onishi et al. (2002), and Warker et al. (2008). Moreton (2008) finds that dependencies between [voice] features of onsets of adjacent syllables and dependencies between height features of vowels in adjacent syllables are easier to learn than dependencies between vowel height and consonant voicing. Since the harder-to-learn voice-height relationship is more phonetically natural and cross-linguistically common (e.g., Canadian Raising) than voice-voice relationship, Moreton argues that certain dimensions of speech sounds are predisposed to interact in learning for non-perceptual reasons (which he assumes to be belonging to the same module within the grammar). Similarly, Finley (2008) argues that height and laxness are predisposed to interact while height and backness are not. Onishi et al. (2002) argue that it is not possible to learn dependencies between speaker voice and the assignment of consonants to syllable positions, while it is easy to learn that the assignment of consonants to syllable positions depends on the intervening vowel.<sup>4</sup>

---

<sup>4</sup> The generality of Onishi et al.'s (2002) argument is questionable because the dependent measure in Onishi et al.'s experiment was the speed of repeating 'legal' vs. 'illegal' syllables after the speaker. The voice of the original speaker is obviously not preserved in the repeated utterance, so it is not clear that it should be expected to interact with phonotactics in the repeater's production. Anecdotally, it appears that one could learn phonotactic constraints that are specific to particular speakers in perception, e.g., if one knows an English-speaking child who restricts velars to the syllable-initial position, one can learn to compensate for this error but would probably not generalize the compensation to other speakers. Experimentally, Eisner & McQueen (2005) find that adjustments to phonemic category boundaries due to exposure to a systematically mispronouncing speaker do not generalize to other speakers.

Warker et al. (2008) observe robust learning of dependencies between whether a certain consonant is restricted to the onset position or the coda position on the identity of some other segment in the same word but no ability to restrict a consonant to the onset or coda only at certain speech rates. Thus Onishi et al. (2002) and Warker et al. (2008) argue for a phonological module that does not contain information on speech rate and speaker identity, making dependencies between indexical and phonological information harder to learn than dependencies within a module.<sup>5</sup> None of the studies address the question of whether the boundaries between modules are innate or acquired on the basis of prior experience. For instance, a listener who acquires language in a bilingual setting where different speakers are likely to have different phonotactic constraints (or is exposed to child speech) may find it easier to learn novel dependencies between speaker voice and phonotactics.

Finally, the issue of the relationship between competing generalizations is addressed by Finley (2008), Guest et al. (2000), and Hudson Kam and Newport (2005). Hudson Kam and Newport (2005) show that, in a particular artificial grammar learning paradigm, children induce a grammar in which the most reliable generalization always wins while adults exposed to the same data engage in probability matching. Guest et al. (2000) have tested the assumption that if learners find out that constraint A outranks constraint B and constraint B outranks constraint C they would infer that constraint A outranks constraint C, which is inherent to the architecture of grammar in Optimality Theory. Finley (2008) shows that certain versions of Optimality Theory predict the existence of harmony systems in which vowels that are in the minority within a word take

---

<sup>5</sup>It is not clear whether module-merging dependencies are predicted to be impossible or merely more difficult than connections within a module.

on the characteristics of vowels that are in the majority and demonstrates that learners tend not to make the predicted generalizations.

## **1.2. The need for grammar**

One very general formal bias that is assumed by all linguistically-oriented work on artificial grammar learning is that the learner generalizes over the stimuli presented to him/her during training rather than simply storing all of the presented stimuli in memory and generalizing to new stimuli based on their similarity to the familiar stimuli, as a pure exemplar-based model would predict (e.g., Goldinger 1998, Hintzman 1986, Nosofsky 1988).

The relative importance of rule-like generalization and retention of the details of presented stimuli is a long-standing debate in learning and categorization (e.g., Albright & Hayes 2003, Brooks 1978, Brooks & Vokey 1991, Daelemans & van den Bosch 2005, Denton et al. 2008, Eisner & McQueen 2005, Erickson & Kruschke 1998, 2002, Goldinger 1998, Hintzman 1986, Marcus et al. 1999, Pinker 1999, Pothos 2005, 2007, Reber 1967, Skousen 1989, 2002). The notion of a rule varies somewhat across domains (see the responses to Pothos, 2005). Pothos (2007:228) defines a rule as ‘a mental operation that allows characterization of a stimulus by examining only a part of it’. Similarly, Marcus et al. (1999) and Pinker (1999) define a rule as an “operation over variables”. According to this definition, one can say that a rule-based account is one in which certain features of the stimulus receive a weight of zero for the purposes of categorization. This is quite different from the notion of a rule in linguistics. In linguistics, a rule is a mental operation that transforms one class of representations into

another class of representations where both classes are classical categories, i.e., categories defined using necessary and sufficient conditions (e.g., Hale and Reiss 2008:195). Pothos's (2007) definition is broader in that it appears to treat analogy to partially specified examples as 'rule-based', although in practice all analogical models are partially underspecified, since it is impractical to consider all properties of a stimulus as having equal potential relevance (Daelemans & van den Bosch 2005). In the present thesis, I will adopt the linguistic definition of the term "rule" and will use the term "generalization" for the less specific notion. I will call the set of generalizations underlying a subject's behavior in generalizing to new stimuli his/her grammar.<sup>6</sup>

There is now an extensive body of data supporting both storage of fine episodic details about the presented stimuli and the storage of generalizations over stimuli. For instance, Palmeri et al. (1993) has shown that repetition priming for a spoken word is enhanced if the word is repeated by the same speaker, regardless of the number of voices presented in the experiment, suggesting automatic storage of voice information. Eisner & McQueen (2005) found that the adjustment of phonemic category boundaries in perceptual learning is speaker-specific. Nonetheless, repetition priming is not reduced when the prime and the target differ acoustically when the prime contains one allophone of a certain phoneme while the target contains another (Darcy et al. 2008, McLennan et al. 2003, Sumner & Samuel 2005). Learners can generalize an artificial grammar across lexicons, where lexicons can even come from separate modalities (Altmann et al. 1995). One can observe long-term priming between sentences that have no lexical overlap as

---

<sup>6</sup> Pothos (2007) opposes rule-based and similarity-based accounts of AGL to the idea that knowledge of an artificial grammar is knowledge of associations. It is unclear why association-based accounts should be opposed to rule-based and similarity-based accounts, since associations can in principle involve both underspecified and fully specified representations, individual features and feature bundles, as well as classical categories, prototypes or even exemplar clouds (Kruschke 1992).

long as they share syntactic structure (Bock 1986). Finally, learners exposed to a rule-plus-exception category structure do not analogize based on the exception, unlike existing exemplar models (Albright & Hayes 2003, Denton et al. 2008, Erickson & Kruschke 1998, 2002). Thus there is a developing consensus that both exemplars and rule-like generalizations are necessary (Denton et al. 2008, Erickson & Kruschke 1998, Nosofsky & Bergert 2007).

If there were no generalizations, as in a pure exemplar-based model, the questions asked by the present thesis would be meaningless; thus it is important to explore whether learners form generalizations over stimuli they are trained on, i.e., whether learners overgeneralize relative to what an exemplar model would predict. When a learner is exposed to a set of stimuli from which s/he is (later) asked to generalize, s/he may memorize the set of stimuli or s/he may fail to do so. If the set of stimuli is memorized, subsequent generalization can be accomplished by comparing the stimuli to which s/he is asked to generalize to the stimuli with whose properties the learner has been familiarized during training. If this is the mechanism behind a subject's generalization behavior, accuracy should be higher with stimuli that have been presented during training than with novel stimuli (unless the novel stimuli are located closer to the average location of familiar stimuli in similarity space than familiar stimuli), and accuracy should be higher with novel stimuli the more similar they are to familiar stimuli belonging to the same category (Nosofsky 1988).



### 1.3. Bayesian learning

Where analogical approaches predict that generalization beyond the training set occurs only when the learner is tested with stimuli that do not belong to the training state, an alternative is presented by the Bayesian approach to perception and learning (Clayards et al. 2008, Goldrick and Larson in press, Kruschke 2008, Levy 2009, Moreton 2008, Norris & McQueen 2008, Perfors et al. 2006). A Bayesian approach to learning (and inference more generally) is explicitly concerned with taking into account the types of biases (or prior probability distributions) that learners bring to the task. It models the bias brought to the task by a learner as a probability distribution over possible hypotheses, which is a natural formal framework for the types of soft constraints on acquisition found in investigations of analytic bias.

The second major feature of a Bayesian approach to learning and inference is that the output of Bayesian inference is a probability distribution, called the posterior distribution, rather than the single most probable point. The posterior distribution is a distribution of posterior probabilities. The posterior probability of a hypothesis is defined as the probability of the hypothesis given the data and is proportional to the product of prior probability of the hypothesis and the probability of the data given the hypothesis.<sup>7</sup> Calculation of the entire probability distribution is necessary for updating the bias for future use in a mathematically sound way (Kruschke 2008, Levy 2009). When the output of inference is the posterior distribution, the distribution can be taken to be the new prior (or bias) on possible hypotheses. When the output is a single most probable hypothesis,

---

<sup>7</sup>  $p(H|D) = \frac{p(H,D)}{p(D)} = \frac{p(H,D)p(H)}{p(D)p(H)} = \frac{p(D|H)p(H)}{p(D)}$  where H stands for 'hypothesis', and D stands for 'data'.

all other hypotheses are implicitly considered equally improbable given the data, which is unrealistic.

In the present study, the hypotheses we are dealing with correspond to the identities of possible stimuli that could be presented. Traditionally, the output of human perception is taken to be a single hypothesis about the identity of the stimulus, thus the only information provided by perception is the identity of the most probable stimulus given the evidence. For instance, Clayards et al. (2008: 804), in a paper arguing for an otherwise Bayesian approach to speech perception, write “the goal of speech perception can be characterized as finding the most likely intended message”. Under a purely Bayesian approach, the output of perception is a probability distribution over possible stimuli. Thus, despite reporting having perceived the single most probable stimulus, the perceiver assigns other similar stimuli a probability, which indicates the degree of belief in the hypothesis that the stimulus has been presented (which is based on the probability of the data given the hypothesis multiplied by the prior probability of the hypothesis). For instance, a subject presented with /bupi/ may report hearing /bupi/ but also (subconsciously) consider it possible but less likely that /buki/ has just been presented. Note that if the learner intends to maximize the probability of being correct, s/he should always report hearing the stimulus s/he considers to be the most probable one (Norris & McQueen 2008) but should update the probability of each possible hypothesis in proportion to how likely s/he believes it to be given the sensory data (Kruschke 2008, Levy 2009).

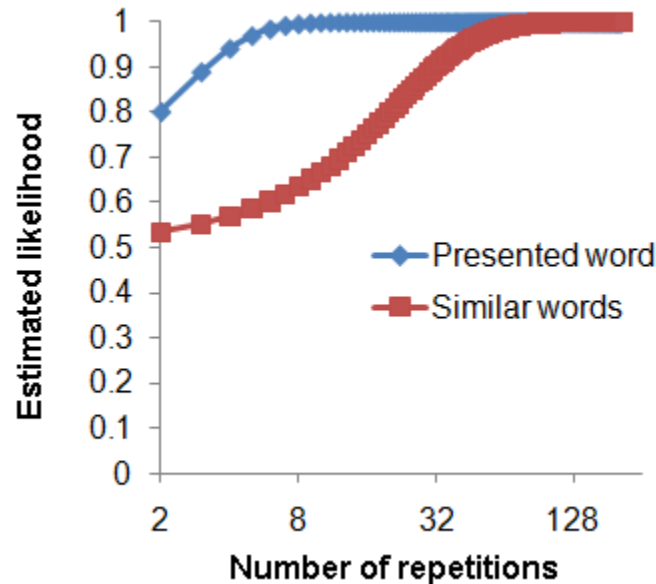
Under the traditional approach, the only stimulus whose probability of future occurrence is incremented as a result of perception is the most probable stimulus (here

/bupi/). Under the Bayesian approach, the probability of each possible stimulus is incremented in proportion to how strongly the perceiver believes it to have been presented (here, the probability of /bupi/ would increase more than the probability of /buki/). If two stimuli are presented together, then their probability of co-occurrence (denoted by connection strength in network models) is increased. On the Bayesian account, the probabilities of co-occurrence of the possible stimuli are increased as well (thus the learner hearing /bupi/ would increase the co-occurrence probability between /p/ and /i/ and, to a lesser extent, the co-occurrence probability between /k/ and /i/). We shall see that this feature of Bayesian learning is extremely helpful to account for the present data.

Some evidence for the incrementing of the estimated probability of occurrence for stimuli that are similar to the presented stimulus (although not necessarily in accordance with Bayes' rule) is provided by the verbal transformation effect (Bashford et al. 2006, Warren 1961, 1996). The verbal transformation effect occurs when a spoken word is presented repeatedly to the subject a large number of times. As repetition continues, other related words begin to be heard in place of the presented word with increasing frequency. Warren (1996) hypothesizes a 'summation effect' that increases the estimated resting activation levels (estimated prior likelihoods of occurrence, which can be converted into prior probabilities by dividing each prior likelihood by the sum of prior likelihoods) of words that are structurally similar to the presented word. Assuming that the increase in resting activation level is sigmoid (a standard assumption based on the common observation of sigmoid learning curves, see Norris and McQueen 2008 for a Bayesian

justification), the words that are similar to the presented word have an opportunity to catch up as the word is being repeated, as shown in Figure 1.1.

Figure 1.1. Estimated prior likelihood of occurrence of a repeated word and a word that is similar to it as a function of the number of repetitions of the repeated word. The probabilities are modeled with a sigmoid function where the probability of the repeated word is equal to  $1/(1+2^{-(N)})$  while the probability of the similar word is  $1/(1+2^{-(N/10)})$  where  $N$  is the number of repetitions.



If the prior likelihoods of words that are similar to the presented word are incremented when the word is presented, a few general theoretical consequences follow. First, generalization of associations of a stimulus to other similar stimuli can now be seen as an inevitable feature of perception and should occur regardless of whether the learner is ever tested on the similar stimuli.<sup>8</sup> Second, exemplars of training stimuli do not have to

<sup>8</sup> An interesting consequence of this prediction is that similar stimuli are likely to be coactivated, leading to the strengthening of connections (in a network model) or, more generally, an increase in the probabilities of

be stored for generalization to be sensitive to similarity to the training stimuli: as long as the testing stimuli are briefly considered during the perception of the training stimuli as candidate percepts, they could acquire the associations of the training stimuli.

#### **1.4. Source-oriented vs. product-oriented generalizations**

Generative rules express source-oriented generalizations. That is, they act on a specific input to change it in well-defined ways into an output of a certain form. Many, if not all, schemas are product-oriented rather than source-oriented. A product-oriented schema generalizes over forms of a specific category, but does not specify how to derive that category from some other.” (Bybee 2001: 128).

One of the major developments in linguistic theory in the past twenty years has been the shift from source-oriented rule-based approaches to constraint-based approaches that incorporate product-oriented generalizations and restrict source-oriented generalizations to relations of identity, called paradigm uniformity constraints (see Benua 1997, Burzio 2002, Kenstowicz 1996, McCarthy 2005 for various formalizations). Thus rule-based phonology has been replaced as the dominant paradigm by Optimality Theory (Prince and Smolensky 1993/2004, McCarthy 2002, Roca 1997), and constraint-based approaches (LFG: Bresnan 2001; HPSG: Pollard and Sag 1994, Sag and Wasow 1999) have come to dominate computational syntax. In phonology, the major motivation for the

---

co-occurrence between them (which provides a theoretical justification for this proposal in Kapatsinski, 2007b).

switch to a constraint-based approach was provided by rule conspiracies (Kisseberth 1970, Prince and Smolensky 1993/2004:1, McCarthy 2002: 53-55, 63, 95-101), in which a number of different input classes are mapped onto the same output, which is somehow less marked than the inputs. In a purely rule-based approach, the fact that the language ‘likes’ a particular type of an output would be unexplained. A constraint saying that some outputs are preferred over others seemed to be required. Since having both rules and constraints in the grammar is unparsimonious, phonologists have opted for a purely constraint-based framework, causing a shift to Optimality Theory (McCarthy 2002: 53-55, Hale and Reiss 2008:216-220), although see Blevins (1997), Paradis (1989), and Roca (1997) for arguments in favor of hybrid rules-plus-constraints approaches. However, arguments against constraint-based approaches continue to be raised (e.g., Bromberger and Halle 1997, Hale and Reiss 2008, Roca 1997), based on alternative views of parsimony as well as the inability of Optimality Theory to account for opaque phonological patterns.

Bybee (2001:126-129) reviews a number of experimental studies purporting to show product-oriented generalization. In all extant cases (Bybee and Slobin 1982, Bybee and Moder 1983, Köpcke 1988, Lobben 1991, Wang and Derwing 1994, Albright and Hayes 2003), the argument is that instead of respecting the input-output mappings present in the lexicon, subjects ‘overuse’ common output patterns deriving them in ways not attested in the lexicon. Unfortunately, the overuse can also be explained by experiment-internal response priming. This hypothesis is supported by the response sequence data in Lobben’s (1991) study on Hausa plurals, which Lobben himself notes:

“the plurals [that don’t obey the rules but all end in ooCii] are appearing concentrated and subsequently... and... this is a typical characteristic of all other plural patterns’ (Lobben 1991:173),

‘[In this example] the second syllable of the singular is left out in the plural form, which never happens with real nouns... The surrounding... plurals, two preceding and seven following... are trisyllabic [in accordance with source-oriented rules]. This... provides an explanation as to why the plural [in this example], which, if produced according to the rule... , would have four syllables, is made to have three syllables in a very unorthodox way’ (Lobben 1991:182)

‘The eight first plurals [the subject produced] are –aayee plurals, allowing us to interpret freely the two irregular [i.e., non-rule-obeying] plurals in this row [of responses]’ (Lobben 1991:202),

‘Because the subjects stick consistently to one one, two, or, more rarely, three or four plural types, the plurals of which often occur in clusters, they can easily be interpreted as being instantiations of the same product-oriented schema’ (Lobben 1991:208).

While Lobben (1991) and Bybee (2001:126-9) interpret the data as supporting the theory that speakers use generalizations they made about Hausa plurals, rather than

singular-plural mappings, the data are also consistent with response priming. While a speaker might reuse a plural pattern after using it nine times in a row with words s/he does not know, this does not mean that such patterns play a role outside of the experimental situation. Other studies purporting to show product-oriented generalization do not report response sequence data but also have the characteristic that a small number of patterns is reused in an experiment with a large number of similar trials (e.g., Wang and Derwing 1994 report three major past tense vowels being reused).

Stronger evidence for product-oriented generalizations is provided by cases of echolalia, in which a morpheme is not attached to a form if the form sounds like it already has the morpheme (Menn and McWhinney 1981, Stemberger 1981, Bybee 2001:128). For instance, *It was thundering and lightning*, not *It was thundering and \*lightninging*. Here speakers of English appear to be using the generalization that the progressive should end in -ing, not that one should add -ing to form the progressive. The stability of the no-change class of English verbs and its apparent resistance to overgeneralization is another possible example of this phenomenon (Menn and McWhinney 1981, Stemberger 1984, Bybee 2001:128). However, phonological factors and checking of the output after the application of the -ing-adding rule (Pinker 1999:61-62) are possible alternative explanations. Furthermore, if it is accepted that these examples involve product-oriented generalizations, it is still not clear whether product-oriented generalizations can trigger changes to the input, as predicted by Optimality Theory (Prince and Smolensky 1993/2004) and Usage-based Phonology (Bybee 2001) or if their influence is restricted to blocking changes that would lead to unacceptable outputs.



Finally, some evidence for product-oriented generalization is provided by artificial grammar learning experiments conducted by Braine (1987), Braine et al. (1990), Brooks et al. (1993), Frigo & McDonald (1998), Gerken et al. (2005), Williams (2003), and Weinert (2009) who find that learners in either explicit or implicit learning paradigms find it difficult to learn and generalize paradigmatic relationships between affixes in an artificial language (e.g., ‘some words take –is, –a, and –al while others take –et, –uk, and –of’ in Weinert 2009) unless there is an accompanying phonological cue that allows learners to distinguish the words taking one class of affixes from words that take the other class of affixes. Thus, in Weinert (2009), learners presented with a novel word bearing –is are at chance at judging the grammaticality of a form of the same word bearing –a (which should be grammatical) or –uk (which should be ungrammatical). The task becomes much easier when the novel form ending in –a or –uk contains a cue to its class, e.g., a round vowel, which is associated with the suffix class {–et;–uk;–of}. The presence of the cue allows the learner to rely on product-oriented generalizations like ‘back vowels are followed by –uk’ rather than relying solely on source-oriented generalizations like  $a \rightarrow uk$ . However, Williams (2003) shows that arbitrary word classes defined solely by paradigmatic mappings between suffixes, implying source-oriented generalizations, can be learned (albeit only in an explicit learning paradigm and with a very small lexicon), suggesting that product-oriented support may be useful but not necessary. Furthermore, Frigo and McDonald (1998) found that learners exposed to a language in which there were phonological cues to gender membership were able to generate correctly suffixed forms given a paradigmatically related form even for novel words that did not bear stem-internal phonological cues to gender membership. This

finding suggests that the learners were able to acquire source-oriented mappings between affixes, although the learning of the mappings was certainly facilitated when the paradigmatically related affixes were syntagmatically associated with co-occurring parts of the stem.

The necessity of supplementing product-oriented generalizations with paradigm uniformity constraints is suggested by the existence of restrictions on the class of inputs that are productively mapped onto a certain class of outputs. For instance, if singulars ending in /k/ correspond to plurals ending in /tʃi/ but singulars ending in /t/ correspond to plurals ending in /ti/, there is no possible relative weighting of ‘plurals must end in /ti/’ and ‘plurals must end in /tʃi/’ that will produce the appropriate singular-plural mappings. Rather, one must somehow indicate that a singular-final velar is less likely to be preserved in the plural than a singular-final alveolar.

Pierrehumbert (2006) presents a particularly interesting case of a productive restriction on the class of inputs mapped onto a certain class of outputs. She shows that when a native English speaker is presented with a novel Latinate adjective ending in /k/ and produces a noun ending in –ity from it, as in ‘interponic’ → ‘interponicity’, the adjective-final singular is changed into an /s/ when followed by –ity. Pierrehumbert argues that English speakers must be using a source-oriented generalization like  $k \rightarrow s\_ity$  and not a product-oriented one like ‘Latinate nouns should end in /siti/’ or ‘Latinate nouns should not end in /kiti/’ for two reasons. First, only adjectives ending in /k/ are mapped onto nouns ending in /siti/ and this generalization cannot be handled using the feature-specific formulations of paradigm uniformity used in Optimality Theory, like

‘keep [velar] features present in the adjective in the noun’, because the set of features that changes when a /k/ is mapped onto /s/ is a superset of features that distinguish /t/ from /s/, yet /t/ is not mapped onto /s/.<sup>9</sup> This argument does not provide decisive evidence against the product-oriented account because the shortcoming is remedied by allowing segment-specific constraints like ‘a /t/ present in the adjective is retained in the noun’.

Pierrehumbert’s second argument presents a much stronger case against the hypothesis that the generalizations the subjects are using are product-oriented.

Pierrehumbert shows that /s/ is not the consonant that most commonly precedes –ity in English. Rather, /l/ precedes –ity much more commonly than /s/ does. Therefore, a learner generalizing over nouns would be expected to believe that –ity should be preceded by /l/ much more often than by /s/. Nonetheless, speakers in Pierrehumbert’s experiment never changed /k/ into /l/ when attaching –ity. Generalization over adjective-noun pairs, on the other hand, would yield the observed pattern of /k/ being mapped onto /s/ and not /l/ because adjectives ending in /k/ never correspond to nouns ending in /liti/ but often correspond to nouns ending in /siti/.

Thus, Pierrehumbert’s data provide evidence for rejecting the strong hypothesis that all generalizations extracted by speakers of natural languages are product-oriented. However, the factors leading learners to prefer source-oriented generalizations over product-oriented ones or vice versa are uncertain. One possibility is that humans have an innate bias in favor of product-oriented (Becker & Fanleib 2009, Braine 1987, Frigo & McDonald 1998) or source-oriented (Chomsky & Halle 1968) generalizations. Another possibility is that the learner favors the most reliable generalizations, whether they are

---

<sup>9</sup> Actually, Pierrehumbert (2006) never presented the subjects with adjectives ending in /t/ so at this point the statement that /t/ does not change into /s/ before –ity is a hypothesis rather than a finding.

source-oriented or product-oriented (Hayes & Wilson 2008). Alternatively, the learner may attempt to restrict the number of generalizations that must be remembered for accurate performance (e.g., Goodman et al. in press). Finally, the type of generalization may depend on the learning situation and the possibilities it affords for noticing paradigmatically related word pairs.

Artificial grammar learning is perfect for testing the distinction between source-oriented and product-oriented generalizations and possible influences on the type of generalizations extracted by the learner. In the present dissertation, I will provide two training paradigms, one explicitly designed to favor source-oriented generalizations and another designed to favor product-oriented ones to test whether the two types of generalizations emerge automatically even when the deck is stacked against them.

The distinction between rules and constraints or source-oriented and product-oriented generalizations, which is prominent in the linguistic debates on the nature of grammar) is absent from the literature on artificial grammar learning (see Pothos 2007 for a review). Unlike natural language grammars, the finite-state artificial grammars typically used in artificial grammar learning experiments do not have different form classes that correspond to different meaning classes and therefore do not have well-defined paradigmatic mappings between form classes.

Furthermore, unlike the learner of a natural language, who needs to create novel sentences, supply plurals to nouns whose plural he's never encountered, and incorporate loanwords into his/her language, the learner of a typical 'artificial grammar' has no need to generate a new form based on his/her knowledge of related forms. Therefore, no input-output mappings are needed; all that is needed is a mapping between inputs and category

labels, e.g., ‘violation’ or ‘no violation’, which according to Hale and Reiss (2008:195) is the definition of a constraint.

The irrelevance of linguistic rules to typical artificial grammars raises an interesting issue, which I shall attempt to address in this dissertation: to what extent does the architecture of the grammar induced from the data depend on the learning task? While the architecture of grammar has been traditionally assumed to be innate, it is quite possible that it is rather a consequence of the way language is typically acquired and of independently motivated characteristics of human languages. This conjecture is supported by results from the categorization literature, where subjects have been shown to generalize in either a rule-like or an exemplar-based manner depending on instructions (Nosofsky et al. 1989).<sup>10</sup> Thus one aim of the present dissertation is to determine if it is possible to manipulate the architecture of the acquired grammar by manipulating the characteristics of the learning task presented to learners in an AGL experiment.

### **1.5. Judgments and production data as windows on the grammar**

It is not clear whether the language user uses the same set of generalizations in perception as in production and, even if this is the case, whether the competing generalizations have the same relative weights in both. Albright and Hayes (2003) suggest that the same generalizations, weighted in the same way, are active in perception and production, but argue that the grammatical generalizations are used only in expanding the lexicon. Stored words, at least ones that feature grammatical irregularities,

---

<sup>10</sup> See also Erickson and Kruschke (1998), who conjecture that the types of generalizations acquired in language acquisition might be different from those acquired in an experimental categorization task because of the high dimensionality of linguistic representations compared to the simple geometric patterns in a typical categorization experiment, which may fundamentally change the nature of the learning task.

are retrieved from the lexicon as wholes. Albright and Hayes (2003) show, following many previous studies (e.g., Berko 1958), that English past tense patterns are extended to novel words in a probabilistic manner. Yet, they argue, an adult native English speaker, almost always produces a particular past tense form for a given verb even if the verb is similar to many other verbs that form the past tense in a different way. Therefore, the weighted generalizations responsible for probabilistic extension of past tense patterns to novel verbs are overridden by lexical information in the case of verbs whose past tense is known to the speaker.

In their study on the English past tense, Albright and Hayes (2003) found that the same set of generalizations and the same set of weights on competing generalizations could account for both elicited production data and naturalness rating data, suggesting that both tasks were performed using the same grammar. A peculiarity of the experimental design in Albright and Hayes (2003) is that alternative past tense forms for a given present tense form were presented in sequence so that the listener would rate them relative to each other. On the other hand, in natural conversation, a listener is unlikely to hear a range of unfamiliar past tense forms for a particular known present tense form. In order to understand the past tense form, the listener may need to determine the present tense form it came from. To determine the likelihood that a past tense form came from the present tense form, s/he may evaluate the past tense form s/he is presented with relative to other past tense forms that could be produced from the same present tense form. This evaluation is necessary if the listener undertakes analysis by synthesis.<sup>11</sup>

However, if the listener does not perform analysis by synthesis, s/he need not consider

---

<sup>11</sup> Which could be expressed as a Bayesian computation of the probability of the singular given the plural based on the prior probability of the singular and the probability of the plural given the singular (cf. Norris & McQueen 2008).

alternative past tense forms that have not been presented. Rather, the listener might evaluate the goodness of a past-present mapping without taking into account other past-present mappings involving the same present tense form, e.g., by implicitly or explicitly testing how readily past tense forms with the same past tense pattern come to mind (a form of the availability heuristic, discovered by Tversky and Kahneman, 1974).

Assuming that past tense forms that are similar to the probe are more likely to come to mind than dissimilar past tense forms, the availability heuristic would also account for the similarity effects in acceptability rating found by Albright and Hayes (2003).

Some evidence for this possibility is presented in Kapatsinski (2007) where Russian speakers were asked to rate novel Russian noun-verb mappings, in which the verb was formed from the noun by attaching one of the Russian stem-extending suffixes (-i or -a). Unlike in Albright and Hayes (2003), alternative plurals for a given singular were not presented adjacently to one another. Kapatsinski (2007) found that a verb bearing a particular stem extension was rated as highly likely to be produced from the presented singular iff it was similar to existing verbs bearing that stem extension. Importantly, novel verbs that were similar to many other verbs bearing a different extension were not rated as less acceptable than novel verbs that were not similar to existing verbs.

Zuraw (2000) follows Bybee (1985, Hooper 1976a) in arguing that high-frequency words are stored in the lexicon and retrieved for production as wholes, thus agreeing with Albright and Hayes (2003) that the grammar is used to expand the lexicon. However, she argues that the generalizations used to perform perceptual tasks may be different from the generalizations used to expand the lexicon by producing novel words.

Zuraw (2000) finds that speakers of Tagalog fail to apply a morphophonemic process (prefix-final nasal substitution) in elicited production, while the same subjects find forms that have undergone the process to sound more natural than forms that have not undergone the process, even though the latter are the forms they are likely to produce. Here it seems that the target input-output mapping, which may be underlain by source-oriented generalizations, has not been learned well enough to be applied in production but that subjects have formed a product-oriented generalization (Bybee and Slobin 1982) stating that prefix-final segments should be nasal. Product-oriented generalizations about the shapes of Russian verbs are also suggested by the data in Kapatsinski (2007). Thus, a possible difference between production and judgment of goodness of a perceived morphophonological mapping, is that judgments are more likely to be influenced by product-oriented generalizations compared to elicited production and may help detect acquisition of product-oriented generalizations that are not evident in production data. Another, not necessarily incompatible, hypothesis is that a speaker must be more confident in a generalization's accuracy to use it in production than to let it influence his/her grammaticality judgments in perception (cf. Kempen & Harbusch's 2005 'production threshold').

## **1.6. Outline of the thesis**

The aim of the present thesis is to test whether characteristics of the training task influence the types of generalizations extracted by the learner. The next chapter introduces the artificial languages that will be used to address these questions throughout the thesis. Chapter 3 introduces the paradigm that is expected to favor source-oriented



generalizations and reports experimental results suggesting that, despite the source-oriented nature of the paradigm, learners extract product-oriented as well as source-oriented generalizations, with source-oriented generalizations dominating production but not perception / judgment. Chapter 4 introduces the product-oriented paradigm, which incorporates many features of natural language learning that the source-oriented paradigm lacks, and reports experimental results showing product-oriented generalization in both perception and production. Chapter 5 examines the processes of affixation and velar palatalization in a natural language, Russian, demonstrating that the account of morphophonology developed in the preceding chapters also works for natural languages and is able to explain an otherwise puzzling phenomenon of language change: the loss of productivity by exceptionless morphophonological generalizations. Chapter 6 concludes the thesis.

## CHAPTER 2

### THE LANGUAGES

In this chapter, I introduce the artificial language types that were presented to subjects in the experiments reported in Chapters 3-4 and describe how the comparison between these languages can shed light on the types of generalizations that are acquired by a learner who is exposed to a lexicon in a particular learning situation.

As discussed in Chapter 1, one major component of the architecture of grammar consists of the types of generalizations that form the grammar. These generalizations can describe typical linguistic structures, militate against atypical or unattested structures and define probable, possible, and impossible mappings between structures. If a generalization is made over paradigmatically related forms like singulars and plurals, and thus defines a mapping between two forms that share the same base, the generalization is said to be ‘source-oriented’ (Bybee & Slobin, 1982; Bybee & Moder, 1983; Bybee 2001: 126-129). One example of a source-oriented generalization is “a singular ending in [k] corresponds to a plural ending in [tʃi]”. If the generalization is made over forms that belong to the same cell of the paradigm, e.g., plural forms, and identifies common or unexpectedly uncommon properties of forms that belong to this paradigm cell, then the generalization is said to be ‘product-oriented’. Bybee (2001:126) points out that ‘Generative rules express source-oriented generalization. That is, they act on a specific input to change it in well-defined ways into an output of a certain form’ and hypothesizes that “Many, if not all, schemas are product-oriented rather than source-oriented.” A major goal of the present thesis is to provide a way to empirically distinguish between product-

oriented and source-oriented generalizations and to identify characteristics of the learning situation that may favor one type of generalization over another. In this section, I describe a set of four artificial languages that feature the process of velar palatalization, whereby velars ([k] and [g]) become alveopalatals ([tʃ] and [dʒ] respectively) when followed by the front vowel [i], and show that differences in the productivity of velar palatalization across these languages can be used to determine whether the learners presented with the languages are extracting product-oriented or source-oriented generalizations.

First, let us consider the artificial languages shown in Table 2.1. Both languages feature two plural suffixes, -i and -a. In all four languages, velar stops obligatorily become alveopalatals in front of the plural suffix -i. In all four languages the  $\{k;g\} \rightarrow \{tʃ;dʒ\}$ i singular-plural mapping is supported by the same number of word types as well as word tokens. The languages differ in how often -i attaches to non-velar-final singulars, which happens thrice as often in Language II as in Language I.

Table 2.1. Two of the four languages used to test the difference between source-oriented and product-oriented generalizations. The variables M and N show the numbers of word pairs exemplifying a particular rule in each of the four languages. M and N can be unequal, and are greater than zero.

	Language I	Language II
$\{k;g\} \rightarrow \{tʃ;dʒ\}$ i	M	
$\{t;d;p;b\} \rightarrow \{t;d;p;b\}$ i	N	3N
$\{t;d;p;b\} \rightarrow \{t;d;p;b\}$ a	3N	N

Throughout this thesis, I will define the productivity of velar palatalization before –i for a particular speaker as the number of word types in the speaker’s output in which a singular ending in [k] or [g] corresponds to a plural ending in [tʃi] or [dʒi] divided by the number of word types in which a singular ending in [k] or [g] corresponds to a plural ending in [i], preceded by either a velar or an alveopalatal, stated mathematically in (1).

$$(1) \quad \text{Productivity of velar palatalization} = N_{\{tʃi;dʒi\}} / (N_{\{tʃi;dʒi\}} + N_{\{ki;gi\}})$$

Whether the productivity of –i with non-velar-final singulars influences the productivity of velar palatalization depends on the kinds of generalizations that are extracted by the learner. First, let us consider the rules that generated the data to which the learner is exposed, shown in the left column of Table 2.1. To recover these generalizations, the learner may search for the most general source-oriented generalizations over singular-plural mappings that can be extracted while minimizing the amount of competition between the extracted generalizations.<sup>23</sup> In the extracted grammar, the rule that triggers velar palatalization,  $\{k;g\} \rightarrow \{tʃ;dʒ\}i$ , is supported by the same number of examples in both languages and does not compete with any other rule. Thus, if the learner extracts these generalizations, the productivity of velar palatalization is expected to be independent of the productivity of –i with non-velar-final singulars, hence the productivity of velar palatalization in Language I is predicted to be identical to the productivity of velar palatalization in Language II.

---

<sup>23</sup> I am assuming that the subjects consider the set of consonants and vowels presented during the experiment to be the full set of segments present in the artificial language.

Alternatively, the learner may not be as concerned with minimizing the competition between rules, extracting the generalization  $C \rightarrow Ci$  or, less formally, ‘just add –i’. One computational model that predicts the extraction of this generalization is the Minimal Generalization Learner (Albright & Hayes, 2003). This generalization competes against the palatalizing generalization  $\{k;g\} \rightarrow \{tʃ;dʒ\}i$  for velar-final singulars and is much more reliable in Language II than in Language I. Thus, palatalization is predicted to be less productive in Language II than in Language I and the more a given subject attaches –i to non-velar-final singulars, the more s/he is expected to attach –i to velar-final singulars without changing the preceding velar into an alveopalatal.

The simplest product-oriented model is one in which the possible generalizations have the form ‘products must have X’ (Bybee 2001). Thus, in the case of our two artificial languages, the relevant palatalizing schemas would have the form ‘plurals end in  $\{tʃ;dʒ\}i$  (in context X)’. This palatalizing schema has the same type frequency in Language I and Language II. However, this does not necessarily mean that the productivity of velar palatalization is the same in the two languages. Suppose that the learner attempts to simultaneously satisfy ‘plurals must end in  $-\{tʃ;dʒ\}i$ ’ and ‘plurals must end in  $-Ci$ ’. The support for the second generalization is greater in Language II than in Language I, thus it will be satisfied more often. The support for the first generalization is the same across the two languages, thus it would be satisfied equally often. Thus the proportion of times a plural ending in –i and derived from a singular ending in [k] or [g] features velar palatalization is expected to be lower in Language II than in Language I due to differences in the number of  $\{k;g\}i$  plurals produced.

The same prediction is made if the learner extracts conditional product-oriented generalizations of the form ‘if the plural ends in –i, the preceding consonant must be {tʃ;dʒ} (in context X)’ (Aslin et al., 1998). The reliability of this generalization (given as the number of plurals that end in {tʃ;dʒ}i divided by the number of plurals that end in –i) differs between the two languages. Since the denominator is much greater in Language II than in Language I, palatalization is correctly predicted to fail more often in Language II than in Language I.

The opposite prediction is made by negative product-oriented generalizations, which militate against unattested sequences, assuming that such generalizations are strengthened whenever the learner expects to hear a sequence but does not in fact hear it. Suppose that the learner develops a preference against [ki], which increases whenever a learner expects to but does not in fact hear it. /C<sub>[-cont]</sub>i/ plurals are more common in Language II than in Language I while /kV/ plurals don’t occur in either language, thus the learner generalizing over plurals would expect and fail to hear /ki/ more often in Language II than in Language I. Thus velar palatalization is predicted to be more productive in Language II than in Language I.

Thus, the comparison between Language I and Language II pits positive product-oriented generalizations and competing weighted rules against negative product-oriented generalizations and non-competing rules.

The comparison between languages I and II on the one hand and languages III and IV on the other hand allows us to distinguish between product-oriented and source-oriented generalizations. As shown in Table 2.2, the difference between languages I-II

and languages III-IV is that languages III-IV have a number of additional singular-plural pairs in which a singular ends in  $\{tʃ;dʒ\}$  and the plural ends in  $\{tʃ;dʒ\}i$ .

Table 2.2. The four languages used to test the difference between source-oriented and product-oriented generalizations. The variables M, N, and K show the numbers of word pairs exemplifying a particular rule in each of the four languages. M, N, and K can be unequal, and are greater than zero.

	Language I	Language II	Language III	Language IV
$\{k;g\} \rightarrow \{tʃ;dʒ\}i$	M			
$\{t;d;p;b\} \rightarrow \{t;d;p;b\}i$	N	3N	N	3N
$\{t;d;p;b\} \rightarrow \{t;d;p;b\}a$	3N	N	3N	N
$\{tʃ;dʒ\} \rightarrow \{tʃ;dʒ\}i$	0		K	

The product-oriented accounts and the source-oriented account crucially differ in the treatment of singular-plural mappings in which the singular ends in  $\{tʃ;dʒ\}$  and the plural ends in  $\{tʃ;dʒ\}i$ , which are found only in languages III and IV. Under the product-oriented account, these mappings exemplify the palatalizing generalizations ‘plurals must end in  $-\{tʃ;dʒ\}i$ ’ or ‘if the plural ends in  $-i$ , the preceding consonant must be  $\{tʃ;dʒ\}$ ’.

Thus, their addition to the training set should increase the productivity of velar palatalization in languages III and IV relative to languages I and II respectively. Under the source-oriented account, these singular-plural pairings exemplify the rule  $C \rightarrow Ci$ , which militates against velar palatalization. Thus, their addition should reduce the

productivity of velar palatalization in languages III and IV relative to languages I and II. In the rest of the thesis, we test these predictions in a training paradigm expected to favor source-oriented generalizations (Chapter 3) and then in a more natural training paradigm expected to favor product-oriented generalizations (Chapter 4).

To summarize, the shape of the grammar is determined in part by the types of generalizations the grammar includes. Three of the dimensions on which generalizations differ are 1) whether they are source-oriented or product-oriented, 2) whether they are positive, defining typical structures, or negative, militating against underobserved structures, and 3) how general they are. In the present thesis, I situate generalizations extracted by learners in an artificial grammar learning paradigm along these three dimensions and examine whether the shape of the grammar depends on the learning situation by manipulating the training paradigm in favor of source-oriented generalizations (Chapter 3) or product-oriented generalizations (Chapter 4). In each case, elicited production data is supplemented by likelihood rating data to examine whether generalizations are weighted equally in production and perception.

Table 2.4 summarizes the predictions of the alternative views on the types of generalizations included in the grammar that a learner extracts from the lexicon s/he experiences. As discussed above, product-oriented generalizations predict that examples in which a singular ending in [tʃ] or [dʒ] corresponds to a plural ending in [tʃi] or [dʒi] favor the conversion of singulars ending in [k] or [g] into plurals ending in [tʃi] or [dʒi] while source-oriented generalizations predict the opposite as long as the palatalizing rule competes with a more general rule that simply adds –i with no changes to the stem.



Table 2.4. Rankings of languages I-IV from the one that favors velar palatalization the most (1) to the one that favors velar palatalization the least according to the various views of grammar.

	Language I	Language II	Language III	Language IV
Source-oriented:				
Minimal competition	1	1	1	1
Reliability-weighted	1	2	2	3
Product-oriented:				
Positive	2	3	1	2
Negative	3	2	2	1

## CHAPTER 3

### GENERALIZATIONS IN A SOURCE-ORIENTED PARADIGM

This chapter describes two experiments in which native English speakers were presented with the four artificial languages described in Chapter 2 in a training paradigm that can be expected to favor source-oriented generalizations (first introduced by Bybee & Newman, 1995). In this paradigm, learners are exposed to singular-plural pairs, which they repeat aloud during training. Exposure to and repetition of singular-plural pairs is expected to favor generalizations over singular-plural pairs, i.e., source-oriented generalizations.

The learners are tested by either being presented with a singular and asked to form the plural or being presented with a singular-plural pair and asked to rate (on a scale from ‘impossible’ to ‘very likely’) how likely they think it is that the presented plural form is the right plural for the presented singular. The two tasks differ in that one task involves perception while the other involves production, and in that the production task requires competition between alternative plural forms, while the perception task does not. For instance, given a singular ending in [t], a learner may rate both a plural ending in [ti] and a plural ending in [ta] as very likely but only one of the plurals would be produced in the production task. Thus comparison between the two tasks allows us to assess the combined impact of modality and the decision rule on the types of grammatical generalizations used in perception and production and the relative weighting of competing generalizations in the two modalities.

The main aim of the present chapter is to examine the types of generalizations that emerge in a clearly source-oriented training paradigm. If, despite training favoring source-oriented generalizations, the learners induce product-oriented generalizations, then strong evidence for a bias in favor of extracting product-oriented generalizations is obtained. On the other hand, if the induced grammar is source-oriented, this result could be due to either the nature of the learning situation or a prior bias in favor of source-oriented grammars. The interpretation of this result would be clarified in Chapter 4, where a product-oriented paradigm is used for training.

### **3.1. The paradigm**



This section describes the experimental paradigm used in the experiments reported in the present chapter. We begin with a description of the training and testing tasks, followed by the exact procedures used, a description of how the training and testing stimuli were generated, and the human participants who took part in the experiments.

#### **3.1.1. Tasks**

The experiment consisted of a training stage, an elicited production test, and a likelihood rating test. During the training stage, a participant would be presented with a series of trials, each of which began with the presentation of a picture of a novel object on the computer screen. Three hundred milliseconds later, the name of the novel object in one of the four artificial languages was presented auditorily over headphones. Once the sound finished playing, the picture was removed and replaced with a picture of two (in Experiment I) or more (in Experiment II) objects of the same type. The picture of


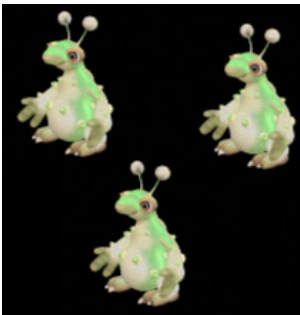
multiple objects was accompanied by the auditory presentation of the plural form of the previously presented noun. Once the sound file finished playing, the participant repeated the singular-plural pair and clicked a mouse button to continue to the next singular-plural pair. The training task is shown schematically in Figure 3.1. Instructions for the training task are shown in Appendix 1.

Figure 3.1. The training task used in Experiments I and II.

Video:						
Audio:		[bouk]			[boutʃi]	
Learner action:	Watch	Watch & listen		Watch	Watch & listen	Repeat aloud, then click
Duration:	300 ms	500-900 ms	500 ms	300 ms	500-900 ms	0-10s



The training stage was followed by the elicited production test, which was exactly like training except instead of hearing the plural form and repeating the singular-plural pair, the learner had to generate the plural and pronounce it aloud. The learner was not required to repeat the singular during the test. The task is shown schematically in Figure 3.2. Instructions for the elicited production task are shown in Appendix 1.

Figure 3.2. The elicited production test.

Video:				
Audio:		[vik]		
Learner action:	Watch	Watch & listen		Say the plural aloud, then click
Duration:	300 ms	500-900 ms	500 ms	0-10s

The elicited production test was followed by the rating task. In the rating task, the subject was presented with a singular-plural pair as s/he would be during training and had to answer “How likely is this plural to be the right plural for this singular?” on a scale from 1=“impossible” to 5=“very likely”. The scale was displayed on the screen, and the learner responded by clicking a numbered rectangle with the mouse. The task is presented schematically in Figure 3.3. Instructions and the rating scale are shown in Appendix 1.

Figure 3.3. The ratings task.

Video:						
Audio:		[fruk]			[fruki]	
Learner action:	Watch	Watch & listen		Watch	Watch & listen	Repeat aloud, then click
Duration:	300 ms	500-900 ms	500 ms	300 ms	500-900 ms	0-10s

### 3.1.2. Stimuli

A given learner was exposed to one of the four languages shown in Table 3.1. All four languages had 30 singular-plural pairs illustrating velar palatalization. Languages I and II had no singulars ending in an alveopalatal, while languages III and IV had 20 singular-plural pairs featuring such a singular. In all 20 cases, the singular corresponded to a plural ending in {tʃ;dʒ}i. In Language I and Language III, -i attached mostly to velar-final singulars, while in Language II and Language IV it frequently attached to singulars ending in a non-velar. In total, learners exposed to Language I or Language II were presented with 62 singular-plural types. Learners exposed to Language III or Language IV were presented with 82 singular-plural types. The full set of stimuli is shown in Appendix 2. Each type was presented twice during training. The large number of

different word types that are presented to subjects and the low token/type ratio is expected to result in generalization across words and lack of memorization of individual wordforms. This feature of the present training paradigm is distinct from the product-oriented paradigm that is presented in Chapter 4, where subjects are presented with a relatively small number of frequently occurring words that they are asked to memorize.

Table 3.1. The four languages presented to learners in Experiments I-II

	Language I	Language II	Language III	Language IV
{k;g} → {tʃ;dʒ}i	30			
{t;d;p;b} → {t;d;p;b}i	8	24	8	24
{t;d;p;b} → {t;d;p;b}a	24	8	24	8
{tʃ;dʒ} → {tʃ;dʒ}i	0		20	

The training stimuli (shown in Appendix 2) were designed in the following manner. First, a number of phonotactically legal English CV, CCV, and sCCV sequences were chosen. Second, each CV or CCV was combined with all possible singular-final codas and the second syllables of plural forms from the four artificial languages. For instance, one English CCV used was [strou], which yielded [strook]-[stroutʃi], [stroug]-[stroudʒi], [strout]-[strouti], [strout]-[strouta], [stroup]-[stroupi], [stroup]-[stroupa], [stroutʃ]-[stroutʃi] and [stroudʒ]-[stroudʒi]. The word pairs featuring velar palatalization were shared across languages. The examples of {t;d;p;b} → {t;d;p;b}a or {t;d;p;b} → {t;d;p;b}i shared singulars, and differed only in the vowel marking the plural.

The singular forms of the plurals that were not shared between the four languages were minimal pairs of singulars ending in a velar. The high similarity between words ending in different consonants is intended to discourage memorization of individual wordforms and encourage abstraction, again in contrast to the product-oriented paradigm introduced in the following chapter. The high similarity between velar-final and non-velar-final singulars is designed to increase competition between the extracted palatalizing and non-palatalizing generalizations.

The auditory stimuli used for elicited production were divided evenly into stimuli presented during training and stimuli that were not presented during training. The novel stimuli used CV, CCV and sCCV sequences that were not used for training. There were 30 test stimuli ending in [k] or [g], 10 test stimuli ending in [tʃ] or [dʒ], and 8 stimuli ending in [p], [b], [t], and [d] each. Half of the stimuli ending in each stop consonant were novel, while the other half had been presented during training. All stimuli ending in alveopalatal affricates were novel. No differences between novel and previously presented stimuli were found, thus novel and previously presented stimuli will be grouped together in the analyses presented throughout the chapter.

There were 72 singular-plural pairs used in the rating task. None of the stimuli had been presented to subjects during training or elicited production. For each singular form, there was a plural form ending in the same consonant as the singular followed by [i], a form ending in the same consonant as the singular followed by [a], a form ending in [tʃi], and a form ending in [tʃa]. Pilot data (the first 18 subjects of Experiment I, who did not participate in the rating task) showed that labial-final singulars were never mapped onto plurals ending in {tʃ;dʒ}V in elicited production, while this sometimes occurred for



alveolar-final singulars. Thus, subjects may consider labial-alveopalatal mappings to be particularly bad, which could lead to all other mappings being rated as relatively good. Therefore, differences between these relatively good mappings would be difficult to detect if labial-alveopalatal mappings were included. Therefore, only alveolar-final, velar-final and alveopalatal-final singulars were included in the rating task. There were six singulars ending in consonants of each of the three remaining places of articulation. The singulars featured minimal triplets differing only in the place of articulation of the final consonant, e.g., [wʌg], [wʌd], and [wʌdʒ], or [kwɛk], [kwɛt], and [kwɛtʃ]. Each triplet had a distinct body that was not presented during training or elicited production and members of a triplet shared the voicing of the final consonant. There were six such triplets, resulting in a total of 18 singulars, each of which had four alternative plurals (shown in Appendix 2).

The singular-plural pairs were presented in random order, which is different from a previous study of naturalness rating of English past tense forms (Albright & Hayes, 2003) where alternative inflected forms for a certain stem were presented one after the other. It was felt that presenting alternative plural forms for the same singular next to each other would encourage competition between the alternative plurals. Such competition does not necessarily occur when a listener evaluates the grammaticality or naturalness of a form s/he hears (as argued in Kapatsinski 2007 for rating novel Russian verbs bearing alternative stem extensions).

The auditory stimuli were recorded by me in a sound proof booth onto a computer.<sup>25</sup> The stimuli were sampled at 44.1 kHz and leveled to have the same mean amplitude. They were presented to the learners at a comfortable listening level of 63 dB.

The visual stimuli differed across the two experiments reported in the present chapter. Experiment I, which compared Language I to Language II used a set of pictures of geometric shapes and patterns created by the experimenter. Experiment II, which compared all four languages shown in Table 3.1, featured a set of made-up creature pictures retrieved from the website [sporepedia.com](http://sporepedia.com), which are exemplified in Figures 3.1-3.3. In addition, Experiment I paired plural forms with a picture of two shapes, while Experiment II paired plural forms with a picture of multiple (5-8) creatures. All pictures were presented in .png format on a black background.

### 3.1.3. Procedures

Learners were tested one a time. The learner was seated in a sound-proof booth. The audio stimuli were delivered via headphones, while the learner's speech was recorded onto a digital audio tape using a head-mounted microphone. The experimenter was seated outside the booth and was able to hear the audio presented to the learner as well as the learner's productions. The learner was unable to see the experimenter. The subject's productions were scored by the experimenter online, as the learner was producing them. The stimuli were presented and ratings recorded using PsyScript

---

<sup>25</sup> While I am not a native speaker of English, I have lived in the US for 10 years (since I was 14) and have studied English in Russia since I was 5. I also avoided vowels for which I do not have a native-like pronunciation (e.g., [ʊ]). The artificial languages presented to subjects were based loosely on Russian (my native language), which features velar palatalization, and are to be compared to Russian in chapter 5, I believe the Russian-influenced pronunciations are not inappropriate. They may also make learners more aware that they are not listening to English and reduce first-language interference. Despite somewhat non-native pronunciations, repetition accuracy was very high during training (97%).

experiment presentation software on Mac OS9.2. The order of presentation of the stimuli was randomized separately for each learner.

#### 3.1.4. Participants

Two separate experiments were conducted using the source-oriented paradigm. In the first experiment, 17 participants were exposed to Language I while 17 other participants were exposed to Language II. In the second experiment, 44 participants were exposed to the four languages in Table 3.1. Each participant was exposed to only one language. Language I and Language IV were presented to 10 participants each, while Language II and Language III were presented to 12 participants each. All of the participants reported being native English speakers with no history of speech, language, or hearing impairments. None reported being fluent in a foreign language. All of the participants were recruited from introductory psychology classes and received a small amount of course credit for participation. Participants were assigned to languages randomly.

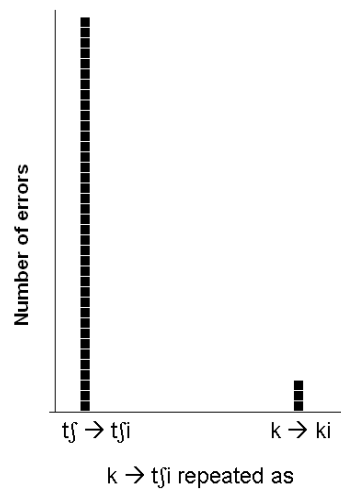
### **3.2. Results I: Errors in perception and memory**

Overall, the training stimuli were perceived and repeated with a high degree of accuracy. The mean error rate across the two experiments was 3%. If a subject made errors in the place of articulation of a crucial consonant on more than 5% of the training trials, s/he was excluded from the experiment. Two participants were excluded from Experiment I on the basis of this criterion.

During training, the subject is asked to repeat the singular-plural pairs s/he is hearing. An interesting type of error that occurred in this task is that the subject would change the final consonant of the singular to be the same as the consonant that preceded the plural suffix. Thus, for instance, a subject exposed to [frouk]-[frouki] might repeat the pair of words as [frouk]-[frouki], thus making the stem of the singular identical to the stem of the plural. When questioned after completing the experiment, the subjects were not aware of making errors of this type, despite the fact that this type of error was the most common error type made during the training. Interestingly, the opposite error pattern, in which the plural stem would become identical to the singular stem (as in [frouk]-[frouki] becoming [frouk]-[frouki]), was very uncommon. The result is shown in Figure 3.4. The differences between the frequencies of the two error patterns is highly statistically significant (38 vs. 3,  $\chi^2(1)=28.9$ ,  $p<.0001$ ).

This result suggests that the learners have a bias against stem changes, which are eliminated in perception or in short-term memory. Furthermore, the plural form is more salient than the singular form. The plural form is more likely to be recalled correctly and to influence the other form in the paradigm. It is important to note that the plural is the ‘product’ while the singular is the source in the present study, thus its greater salience relative to the singular form suggests that generalization over products may be easier than generalization over sources. While the plural follows the source in the present task, hence the finding may well be a recency effect, the product also follows the source in the normal timecourse of production and in naturalness rating, thus the present result should be taken into account in interpreting production and naturalness judgment behavior.

Figure 3.4. The learners are more likely to change the singular to be identical to the stem of the plural than they are to change the plural to fit the singular during training.



The result in Figure 3.4 indicates that characteristics (or sublexical chunks) of the plural form can persist and spread to influence the source form. Potentially, the persisting chunks and patterns can also influence other product forms that are produced in close temporal proximity, generating a type of response priming, which could account for apparently product-oriented behavior in previous elicited production studies. Bybee (2001) reviews a number of studies purporting to show product-oriented generalization in an elicited production task (Bybee and Slobin 1982, Bybee and Moder 1983, Köpcke 1988, Lobben 1991, Wang and Derwing 1994, Albright and Hayes 2003). In all cases, the argument is that instead of respecting the input-output mappings present in the lexicon, subjects ‘overuse’ common output patterns deriving them in ways not attested in the lexicon. As discussed in Chapter 1, the overuse can also be explained by experiment-internal response priming, which is supported by the response sequence data in Lobben’s (1991) study on Hausa plurals and is also consistent with the data in Figure 3.4.

A more positive way to view this result is that this persistence of chunks and patterns found in product forms may be the mechanism that leads to the emergence of product-oriented generalizations. As long as this persistence can be shown to be long-lasting, it can potentially lead to the continual reuse of the same product-oriented patterns regardless of the shape of the source.

One caveat is that deriving the singular from the plural is easier than deriving the plural from the singular in languages I and II because the plural form can have either –i or –a but the form of the singular is certain given the plural. Albright (2005, 2008) proposes that speakers acquiring morphological paradigms choose the source form to be the one that allows for the most reliable rules to be formulated. Thus, he would predict the plural to be the chosen source form in Languages I and II. However, this appears to be highly unlikely in the present paradigm where subjects are asked to repeat singular-plural pairs, and the plural always follows the singular. In Languages III and IV, the shape of the singular is not predictable from the plural because a plural ending in [tʃi] can correspond to a singular ending in either [k] or [tʃ]. It is impossible to determine if Languages III and IV differ from Languages I and II with respect to this effect because of the low error rate.

### **3.3. Results II: Types of generalizations extracted from the input**

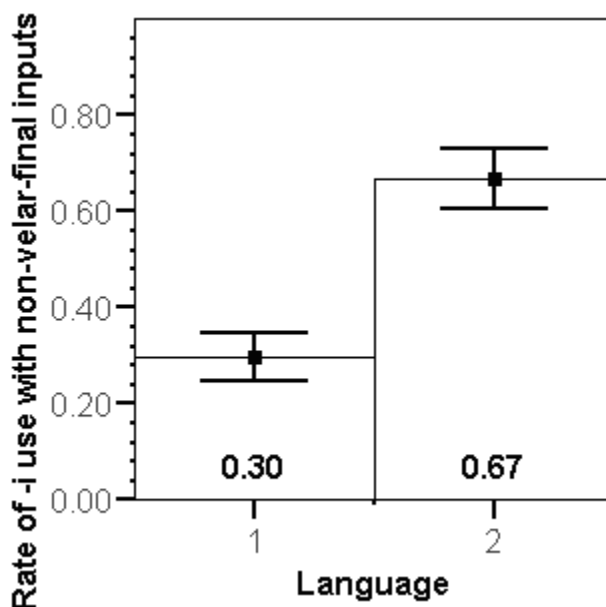
We now turn to the results of elicited production and likelihood rating. First, we examine the types of generalizations that are extracted from the input in the present training paradigm using the data from elicited production, followed by likelihood rating data.

### 3.3.1. Elicited production

#### 3.3.1.1. Experiment I: Language I vs. Language II

In the first experiment, learners were exposed either to Language I, in which the suffix *-i* attaches mostly to the segments it palatalizes (velars), or Language II, in which the suffix *-i* is also likely to attach to non-velar-final singulars. In the lexicon of Language I presented to learners, labial-final and alveolar-final singulars take *-i* 75% of the time, while they only take *-i* 25% of the time in Language II. Figure 3.5 shows that the learners were able to match the proportions in the training data, attaching *-i* to labials and alveolars 30% of the time in Language I and 67% of the time in Language II ( $t(28)=4.4, p<.001$ ). Thus the training was successful in making *-i* more productive with non-velar-final singulars in Language II than in Language I.

Figure 3.5. The learners successfully matched the probability with which *-i* and *-a* attach to labial-final and alveolar-final singulars (25% *-i* vs. 75% *-a* in the input).



More interestingly, Figure 3.6 shows that participants exposed to Language I, the language predicted to favor velar palatalization by virtue of disfavoring the use of *-i* with non-velar-final singulars, palatalized the velar before *-i* 67% of the time, while participants exposed to Language II palatalized the velar before *-i* only 38% of the time ( $t(28)=2.316, p<.05$ ).<sup>26</sup> This result is consistent with reliability-weighted competition between the palatalizing rule  $\{k;g\} \rightarrow \{tʃ;dʒ\}i$  and the more general consonant-retaining rule  $C \rightarrow Ci$ . It is also consistent with positive product-oriented generalizations, ‘plurals must end in  $-\{tʃ;dʒ\}i$ ’ and ‘plurals must end in *-i*’, or ‘if the plural ends in *-i*, the preceding consonant must be  $\{tʃ;dʒ\}$ ’.

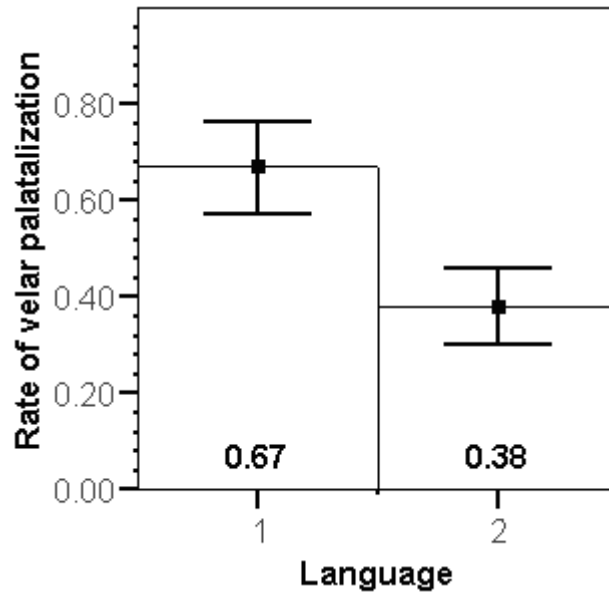
The result in Figure 3.6 is inconsistent with error-driven learning of negative product-oriented generalizations like ‘plurals cannot end in  $[ki]$ ’, assuming that a negative product-oriented generalization is strengthened in proportion to the difference between observed probability of a product and its expected probability. Since there are more plurals ending in  $[i]$  in Language II than in Language I,  $[ki]$  is expected to occur more often in Language II than in Language I. Therefore, its absence should be more salient in Language II than in Language I. If the constraint against  $[ki]$  is strengthened every time it is expected but not observed, this constraint should be stronger in Language II. Therefore, velar palatalization should be more productive in Language II but the exact opposite pattern is observed.

---

<sup>26</sup> Four subjects were excluded from this analysis because they attached *-i* to a velar-final singular in only five words or fewer, making the estimate of velar palatalization rate unreliable.



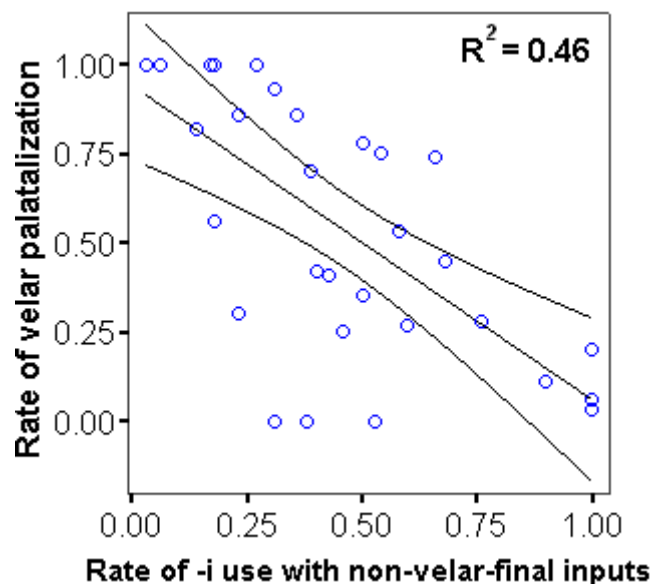
Figure 3.6. Learners exposed to Language 2 are less likely to palatalize the velar before –i than learners exposed to Language 1.



Not all learners who are exposed to a particular language actually acquire the same grammar. There is variability in how much a given generalization is weighted by an individual. Reliability-weighted rules and positive product-oriented generalizations claim that the rate at which a given individual changes the velar into an alveopalatal when s/he attaches –i to it should be predictable from how often the individual attaches –i (as opposed to –a) to non-velars. Namely, the more productive –i is with non-velars, the more likely velars should be retained before –i. Figure 3.7 shows that this prediction is borne out by the data. There is a strong and significant negative correlation ( $r(28)=-.68$ ,  $p<.001$ ) between how much a subject uses –i with non-velar-final inputs and how likely s/he is to palatalize a velar before –i. Interestingly, when the Rate of –i use with non-velars and Language (I vs. II) are entered into an ANCOVA as predictors of the rate of velar palatalization, only the rate of –i use with non-velars remains significant

( $F(1,27)=14.23$ ,  $p<.001$ , for rate of  $-i$  use with non-velars;  $F(1,27)=.082$ ,  $p>.5$  for Language). Thus, the variable that is predicted to account for the productivity of velar palatalization by reliability-weighted rules and positive product-oriented generalizations in fact accounts for all of the variance in the productivity of velar palatalization that is attributable to the artificial language to which a learner is exposed.

Figure 3.7. Subjects for whom  $-i$  is productive with inputs that cannot undergo velar palatalization are the subjects for whom velar palatalization is unproductive. Curves show the 95% confidence region for the regression line.

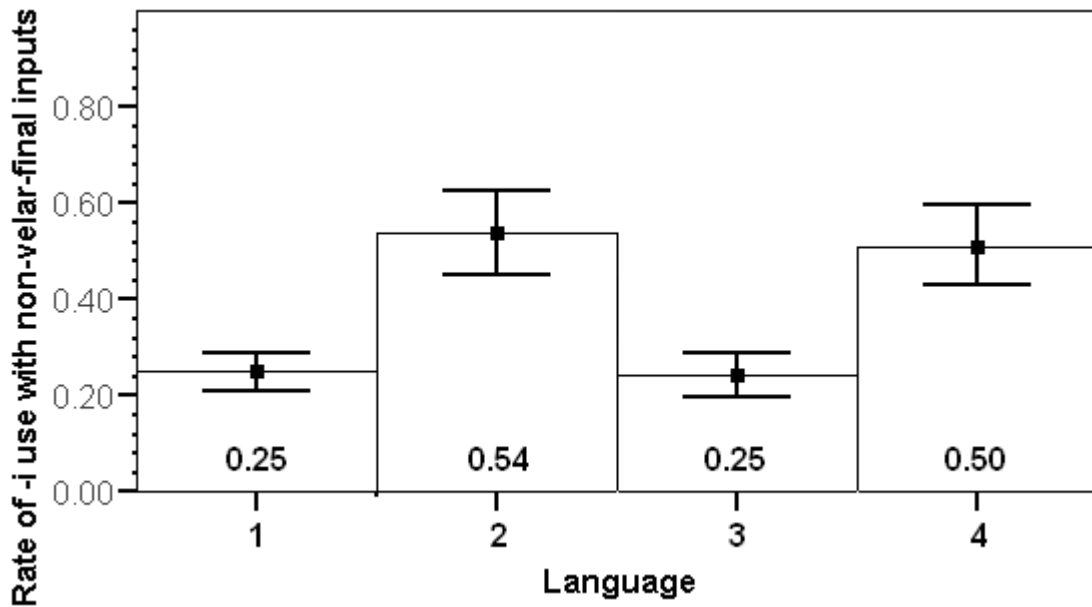


### 3.3.1.2. Experiment II: Languages I-IV

In the second experiment, participants were exposed to one of four languages: Language I, Language II, Language III, or Language IV. In Language I and Language III,  $-i$  tends not to attach to labial-final and alveolar-final singulars, while in Language II and Language IV it attaches to such singulars often. Figure 3.8 shows that the subjects

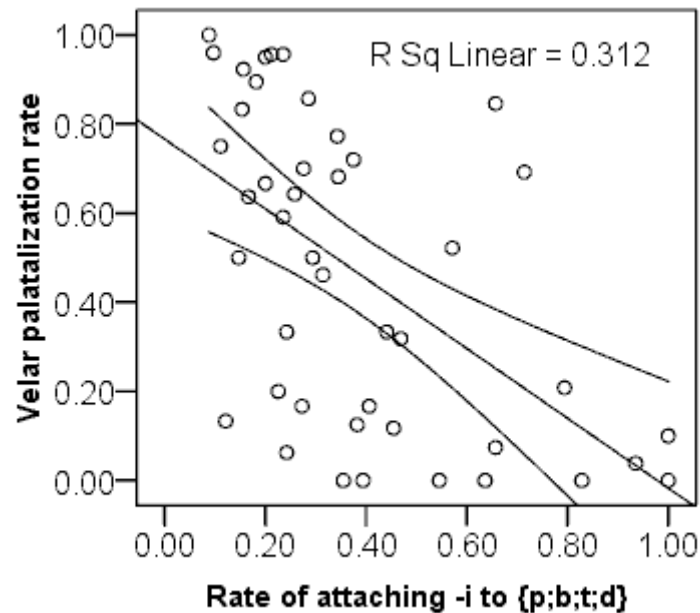
exposed to Languages II and IV attached -i to singulars ending in a labial or a coronal more than did subjects exposed to Language I or Language III.

Figure 3.8. The learners successfully learned that -i tends not to attach to labial-final and alveolar-final singulars in Languages I and III (25% in the input) but underestimated the use of -i with alveolar-final and labial-final singulars in Languages II and IV (75% in the input).



Like in Experiment I, there is a significant negative correlation between the rate of velar palatalization and the probability of attaching -i to coronals and palatals ( $r=.56$ ,  $p<.001$ , shown in Figure 3.9).

Figure 3.9. Subjects for whom  $-i$  is productive with inputs that cannot undergo velar palatalization are the subjects for whom velar palatalization is unproductive. The y-axis indicates the proportion that examples of  $\{k;g\} \rightarrow \{tʃ;dʒ\}i$  mappings form out of all examples of  $\{k;g\} \rightarrow Ci$  mappings produced by a learner. Curves show the 95% confidence region for the regression line.



The negative correlation shown in Figure 3.9 can have two possible causes: 1) learners who often attach  $-i$  to labials and alveolars do not produce velar-alveopalatal mappings as often as learners who rarely attach  $-i$  to labials and alveolars, or 2) learners who often attach  $-i$  to labials and alveolars are also likely to often attach  $-i$  to velars without changing the velar. Both positive product-oriented generalizations and weighted rules predict that the correlation should have the latter cause: the generalization whose reliability is responsible for differences in the rate of velar palatalization is the rule  $C \rightarrow Ci$  under the source-oriented account and ‘plurals must end in  $-i$ ’ under the product-oriented

account. Figures 3.10-3.11 shows that this is indeed the case: the correlation between the probability of attaching -i to {p;b;t;d} and attaching -i to velars without changing the velar is extremely strong ( $r=.81$ ,  $p<.001$ ) while the correlation between the probability of attaching -i to {p;b;t;d} labials and mapping velars onto {tʃ;dʒ}i is weak and insignificant ( $r=.34$ , n.s.; as can be seen in the graph, the 95% confidence region for the regression line includes flat lines).

Figure 3.10. The rate of attaching -i to non-velars influences the rate of velar palatalization by influencing how often -i is attached to the velar without changing the velar. Curves show the 95% confidence region for the regression line.

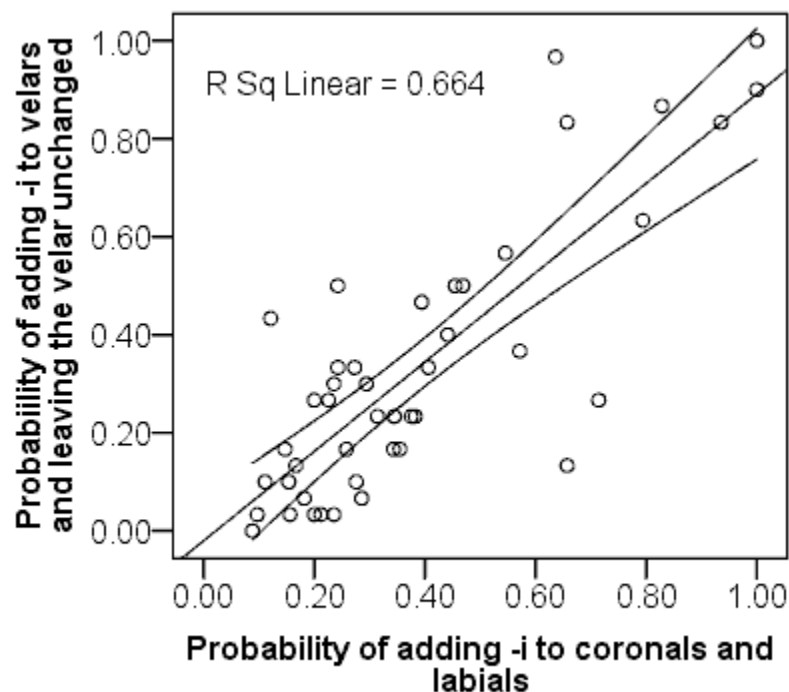
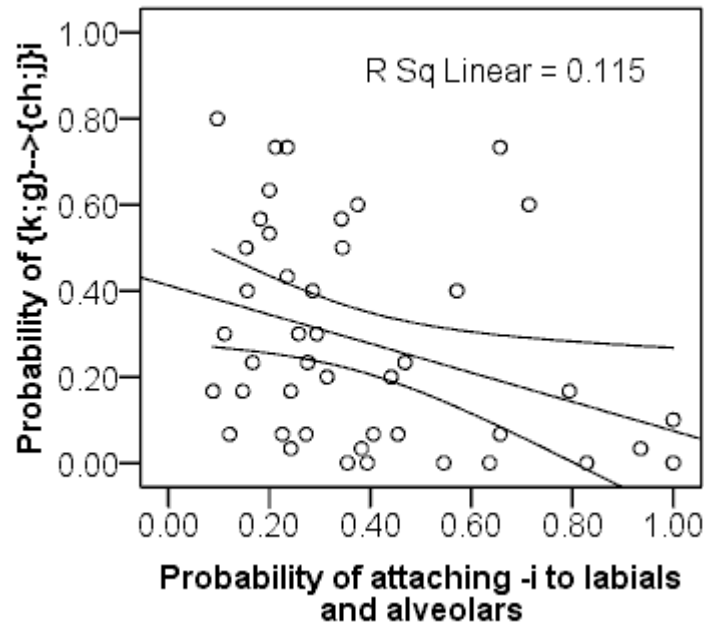


Figure 3.11. The rate of attaching –i to non-velars does not influence how often –i is attached to the velar changing the velar to an alveopalatal as much. The y-axis indicates the proportion that examples of {k;g}→{tʃ;dʒ}i mappings form out of all examples of {k;g}→CV mappings produced by a learner. Curves show the 95% confidence region for the regression line.



In Language I and Language II, there are no examples in which –i attaches to a singular ending in an alveopalatal, while Language III and Language IV feature 20 examples of this type. However, subjects exposed to any language tended to attach –i to singulars ending in an alveopalatal with approximately equal high probability (.74 for Language I, .72 for Language II, .8 for Language III, and .78 for Language IV), an effect that is discussed in greater detail in the next section. Nonetheless, Figures 3.12-3.13 show that the addition of {tʃ;dʒ} examples reduced the rate of velar palatalization. When the presence of {tʃ;dʒ}→{tʃ;dʒ}i examples and the probability of attaching –i to alveolars

and labials are entered into an ANCOVA on ranks, both are significant ( $F(1,41) = 29.14$ ,  $p < .00001$  for rate of attaching to alveolars and labials;  $F(1,41) = 7.8$ ,  $p = .008$ , for presence of  $\{tʃ;dʒ\} \rightarrow \{tʃ;dʒ\}i$  examples). Thus velar palatalization rate is reduced if  $-i$  often attaches to non-velars, which may be labials and coronals (Languages 2 and 4 vs. Languages 1 and 2) or alveopalatals (Languages 3 and 4 vs. Languages 1 and 2).

Figure 3.12. Rate of velar palatalization across the four languages:  $1 > 2 = 3 > 4$

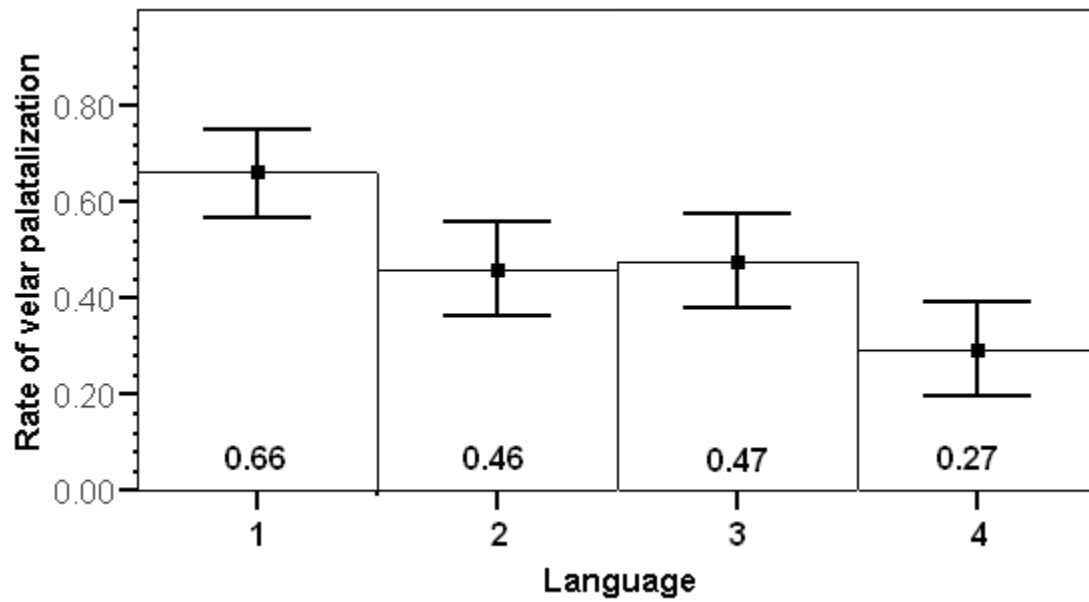
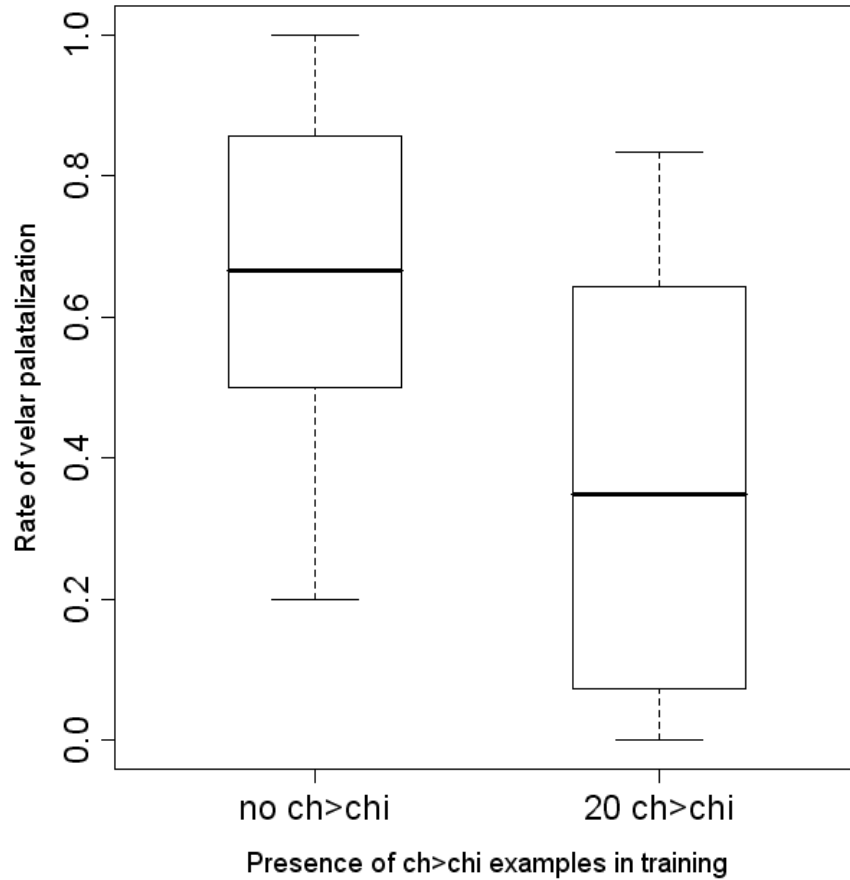


Figure 3.13. The addition of 20 singular-plural pairs exemplifying  $\{tʃ;dʒ\}i$  reduces the productivity of velar palatalization.



To summarize, the addition of examples in which a singular ending in an alveopalatal corresponds to a plural ending in  $\{tʃ;dʒ\}i$  reduces the probability that a velar-final singular will give rise to a plural ending in  $\{tʃ;dʒ\}i$  in production. This result directly contradicts the hypothesis that the learners are extracting product-oriented generalizations, where the generalization responsible for velar palatalization is ‘plurals must end in  $\{tʃ;dʒ\}i$ ’. The examples whose addition reduces the productivity of velar palatalization in the present paradigm exemplify the product-oriented generalization that

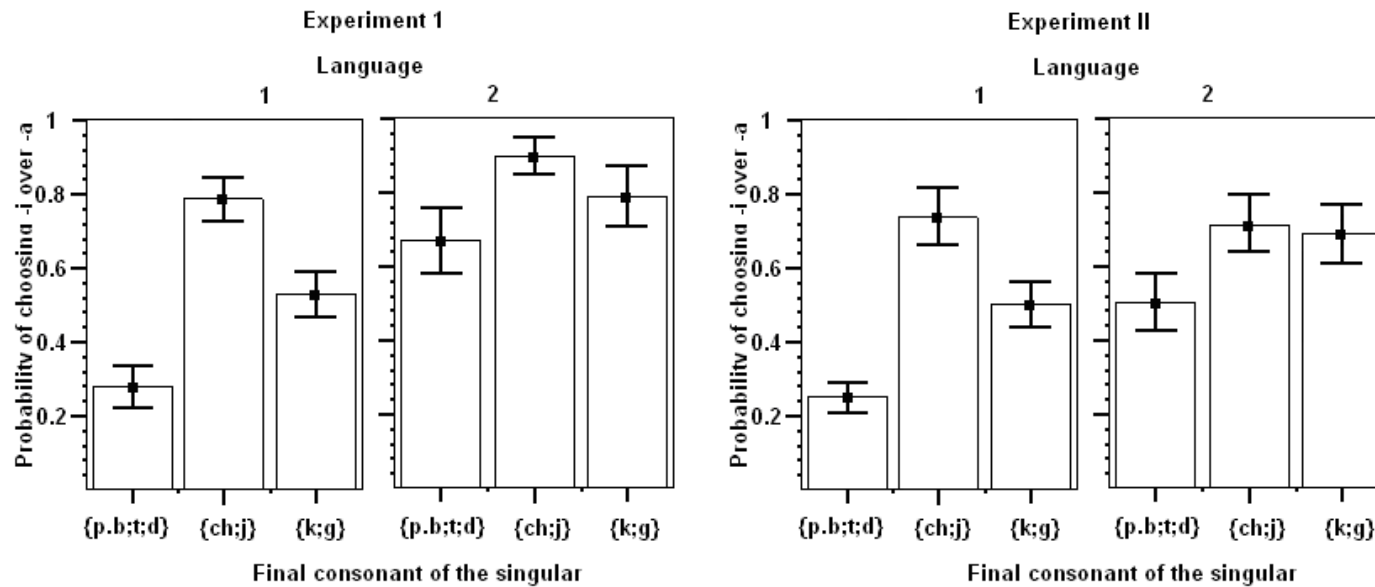


is supposed to favor velar palatalization. By contrast, the results are expected under the hypothesis that the learners are using source-oriented generalizations. The examples in which an alveopalatal is mapped onto an alveopalatal followed by  $-i$  are examples of the rule ‘just add  $-i$ ’ ( $C \rightarrow Ci$ ), which disfavors velar palatalization. Thus, the present training paradigm gives rise to the use of source-oriented generalizations in production.

### 3.3.1.3. An apparently product-oriented effect

Figure 3.14 shows how often the learners exposed to Language I or Language II in either Experiment I or Experiment II attach  $-i$  to singulars ending in labials, alveolars, alveopalatals, and velars. It is interesting to note that in both Language I and Language II singulars ending in an alveopalatal take  $-i$ , despite the fact that the learners are not exposed to any singulars ending in an alveopalatal during training. Furthermore, singulars ending in a velar are less likely to take  $-i$  than singulars ending in an alveopalatal in Language I, despite the fact that learners are exposed to no examples in which a singular ending in an alveopalatal takes  $-i$  but are presented with 30 examples in which a singular ending in a velar takes  $-i$  (and no examples in which a singular ending in a velar takes  $-a$ ). Thus, we are faced with a puzzle: why do the learners infer that alveopalatal-final singulars take  $-i$  and fail to attach  $-i$  to velars as often as the input data would suggest?

Figure 3.14. The probability of choosing –i over –a as the plural marker depending on the language to which a subject is exposed and the place of articulation of the final consonant of the singular.



One generalization that would predict that a singular ending in {tʃ;dʒ} should correspond with a plural ending in {tʃ;dʒ}i is the product-oriented generalization ‘plurals must end in {tʃ;dʒ}i’. This generalization has equal support in the two languages and would therefore predict that a plural ending in {tʃ;dʒ}i is a likely output even in Language I. Importantly, this is the same generalization that is held responsible for velar palatalization on the product-oriented account. Therefore, the more reliable this generalization is for a particular learner, the more s/he should palatalize velars and the more s/he should add –i to alveopalatals. Therefore, there is predicted to be a positive correlation between how often a learner produces a plural ending in {tʃ;dʒ}i in response to a singular ending in a velar and how often the same learner produces a plural ending in {tʃ;dʒ}i in response to a singular ending in an alveopalatal. However, this correlation is not reliably observed in the data (Experiment I:  $r(28) = -.03$ , n.s.; Experiment II:  $r(20) = .3$ , n.s.). Instead, as shown in Figures 3.15-3.16 the probability of attaching –i to an alveopalatal is strongly correlated with the probability of attaching –i to a velar regardless of whether the velar changes into an alveopalatal once –i is attached (Experiment I:  $r(32) = .61$ ,  $p < .001$ ; Experiment II:  $r(20) = .70$ ,  $p < .001$ ). Finally, there is no significant correlation between the probability of attaching –i to alveopalatal-final singulars and attaching –i to labial-final singulars when the probability of attaching –i to velar-final singulars is partialled out (Experiment I:  $r(31) = .03$ , n.s., Experiment II:  $r(19) = -.05$ , n.s.).

Figure 3.15. The probability of attaching –i to an alveopalatal-final singular correlates with the probability of attaching –i to a velar-final singular, regardless of whether the velar is changed into an alveopalatal. The data are from Experiment I.

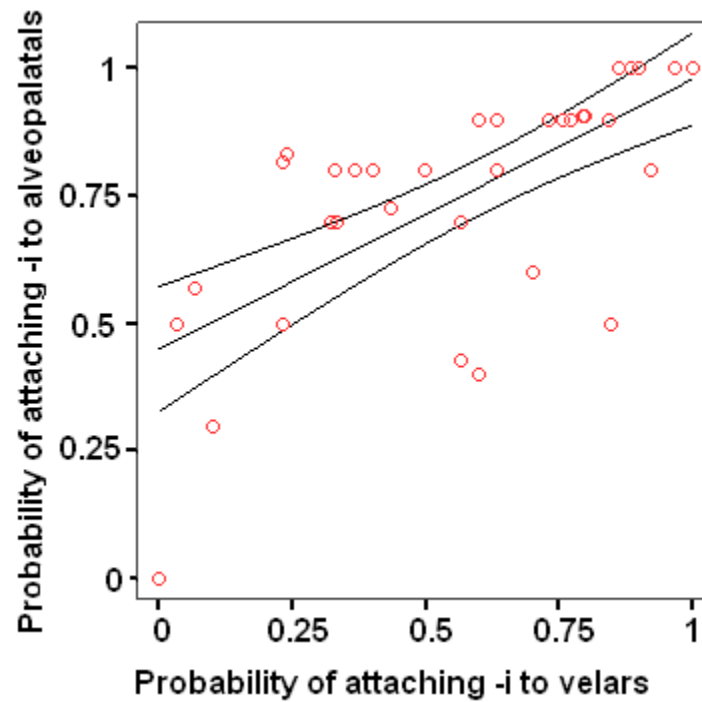
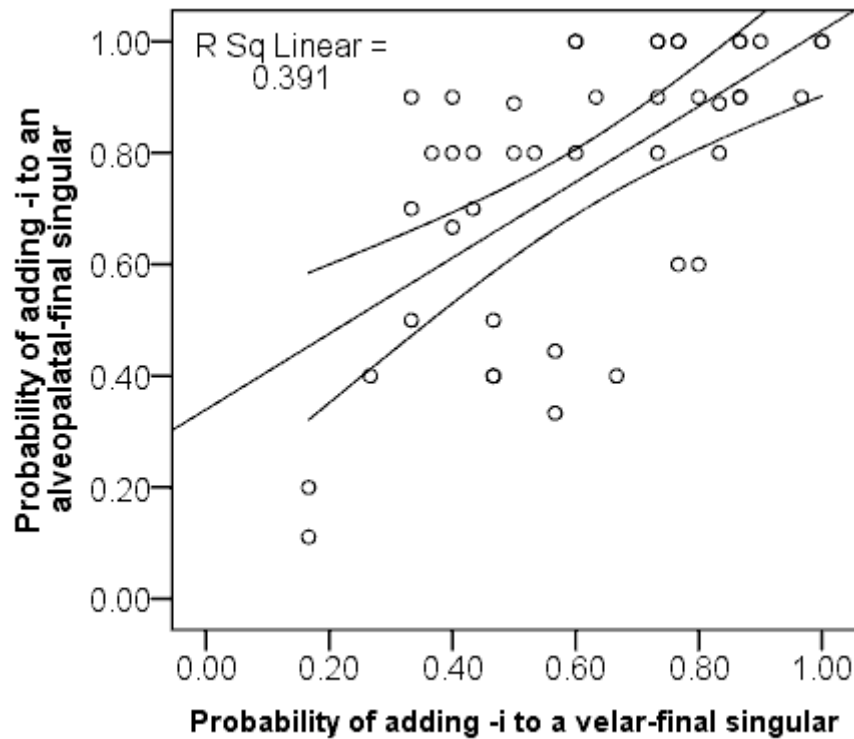


Figure 3.16. The probability of attaching –i to an alveopalatal-final singular correlates with the probability of attaching –i to a velar-final singular, regardless of whether the velar is changed into an alveopalatal. The data are from Experiment II (all points are shown, including points from subjects Exposed to Language III or Language IV).



Thus, if a subject is likely to attach –i to velars, s/he is also likely to attach –i to alveopalatals. Alveopalatal-final singulars appear to be categorized into the same category as velar-final singulars. Thus the generalization responsible for the high rate of attachment of –i is neither ‘plurals must end in {tʃ;dʒ}i’ nor ‘plurals must end in –i’. Rather, the relevant generalization defines a class of source or product forms that take –i. In a source-oriented grammar, this would be the rule ‘[ ] → i / {tʃ;dʒ;k;g}\_\_’, which may be followed in production by the rule {k;g} → {tʃ;dʒ} / \_i . In a product-oriented grammar, the facts could be accounted for with the conditional schema ‘if the last consonant of the

plural is {k;g;tʃ;dʒ}, the plural ends in -i'. Under either account, the crucial point is that the learner does not infer the most specific generalization possible (contra the Subset Principle for phonology, e.g., Berwick 1986, Dell 1981, Hale & Reiss 2003, 2008, Langacker 1987). Under the product-oriented account, the learner notices that alveopalatals are followed by -i in the plural and generalizes that velars would as well. Under the source-oriented account, s/he notices that velars take -i and generalizes that so would alveopalatals.

To summarize, alveopalatals take -i regardless of the language to which the subjects are exposed because alveopalatals and velars are grouped into the same category. These data do not provide exclusive support for product-oriented generalizations and are compatible with a source-oriented grammar. The data are incompatible with the Subset Principle. Whether the grouping of alveopalatals into the same category as velars is due to prior similarity between velars and alveopalatals in the minds of English speakers or to the fact that velars and alveopalatals are allophones in the artificial languages used in the present experiment remains a matter for future research.

The underuse of -i with velar-final singulars can be attributed to a bias against stem changes. Attaching -i to a velar requires the subject to face a choice between changing the stem or producing a form ending in [ki] or [gi], which s/he knows to be suboptimal. Therefore, attaching -a may be the safest option. On the other hand, attaching -i to an alveopalatal does not require stem changes and is thus preferred over attaching -a.

#### 3.3.1.4. A model of source-oriented generalization

The participants in the present experiment seem to have extracted a set of source-oriented generalizations that can be captured by a model that would induce rules from a lexicon and would weight the rules depending on how much statistical support the lexicon provides for each rule. One such model is the Minimal Generalization Learner developed by Albright & Hayes (2002, 2003). The model starts with a set of morphologically related word pairs as in (1).

- (1)    mot    mota  
       mok    moka  
       drug   drudzi  
       krug   krudzi

For each word pair, the model creates a word-specific rule as in (2).

- (2)    []→a/mot\_  
       []→a/mok\_  
       g→dzi/dru\_  
       g→dzi/kru\_

Then, rules that involve the same change are combined. Contexts in which the same change, e.g., []→i, happens are compared by matching segments starting from the location of the change. If segments match, they are retained in the specification of the

context for the change and the pair of segments further away from the change is compared. When this comparison process reaches the nearest pair of segments that do not match, the phonological features they share are extracted and retained in the specification of the context. Segments that are further away from the location of the change than the closest pair of non-matching segments are not compared and are replaced by a free variable in the specification of context.

For instance, the rules in (3) are combined into the rule in (4). Since the change involves the end of the stem, comparison starts from the end. The last segments in the context are both /u/, so they are retained and preceding segments are compared. Since the preceding segments are both /r/, they are retained as well and comparison proceeds to the preceding segment. These segments do not match but they are the closest pair of segments to the change that doesn't match, so the matching features are retained in the rule.

(3)  $g \rightarrow d\zeta i / dru\_$

$g \rightarrow d\zeta i / kru\_$

(4)  $g \rightarrow d\zeta i / [+cons; -cont; -son; -Labial] ru\_$

The resulting more general rules are then compared to each other and even more general rules derived if the same change occurs in multiple contexts, eventually resulting in quite general rules, such as  $[] \rightarrow i / C\_$ . However, all rules are retained in the grammar. Instead of removing non-maximally-general rules from the grammar, the RBL weights each rule by its reliability. Reliability is defined as the number of words to which the rule



applies divided by the total number of words to which it *could* apply. For instance, the reliability of the rule in (4) is the number of words of the form in (5) that are derived from words with the shape in (6) divided by the total number of words with the shape in (6) in the lexicon.

(5) [+cons;-cont;-son;-Labial]rudzi

(6) [+cons;-cont;-son;-Labial]rug

A reliable rule is more likely to apply to a novel word than a less reliable rule. For instance, if the rule in (7) is more reliable than the rule in (8), and these are the only rules that can apply to the singular [dig], the plural is more likely to be /didzi/ than /diga/.

(7)  $Vg \rightarrow Vd_{zi}$

(8)  $Vg \rightarrow Vga$

After being applied to the two artificial languages used in both experiments that were presented to the human participants in the present experiments, the model extracted the expected generalizations and assigned them the reliabilities shown in Table 3.2. As expected, the reliability of ‘just add –i’ ( $[ ] \rightarrow i / C_{[-cont]} \_$ ) is higher in Language II than in Language I. Therefore, the rate of velar palatalization before –i is correctly predicted to be lower in Language II than in Language I and lower in Language IV than in Language III under this source-oriented model. In addition, it is important to note that this difference in palatalization rates comes about because a learner exposed to Language II is

expected to attach –i to velars without changing the velar more often than a learner who is exposed to Language I, not because a learner exposed to Language II is expected to change velars into alveopalatals less often than a learner exposed to Language II hence the results in Figures 3.10-3.11.

Table 3.2. The generalizations extracted by the Minimal Generalization Learner from the first two languages presented to human participants in the present study.

	Language I	Language II
k→tʃi/V_	0.85	
g→dʒi/V_		
[ ]→i/C <sub>[-cont]</sub> _	0.18	0.57
[ ]→a/C <sub>[-cont]</sub> _	0.57	0.18

In the language with which the subjects are presented, –a is never attached to a velar-final singular. However, it is attached to a velar around 45% of the time in the output of participants exposed to Language I, and 20-25% of the time in the output of participants exposed to Language II. This apparent discrepancy between training data and the learned system is predicted by the model, although at a rate that is lower than the rate observed in the data (the predicted rate is 37% for Language I but only 10% for Language II). The difference between observed and predicted rates appears to be explainable by the bias against stem changes observed during training: if the subject chooses to attach –i, s/he is faced with the choice of changing the stem, which is difficult, or leaving it unchanged, resulting in a suboptimal product, hence –i may be avoided in favor of –a.

It should be noted that the probability of applying a rule in this model is taken to be the ratio of its reliability to the sum of reliabilities of all applicable rules. This is not the only possible choice rule, nor the choice rule that maximizes the probability of matching the input (Hudson Kam & Newport 2005, Norris & McQueen 2008).

Alternatively, the learner who has extracted the rules in Table 3.2 and weighted them as shown could always apply the most reliable applicable rule. In that alternative model, the rate of velar palatalization would be predicted to be identical across the two languages because the most reliable rule that can apply to a velar-final singular is the palatalizing rule in both languages. Therefore, for the source-oriented model to predict a difference between Language I and Language II (or between Language III and Language IV), the choice between competing rules must be stochastic in nature.

The Minimal Generalization Learner predicts no difference between Languages I and II on the one hand, and Languages III and IV on the other. The reason for this null prediction is minimal generalization. The addition of these examples leads to the development of a specific rule that attaches –i to alveopalatals and no other final segments. However, as we have seen in Section 3.4, alveopalatals and velars are classified into a single category by the subjects in the present study. From the perspective of the Minimal Generalization Learner, the subjects overgeneralize from velars to alveopalatals and from alveopalatals to velars. It is only when alveopalatals and velars are coded as being identical that the observed difference between Languages I and II on the one hand and Languages III and IV on the other is produced.

An alternative to considering velars and alveopalatals identical is presented by the emerging framework of Bayesian learning (see Kruschke, 2008, for a review). The

Minimal Generalization Learner assumes that the output of word recognition is a single word, about whose identity the learner is certain. Thus the learner compares across pairs of words that are perceived perfectly. In this framework, it is not clear why the learner should overgeneralize, rather than coming up with the narrowest generalization possible (in fact, narrowest possible generalizations have important theoretical advantages with regard to language learnability, see Berwick 1986, Dell 1981, Hale & Reiss 2003, 2008).

In a Bayesian framework, however, the output of perception is a distribution of probabilities across possible percepts. Thus, one may not be completely certain that the presented word was, for instance [bupi] and not [buki] even if one reports hearing [buki]. In order to report what one has just heard, one needs to make a decision and choose one possible percept. It is reasonable to assume that the perceiver reports the percept s/he considers to be the most probable given the evidence and his/her prior expectations (e.g., Norris & McQueen 2008). However, other percepts may only differ in probability from the most probable percept by a little. Traditional approaches to learning assume that the only percept whose probability level is incremented as a result of perception and the only percept that is associated with other stimuli occurring in the environment at the same time is the most probable percept, the percept reported by the learner. Bayesian learning assumes that the probabilities (and thus possibly acceptability values) of less probable percepts are incremented as well and that these less probable percepts can be associated with stimuli that are perceived at the same time.<sup>27</sup>

---

<sup>27</sup> It is important to note that the probability of the hypothesis given the data is equal to the prior probability of the hypothesis (i.e., the prior bias for or against the hypothesis) multiplied by the probability of the data given the hypothesis. For the probability of the hypothesis given the data to be above zero, it is necessary for the probability of the data given the hypothesis to be above zero, no matter how strong the bias in favor of the hypothesis is. Therefore, for the learner to believe that the probability of [ki] given the training data to be above zero, s/he must believe that the probability of hearing some of the acoustic signals presented during training given that the intended production was [ki] is above zero.

Thus, a learner who hears [bupi] and briefly considers [buki] as a possible lexical match for the acoustic stimulus s/he is hearing may increment the goodness of [buki] as a result of this experience despite not having settled on it as the most probable percept. Since [VpV] and [VbV] are acoustically similar to [VkV] and [VgV] respectively, it is expected that the learner will generalize from hearing examples of [Vpa] and [Vba] or [Vpi] and [Vbi], and consider the never presented plurals ending in [ki] and [ka] to be partially acceptable, as we have observed. On the other hand, [Vpa] and [Vpi] are relatively dissimilar, which means that [Vpi] would not be considered a probable stimulus when [Vpa] is presented and vice versa. Thus, as we observed, the learner should be able to match the probabilities of –i and –a after labials and coronals relatively easily but matching the distribution of –i and –a across consonants with different place values should be relatively difficult. This is consistent with the apparent overuse of –a with velar-final singulars.

Finally, we have noted that the stem of the singular is sometimes made to fit the stem of the plural during training. It is likely that these overt errors are just the tip of the iceberg and that subconsciously the learner is often not completely certain of having perceived, e.g., [buk]-[butʃi] and not [butʃ]-[butʃi]. That is, while hearing [buk]-[butʃi], the learner allocates some probability mass to [butʃ]-[butʃi], increasing the estimated goodness of [butʃ]-[butʃi]. In this way, alveopalatal-final singulars may become associated with –i as a result of presentation of velar-final singulars that correspond to plurals ending in an alveopalatal followed by –i. On the other hand, hearing [butʃ]-[butʃi] should not improve the goodness of [buk]-[butʃi] as much because the singular stem is unlikely to dissimilate from the plural and word-final [k] and [tʃ] are not acoustically

similar (although they do appear to be categorized into the same category by the learners). In fact, hearing [butʃ]-[butʃi] should prime the pair, making it a stronger competitor. When [buk]-[butʃi] is subsequently presented, it would then receive a lower probability relative to [butʃ]-[butʃi] than if [butʃ]-[butʃi] had not been presented. Thus, presentation of [buk]-[butʃi] would lead to a smaller increment in its acceptability if [butʃ]-[butʃi] has been recently experienced.

This hypothesis provides a possible explanation for the difference between Languages I and II and Languages III and IV as well as for the apparent grouping of velars and alveopalatals into a single category. Under the Bayesian approach, the generalization from one segment to another happens not because the segments are perceived to be identical but because they are similar enough for the ‘borrowing’ segment to be a plausible alternative to the ‘lender’ segment at a time when the borrowed property and the lender segment are perceived.

### 3.3.2. Plural likelihood rating

We now turn to the results of the rating task. In this task the subjects were presented with a singular-plural pair and asked “how likely is this plural to be the right plural for this singular?” The subject rated the likelihood on a five-point scale ranging from “impossible” (1) to “very likely” (5).

Figure 3.17 shows mean ratings given to the various singular-plural mappings by subjects exposed to one of the four possible languages. The error bars show 95% confidence intervals. Thus if the error bars on two columns do not overlap, we can be confident with  $p < .05$  that the mean ratings shown by the columns are different.

Figure 3.17 indicates that in Languages I, III, and IV the subjects judged velar palatalization to be significantly more likely than alveolar palatalization (in Language II the numerical trend is in the same direction). In all languages,  $\{tʃ;dʒ\} \rightarrow \{k;g\}a$  was judged to be less acceptable than  $\{k;g\} \rightarrow \{k;g\}a$ , and in Language II and Language IV  $\{tʃ;dʒ\} \rightarrow \{k;g\}i$  was judged to be less acceptable than  $\{k;g\} \rightarrow \{k;g\}i$ . These findings indicate that subjects did not judge the plural forms with no reference to the singular. Rather the subjects appear to be aware at least of the fact that singulars ending in a velar are not as likely to retain the consonant in the plural as singulars ending in an alveolar or an alveopalatal.

In Language II and Language IV, plurals ending in  $\{t;d\}i$  are rated higher than plurals ending in  $\{t;d\}a$ . The reverse is true for Language III and Language IV. This finding mirrors the effect in elicited production and reflects the statistics of the training data. In addition, in agreement with the finding that learners in all conditions tend to attach  $-i$  to alveopalatals, learners exposed to any language rate  $\{tʃ;dʒ\} \rightarrow \{tʃ;dʒ\}i$  mappings higher than  $\{tʃ;dʒ\}a$  mappings. Plurals ending in  $\{tʃ;dʒ\}i$  that are derived from alveolars or velars are also rated higher than plurals ending in  $\{tʃ;dʒ\}a$  that are derived from the same source. Finally, in Language II and Language IV forms featuring velar palatalization are rated higher than forms featuring the attachment of  $-a$  to a velar-final singular.

Figure 3.17. Mean ratings of singular-plural mappings across the four languages. The mappings are (from left to right):  $\{tʃ;dʒ\} \rightarrow \{k;g\}a$ ,  $\{tʃ;dʒ\} \rightarrow \{k;g\}i$ ,  $\{tʃ;dʒ\} \rightarrow \{tʃ;dʒ\}a$ ,  $\{tʃ;dʒ\} \rightarrow \{tʃ;dʒ\}i$ ,  $\{k;g\} \rightarrow \{tʃ;dʒ\}a$ ,  $\{k;g\} \rightarrow \{tʃ;dʒ\}i$ ,  $\{k;g\} \rightarrow \{k;g\}a$ ,  $\{k;g\} \rightarrow \{k;g\}i$ ,  $\{t;d\} \rightarrow \{tʃ;dʒ\}a$ ,  $\{t;d\} \rightarrow \{tʃ;dʒ\}i$ ,  $\{t;d\} \rightarrow \{t;d\}a$ ,  $\{t;d\} \rightarrow \{t;d\}i$ . The error bars show 95% confidence intervals.

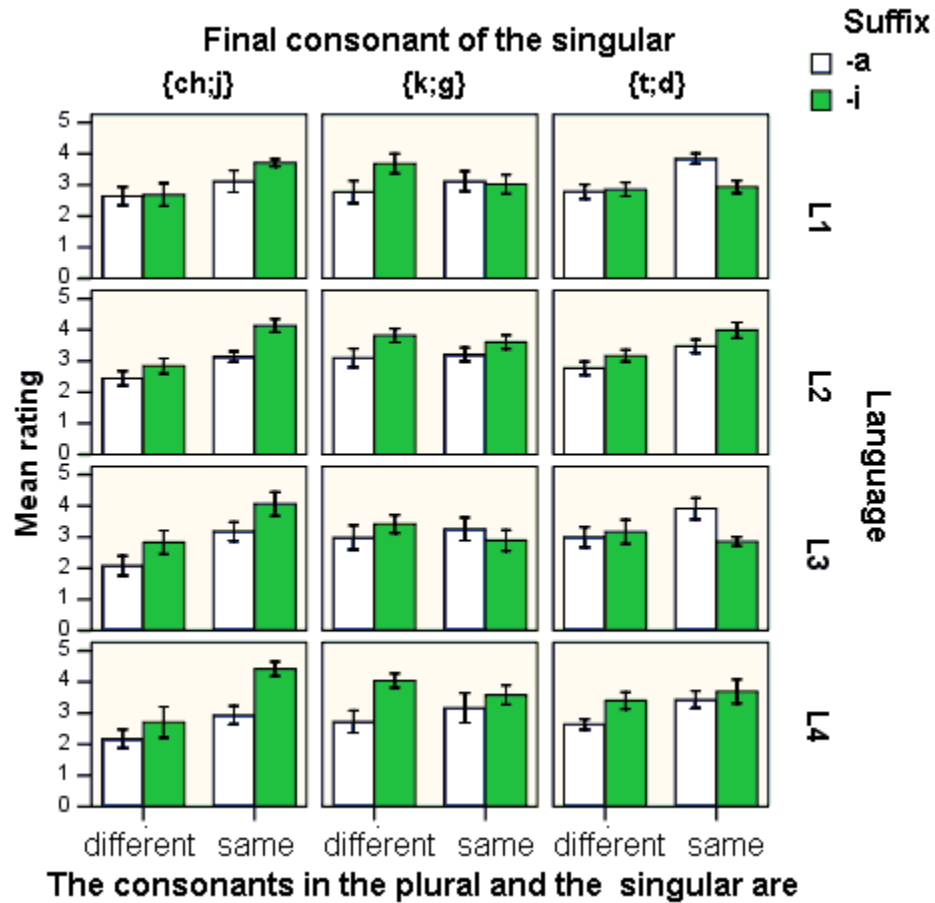


Figure 3.18 shows the major significant difference between subjects exposed to different languages. The addition of  $\{tʃ;dʒ\} \rightarrow \{tʃ;dʒ\}i$  examples to training, which distinguished languages 3 and 4 from languages 1 and 2 significantly decreased likelihood judgments for  $\{t;d\} \rightarrow \{t;d\}i$  mappings relative to  $\{t;d\} \rightarrow \{tʃ;dʒ\}i$  mappings



( $p=.01$  according to the Wilcoxon test). The increase in the number of  $\{t;d\} \rightarrow \{t;d\}i$  examples, which distinguished languages 2 and 4 from languages 1 and 3 increased likelihood judgments for  $\{t;d\} \rightarrow \{t;d\}i$  mappings relative to  $\{t;d\} \rightarrow \{tʃ;dʒ\}i$  mappings ( $p=.02$  according to the Wilcoxon test). Thus, examples of alveolar palatalization are most acceptable relative to retaining the alveolar in the plural in perception in Language 3, followed by Language 2 and Language 4, followed by Language 1. This is the exact opposite ranking from the one observed for velar palatalization in production and strongly suggests the use of product-oriented generalizations. While it is tempting to conclude that there is a difference between production and perception, with perception being more product-oriented than production, there is in fact a non-significant trend for alveolar palatalization to be favored by the addition of examples of  $tʃ \rightarrow tʃi$  in production as well ( $p=.11$ ), thus it is also possible that there is a difference between the effect of examples of  $tʃ \rightarrow tʃi$  on velar and alveolar palatalization.<sup>28</sup>

---

<sup>28</sup> In fact, as we shall see, the effect on alveolar palatalization reaches significance if the data from both training paradigms are pooled, thus this is the more likely explanation.

Figure 3.18. Mean standardized likelihood judgment for mappings  $\{t;d\} \rightarrow \{t;d\}i$  minus mean standardized likelihood judgment for  $\{t;d\} \rightarrow \{tʃ;dʒ\}i$  mappings across the four languages.

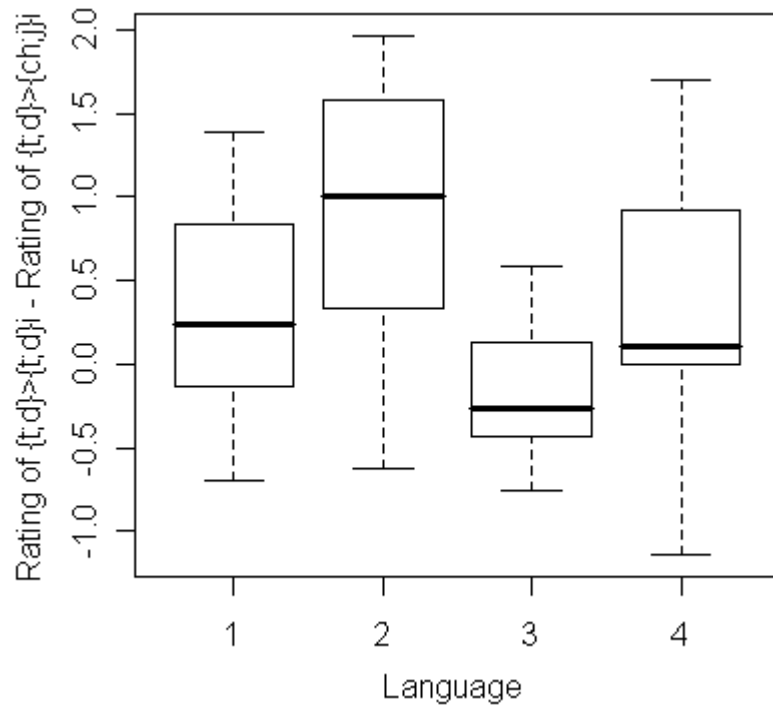


Table 3.3 shows correlations between standardized ratings of plural forms that simply attach  $-i$  to a velar, plural forms that display velar palatalization, plural forms that show attachment of  $-i$  to an alveopalatal, and plural forms featuring attachment of  $-i$  to an alveolar, with or without palatalization. All significant correlations are positive. Since the ratings were standardized (converted to Z scores by subtracting the subject's mean from each score and dividing the difference by that subject's standard deviation), this result is not due to differences in means or standard deviations across subjects. Unlike in elicited production, favoring the attachment of  $-i$  to alveolars does not correlate with disfavoring velar palatalization relative to simply attaching  $-i$  to a velar. Rather subjects

tend to favor or disfavor all (legal) plurals ending in –i (although plurals ending in –i do appear to be subdivided into plurals ending in a stop followed by –i and plurals ending in {tʃ;dʒ}i). This finding suggests that input-output mappings differing in palatalization do not compete with each other in perception/rating as much as they do in production.

Table 3.3. The correlations between ratings of singular plural mappings featuring the suffix –i.

		{t;d} → {t;d}i	{k;g} → {tʃ;dʒ}i	{t;d} → {tʃ;dʒ}i	{tʃ;dʒ} → {tʃ;dʒ}i
{k;g} → {k;g}i	R	.32	.00	-.16	.17
	p	.04	.98	.31	.28
{t;d} → {t;d}i	R		.35	.01	.12
	p		.02	.95	.64
{k;g} → {tʃ;dʒ}i	R			.3	.45
	p			.05	.002
{t;d} → {tʃ;dʒ}i	R				.29
	p				.05

Tables 3.4-3.6 show all correlations between likelihood ratings of individual input-output mappings that may be expected to compete. The tables show that negative correlations between ratings of singular-plural mapping types tend to occur when the mappings do not share the plural suffix. Table 3.5 shows that acceptability of plurals ending in –{k;g}a positively correlates with acceptability of plurals ending in –{tʃ;dʒ}a

and the acceptability of plurals ending in  $-\{k;g\}i$  correlates positively with the acceptability of plurals ending in  $-\{tʃ;dʒ\}i$ . However, if a subject considers  $-i$  likely, s/he tends to consider  $-a$  less likely.

Table 3.4. Correlations between likelihood ratings of  $\{t;d\} \rightarrow X$  mappings.

		$\{t;d\} \rightarrow \{t;d\}i$	$\{t;d\} \rightarrow \{tʃ;dʒ\}a$	$\{t;d\} \rightarrow \{tʃ;dʒ\}i$
$\{t;d\} \rightarrow \{t;d\}a$	R	-.48	.02	-.6
	p	.009	.9	.00002
$\{t;d\} \rightarrow \{t;d\}i$	R		-.54	.01
	p		.0002	.95
$\{t;d\} \rightarrow \{tʃ;dʒ\}a$	R			.18
	p			.23

Table 3.5. Correlations between likelihood ratings of  $\{k;g\} \rightarrow X$  mappings.

		$\{k;g\} \rightarrow \{k;g\}i$	$\{k;g\} \rightarrow \{tj;d3\}a$	$\{k;g\} \rightarrow \{tj;d3\}i$
$\{k;g\} \rightarrow \{k;g\}a$	R	-.15	.30	-.55
	p	.35	.04	.00009
$\{k;g\} \rightarrow \{k;g\}i$	R		-.29	.005
	p		.06	.98
$\{k;g\} \rightarrow \{tj;d3\}a$	R			-.4
	p			.007

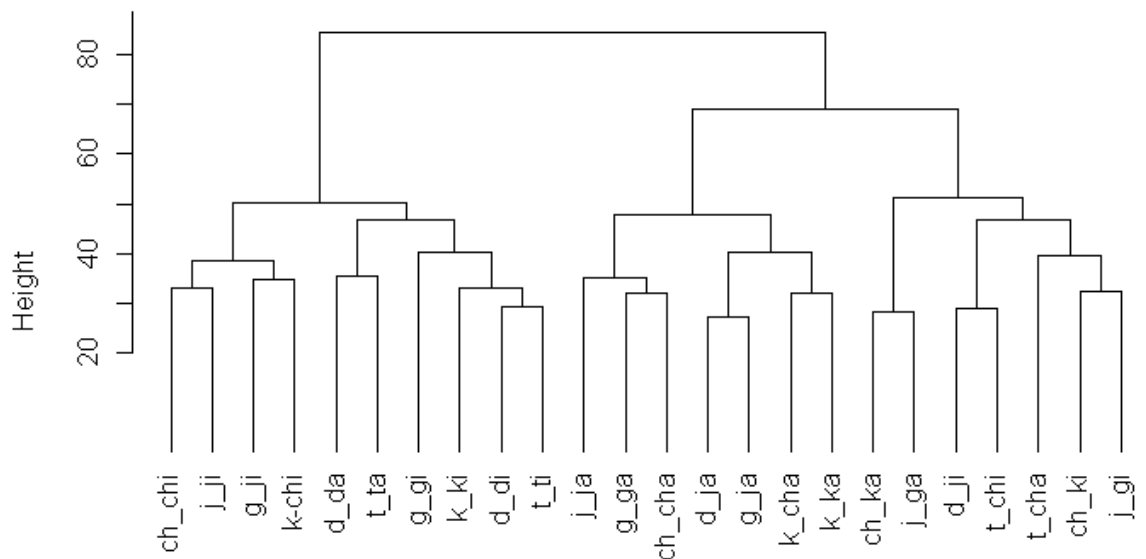
Table 3.6. Correlations between likelihood ratings of  $\{tj;d3\} \rightarrow X$  mappings.

		$\{tj;d3\} \rightarrow \{tj;d3\}i$	$\{tj;d3\} \rightarrow \{k;g\}a$	$\{tj;d3\} \rightarrow \{k;g\}i$
$\{tj;d3\} \rightarrow \{tj;d3\}a$	R	-.44	.11	-.55
	p	.003	.49	.0001
$\{tj;d3\} \rightarrow \{tj;d3\}i$	R		-.59	-.03
	p		.00002	.82
$\{tj;d3\} \rightarrow \{k;g\}a$	R			.2
	p			.19

Figure 3.19 shows a hierarchical clustering solution based on the similarities in the subjects' reactions to the various singular-plural mappings. In this graph, the higher the horizontal connection between two singular-plural mappings, the more similarly they

were treated by the subjects. The left cluster at the top level contains mappings that are legal in all four languages plus the  $\{k;g\} \rightarrow \{k;g\}i$  mapping, which violates velar palatalization and tends to be accepted by subjects who accept  $\{t;d\} \rightarrow \{t;d\}i$  mappings. This is consistent with the results of elicited production where the subjects who attached  $-i$  to nonvelars were the subjects who were likely to violate velar palatalization. Interestingly, the mappings that carry out velar palatalization ( $\{k;g\} \rightarrow \{tʃ;dʒ\}i$ ) are grouped together with mappings that result in the same product without changing the consonant ( $\{tʃ;dʒ\} \rightarrow \{tʃ;dʒ\}i$ ).

Figure 3.19. A hierarchical clustering analysis of the possible singular-plural mappings based on the subjects' ratings of the mapping likelihoods.<sup>29</sup>



To summarize the results, subjects exposed to singular-plural mappings during training extract both source-oriented and product-oriented generalizations. Thus, despite

<sup>29</sup> The clustering solution is based on the coordinate matrix of the output of principal components analysis with centering and scaling that used subjects as dimensions. Clustering was done using manhattan distance, since subjects are independent non-interacting dimensions, and average clustering method; Ward clustering, Mcquitty clustering, and complete clustering yield the same solution.

the source-oriented training, the subjects use product-oriented generalizations to evaluate the acceptability of a plural given a singular in perception. In the perceptual task of likelihood judgment, the likelihood that a given plural is the right plural for a given singular is evaluated by not only 1) examining whether the singular was likely to be mapped onto this plural, resulting in restrictions on singulars that can be mapped onto a certain plural, but also 2) judging the shape of the plural as typical or atypical in the language. Thus, the addition of  $-i$  to singulars ending in  $\{tʃ;dʒ\}_i$  increases likelihood ratings of all plurals ending in  $\{tʃ;dʒ\}_i$  regardless of the shape of the singular they are derived from. In production, there is a trend for the addition of examples of  $\{tʃ;dʒ\} \rightarrow \{tʃ;dʒ\}_i$  to increase likelihood ratings of  $\{t;d\} \rightarrow \{tʃ;dʒ\}_i$  mappings but the likelihood ratings of  $\{k;g\} \rightarrow \{tʃ;dʒ\}_i$  are significantly decreased. Thus there is some evidence that learners rely on source-oriented generalizations in production more than they do in perception / likelihood rating.

### 3.4. Conclusion

In the experiments reported in this chapter, participants were presented with one of four artificial languages in a learning situation that was expected to favor source-oriented generalizations: the participants were presented with singular plural pairs, and the number of words was so high that individual plural forms were not learned (no differences in accuracy between forms presented during training and forms not presented during training were found). The participants were found to extract source-oriented generalizations that competed with each other and were weighted by the evidence that supported them in the lexicon. The choice between the extracted generalizations was

shown to be stochastic in nature, with the participants obeying the competing generalizations in proportion to their reliability or type frequency values. Finally, participants were shown not to obey the Subset Principle in that they did not extract the most specific generalizations possible. This behavior was shown to be explainable if the output of perception is not a single percept but rather a distribution of probabilities across possible percepts, with the acceptability of each of the possible percepts being incremented in proportion to the estimated probability that the percept has been presented.

In perception, the subjects were found to pay more attention to the shape of the plural than to the shape of the singular. In training, the subjects were much more likely to make the stem of the singular be identical to the stem of the plural than to make the stem of the plural fit the stem of the singular. In rating how likely a certain plural was to be the right plural for a given singular, the listeners evaluated the typicality of the shape of the plural for the language they were exposed to. An increase in plural form typicality outweighed an equivalent decrease in the typicality of the singular-plural mapping for the purposes of likelihood rating but not for production. I would like to argue that perception is inherently more product-oriented than production. In perception, the listener need not recreate the unfamiliar perceived form from some morphologically related base, applying the source-oriented generalizations s/he would use to produce the novel form. Rather, s/he needs only to activate morphologically related forms in the lexicon. It is tempting to speculate that the product-oriented nature of perception might lead to the evolution of an originally source-oriented system into a product-oriented one, since violations of source-oriented generalizations are likely to be accepted by the listener as long as they obey



product-oriented generalizations. Thus, over time, restrictions on the types of inputs that are mapped onto an output may be relaxed, which would make product-oriented generalizations even stronger relative to source-oriented generalizations.<sup>30</sup>

In the following chapter, I turn to a product-oriented paradigm in which participants are presented with a small number of forms occurring one at a time where each form is presented a large number of times and remembered by subjects. In the final chapter preceding the general conclusion, velar palatalization in a natural language, Russian is considered.

---

<sup>30</sup> To pursue this idea further, it would be interesting to conduct an iterated learning experiment (Kalish et al. 2007, Kirby et al. 2008) where the singular-plural mappings produced by  $\text{Generation}_{n-1}$  would be filtered by eliminating the forms rated as relatively unnatural and given to  $\text{Generation}_n$  for learning, with the process iterated as needed.

## APPENDIX 1. INSTRUCTIONS FOR THE SOURCE-ORIENTED PARADIGM.

### Instructions for the training

You will be presented with singular-plural pairs from a foreign language.

For instance,



bup



bupi

Please, repeat the pair aloud.

In this example, you would say 'bup-bupi'.

Then click the mouse to continue to the next pair.

Click anywhere to begin.

## Instructions for elicited production

You will now see some more creatures and hear their names.

For instance, you may see that



is called 'wug'.

You will then see a picture of two or more of such creatures.



Say what they are called in the language you have been learning.

Use your knowledge of the made-up language to form the plural.  
Do NOT use English plurals.

Once you make your response click anywhere to continue.  
Click anywhere to begin.

## Instructions for naturalness rating

You will now hear singular-plural pairs.

Some of them contain the correct plural for the word.  
Some do not.

Your task is to tell whether the plural you've just heard  
is likely to be the right plural for the word.

Please respond by clicking the green squares.

Click to continue.

Display during the rating task

**How likely is this to be the right plural for this singular?**

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
<b>1=impossible</b>				<b>5=very likely</b>

## APPENDIX 2: LIST OF STIMULI

Stimuli presented for training had the shape C(C)(C)VC  $\rightarrow$  C(C)(C)VC{i;a}.

There was a limited set of bodies, from which the full stimuli were formed. The table below shows the bodies and the singular-plural ending pairs that the bodies could carry in the four languages ('ALL' means that the singular-plural pair was featured by all languages).

C(C)(C)V	k $\rightarrow$ tʃi g $\rightarrow$ dʒi	t $\rightarrow$ ti d $\rightarrow$ di p $\rightarrow$ pi b $\rightarrow$ bi	t $\rightarrow$ ta d $\rightarrow$ da p $\rightarrow$ pa b $\rightarrow$ ba	tʃ $\rightarrow$ tʃi dʒ $\rightarrow$ dʒi
skrou	ALL	L1, L3	L2, L4	L3, L4
strou	ALL	L2, L4	L1, L3	L3, L4
drɪ	ALL	L1, L3	L2, L4	L3, L4
pɪ	ALL			L3, L4
flɛ	ALL	L2, L4	L1, L3	L3, L4
frɑ	ALL	L2, L4	L1, L3	L3, L4
spri	ALL	L2, L4	L1, L3	L3, L4
kæ	ALL			L3, L4
bu	ALL	L2, L4	L1, L3	L3, L4
dʒɑ	ALL			L3, L4
deɪ	ALL	L2, L4	L1, L3	
stou	ALL			
dræ	ALL			
kru	ALL			
streɪ	ALL	L2, L4	L1, L3	

Testing stimuli that did not appear during training:

	k/g	t/p/b/d	tʃ/dʒ
dre	+		
bli	+		
ma	+		
spraɪ	+		
spæ	+		
waɪ	+		
sni	+		
plou	+		
dæ	+		
gleɪ	+		
drou		+	
ble		+	
mi		+	
spru		+	
spi		+	
sai			+
wɪ			+
pli			+
fre			+
gla			+

# Rating stimuli

	k → ki	g → gi
	k → ka	g → ga
	k → tʃi	g → dʒi
	k → tʃa	g → dʒa
	t → ti	d → di
	t → ta	d → da
	t → tʃi	d → dʒi
	t → tʃa	d → dʒa
	tʃ → ki	dʒ → gi
	tʃ → ka	dʒ → ga
	tʃ → tʃi	dʒ → dʒi
	tʃ → tʃa	dʒ → dʒa
dæ	+	
fru		+
kwa		+
lu	+	
sleɪ	+	
wæ		+

### APPENDIX 3. FEATURES USED FOR TRAINING THE MODEL

Segment	son	lab	pal	dors	voice	cont	lax
b	0	1	0	0	1	0	-1
d	0	0	0	0	1	0	-1
g	0	0	0	1	1	0	-1
p	0	1	0	0	0	0	-1
t	0	0	0	0	0	0	-1
k	0	0	0	1	0	0	-1
tʃ	0	0	1	1	0	0	-1
dʒ	0	0	1	1	1	0	-1
ʃ	0	0	1	1	0	1	-1
s	0	0	0	0	0	1	-1
f	0	1	0	0	0	1	-1
v	0	1	0	0	1	1	-1
w	2	1	0	0	1	1	-1
z	0	0	0	0	1	1	-1
r	2	0	0	0	1	1	-1
l	2	0	0	0	1	0	-1
m	1	1	0	0	1	0	-1
n	1	0	0	0	1	0	-1
ɑ	5	0	0	1	1	1	1
ɛ	4	0	1	1	1	1	1
i	3	0	1	1	1	1	0
ɪ	3	0	1	1	1	1	1
oʊ	4	1	0	1	1	1	0
u	3	1	0	1	1	1	0
eɪ	4	0	1	1	1	1	0
ʊ	3	1	0	1	1	1	1
ʌ	4	0	0	1	1	1	1
æ	5	0	1	1	1	1	1



## CHAPTER 4

### GENERALIZATION IN A PRODUCT-ORIENTED PARADIGM

Experiments I-II presented in the previous chapter have employed a training paradigm that has several features that distinguish it from natural language acquisition that may be expected to favor source-oriented generalizations: 1) the learners are presented with singular-plural pairs, 2) the learners are told that they will be tested on plural formation, and 3) each word is presented only twice, hence the learners are unlikely to memorize the words. All of these features favor source-oriented generalizations.

In natural language acquisition, 1) most plurals are memorized and can be retrieved from memory as wholes (Albright and Hayes 2003, Bybee 1985, 2001, Bybee and Moder 1983, Bybee and Slobin 1982, Dabrowska 2008, Halle 1973, Hay 2003, Hooper 1976a, Zuraw 2000), 2) learners are exposed to individual wordforms in sentences, rather than singular-plural pairs, and 3) a small number of words are experienced often (Zipf 1935). These features may make real languages favor product-oriented generalizations through making individual plural forms more prominent and relationships between singulars and plurals less salient (cf. Morgan et al. 1989 for transformational rules in syntax).

This chapter replicates the experiments reported in the preceding chapter using a paradigm that is closer to a natural language learning situation. In the remaining experiments, learners were trained by 1) presenting individual wordforms (singular or plural) in sentences, 2) asking subjects to learn words, without telling them that they will

be tested on plural formation, 3) presenting each word multiple times so that individual plural forms are remembered, and 4) testing lexical knowledge prior to testing plural formation.

If these modifications lead to a switch from the use of source-oriented generalizations in production to the use of product-oriented ones, then we can say that characteristics of the learning situation that are typical for natural language acquisition favor the extraction of a product-oriented grammar from lexical data. If the modifications have no effect, and subjects still extract source-oriented generalizations, then further support for source-oriented grammar models such as Albright and Hayes (2003) is provided.


#### **4.1. The paradigm**

The present paradigm is based loosely on the paradigm developed by Peperkamp et al. (2006). In their study, subjects were exposed to short phrases paired with pictures. The first word in the phrase was one of two novel determiners, which the subjects are told mean ‘two’ and ‘three’. The second word was a noun. The subjects were asked to memorize the nouns. The test consisted of two stages. In the first stage, subjects were presented with pictures of nouns they were trained on and asked to supply the name. The second stage was an elicited production test of the same type as in our Experiments I-II. For the purposes of the present study, this paradigm was modified in two ways. First, the training-recall sequence was repeated twice so that subjects were certain they were being tested on recalling the nouns by the time of the elicited production test. Second, the elicited production task was followed by the likelihood rating task.

#### 4.1.1. Tasks

As in the previous experiments, each singular-plural pair was matched with a picture pair. However, pairings of singular nouns with objects and pairings of plural nouns with objects appeared in random order. The learner was asked to learn the names for the objects. The learner repeated the noun forms they were presented with. If the noun appeared in a sentential frame, only the noun needed to be repeated. The training task is shown schematically in Figure 4.1. Instructions are presented in Appendix 1.


Figure 4.1. The training task

Video:			
Audio:		[boutʃi]	
Learner action:	Watch	Watch & listen	
Duration:	600 ms	500-900 ms	500 ms

After going through the training set once, the learners were tested on recalling the object names by being presented with an object or a set of identical objects and asked for the corresponding noun form. They were instructed to produce the right form of the noun

(whether singular or plural). The instructions are shown in Appendix 1. The task is shown schematically in Figure 4.2.

Figure 4.2. The recall task.

Video:		
Audio:		
Learner action:	Say the right form of the noun, then click	
Duration:	0-10 s	500 ms

The training-recall sequence was repeated twice and then followed by the same generalization and rating tasks used in the previous experiments.

#### 4.1.2. Stimuli

The tricky part of conducting the present experiment is deciding on the number of nouns to be used and the number of times often each noun should be presented so that some nouns are learned and there is generalization to novel nouns and even nouns with new final segments. Peperkamp et al. (2006) used 12 nouns, each of which was presented either 8 or 16 times during training but they do not present any data on whether the nouns

were learned correctly.<sup>37</sup> However, the number of different words needed for generalization has been recently addressed in artificial grammar by Gerken (2006), Gerken and Bollt (2008), Needham et al. (2005), Quinn and Bhatt (2005), and Xu and Tenenbaum (2007), all of whom suggest that at least three different types are needed to generalize to new types with both sound (AAB vs. ABA syllable patterns, Gerken 2006, heavy syllables, Gerken and Bollt 2008), and object (Needham et al. 2005, Quinn and Bhatt 2005, Xu and Tenenbaum 2007) categories (see also Bybee (2001:121-124), Bybee and Pardo (1981) for natural language data). Therefore, I decided that at least three different words exemplifying each of the to-be-learned generalizations are needed, although, given the preference for the suffix that never triggers stem changes (-a) demonstrated by the subjects in Experiments I-II, I considered it unnecessary to present many different examples of the suffix -a. The resulting languages are shown in Table 4.1. The individual stimuli are shown in Appendix 2.

Table 4.1. The four languages presented to learners in Experiment IV

	Language I	Language II	Language III	Language IV
$k \rightarrow t\text{fi} / V\_$	4			
$[ ] \rightarrow i / \{p;t\}\_$	2	6	2	6
$[ ] \rightarrow a / \{p;t\}\_$	6	2	6	2
$[ ] \rightarrow i / t\text{f}\_$	0		4	

<sup>37</sup> Leaving aside for now the question of whether correct recall of presented nouns is necessary or even helpful for internalizing the regularities implicit in the set of nouns. I will come back to this issue in the discussion.

Goldberg et al. (2004) have shown that the learning of novel argument structure constructions is facilitated if a few of the verbs associated with a construction occur very often while the majority occur infrequently compared to a condition in which all verbs occur equally often. Goldberg (2006:85-89) reports that the same result also holds for dot pattern classification, suggesting that it is not a peculiarity of syntax (where the meaning of the construction might be gleaned off the meaning of the most frequent verb) and thus may also hold for morphophonology. Therefore, one word exemplifying  $k \rightarrow t\text{f}i$ , one word exemplifying the most frequent  $p \rightarrow pV$  pattern in each language, one word exemplifying the most frequent  $t \rightarrow tV$  pattern in each language, and one word exemplifying  $t\text{f} \rightarrow t\text{f}i$  were presented 42 times each, while the other words were presented 14 times each.

Recchia et al. (2008) exposed human learners to an artificial lexicon in which words differed in frequency and the number of different sentences and pictorial scenes they appeared in. They found that frequency of presentation influenced lexical decision only if the word appeared in multiple different contexts, i.e., it had high contextual diversity. Contextual diversity was increased in the present experiment by combining each word with multiple frames: each word could be inserted in the sentences ‘{That’s a; Those are the} \_\_\_\_’ and ‘{I am a; We are the} \_\_\_\_’, and also appeared on its own and produced in a scared voice, a normal voice, or a touched voice. In addition, a voice was created for each individual creature by manipulating the speed, shifting the formant ratio, the pitch median, and the pitch range of the original speaker (me) using the ‘Change gender’ function in Praat (Boersma & Weenink 2005). The individual creature voices were used for producing the utterances fitting the schema ‘{I am a; We are the}\_\_’. In

addition, for the frequent words, the isolated word productions were produced in four different creature voices each.

#### 4.1.3. Procedures

The procedures were the same as in the previous experiments.

#### 4.1.4. Participants

All of the participants reported being native English speakers with no history of speech, language, or hearing impairments. None reported being fluent in a foreign language. All participants were recruited from introductory psychology classes and received course credit for participation. None participated in the experiments reported in Chapter 3. Participants were assigned to languages randomly. There were 11 participants assigned to learn each language. One participant assigned to Language IV was subsequently excluded because of forming plurals using a pattern that was not presented in training (adding /tʃa/). One participant assigned to learn Language 1 was excluded from analyses of ratings because of computer error resulting in his ratings being lost.

## 4.2. Results I: Errors in recall

### 4.2.1. Overall accuracy of recall across languages

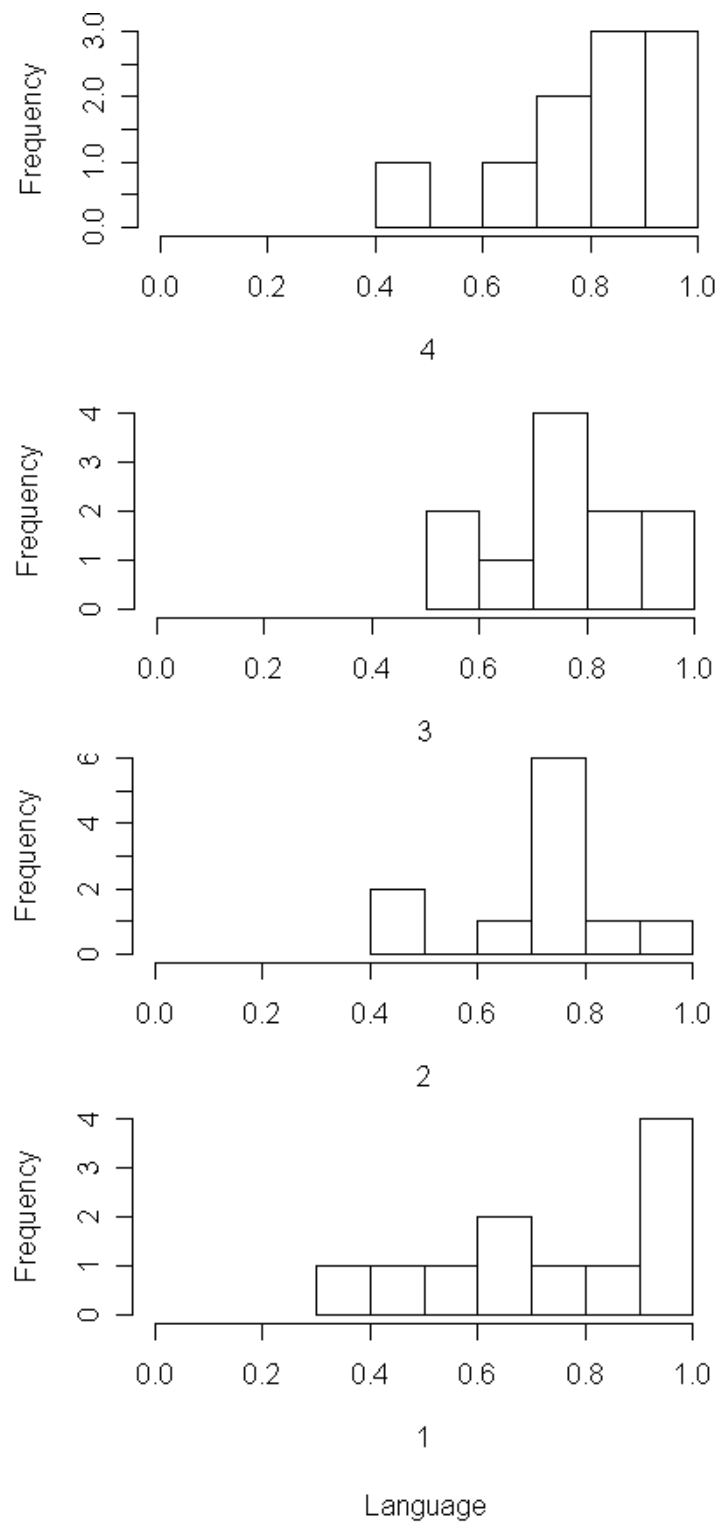
The distributions of recall accuracies for each of the four languages are shown in Figure 4.3. The groups of learners exposed to different languages do not differ significantly in overall accuracy of recall (all  $p > .3$  according to the Wilcoxon test). However, since learners presented with Language I or Language II were asked to recall 12 words while those exposed to Language III or Language IV were asked to recall 16 words, learners exposed to Language III or Language IV recall significantly more words than learners exposed to Language I or Language II ( $p = .0004$  according to the Wilcoxon test). Thus each word presented to a learner appears to have a certain probability of being recalled that is independent of how many words the learner is asked to remember, at least for the limited range of lexicon sizes examined in the present study.<sup>38</sup>

---

<sup>38</sup> The advantage of languages III and IV over languages I and II in number of words learned may be a spacing effect, since, on average, exposures to a wordform are further from each other in time in languages III and IV than in languages I and II.



Figure 4.3. The distributions of recall accuracies for each of the four languages. The individual accuracy values have been rounded to the nearest decimal.

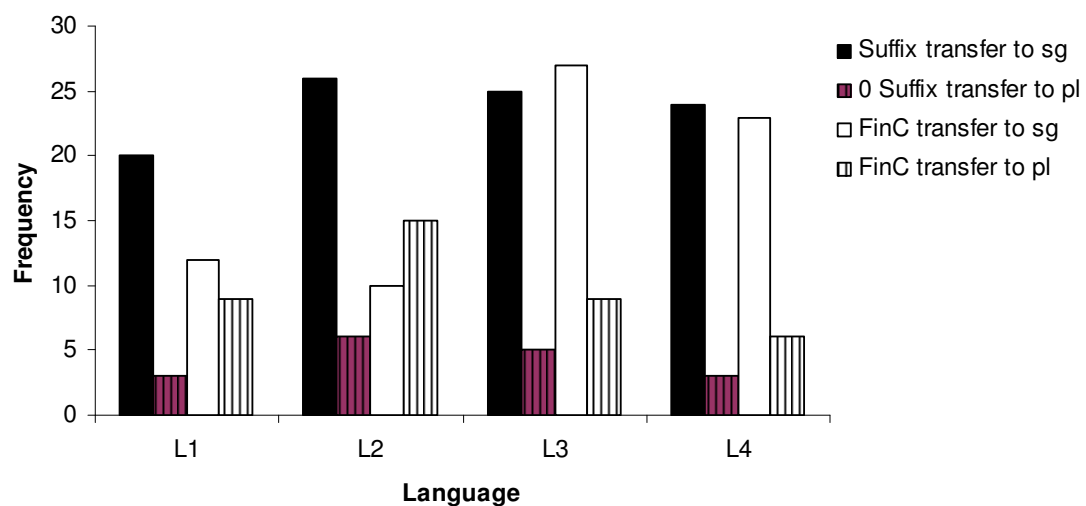


#### 4.2.2. Memorizing singulars vs. plurals

One of the things that the learner needs to learn is which wordforms are singular, and which are plural. Thus in recalling the name for a picture of a single object, the learner may erroneously use a wordform bearing a plural suffix and/or having a stem-final [tʃ] despite the fact that the actual wordform ended in [k] and had no suffix.

Similarly, in recalling the name for a picture of multiple objects, the learner may erroneously omit the plural suffix and/or produce a stem ending in [k] despite the fact that the plural form of the stem ends in [tʃ]. The actual numbers of occurrences of each of these error patterns across the four artificial languages are shown in Figure 4.4. As the figure shows, learners erroneously attach plural affixes to singular forms much more often than they erroneously fail to attach a plural affix to a plural form ( $\chi^2(1)=54.3$ ,  $p<.00001$ ). There is no significant effect of language on the advantage of erroneous suffix addition over erroneous suffix omission ( $\chi^2(3)=.46$ ,  $p=.93$ ).

Figure 4.4. Numbers of errors in recall where the suffix from the plural form is erroneously attached to the singular form (solid black bars), the plural suffix is erroneously omitted from the plural form (striped dark bars), the final consonant of the plural stem ([tʃ]) is erroneously produced in the singular (solid white bars), or the final consonant of the singular stem ([k]) is erroneously produced in the plural (striped white bars).



The finding that plural affixes are more likely to be attached to singular forms than be omitted from plural forms might be interpreted as suggesting that plural forms are more likely to be remembered correctly and serve as the base of the paradigm than singular forms. Albright (2005, 2008) predicts that this should be the case in Language I and Language II because derivation of singulars from plurals would be more successful in these languages than derivation of plurals from singulars: when deriving a plural from a singular in Language I or Language II, one faces uncertainty in whether the plural should end in *-i* or *-a*, while there is no uncertainty when the singular is derived from a plural (for plurals ending in a stop followed by a vowel, the singular can be formed by

simply deleting the vowel, while plurals ending in [tʃi] correspond to singulars ending in [k]). However, this hypothesis is contradicted by lack of a consistent preference for plural-to-singular stem-final consonant transfer. While such a tendency appears to be present in Language III and Language IV, it can be as easily explained by transfer from singulars ending in [tʃ], which are not presented in Language I and Language II, the two languages in which no preference for plural-to-singular final consonant transfer is observed.

Albright's (2005, 2008) hypothesis is also inconsistent with the presence of a preference for erroneous affix retention over erroneous affix omission in languages III and IV, because in these languages rule-based derivation of plurals from singulars should be easier than derivation of singulars from plurals. In these languages, the learner who decides to remember plurals and produce singulars by rule would face uncertainty in 8 words that end in [tʃi] and can be mapped onto singulars ending in [k] or [tʃ] with equal probability. The learner who decides to remember singulars and produce plurals from singulars would face uncertainty in the 8 remaining words that can be mapped onto plurals ending in either [i] or [a] but the probabilities of [i] and [a] are not equal in either language: there is a 75% probability of [i] in language IV and [a] in Language III. Thus, the formation of plurals from singulars would result in fewer errors than the formation of singulars from plurals. Thus, singulars should be more salient than plurals, and suffix omission should be more common than suffix insertion in Language III and Language IV.

Albright (2005, 2008) hypothesizes there should be a single base form in a paradigm. Thus, the learner may choose either plurals or singulars as bases to memorize but s/he may not memorize plurals for words that have a plural form that is more

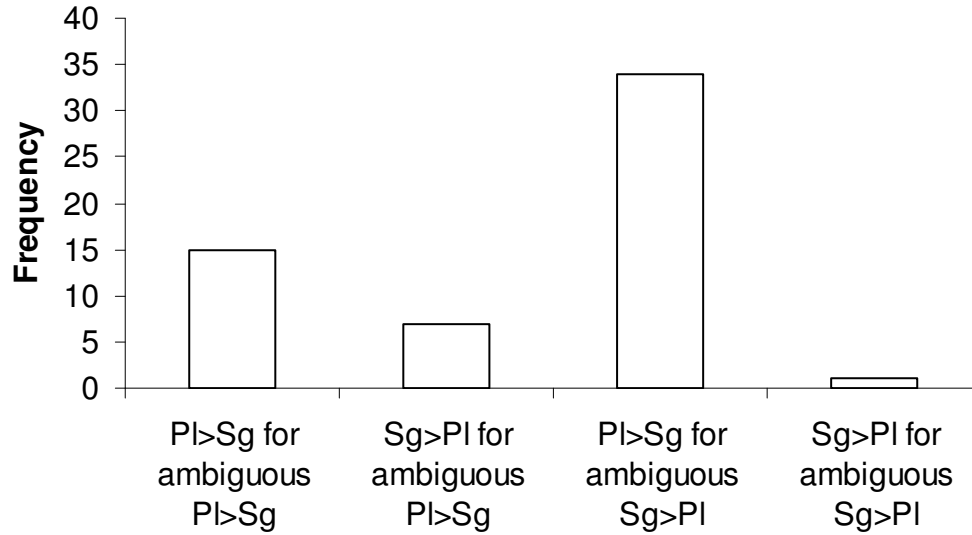
informative than the singular form and singulars for words that have a singular form that is more informative than the plural. This alternative strategy would be much more optimal for learners of languages III and IV than the strategy Albright favors. Thus, the learners could memorize [k]-final and [tʃ]-final singulars and {p;t}{a;i}-final plurals.<sup>39</sup> If the learners adopt this strategy, they should be most likely to erroneously add plural affixes to singulars ending in [p] or [t] and to erroneously omit affixes from plurals ending in {k;tʃ}i. This hypothesis is confirmed by the data in Figure 4.5: transfer from the plural form to the singular form is most likely when the formation of the singular from the plural is more likely to be accurate than the formation of the plural from the singular (comparing the difference between plural-to-singular transfers vs. singular-to-plural transfers for ambiguous singular-plural mappings, i.e., {p;t} → {p;t}{a;i} to the corresponding difference for ambiguous plural-singular mappings, i.e., tʃi → {k;tʃ},  $\chi^2(1)=15.2, p<.0001$ ).<sup>40</sup> Thus, the learners appear to pay most attention to the most informative wordforms within a morphological family.

---

<sup>39</sup> In fact, they could also just memorize the plurals that are exceptional in the language, i.e., {p;t}i plurals in Language III and {p;t}a plurals in Language IV and derive other plurals using the ‘add -i’ rule.

<sup>40</sup> This does not appear to be simply due to a difference between -i and -a, since -i is transferred from a {p;t}i plural to the singular 20 times, with -a being transferred from a {p;t}a plural 14 times ( $\chi^2(1)=1.06, p=.3$ ), and -i being transferred from a {k;tʃ}i plural 15 times. Thus, -i alone is transferred (non-significantly) more along the unambiguous mapping where -a has on average the same chance of being transferred at random than along the ambiguous mapping where -i is the only affix that could be transferred. Furthermore, -i is erroneously replaced by -a much more often than -a is replaced by -i (40 times vs. 12 times,  $\chi^2(1)=.0001$ ), suggesting that the suffix -a is not preferred by the learners over -i in general.

Figure 4.5. Transfer of plural affixes to singular forms and omission of plural affixes from plural forms in languages III and IV for stems in which the plural-to-singular mapping is ambiguous (stems ending in [k] or [tʃ]) and stems in which the singular-plural mapping is ambiguous (stems ending in [p] or [t]).<sup>41</sup>



Overall, there appears to be a bias in favor of adding affixes, rather than deleting them, an “addition bias” (see Hartsuiker 2002 for a connectionist model of such a bias in speech production). This seems to be contrary to existing morphological speech error studies, in which morphological omission errors seems to be much more common than addition errors (e.g., del Viso et al. 1991, Perez et al. 2007). Furthermore, language

<sup>41</sup> Languages I and II are not considered because plural-singular mappings are unambiguous in these languages and the numbers of singulars ending in {k;tʃ} and thus having unambiguous singular-plural mappings and singulars ending in {p;t} and thus having ambiguous singular-plural mappings are unequal (the latter are twice as numerous as the former). Given this inequality, the baseline against which the observed numbers of errors are to be compared is uncertain. If the reader believes in a certain baseline s/he can analyze the following numbers: plural suffixes were omitted from [tʃi]-final plurals 3 times; they were added to [k]-final singulars 13 times; they were omitted from {t;p}{i;a}-final plurals 33 times; and they were added to {t;p}-final singulars 6 times. In any case, these data appear to point in the same direction as the data for languages III and IV.

learners appear to omit affixes much more often than they erroneously insert affixes into base forms. However, this asymmetry in natural languages may be a relative frequency effect, since affixed forms tend to be less frequent than unaffixed forms, which is expected to make their mental representations develop later and remain weaker than representations of unaffixed forms (Hay 2003). An addition bias does appear to exist for purely phonological speech errors (e.g., Hartsuiker 2002, Nootboom 1969, Stemberger 1990, although all of these studies involve consonant deletion vs. addition), and learners who insert plural suffixes are likely not to realize they are suffixes. Thus the possibility of an addition bias cannot be ruled out, especially in situations like the present one where 1) the affixed form is as frequent as the unaffixed form in the input; 2) the phonotactics of affixed and unaffixed forms are similar in familiarity and simplicity; and 3) the affixed and unaffixed forms are similar in informativeness. Alternatively, the plural (or affixed) form may be more salient than the singular (or unaffixed) form for the learners, which is also supported by the tendency to modify the stem of the singular to fit the stem of the plural in Chapter 3 (Figure 3.4).

To summarize the findings on recall errors, learners of all languages are similar in overall accuracy of recall. That is, they tend to recall the same proportion of words they are presented with. However, since learners of Language III and Language IV are presented with 4 more words than learners presented with the other two languages, learners of Language III and Language IV recall more words than learners of Language I or Language II. Learners tend to best remember the form of a word that allows them to project the other form of the same word with maximal accuracy. This is the plural form for stems ending in [p] or [t] and the singular form for stems ending in [k] or [tʃ].

### 4.3. Results II: Types of generalizations extracted from the input

#### 4.3.1. Elicited production

##### 4.3.1.1. Velar-final singulars

Figure 4.6 shows that the probability of a [k] being mapped onto [tʃi] is significantly higher in languages III and IV than in languages I and II ( $p < .05$  according to the Wilcoxon test). This difference is in the opposite direction from the corresponding finding in the source-oriented paradigm. If we combine the results from both experiments (entering training paradigm, whether –i is attached to [tʃ] in training, whether –i often attaches to [p] and [t] in training, and all interactions into a Friedman test as predictors of rank-transformed production probability) the only significant effect is an interaction between experiment and whether or not examples of [tʃ] being mapped onto [tʃi] are presented to the learner ( $F(1,79) = 6.25, p = .01$ ).<sup>42</sup> If we take the probability of [k] being mapped onto [ki] as the dependent variable, there is also a significant interaction in the same direction (shown in Figure 4.7): the additional examples of [tʃ] → [tʃi] presented during training increase the probability of eliciting the production of [k] → [ki] in the source-oriented training paradigm while decreasing the probability of eliciting [k] → [ki] in the product-oriented training paradigm ( $F(1,79) = 4.02, p < .05$ ).

Together these results are strongly indicative of the [tʃ] → [tʃi] examples providing more support for the product-oriented generalization ‘plurals must end in [tʃi]’ rather than

---

<sup>42</sup> Note that here I am considering the absolute productivity of the mapping carrying out velar palatalization relative to all other options, rather than productivity of velar palatalization given that –i is chosen as the suffix.



the source-oriented generalization ‘just add –i’ for the learners in the product-oriented training paradigm and providing more support for the source-oriented generalization in the source-oriented training paradigm. Thus, the characteristics that distinguish the two training paradigms jointly are able to influence how much support a lexical item is inferred to provide for source-oriented vs. product-oriented generalizations by the human learner.

Figure 4.6. Productivity of  $[k] \rightarrow [tʃi]$  in elicited production as a function of whether the training is product-oriented (left two boxes) or source-oriented (right two-boxes) and whether or not learners are presented with examples of  $[tʃ] \rightarrow [tʃi]$ .

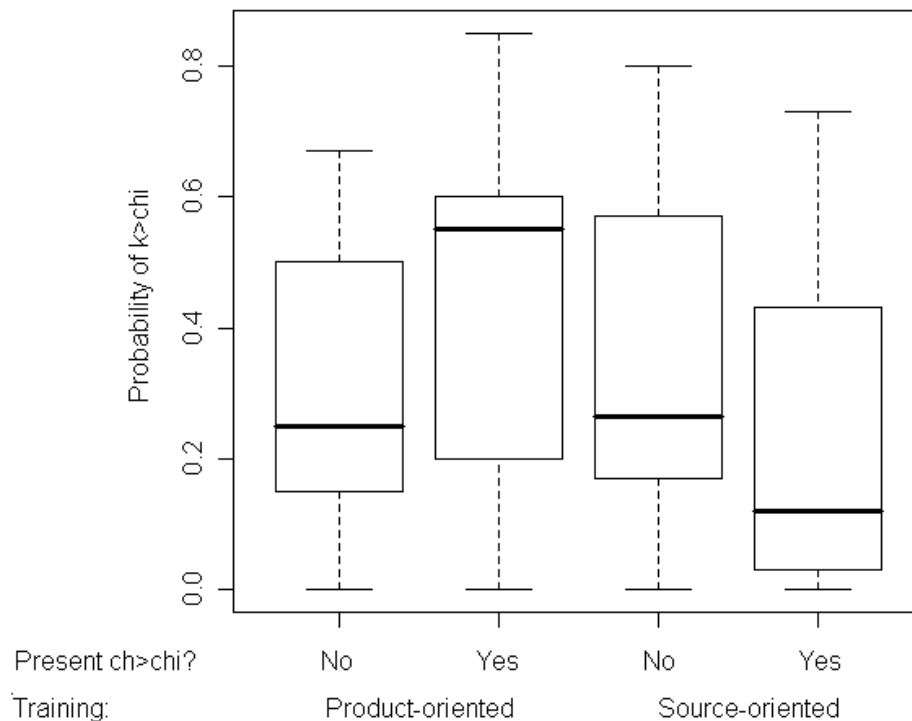
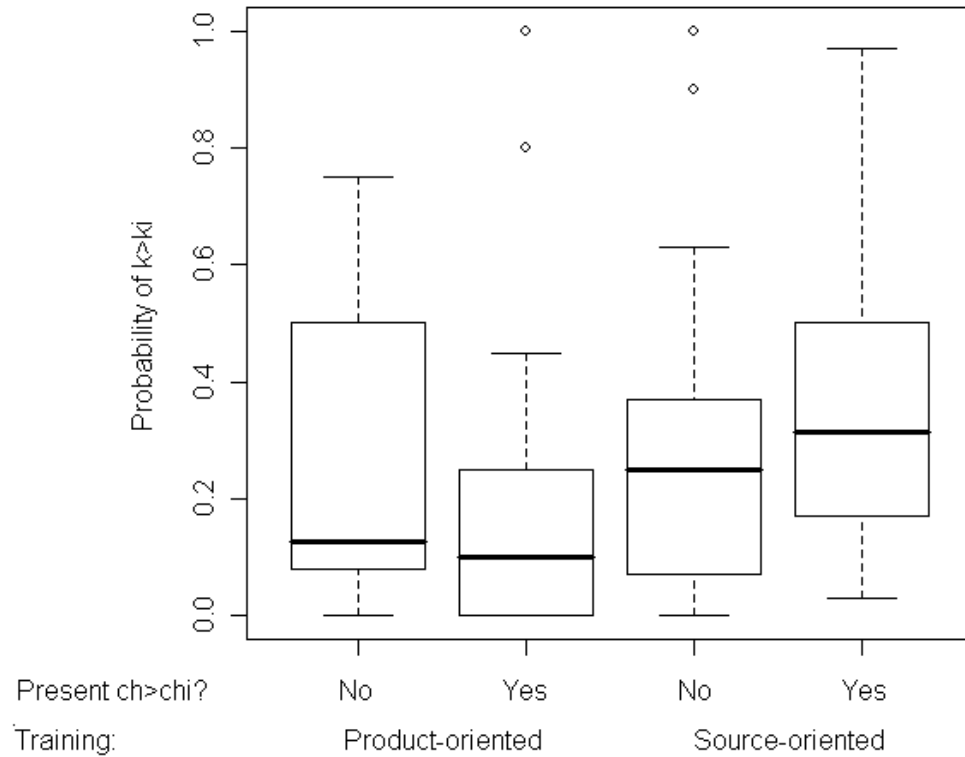


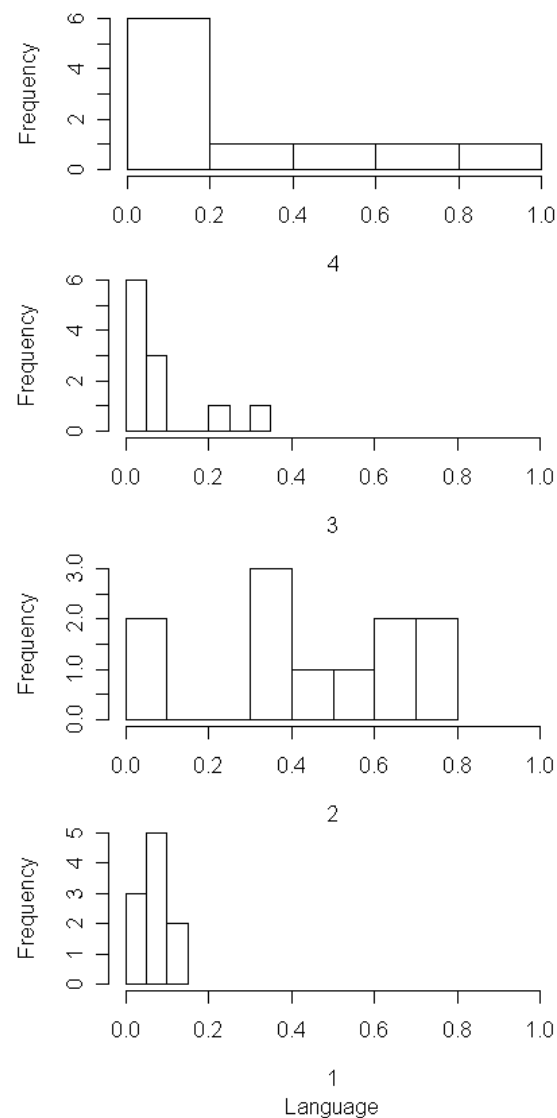
Figure 4.7. Productivity of  $[k] \rightarrow [ki]$  in elicited production as a function of whether the training is product-oriented (left two boxes) or source-oriented (right two-boxes) and whether or not learners are presented with examples of  $[tʃ] \rightarrow [tʃi]$ .



In addition, like in Chapter 3, languages II and IV favor attaching  $-i$  to velars relative to languages I and III ( $p < .0001$  according to the Wilcoxon test). As shown in Chapter 2, this result is consistent with both positive product-oriented and source-oriented generalizations because the examples of  $-i$  attaching to stops favor the source-oriented generalization ‘just add  $-i$ ’ and the product-oriented generalization ‘plurals must end in  $-[i]$ ’. Furthermore, as predicted, the additional examples influence the probability of simply attaching  $-i$  without significantly influencing the probability of the  $k \rightarrow tʃi$  mapping. This result is not consistent with product-oriented constraints weighted by the

difference between observed and expected frequencies of product classes, since the expected frequency of [ki] is much higher in languages II and IV than in languages I and III while the observed frequency is the same (0).

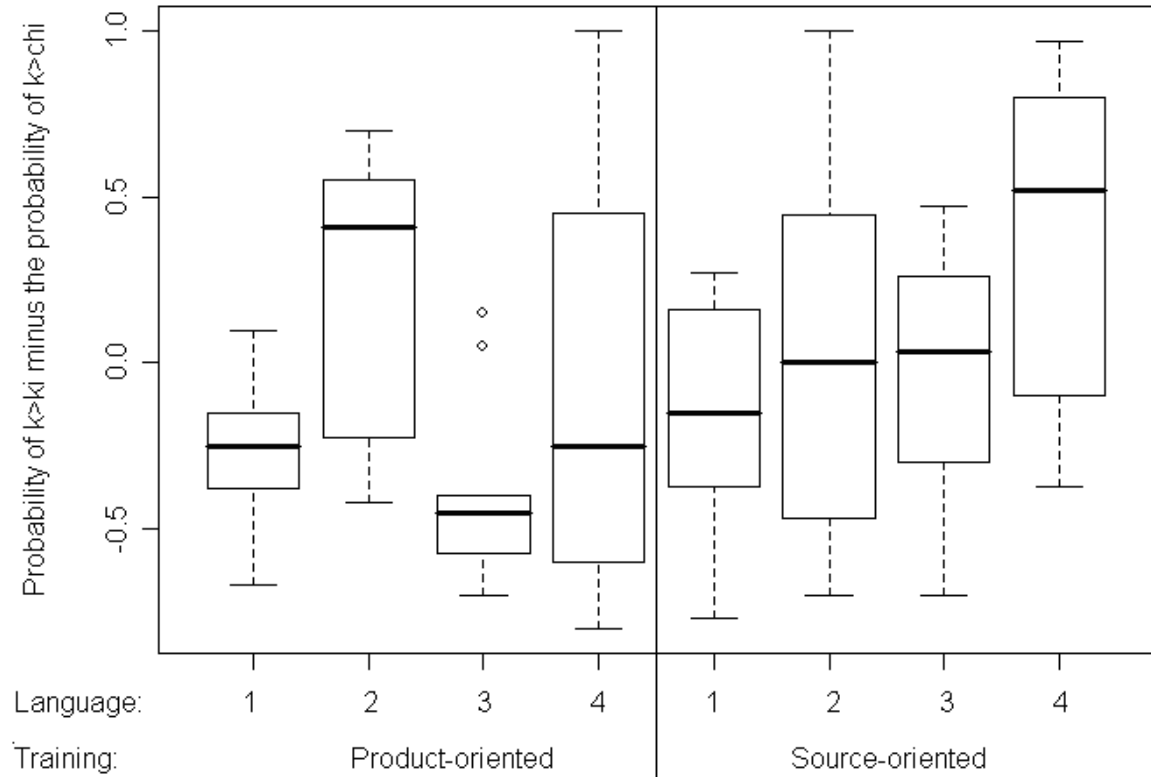
Figure 4.8. Adding –i without changing the stem-final consonant is more productive in languages II and IV than in languages I and III.



There is an overall decrease in the probability of  $[k] \rightarrow [ki]$  in the product-oriented paradigm relative to the source-oriented paradigm ( $F(1,79)=8.48$ ,  $p=.005$ ). Thus, the productivity of velar palatalization before  $-i$  increases when we switch from the source-oriented paradigm to a product-oriented paradigm. In other words, the source-oriented paradigm appears to disfavor stem changes, which is perhaps not surprising given that the learners in the source-oriented paradigm are presented with pairs of words that share the stem. This finding has methodological implications for studies that attempt to determine whether learners are biased in favor of phonological generalizations involving a change in the stem vs. a change in the affix (e.g., Bybee & Newman 1995, Zaba 2009): if one presents the learners with pairs of words that share the stem during training, one is likely to find a bias against learning rules requiring stem changes (e.g., Finley 2008, Zaba 2009) but such a bias may not exist in a more natural training paradigm.

The overall productivity of  $k \rightarrow tʃi$  relative to  $k \rightarrow ki$  across the two training paradigm and the four languages is shown in Figure 4.9. The difference between languages 1 and 3 on the one hand and languages 2 and 4 on the other illustrates the influence of whether  $-i$  often attaches to  $[p]$  and  $[t]$  (it does in languages 2 and 4), which makes it also attach to  $[k]$  without triggering palatalization. The difference between languages 1 and 2 on the one hand and languages 3 and 4 on the other illustrates the influence of adding examples of  $-i$  attaching to  $[tʃ]$ , which has the opposite effects in the two training paradigm, favoring  $[k] \rightarrow [tʃi]$  relative to simply attaching  $-i$  to  $[k]$  in the product-oriented paradigm and having the opposite effect in the less natural source-oriented training paradigm.

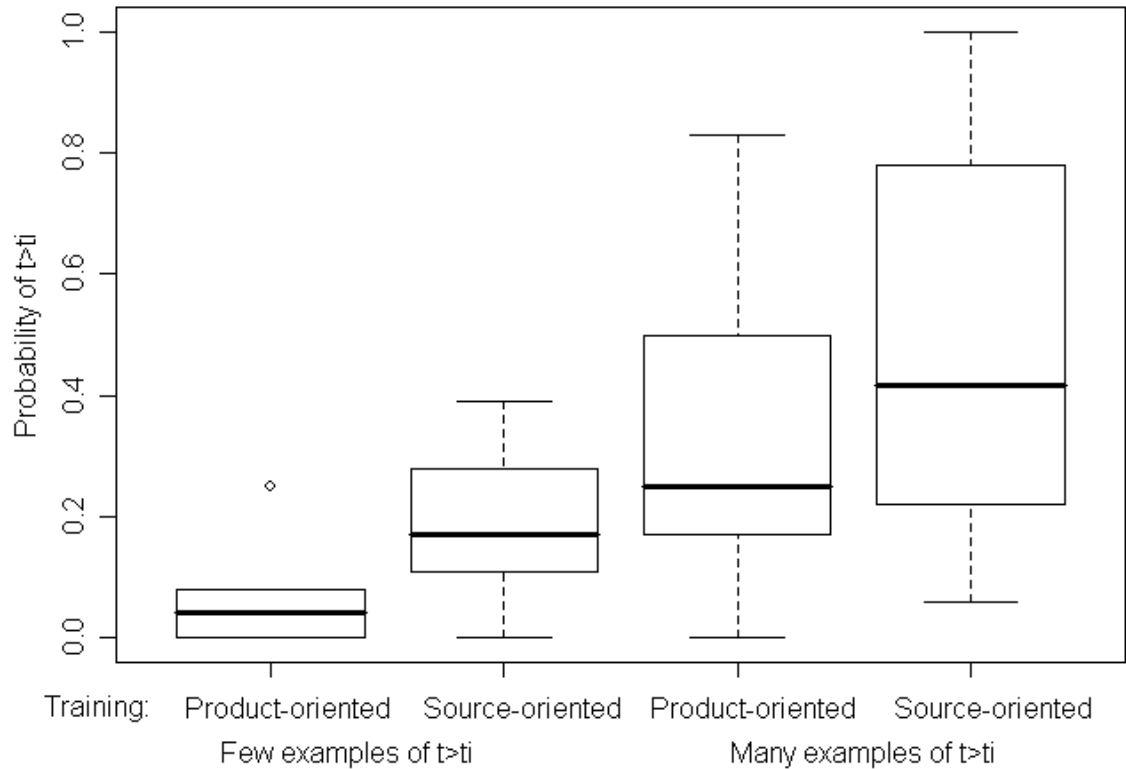
Figure 4.9. Number of word pairs instantiating  $[k] \rightarrow [ki]$  produced by each subject in elicited production minus the number of word pairs instantiating  $[k] \rightarrow [tʃi]$  produced by the same subject.



#### 4.3.1.2. Alveolar-final singulars

Across training paradigms, the addition of examples of  $[t] \rightarrow [ti]$  increases the probability of  $[t] \rightarrow [ti]$  (Figure 4.10;  $p < .0000001$  according to the Wilcoxon test). This is hardly a surprising result: the learners appear to be picking up on the distribution of  $-i$  vs.  $-a$  following  $[t]$  in the input.

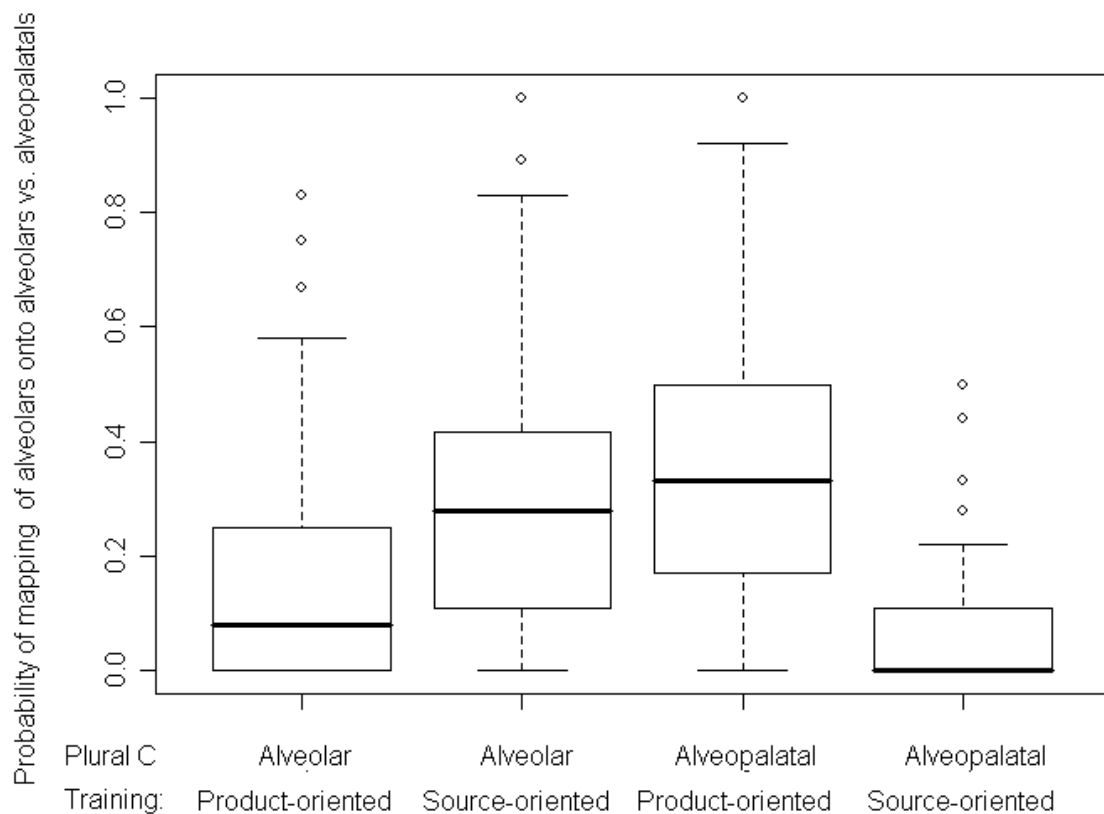
Figure 4.10. Additional examples of [t]→[ti] increase the probability that [t] is mapped onto [ti].



More intriguingly, Figure 4.10 shows that the probability of [t] being mapped onto [ti] is much lower following product-oriented training than following source-oriented training. One of the pieces of evidence for the source-oriented nature of generalization in Chapter 3 was the finding that the only singular-final consonants that are mapped onto alveopalatals in the plural are the velars. This is not the case in the present training paradigm. The product-oriented nature of generalization in the product-oriented training paradigm is also indicated by the fact that velar palatalization is frequently extended to alveolar sources in this paradigm (as shown in Figure 4.11), while such overgeneralization over sources was rare in the experiments presented in Chapter 3.

The differences in probabilities of [t]→[ti] mappings and [t]→[tʃi] mappings across training paradigms is highly significant ( $p=.001$  and  $p<.0000001$  respectively according to the Wilcoxon test) and the probability of palatalizing an alveolar is not significantly different from the probability of palatalizing a velar following product-oriented training ( $p=.48$  according to the Wilcoxon test; the median difference is 0), while velar palatalization is vastly more probable following source-oriented training ( $p<.00001$  according to the Wilcoxon).

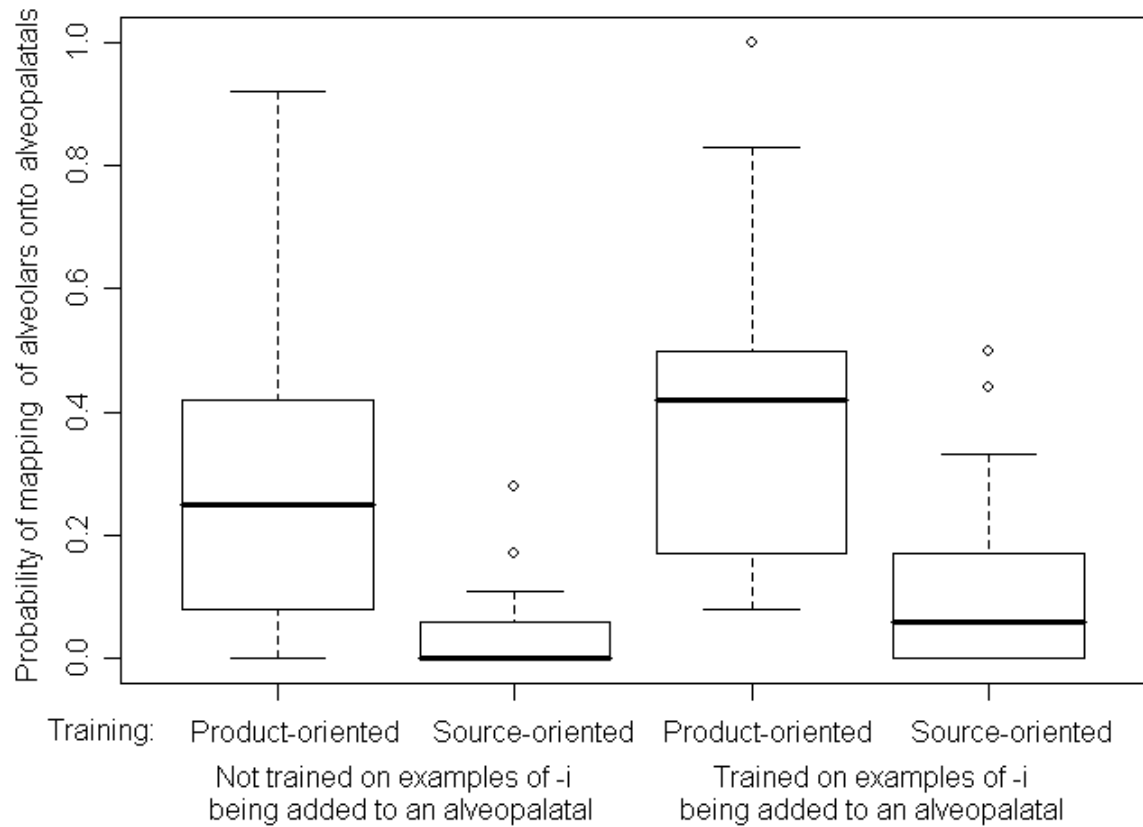
Figure 4.11. Velar palatalization is often overgeneralized to alveolar sources in the product-oriented paradigm but not in the source-oriented paradigm.



When the data from both experiments are combined, languages III and IV significantly favor alveolar palatalization relative to languages I and II (see Figure 4.11), i.e., the addition of [tʃ]→[tʃi] examples favors the mapping of [t] onto [tʃi] ( $p=.03$  according to the Wilcoxon test), and there is no significant interaction between training paradigm and whether or not examples of [tʃ]→[tʃi] are presented ( $F(1,79)<1$ ,  $p=.77$  according to an Friedman test). This finding suggests that product-oriented generalizations may play a role even in elicited production (and not just ratings) following source-oriented training (although when only the data from only one training paradigm are considered, the effect fails to reach significance:  $.15>p>.1$  according to the Wilcoxon for both paradigms).



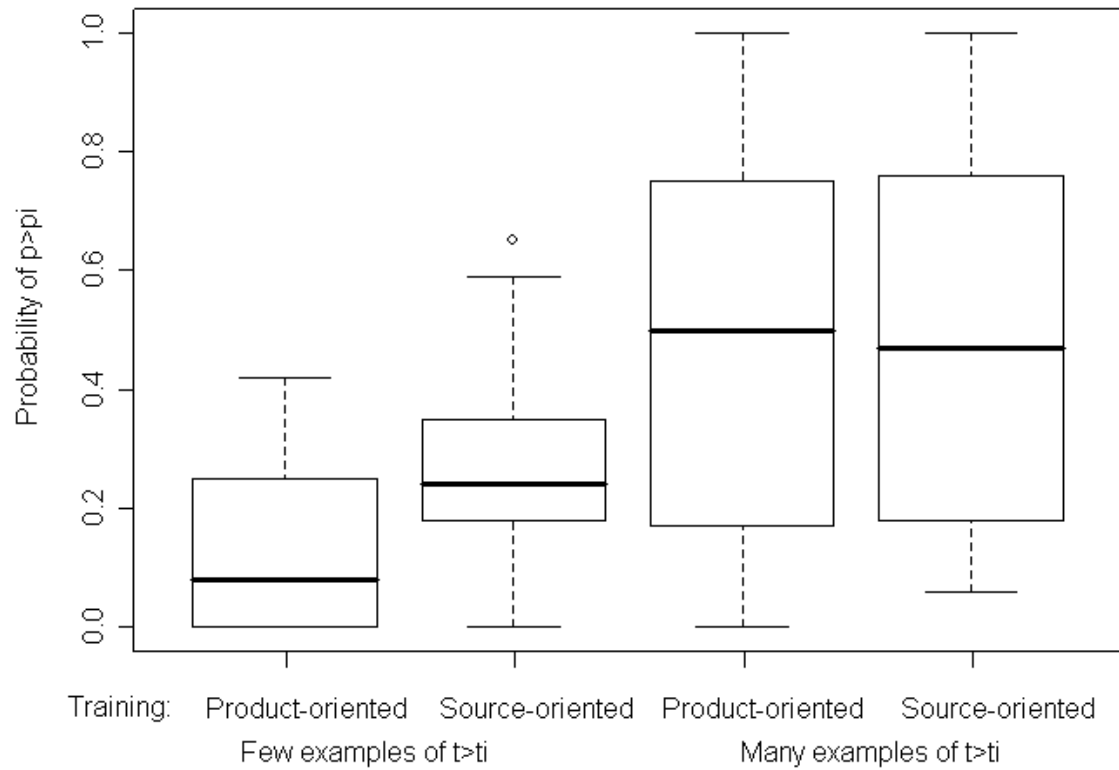
Figure 4.12. Examples of [tʃ]→[tʃi] favor [t]→[tʃi].



#### 4.3.1.3. Labial-final singulars

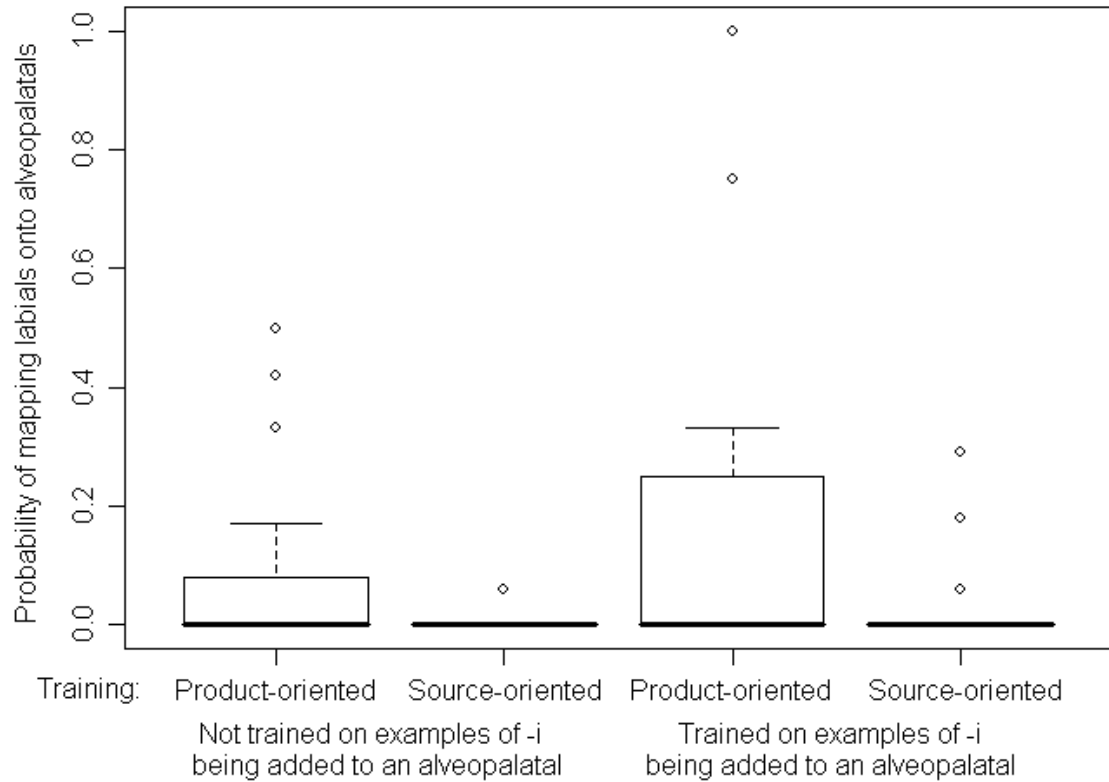
Across training paradigms, the addition of examples of [p]→[pi] leads to increased probability of mapping a singular [p] onto a plural [pi] in elicited production (Figure 4.13;  $p < .00001$  according to the Wilcoxon test; the interaction between training paradigm and the abundance of examples of [p]→[pi] is only marginal:  $F(1,79)=2.89$ ,  $p=.09$  according to an Friedman test).

Figure 4.13. Additional examples of  $[p] \rightarrow [pi]$  increase the probability that  $[p]$  is mapped onto  $[pi]$ .



Just as with alveolar-final singulars, product-oriented training favors unfaithful (stem-changing) stem mappings over source-oriented training: the incidence of a  $[p]$  being mapped onto  $[(p)t]i$  is much higher after product-oriented training than after source-oriented training (Figure 4.14,  $p=.0003$  according to the Wilcoxon test).

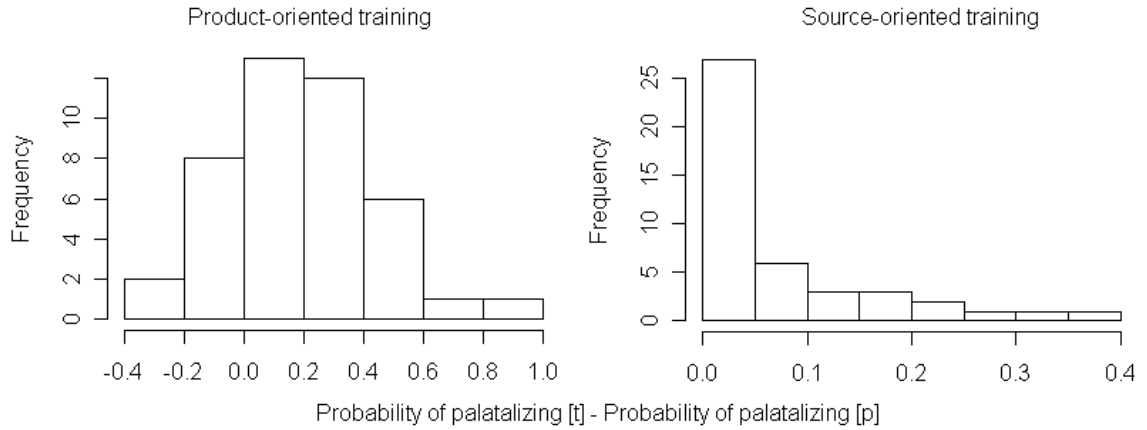
Figure 4.14. Velar palatalization is sometimes overgeneralized to labial sources in the product-oriented paradigm but not in the source-oriented paradigm.



While Figure 4.14 suggests a trend for the addition of examples of  $[tʃ] \rightarrow [tʃi]$  to favor  $[p] \rightarrow [(p)tʃi]$ , most learners never overgeneralize palatalization to labials, showing a zero probability of  $[p] \rightarrow [(p)tʃi]$  and the apparent trend is not significant ( $p=.6$  according to the Wilcoxon).

Figure 4.15 shows that overgeneralization of velar palatalization to labial sources following either kind of training is much less likely than overgeneralization to alveolar sources ( $p<.00001$  for product-oriented training,  $p=.0002$  for source-oriented training, according to the Wilcoxon test).

Figure 4.15. The distribution of differences between a subject's probability of palatalizing alveolars and his/her probability of palatalizing labials.



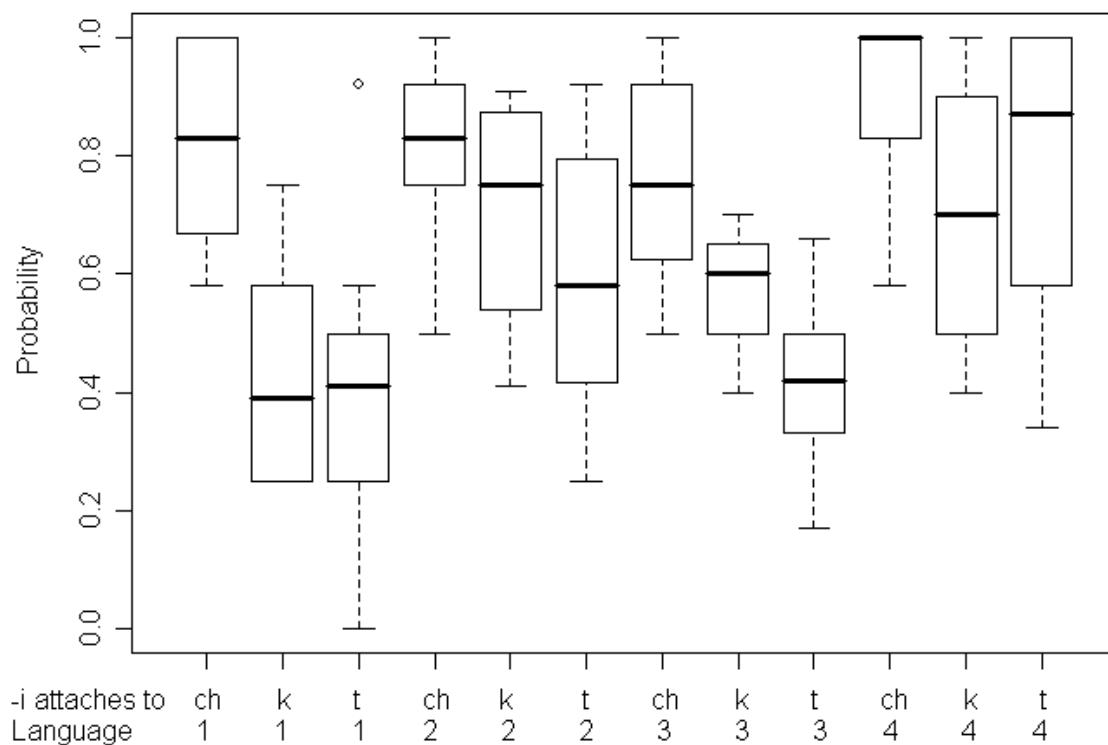
#### 4.3.1.4. Alveopalatal-final singulars

In Chapter 3, we saw that learners exposed to any one of the artificial languages had a tendency to attach  $-i$  rather than  $-a$  to singulars ending in  $[tʃ]$ . I argued that the learners did not do this due to obeying the product-oriented generalization “plurals must end in  $[tʃi]$ ” because the probability of attaching  $-i$  to  $[tʃ]$ , thus producing  $[tʃi]$ , did not correlate with the probability of producing  $[tʃi]$  from  $[k]$  but rather with the probability of attaching  $[i]$  to  $[k]$ , regardless of whether this produced  $[tʃi]$  or  $[ki]$ . Thus, the learners appeared to make a generalization over source forms that  $[k]$  and phones like it, particularly  $[tʃ]$ , take  $-i$ .

The dominance of  $[tʃ] \rightarrow [tʃi]$  over  $[tʃ] \rightarrow [tʃa]$  holds after the product-oriented training as well (the median rate of  $-i$ -attachment to  $[tʃ]$  is 83%). Like after the source-oriented training, the rate of  $-i$ -attachment to  $[tʃ]$  is significantly higher than the rate of  $-i$

attachment to [k] or [t], even in Language 2 and Language 4 ( $p=.009$  comparing to  $[k] \rightarrow \{ki;tfi\}$ ;  $p=.02$  comparing to  $[t] \rightarrow \{t;tfi\}$ ). This result is shown in Figure 4.16. The high rate of  $-i$  attachment to  $[t]$  relative to  $[k]$  suggests that  $[t]$  is either further away for the learners from  $[t]$  and  $[p]$ , the segments that are heard taking  $-a$ , or that there is an avoidance of the observed stem change that depresses the rate of  $[k] \rightarrow [tfi]$ .

Figure 4.16. Probability that  $-i$  is attached to  $[t]$  is higher than the probability that it is attached to  $[k]$  or  $[t]$  (whether or not  $[k]$  or  $[t]$  change as a result).



If the same generalization is responsible for two singular-plural mappings, then a subject who has assigned a high weight to the generalization should use both singular-plural mappings productively while a subject who has assigned a low weight to the

generalization should not. In other words, probabilities of the mappings for individual subjects should correlate. Figures 4.17-4.18 show clustering solutions for correlation matrices based on the data from the source-oriented training paradigm (Figure 4.17) and the product-oriented training paradigm (Figure 4.18). The major difference between the two solutions is that the mapping  $[tʃ] \rightarrow [tʃi]$  is clustered together with  $[k] \rightarrow [k;tʃ]i$  following source-oriented training but it is clustered together with  $\{k;t;p\} \rightarrow [tʃi]$  following product-oriented training. These data are consistent with the hypothesis that the more natural product-oriented paradigm favors product-oriented generalizations: the product-oriented training places stronger restrictions on the product while source-oriented training places stronger restrictions on the source.

Figure 4.17. Clustering of mappings after source-oriented training. The variables ‘paddi’, ‘taddi’ and ‘kaddi’ are probabilities of adding –i to [p], [t], or [k] respectively regardless of whether the source consonant is changed.

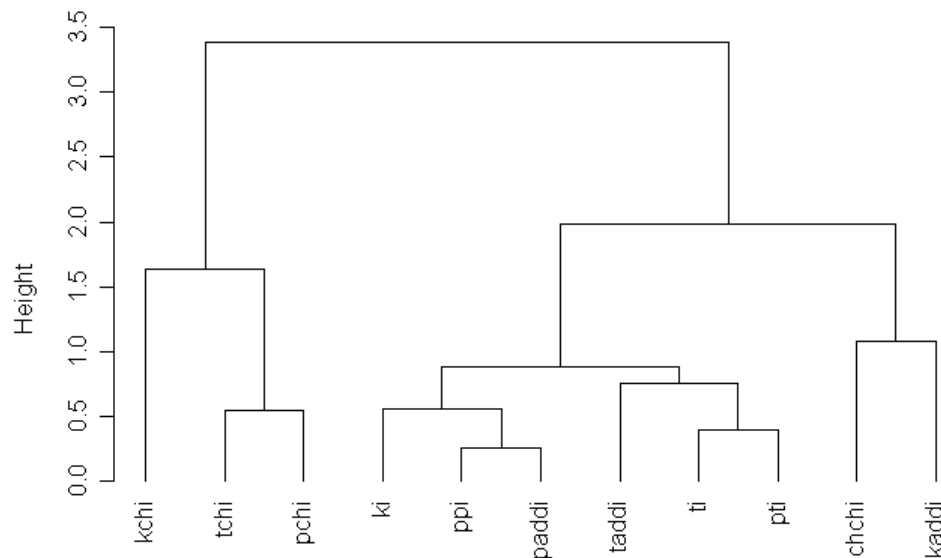
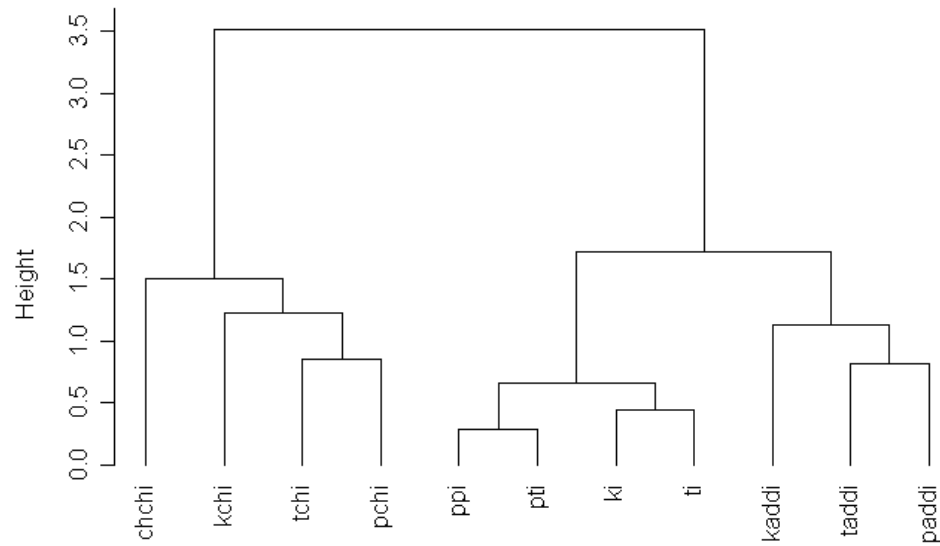


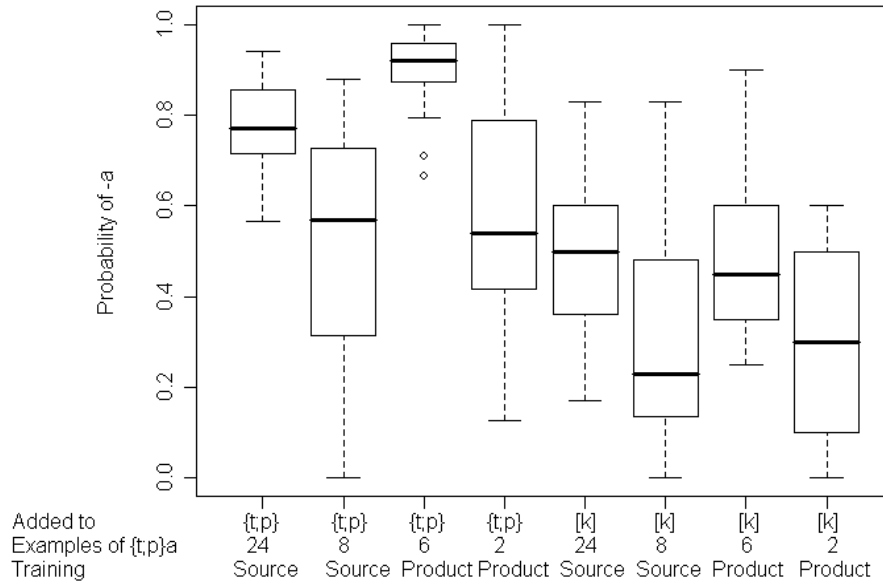
Figure 4.18. Clustering of mappings after product-oriented training.



#### 4.3.1.5. The two suffixes

Figure 4.19 shows the probability of attaching –a to stops following source-oriented and product-oriented training. The probability of attaching –a to either alveolars and velars is above zero in all languages in both training paradigms despite the fact that the addition of –a is exemplified by only two word types in languages II and IV in the product-oriented paradigm. Thus this result provides evidence against the hypothesis that at least three exemplifying word types are necessary for the formation of a generalization (contra Bybee & Pardo 1981, Bollt & Gerken 2008, Xu & Tenenbaum 2007). Whether this is due to the fact that each word type is repeated by a number of different “creature voices” and in a number of different syntactic contexts is a question for future research.

Figure 4.19. The probability of adding –a to singular consonants ending in [p], [t], or [k] following source-oriented or product-oriented training on many or few examples of  $\{p;t\} \rightarrow \{p;t\}a$  and accordingly few or many examples of  $\{p;t\} \rightarrow \{p;t\}i$ .

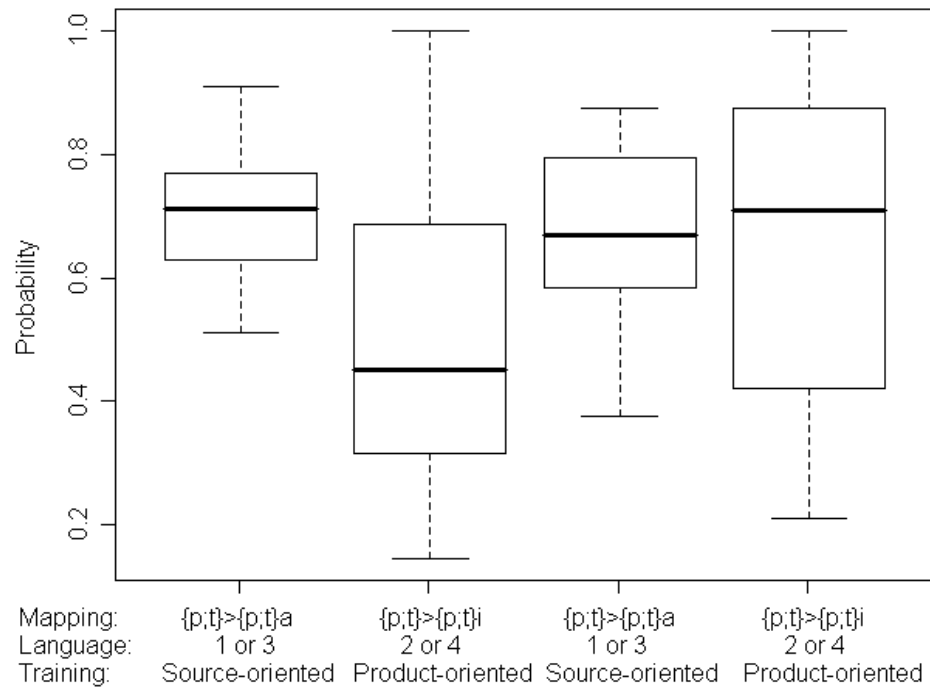


The addition of –i is exemplified by more examples in languages II and IV than the addition of –a is exemplified by in languages I and III, and the fact the number of examples of –i being added to [p] or [t] in languages I and III is equal to the number of examples of –a being added to [p] or [t] in languages II and IV. Thus, given no preference for –a over –i, learners of languages I and III should apply –a to [p] or [t] as often as or less often than learners of languages II and IV should apply –i to [p] or [t]. However, Figure 4.20 shows that following source-oriented training -a is used by learners of languages I and III more often than –i is used by learners of languages II and IV ( $p=.008$  for the source-oriented paradigm) while after product-oriented training there is no significant difference ( $p=.86$ ). Thus the learners appear to disprefer the suffix that triggers stem changes and overuse the suffix that does not trigger stem changes following



source-oriented training while product-oriented training, which makes unfaithful (stem-changing) mappings more likely also leads to a loss of dispreference for the suffix that triggers stem changes.

Figure 4.20. Learners exposed to many examples of  $\{p;t\} \rightarrow \{p;t\}a$  produce the mapping more often than learners exposed to many examples of  $\{p;t\} \rightarrow \{p;t\}i$  produce  $\{p;t\} \rightarrow \{p;t;t\}i$ .

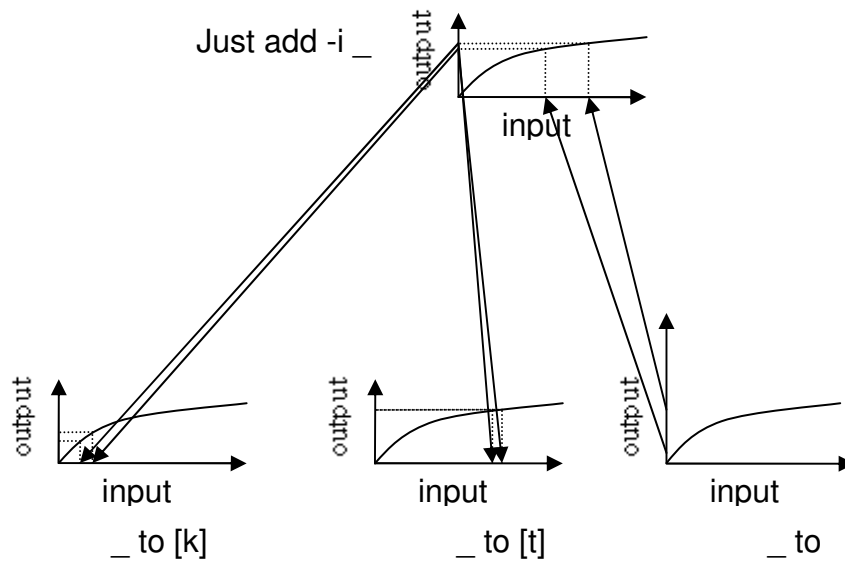


#### 4.3.1.6. Discussion of elicited production

The characteristics of the training paradigm used in the present chapter that distinguish it from the less natural training paradigm used in the previous chapter appear to favor product-oriented generalizations over source-oriented generalizations. First, the grammar extracted by the learners in the present paradigm does not restrict sources that

can be mapped onto [tʃi] as much as the grammar extracted by the learners in the source-oriented paradigm. Second, the addition of examples of [tʃ] → [tʃi] appears to favor the product-oriented generalization ‘plurals must end in [tʃi]’ over the source-oriented generalization ‘just add –i’ while the data for the source-oriented training paradigm are more ambiguous. There, the addition of [tʃ] → [tʃi] appears to favor [t] → [tʃi] over [t] → [ti] and [k] → [ki] over [k] → [tʃi]. This may perhaps be explained by the fact that [t] → [tʃi] is much weaker than [t] → [ti] while [k] → [ki] is weaker than [k] → [tʃi]. If we conceptualize generalizations responsible for mappings as units with a decelerating activation function, then the same amount of additional support would increase the activation of a weak mapping more than it would increase the activation of a strong mapping, as shown in Figure 4.21.

Figure 4.21. A possible explanation for why examples of  $[tʃ] \rightarrow [tʃi]$  favor the weak mapping  $[k] \rightarrow [ki]$  over the stronger mapping  $[k] \rightarrow [tʃi]$  and the weak mapping  $[t] \rightarrow [tʃi]$  over the stronger mapping  $[t] \rightarrow [ti]$ . The same amount of support increases the strength of a weak generalization more than it increases the strength of a strong generalization. Arrows indicate direction of inference where the presentation of an example of  $[tʃ] \rightarrow [tʃi]$  increases the inferred support for each of the depicted generalizations by the distance between the dotted lines on the vertical axis corresponding to the generalization.



Learners in both experiments display certain restrictions on sources. Thus, even following product-oriented training, learners tend not to map  $[p]$  onto  $[tʃi]$ . Thus learners must know at least which source consonants change and which ones do not, a type of source-oriented generalization, since it can only be learned by noting which consonants participate in unfaithful (stem-changing) singular-plural mappings.

Why do learners overgeneralize velar palatalization to alveolar but not (so much) to labial sources. One possibility is a substantive bias provided by Universal Grammar.

That is, the learners do not need to learn that [p] is never mapped onto [tʃi] because they know that this process is completely unnatural and typologically rare to non-existent (Bhat 1974, Guion 1998, Wilson 2006) and thus may be part of Universal Grammar.<sup>43</sup> An alternative is that the mapping [t]→[ti] and [k]→[ki], which feature some acoustic palatalization of the [t] and [k], and are perhaps better represented as [t]→[tʲi] and [k]→[kʲi] provide some support for [t]→[tʃi] because the learner presented with [tʲi] or [kʲi] assigns some probability to having heard [tʃi]. Further experimentation is necessary to determine which explanation is correct.

Thus I would suggest that overgeneralization to [t]→[tʃi] is in part due to assigning some probability to [tʃi] while perceiving [ti]; this weight would be somewhat larger when there are many examples of C→[tʃi] in the environment but would also be dependent on how much [tʲi] sounds like [tʃi] to the learner. Overgeneralization to [p]→[tʃi] appears to be dependent on generalization to [t]→[tʃi] such that there are no subjects who have a higher [p]-palatalization probability than [t]-palatalization probability before –i.<sup>44</sup> Thus, overgeneralization to [p] may happen when the subject generalizes that anything can be mapped onto [tʃi], despite having no perceptual evidence of ever being presented with [p]→[tʃi]. Thus, unlike overgeneralization to [t],

---

<sup>43</sup> The learners could assume the much more natural process [p]→[ptʃi], making the generalization that [tʃi] is a suffix. Some learners do use this (and the corresponding [k]→[ktʃi] mapping), although only one learner used these mappings consistently. However, the resulting form is not very typical of the artificial languages, containing a consonant cluster.

<sup>44</sup> One subject has a higher [p]→[ptʃi] rate than [t]→[tʃi] rate, while two subjects have a higher absolute [p]→[tʃi] rate than [t]→[tʃi] rate but not if one looks at the probability of palatalizing [t] vs. [p] before –i.

overgeneralization to [p] is a *genuine* overgeneralization, rather than an inevitable result of perceptual processes.

Finally, it is as of yet unclear which features that distinguish the product-oriented paradigm (if any) are singly necessary and/or singly sufficient to produce the observed difference between the two training paradigms. For instance, the greater tolerance of unobserved source-product mappings seen in the product-oriented paradigm may be due to either the lower number of word types that exemplify a given source-product mapping in the product-oriented paradigm, or the fact that learners in the source-oriented paradigm are exposed to source-product pairs, or both. Again, this question is a matter for future research.

#### 4.3.2. The relationship between elicited production and recall

In order for a subject to recall a word pair exemplifying velar palatalization, s/he needs to recall a singular ending in [k] and a corresponding plural ending in [tʃi]. In each language, there are four such word pairs that can be recalled. Figure 4.22 shows how often each possible number of exemplifying word pairs is actually recalled by learners. Overall the numbers are quite low, and there are no significant differences across languages (all  $p > .4$  according to the Wilcoxon test). Interestingly, Figure 4.23 shows that the correlation between the number of examples of velar palatalization recalled by a subject and the productivity of velar palatalization for that subject in elicited production is very weak ( $t(41) = 1.53$ ,  $p = .13$  overall, with all  $p > .1$  within-language as well). Thus it appears that remembering examples of velar palatalization well enough to be able to recall them on cue is not necessary for velar palatalization to be productive. For instance,

Subject 8 in Language 4 group regularized all [k]→[tʃi] singular-plural mappings in recall to [tʃ]→[tʃi] but when presented with a [k]-final singular produced a [tʃi]-final plural 60% of the time in elicited production.

Figure 4.22. The distributions of numbers of examples of velar palatalization correctly recalled across languages.

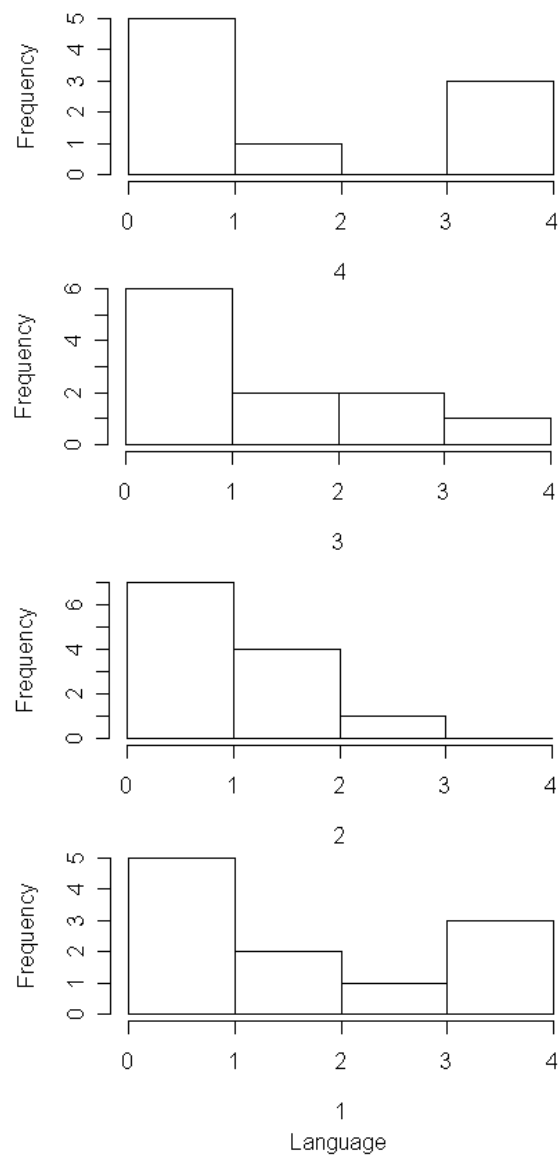
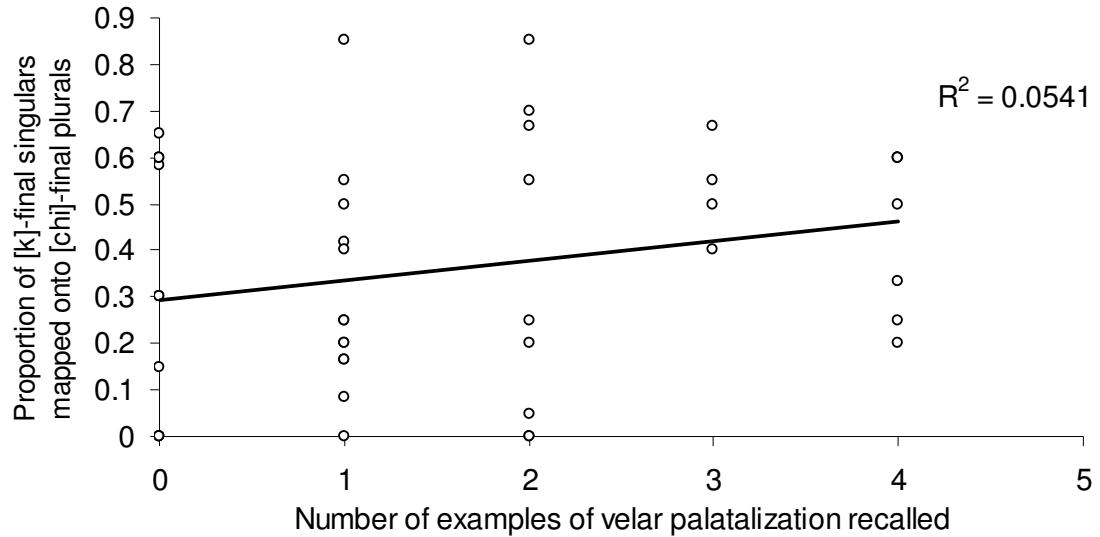


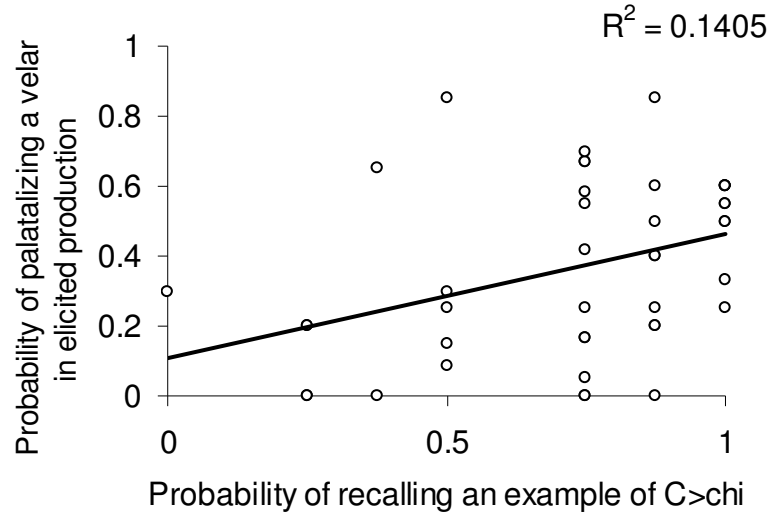
Figure 4.23. Productivity of  $k \rightarrow tʃi$  mappings as a function of the number of examples of such mappings recalled.



There are two possible conclusions from this result. First, it may be the case that the productivity of a rule does not depend much on the support for it in the learner's active lexicon. Rather, it depends on the support for it in the lexicon to which the learner is exposed. The other possibility is that rather than looking at the number of recalled examples of the particular source-product mapping, we should be examining the probability of recalling examples of all source-product mappings that result in the same product.<sup>45</sup> This correlation is shown in Figure 4.24 and it is significant ( $t(41)=2.59$ ,  $p=.01$  according to an ANCOVA on ranks controlling for Language). This finding provides further support for the largely product-oriented nature of generalization in this training paradigm.

<sup>45</sup> We could also examine the *number* of recalled examples resulting in the same product but that would confound how many examples are recalled and how many are presented.

Figure 4.24. The correlation between the probability of recall for examples of  $C \rightarrow [tʃi]$  and the probability of  $[k]$  being mapped onto  $[tʃi]$  in elicited production.



The only other relationship between the number of examples recalled and production probability is that if a subject ‘recalls’ examples of  $k \rightarrow ki$ , which are of course never presented during training, s/he is also likely to produce  $k \rightarrow ki$  mappings in elicited production. This relationship is significant even when Language is controlled ( $p=.03$  according to ANCOVA) and is not surprising given that the ‘recalled’ examples are likely to be produced by the subject based on the singular form of the ‘recalled’ noun, just as in elicited production.

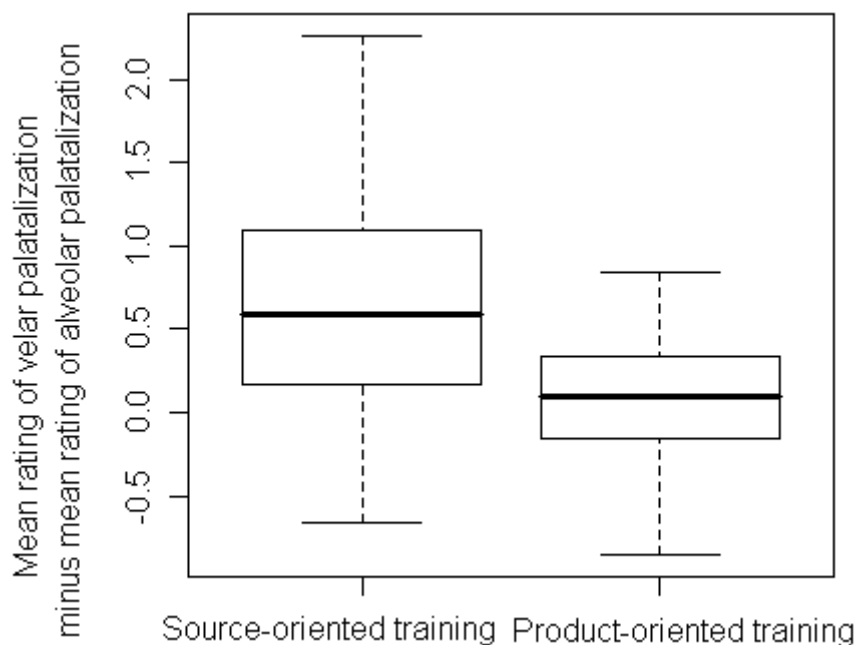
Correlations of recall with probability of palatalizing  $[t]$  and  $[p]$  are weak and insignificant ( $r=.12$ ,  $p=.4$  for  $[t]$ ;  $r=-.01$ ,  $p=.96$  for  $[p]$ ), thus it appears that overgeneralization of velar palatalization is driven largely not based on analogy to remembered examples of  $C \rightarrow [tʃi]$  mappings. Rather, it appears to depend much more strongly on how often  $[tʃi]$  is perceived (or assigned a non-zero probability of having been presented) during training (i.e., the language to which the learners are exposed).



#### 4.3.3. Plural likelihood rating

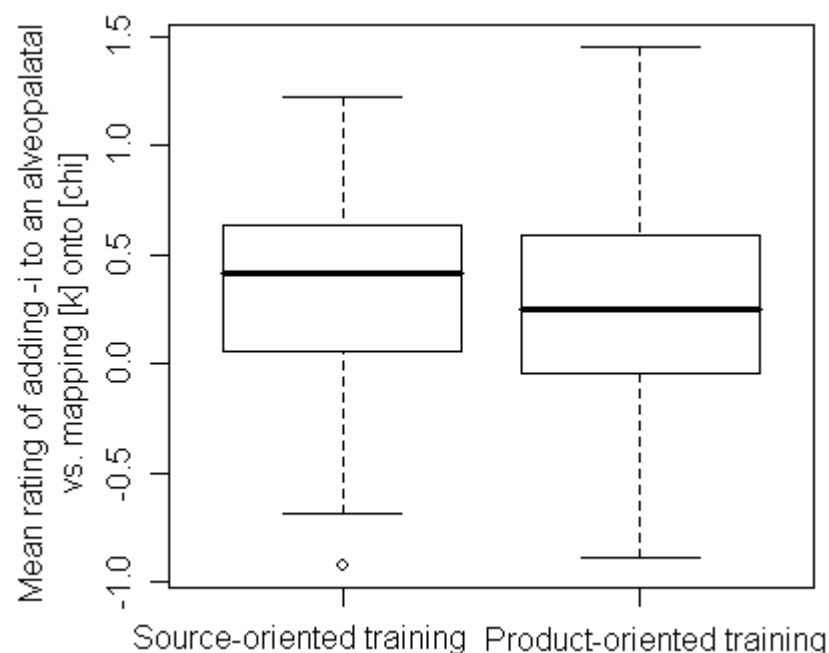
A major difference between the source-oriented and product-oriented training paradigms is that learners in the source-oriented paradigm rate  $[k] \rightarrow [tʃi]$  mappings higher than  $[t] \rightarrow [tʃi]$  mappings ( $p < .00001$  according to the Wilcoxon) while learners in the product-oriented paradigm do not ( $p = .16$ ). Thus there is a significant influence of the training paradigm on the difference between ratings given to  $[k] \rightarrow [tʃi]$  mappings vs.  $[t] \rightarrow [tʃi]$  mappings ( $p < .0001$  according to the Wilcoxon). The data are shown in Figure 4.25. Thus there are fewer restrictions on sources that can be mapped onto an observed product following product-oriented training than following source-oriented training.

Figure 4.25. Learners rate examples of velar palatalization higher than examples of alveolar palatalization following source-oriented but not product-oriented training.



Nevertheless, Figure 4.26 shows that after either kind of training  $[k] \rightarrow [tʃi]$  mappings are rated lower than  $[tʃ] \rightarrow [tʃi]$  mappings ( $p=.0001$  following source-oriented training,  $p=.0009$  following product-oriented training, with no significant effect of training on the difference scores,  $p=.29$  according to the Wilcoxon), even though the number of examples of  $[k] \rightarrow [tʃi]$  presented to learners is equal to or higher than the number of presented examples of  $[tʃ] \rightarrow [tʃi]$ .

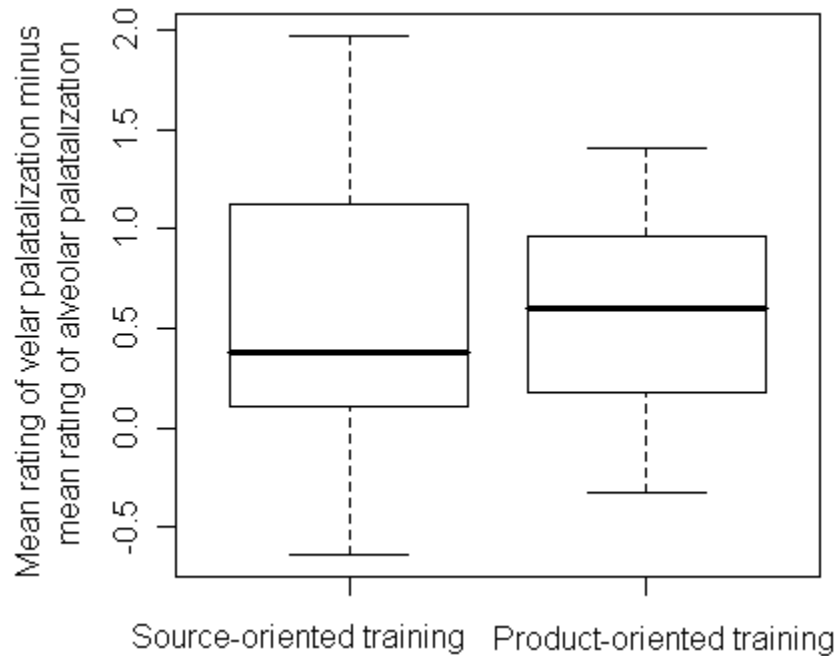
Figure 4.26. After either kind of training,  $[k] \rightarrow [tʃi]$  mappings are rated lower than  $[tʃ] \rightarrow [tʃi]$  mappings.



Similarly,  $[k] \rightarrow [ki]$  and  $[k] \rightarrow [ka]$  are rated higher than  $[tʃ] \rightarrow [ki]$  and  $[tʃ] \rightarrow [ka]$  following either training type ( $p<.00001$  following either training type, with no significant effect of training on the difference scores,  $p=.95$ ). The data are shown in

Figure 4.27. There are also no significant differences between the relatively low-rated  $[k] \rightarrow [tʃa]$ ,  $[tʃ] \rightarrow [tʃa]$ , and  $[t] \rightarrow [tʃa]$  (all  $p > .1$ ). Thus product-oriented training does not simply increase tolerance of stem changes. It increases tolerance of unobserved stem changes that result in observed products, at least those that have been observed to result from stem changes.<sup>46</sup>

Figure 4.27. After either kind of training,  $[k] \rightarrow [k]\{i;a\}$  examples are rated higher than  $[tʃ] \rightarrow [k]\{i;a\}$  examples.



The faithful mappings  $[k] \rightarrow [k]\{i;a\}$  is preferred over the unfaithful (stem-changing) mappings  $[tʃ] \rightarrow [k]\{i;a\}$  despite the fact that the outputs of both mappings are never observed even in languages III and IV where  $[k]$  and  $[tʃ]$  are equally common

<sup>46</sup> It seems unlikely that ratings for unfaithful mappings of  $[t] \rightarrow [pi]$ ,  $[k] \rightarrow [pi]$ ,  $[p] \rightarrow [ti]$ , and  $[k] \rightarrow [ti]$  would increase under product-oriented training if these were tested. Hence the qualification.

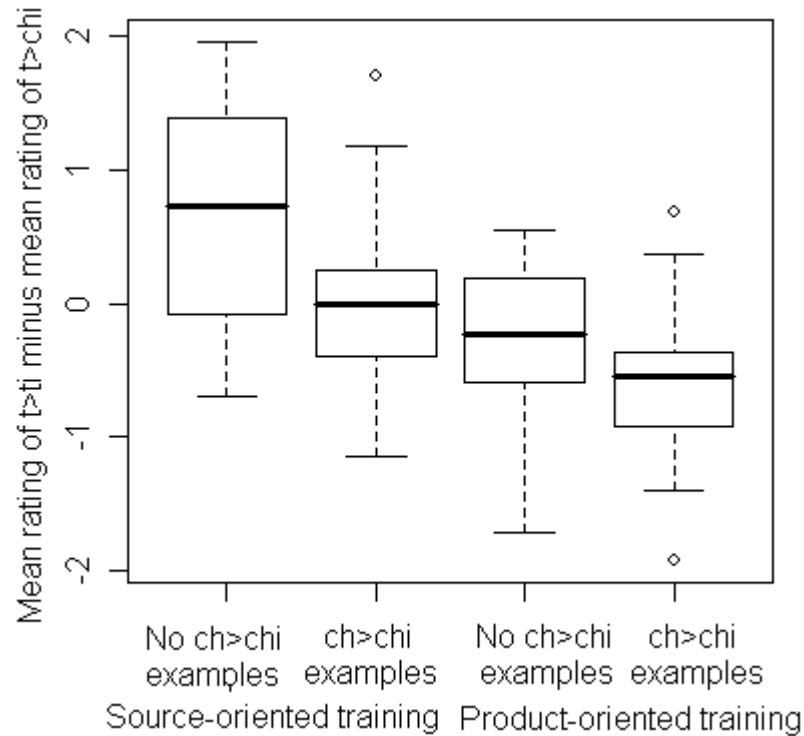
singular-final consonants ( $p < .000001$ , according to the Wilcoxon) and even following product-oriented training ( $p = .0002$  for just languages 3 and 4 following product-oriented training). Similarly, the faithful mapping  $[tʃ] \rightarrow [tʃi]$  is preferred over the unfaithful (stem-changing) mappings  $\{k;t\} \rightarrow [tʃi]$  even in languages I and II where examples of  $[k] \rightarrow [tʃi]$  are presented and examples of  $[tʃ] \rightarrow [tʃi]$  are not ( $p = .005$ , according to the Wilcoxon).

Thus, it appears that learners prefer faithful mappings over unfaithful mappings, and generalization is not purely product-oriented even after product-oriented training.

Whether the bias against changes is innate (Prince & Smolensky 1993/2004) or something learners have acquired during first language acquisition is a matter for future research.

While both training paradigms provide evidence for restrictions on source-product mappings, both training paradigms also show evidence of product-oriented generalizations in that the addition of examples of  $[tʃ] \rightarrow [tʃi]$  increases the ratings of  $[t] \rightarrow [tʃi]$  relative to  $[t] \rightarrow [ti]$  ( $p < .001$  in both paradigms). The data are shown in Figure 4.28. Thus, both source-oriented and product-oriented generalizations appear to be extracted by learners in both paradigms: the learners learn both what typical plural forms sound like and which segments must be preserved in the plural form if they occur in the singular form.

Figure 4.28. In both training paradigms, the addition of examples  $[tʃ] \rightarrow [tʃi]$  increases ratings of  $[t] \rightarrow [tʃi]$  relative to  $[t] \rightarrow [ti]$ .



The addition of extra examples of  $[p] \rightarrow [pa]$  and  $[t] \rightarrow [ta]$  at the expense of  $[p] \rightarrow [pi]$  and  $[t] \rightarrow [ti]$  has a stronger effect in product-oriented training where there are significant effects of this manipulation on ratings of  $[t] \rightarrow [ta]$  ( $p < .0001$ ) and  $[k] \rightarrow [ka]$  ( $p = .007$ ), and a trend in the same direction on  $[tʃ] \rightarrow [tʃa]$  ( $p = .08$ ) compared to no significant effects on any  $C \rightarrow Ca$  mappings following source-oriented training (with a trend on  $[t] \rightarrow [ta]$ ,  $p = .07$ ). The manipulation appears to be stronger in product-oriented training, where languages II and IV have 6 word pair types exemplifying  $\{t;p\} \rightarrow \{t;p\}a$  and languages I and III have 2 word pair types than in source-oriented training where the difference is between 8 examples in languages I and III and 24 examples in languages II and IV. Thus the difference between 2 and 6 appears to be greater than the difference

between 8 and 24, despite the distances being equal on the log scale. This provides support for the idea that the dependence of productivity on type frequency is a decelerating one (Bybee & Pardo 1981, Gerken & Bollt 2008, Xu & Tenenbaum 2007) and suggests that the deceleration is higher than in the logarithmic function. However, it is important to remember that even with 2 word types exemplifying the suffix -a, there is some generalization of -a beyond the training set, as shown by Figure 4.16, thus we do not have support for the idea that productivity is zero for type frequency below 3 or 4 (Bybee & Pardo 1981, Gerken & Bollt 2008, Xu & Tenenbaum 2007).

Table 4.2 shows correlations between standardized ratings of plural forms that simply attach -i to a velar, plural forms that display velar palatalization, plural forms that show attachment of -i to an alveopalatal, and plural forms featuring attachment of -i to an alveolar, with or without palatalization, following product-oriented training. It can be compared to Table 3.3 in Chapter 3, which shows correlations between the same variables following source-oriented training. There is one strongly significant interaction with training type. Product-oriented training results in a stronger correlation between ratings of velar and alveolar palatalization ( $r=.7$ ,  $p=.0000002$ ) than does source-oriented training ( $r=.3$ ,  $p=.05$ ); the interaction is significant at  $p=.008$  according to the Friedman test.<sup>47</sup> This finding is consistent with product-oriented training allowing velar palatalization to generalize to the alveolars more than the source-oriented training.

---

<sup>47</sup> Product-oriented training results in a weak barely significant positive correlation between ratings of examples of  $\{t;d\} \rightarrow \{tʃ;dʒ\}i$  and ratings of  $\{k;g\} \rightarrow \{k;g\}i$  ( $r=.31$ ,  $p=.04$ ) while source-oriented training results in a weak (non-significant) negative correlation ( $r=-.16$ ,  $p=.31$ ); the interaction itself is barely significant ( $p=.03$ ) according to the Friedman test. There is also a weak interaction between training type and the correlation between ratings of  $\{tʃ;dʒ\} \rightarrow \{tʃ;dʒ\}a$  and  $\{tʃ;dʒ\} \rightarrow \{k;g\}i$  ( $p=.04$ ). The negative correlation is weaker after product-oriented training ( $r=-.55$ ,  $p=.0001$  vs.  $r=-.14$ ,  $p=.4$ ).

Table 4.2. The correlations between ratings of singular plural mappings featuring the suffix -i.

		$\{t;d\} \rightarrow \{t;d\}i$	$\{k;g\} \rightarrow \{tʃ;dʒ\}i$	$\{t;d\} \rightarrow \{tʃ;dʒ\}i$	$\{tʃ;dʒ\} \rightarrow \{tʃ;dʒ\}i$
$\{k;g\} \rightarrow \{k;g\}i$	R	.48	.39	.31	.24
	p	.001	.01	.04	.12
$\{t;d\} \rightarrow \{t;d\}i$	R		.15	.22	.06
	p		.36	.17	.7
$\{k;g\} \rightarrow \{tʃ;dʒ\}i$	R			.7	.48
	p			.0000002	.001
$\{t;d\} \rightarrow \{tʃ;dʒ\}i$	R				.39
	p				.01

Tables 4.3-4.5 present correlations between ratings of singular-plural mappings that may compete for the same singular following product-oriented training. Just like after source-oriented training, the correlations between singular-plural mappings are most likely to be negative when they result in very different products ( $[tʃi]$  vs.  $C_{[-cont]a}$ ). When the consonant or (especially) the suffix are shared, correlations of ratings are sometimes positive, suggesting that source-product mappings can support each other if they result in similar products, which is expected only if the learners are relying on product-oriented generalizations for rating the likelihood / acceptability of a mapping.

Table 4.3. Correlations between likelihood ratings of  $\{t;d\} \rightarrow X$  mappings.

		$\{t;d\} \rightarrow \{t;d\}i$	$\{t;d\} \rightarrow \{t\bar{f};d\bar{3}\}a$	$\{t;d\} \rightarrow \{t\bar{f};d\bar{3}\}i$
$\{t;d\} \rightarrow \{t;d\}a$	R	-.28	.24	-.46
	p	.07	.12	.002
$\{t;d\} \rightarrow \{t;d\}i$	R		-.54	.22
	p		.0002	.17
$\{t;d\} \rightarrow \{t\bar{f};d\bar{3}\}a$	R			-.22
	p			.15

Table 4.4. Correlations between likelihood ratings of  $\{k;g\} \rightarrow X$  mappings.

		$\{k;g\} \rightarrow \{k;g\}i$	$\{k;g\} \rightarrow \{t\bar{f};d\bar{3}\}a$	$\{k;g\} \rightarrow \{t\bar{f};d\bar{3}\}i$
$\{k;g\} \rightarrow \{k;g\}a$	R	-.35	.38	-.53
	p	.02	.01	.0003
$\{k;g\} \rightarrow \{k;g\}i$	R		-.3	.39
	p		.05	.01
$\{k;g\} \rightarrow \{t\bar{f};d\bar{3}\}a$	R			-.24
	p			.12



Table 4.5. Correlations between likelihood ratings of  $\{tʃ;dʒ\} \rightarrow X$  mappings.

		$\{tʃ;dʒ\} \rightarrow \{tʃ;dʒ\}i$	$\{tʃ;dʒ\} \rightarrow \{k;g\}a$	$\{tʃ;dʒ\} \rightarrow \{k;g\}i$
$\{tʃ;dʒ\} \rightarrow \{tʃ;dʒ\}a$	R	-.35	.31	-.13
	p	.02	.05	.4
$\{tʃ;dʒ\} \rightarrow \{tʃ;dʒ\}i$	R		-.52	-.17
	p		.0004	.26
$\{tʃ;dʒ\} \rightarrow \{k;g\}a$	R			.2
	p			.21

To summarize, product-oriented training increases the likelihood of velar palatalization being generalized to alveolar sources, resulting in alveolar palatalization, examples of which are never presented. Nonetheless both training paradigms lead to the extraction of both source-oriented generalizations in the form of paradigm uniformity constraints (Downing et al. 2005), the knowledge of which segments change, and product-oriented generalizations, the knowledge of what typical plural forms are like. The productivity of a generalization depends on how much support for it there is in the training data with the dependency between amount of support and productivity being a decelerating one with a stronger deceleration rate than the logarithmic function.

#### 4.3.4. Combining results from rating and production

Figure 4.29 shows the clustering solution for correlations of all mappings used or rated in production and perception following source-oriented training. Figure 4.27 shows

the corresponding clustering solution following product-oriented training. The two figures both contain a cluster in which any source is mapped onto [tʃi] and a cluster in which -i is simply added to the singular. Furthermore, [tʃ] → [tʃi] is unambiguously classified as an instance of  $C \rightarrow [tʃi]$  rather than an instance of ‘just add -i’ in both perception and production and even in the source-oriented training paradigm (although if the more abstract mapping  $k \rightarrow \{k; tʃ\}i$  is added to the clustered matrix,  $tʃ \rightarrow tʃi$  switches cluster membership following source-oriented training but not product-oriented training, as was shown in Figures 4.17-4.18). These results provide support for the overall primacy of product-oriented generalizations over source-oriented generalizations (Bybee 2001) and suggest a similar weighting of source-oriented and product-oriented generalizations in perception and production.

Figure 4.29. The clustering of the correlation matrix between ratings and production probabilities of various mappings following source-oriented training. ‘R’ stands for ‘ratings’, while ‘P’ stands for production probabilities.

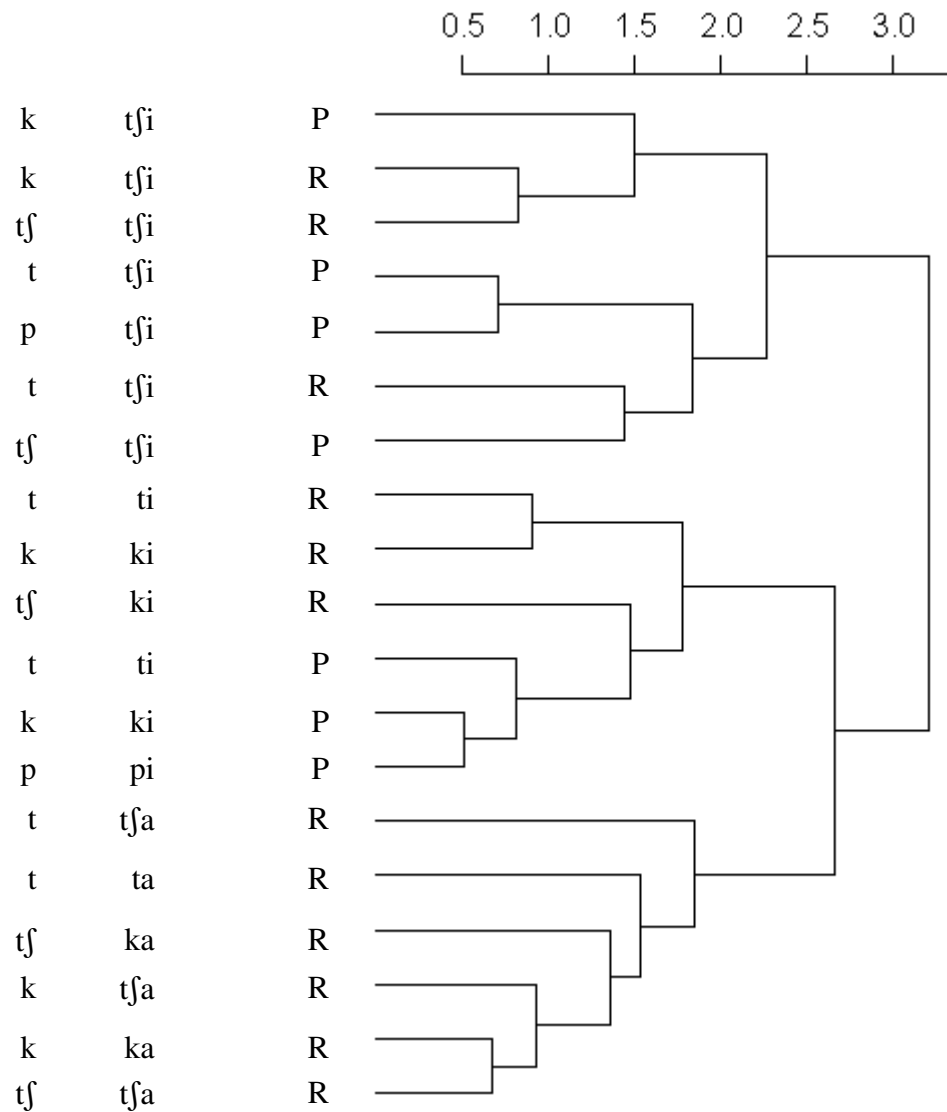
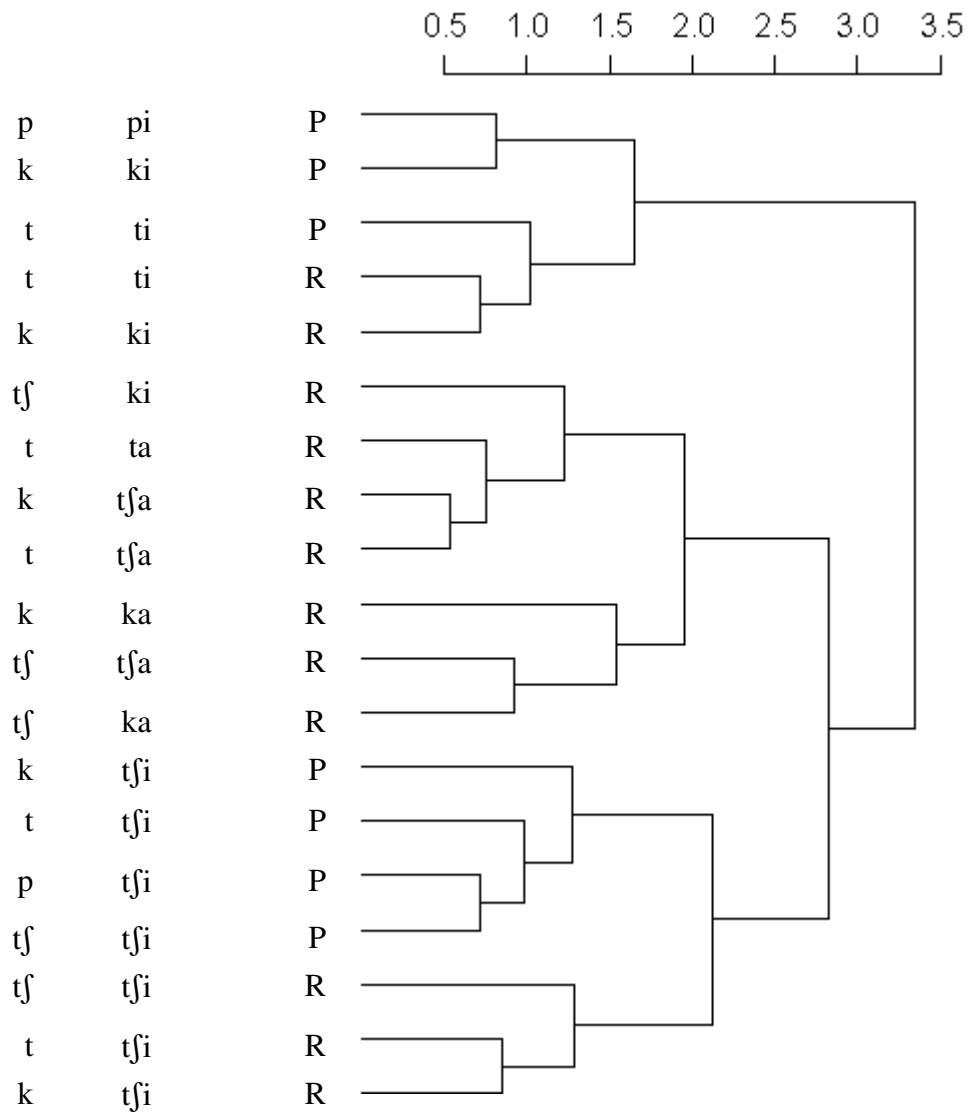


Figure 4.30. The clustering of the correlation matrix between ratings and production probabilities of various mappings following product-oriented training. ‘R’ stands for ‘ratings’, while ‘P’ stands for production probabilities.



#### 4.3.5. The relationship between recall and ratings

There is a significant positive correlation between the number of examples of  $\{k;tf\} \rightarrow tfi$  recalled and the ratings of examples of  $tf \rightarrow tfi$  relative to examples of  $tf \rightarrow ki$  ( $r=.31$ ,  $p=.03$ ). However, all other correlations between numbers of words exemplifying a

certain mapping or product-oriented generalization recalled by a subject and the ratings assigned by the same subjects to words exemplifying the same generalization are quite low ( $.1 < r < .25$ ), suggesting that the acceptability of a mapping is not determined by the number of known words exemplifying it, at least if only wordforms whose meanings are known to the learner are counted.

#### **4.4. Conclusion**

This chapter introduced a training paradigm in which, unlike in the source-oriented training paradigm presented in Chapter 3, the learners were asked to learn and recall words, rather than learning to form plurals, they were presented with individual wordforms in random order, rather than singular-plural pairs, and the number of words in the presented lexicon was small enough for most subjects to memorize most of the words. In both perception and production, the addition of examples of  $tʃ \rightarrow tʃi$  increased the productivity of  $k \rightarrow tʃi$ ,  $t \rightarrow tʃi$  and  $p \rightarrow tʃi$  mappings. Thus examples of  $tʃ \rightarrow tʃi$  were taken to provide greater support for the product-oriented generalization that plurals end in  $[tʃi]$  than for the source-oriented generalization ‘just add  $-i$ ’.

Learners in the product-oriented paradigm extended palatalization from velars to alveopalatals and sometimes labials, while learners in the source-oriented paradigm restricted velar palatalization to velars, indicating that source-oriented training facilitates restricting the class of sources that can be mapped onto a product. This may be due to both the fact that the source-oriented paradigm involves presentation of forms sharing the same stem next to each other and/or the fact that the number of different word pairs exemplifying a source-product mapping (i.e., its type frequency) is lower in the product-

oriented paradigm than in the source-oriented paradigm: the generalization that alveolar consonants from the source must be retained unchanged is exemplified by 16 word pairs in the source-oriented paradigm and only 4 word pairs in the product-oriented paradigm. It is, however, interesting to note that velar palatalization is more productive in the product-oriented paradigm than in the source-oriented paradigm despite being exemplified by only 4-8 word pairs in the product-oriented paradigm and 30-50 in the source-oriented paradigm. It appears that the increase in type frequency from 4 to 30 or from 8 to 50 does not benefit palatalization as much as the increase in type frequency from 4 to 16 benefits alveolar retention. Thus, the relationship between type frequency and productivity appears to be different for positive product-oriented generalizations and paradigm uniformity constraints (or source-oriented generalizations more generally) with some minimum type frequency perhaps being necessary for a paradigm uniformity constraint to become productive. Alternatively (or additionally), reoccurrence of the stem in close temporal proximity may enforce paradigm uniformity.

The finding that the change of a training paradigm influences the degree to which learners rely on source-oriented or product-oriented generalizations suggests that learning biases in favor of particular kinds of generalizations are not innate but rather arise due to the nature of the learning task faced by the learners of a language. The present results suggest that product-oriented generalizations are favored especially when the lexicon exemplifying the set of generalizations to be learned is small, e.g., the set of English irregular verbs examined by Bybee & Moder (1983), Bybee & Slobin (1982), and Wang & Derwing (1994), and when different forms of the same word are unlikely to co-occur in close proximity.

Learners in both training paradigms acquire a grammar that contains both the knowledge of what typical plural forms sound like, as indicated by the finding that  $t \rightarrow t\text{f}i$  mappings are facilitated by the addition of examples of  $t\text{f} \rightarrow t\text{f}i$  in both paradigms, and the knowledge of which source consonants change or remain the same when the plural is produced from a singular, as indicated by the fact that  $p \rightarrow (p)t\text{f}i$  is rarely produced by the learners after either kind of training. Thus the resulting grammar contains both source-oriented and product-oriented generalizations. There are at least two possible reasons for why palatalization is restricted to [t] but not [p]. First, learners may come to the experiment with the knowledge that  $[p] \rightarrow [t\text{f}]$  is an unnatural mapping (e.g., Wilson 2006). Alternatively, the amount of perceptual evidence provided by  $[t] \rightarrow [ti]$  for  $[t] \rightarrow [t\text{f}i]$  may be greater than the amount of perceptual evidence provided by  $[p] \rightarrow [pi]$  for  $[p] \rightarrow (p)[t\text{f}i]$ . Under the first explanation, the source-oriented generalizations obeyed by the learners are not learned from exposure to the artificial language data while under the second explanation they are. Manipulation of the training data, including the degree of palatalization of /t/ before [i] and the number of  $[t^{(i)}i]$  examples, would help disentangle the two explanations.

The exact nature of source-oriented generalizations extracted by the learners is uncertain. One possibility is that the acquired source-oriented generalizations are rules (i.e., source-product mappings, as in Albright & Hayes 2003, or Chomsky & Halle 1968). One piece of evidence that can be interpreted as evidence for rules is that the learners in either paradigm do not produce  $k \rightarrow ti$ ,  $k \rightarrow pi$ ,  $k \rightarrow ta$ , and  $k \rightarrow pa$  mappings, even when they are exposed to Language I or Language II where the number of presented plurals ending

in [ti], [pi], [ta], or [pa] is almost equal to the number of plurals ending in [tʃi], and, furthermore, plurals ending in [pi], [ti], [pa], and [ta] are recalled better than [tʃi] plurals, at least in the product-oriented paradigm. However, an alternative explanation is that the learners tolerate only unfaithful (stem-changing) mappings that result in a product that is observed to result from unfaithful mappings during training.

In addition to knowledge about which products result from unfaithful mappings, learners must also have knowledge about which source segments must be retained in the product (i.e., paradigm uniformity constraints, see Downing et al. 2005 for various formalizations). Thus the learners in the source-oriented paradigm may be learning that a stem-final [t] in the singular is retained in the plural while a stem-final [k] is not while also learning that plurals should end in [tʃi]. Thus, examples of  $tʃ \rightarrow tʃi$  increase the evidence for ‘retain [k]’, ‘retain [tʃ]’, ‘retain [t]’, and ‘plurals must end in [tʃi]’.

Examples of  $tʃ \rightarrow tʃi$  favor  $t \rightarrow tʃi$  over  $t \rightarrow ti$  but do not favor  $k \rightarrow tʃi$  over  $k \rightarrow ki$  as consistently. Thus, we might say that examples of  $tʃ \rightarrow tʃi$  provide more evidence for  $t \rightarrow tʃi$  than for  $k \rightarrow tʃi$  and/or they provide more evidence for  $k \rightarrow ki$  than for  $t \rightarrow ti$ .

Interestingly, learners in the source-oriented paradigm, for whom  $tʃ \rightarrow tʃi$  favors  $k \rightarrow ki$  over  $k \rightarrow tʃi$ , appear to classify source [tʃ] with [k] rather than [t], since rate of attaching –i to [tʃ] is predicted by the rate of attaching –i to [k] but not the rate of attaching –i to [t]. It is thus tempting to conclude that examples of retaining [tʃ] provide more evidence for ‘retain [k]’ than for ‘retain [t]’, for the learners in the source-oriented paradigm. As a result ‘retain [k]’ benefits from the examples more than does ‘plurals must end in [tʃi]’



while ‘retain [t]’ benefits less. Another factor that may make ‘retain [k]’ benefit from the additional examples more than ‘retain [t]’ is that ‘retain [k]’ is a weaker generalization to begin with, since there is abundant evidence against it presented during training.

In order to perform well in recall, learners have a choice of remembering both forms of each word that they are presented with during training, or remembering one of the forms and deriving the other form from it using the grammar. If the second option is chosen, one of the forms must be chosen for memorization. This is probably not a conscious process but rather a variety of selective attention. It appears that the learners in the present experiment choose to memorize the forms that are most informative for deriving the other form (cf. Albright 2005, 2008). Interestingly, they do not appear to choose either the singular forms or the plural forms as a single class but rather memorize singulars for the subclass of nouns for which deriving the plural from a singular is easier than deriving the singular from the plural and memorize plurals for the subclass of nouns for which deriving the singular from the plural is easier.

Finally, as predicted by both positive product-oriented generalizations that are supported by the experiments reported in the present thesis thus far and competing weighted rules of the type proposed by Albright & Hayes (2003), additional examples of  $t \rightarrow ti$  and  $p \rightarrow pi$  increase the production probability and ratings of  $k \rightarrow ki$ . In the next chapter, I show that this hypothesis also holds for velar palatalization in a natural language, Russian, and show that it is able to explain some otherwise puzzling effects in language change.

## APPENDIX 1: INSTRUCTIONS

### Instructions for the training

You will hear names for strange creatures in a made-up language.

For instance, you might hear that this



is called 'gwink'

or that these



are called 'sworpi'

Please repeat the names you hear aloud.

You will be tested on remembering the creature names.

Click to begin.

### Instructions for the recall test

You will now be shown pictures of creatures you've already seen.

Say what they are called aloud.

**Remember, this is not English!**

**Be sure to use the right form of the name.**

**If you see more than one creature, use the plural form of the name  
and if you see just one creature use the singular.**

**DO NOT use the English plural.**

**Say the plural form from the made-up language.**

**Click to start.**

Instructions for generalization and rating tests were the same as in Chapter 3.

## APPENDIX 2: STIMULI USED FOR TRAINING AND ELICITED PRODUCTION

### Training stimuli

The forms of /blaɪk/, /zaɪt/, and /bup/ and /slaɪtʃ/ were presented 21 times each (albeit forms of /slaɪtʃ/ were presented only to subjects exposed to Language III or Language IV), with all other forms being presented 7 times each.

Singular	Plural			
	Language I	Language II	Language III	Language IV
blaɪk	blaɪtʃi			
truk	trutʃi			
swik	switʃi			
vork	vortʃi			
blort	blorta	blorti	blorta	blorti
hit	hita	hiti	hita	hiti
zaɪt	zaɪta	zaɪti	zaɪta	zaɪti
flort	florti	florta	florti	florta
bup	bupa	bupi	bupa	bupi
floop	floupa	floopi	floupa	floopi
gwip	gwipa	gwipi	gwipa	gwipi
klup	klupi	klupa	klupi	klupa
bortʃ	Neither singular nor plural		bortʃi	

	presented	
dwitʃ	Neither singular nor plural presented	dwitʃi
frutʃ	Neither singular nor plural presented	frutʃi
slaɪtʃ	Neither singular nor plural presented	slaɪtʃi

#### Elicited production stimuli

Body of the stem	k	tʃ	t	p
blaɪ		+	+	+
swi		+	+	+
tru		+	+	+
vor		+	+	+
blor	+	+		+
flor	+	+		+
hi	+	+		+
zai	+	+		+
bu	+	+	+	
floo	+	+	+	
gwi	+	+	+	
klu	+	+	+	
bor	+		+	+
dwi	+		+	+
fru	+		+	+
slaɪ	+		+	+
fla	+			

fli	+			
kra	+			
kri	+			
wa	+			
wi	+			
skla	+			
skli	+			

For stimuli used in rating, see Appendix 2 in Chapter 3.

## CHAPTER 5

### A NATURAL LANGUAGE: RUSSIAN

In the present chapter, we turn to velar palatalization in a natural language, Russian, examining whether the productivity of velar palatalization in spontaneous loanword adaptation (from English into Russian) depends on the probability of the palatalizing suffix attaching to non-velars and whether morphological adaptation of verbs and nouns borrowed into Russian is underlain by one processing stage or two. The present chapter aims to lend the experimental data presented in the preceding chapters some ecological validity.

#### **5.1. Velar palatalization in Russian**

If one looks at a dictionary of modern Russian, velar palatalization appears to involve several exceptionless morphophonological rules, which can be stated simply as “velars become alveopalatals before the derivational suffixes X” where the relevant derivational suffixes either begin with a front vowel or used to begin with a front vowel historically. For the purposes of this chapter, we will be concentrating on Russian verbs with the highly productive stem extension –i, and the diminutive suffixes for masculine nouns, –ik/ek and –ok, which obligatorily trigger velar palatalization in the lexicon as depicted by dictionaries (e.g., Levikova 2003, Sheveleva 1974).

Example (1) shows that in Russian verbs are derived from consonant-final nouns by adding the stem extension (in this case /i/) followed by verbal inflection (in this case the infinitival marker *tʲ*). As shown in (1), velars at the ends of noun roots change into

alveopalatals when a verb is derived from the root. This does not happen with all stem extensions, as evidenced by the existence of Russian verbs like  $n^iux+a+t^j$ ,  $plak+a+t^j$ , and  $stalk+iva+t^j$ , but it always happens with the stem extension -i.

(1)

$k \rightarrow tʃ^i$

klok            klotʃ+i+t<sup>j</sup>

durak           duratʃ+i+t<sup>j</sup>

polk            poltʃ+i+t<sup>j</sup>

jamʃtʃik        jamʃtʃitʃ+i+t<sup>j</sup>

$g \rightarrow ʒ^i$

flag            flaʒ+i+t<sup>j</sup>

dolg            dolʒ+i+t<sup>j</sup>

$x \rightarrow ʃ^i$

grex            greʃ+i+t<sup>j</sup>

The mappings between velar consonants and the corresponding alveopalatals are constant across Russian. Thus, if velars change into alveopalatals in some context, /k/

always becomes [tʃ], /g/ becomes [ʒ], and /x/ becomes /ʃ/. The Russian phone inventory does not contain [dʒ]. The phone [ɨ] cannot follow velars or [tʃ] while the phone [i] cannot follow [ʃ] or [ʒ]. Whether [i] and [ɨ] are allophones of /i/ and chosen during a separate allophone selection stage or separate stem extensions does not influence the qualitative results presented here. The reported graphs are based on a model that treats the choice between [i] and [ɨ] as happening after the morphophonological competition modeled.

In the Russian lexicon, -a is favored over -i by velar-final roots while -i is favored elsewhere. The distribution in the diminutive system is quite different. Only masculine diminutive suffixes will be considered for the purposes of this chapter, because the loaned English nouns end in a consonant, consequently being adopted into the masculine gender. There are three highly productive masculine diminutive suffix morphs, -ik, -ek, and -ok. The morphs -ek and -ik are in complementary distribution in the established lexicon and thus can be considered allomorphs of a single morpheme. The suffixes that trigger palatalization in the lexicon, -ok and -ek, are heavily favored by velar-final nouns, with -ek attaching only to velar-final bases. The suffix -ik, on the other hand, does not attach to velar-final bases, thus one could argue that the Russian lexicon provides no evidence in whether -ik would trigger or fail to trigger velar-palatalization if it were to be attached to a velar-final base, although I will argue that the lexicon does in fact provide



the relevant information and Russian speakers use this information in loanword adaptation.<sup>59</sup> Examples are shown in (2).

(2) lug	→	lu'ʒok	
luk	→	lu'tʃok	
lutʃ	→	'lutʃik	
'fartuk	→	'fartutʃek	
ka'bluk	→	kablu'tʃok	
tʃelo'vek	→	tʃelo'vetʃek	
rog	→	ro'ʒok	
noʒ	→	'noʒik	→ 'noʒitʃek
tʃas	→	tʃa'sok	
tʃas	→	'tʃasik	
ʃag	→	ʃa'ʒok	→ ʃa'ʒotʃek

---

<sup>59</sup> Since -ek and -ik are unstressed, they have the same phonetic realization, the choice between them may be part of orthography. However, the answer to the question of whether the choice is made in orthography or in phonology is not relevant to the modeling of output stem shape as long as the choice of the allomorph follows the decision on whether to change the stem.

## 5.2. Methods

### 5.2.1. Data collection

When an English verb is borrowed into Russian, it must be assigned a stem extension. In order to get a sample of such borrowings, I took all verbs found in the British National Corpus retrieved by searching for “\*x.[vvi]” in the online interface provided by Mark Davies (<http://corpus.byu.edu/bnc/>) where ‘x’ is any letter. The resulting verbs were transliterated into Cyrillic.

For each verb, possible Russian infinitival forms were derived. For instance, if the English verb is *lock*, some possible Russian infinitives are /lotʃitʲ/, /lokitʲ/, /lokatʲ/, /lokovatʲ/ and /lokirovatʲ/. Verbs for which an established Russian form already existed (e.g., *format* > /formatirovatʲ/) were excluded. Existence was determined by the occurrence in either the Reverse Dictionary of Russian (Sheveleva 1974), Big Dictionary of Youth Slang (Levikova 2003), or the present author’s memory. This yielded 472 different verbs. For 56 of them, the final consonant of the English form was a velar, for 99 it was a labial, and for 317 it was a coronal. In the case of the nouns, all possible English monosyllables ending in /k/ or /g/ were created and transliterated into Russian manually. Then possible diminutive forms were created from them and submitted to Google. An additional sample of non-velar-final nouns was then created by matching the distribution of final consonant types (in terms of manner and voicing) and preceding vowels in the sample of velar-final nouns.

The frequencies of the possible infinitives and nominative diminutives on the web were determined by clicking through the pages of results returned by Google to eliminate identical tokens and to allow Google to ‘eliminate similar pages’, which increases

speaker diversity by eliminating results that come from the same server, e.g., different pages from the same bulletin board. In addition, clicking through is necessary when one of the possible forms has a homonym.

Finally, to have a reasonably reliable estimate of the likelihood of failure of velar palatalization before *-i* for each verb, velar-final verbs and nouns that had 10 or fewer tokens containing the palatalizing suffixes were excluded from the sample. This yielded 36 velar-final verbs and 19 velar-final nouns that could undergo velar palatalization and had a reasonably large number of tokens containing the relevant suffixes.

#### 5.2.2. Modeling

The results of the corpus study were modeled using the Minimal Generalization Learner (Albright & Hayes 2003), a model of rule induction and weighting introduced in Chapter 3. While the model is source-oriented, the crucial prediction that velar palatalization should fail before the suffixes that attach to consonants that cannot change is, as we have seen shared between the hypothesis that learners induce competing reliability-weighted rules embodied by the Minimal Generalization Learner and the hypothesis that learners induce positive product-oriented generalizations. One Minimal Generalization Learner was trained to form verbs from bases by attaching the appropriate stem extension and changing the stem as needed while another was trained to form diminutive forms from nouns.

#### 5.2.2.1. Training the model

The model of the stem extension process was presented with the set of stem-verb pairings found in the Reverse Dictionary of Russian (Sheveleva 1974) and/or the Big Dictionary of Youth Slang (Levikova 2003). The Reverse Dictionary contains 125,000 words extracted from the four major dictionaries of Russian that existed in 1965 (Sheveleva 1974: 7). The Slang Dictionary is much smaller, containing 10000 words. The main results presented below held regardless of whether the Reverse Dictionary, the Slang Dictionary, or both were used. Only the results based on the full training set will be presented. Only stems that occurred independently as separate words (or, in the case of feminine nouns, with an inflectional affix) were included but no attempt was made to exclude deverbal nouns, since it is not clear that deverbal nouns are identified as such synchronically. No stem extensions were excluded from the training set. Thus, aside from verbs featuring the highly productive *-i* and *-a*, verbs having *-ova*, *-irova*, *-izirova*, and *-e* were also included. The full training set consisted of 2396 verb-stem pairs, of which 286 stems had final /k/ and 85 had final /g/. There were 22 examples of  $g \rightarrow ʒi$  and 62 examples of  $k \rightarrow tʃi$ . The model of diminutive formation was trained on a set of 1154 diminutive nouns extracted from the Reverse Dictionary of Russian. All diminutive nouns whose base ends in a consonant were extracted regardless of the diminutive suffix used. The Slang Dictionary contains only a very small number of diminutives and thus was not used.

The model requires a feature set to generalize across segments. Two feature sets were devised, the compact set and the redundant set (shown in the appendix). In the compact set, vowels and consonants were defined over the same set of features. Vowels

differed from consonants on sonority (a multivalued feature that also distinguished between low and high vowels) and used the consonantal [place] values to denote frontness (coronality) and roundness (labiality). In the redundant set, featural distance between vowels and consonants was maximized. The results were not significantly affected by the choice of the feature set.

The learner models competition between input-output mappings. Therefore it is crucial to define what is meant by the input and the output. For the present chapter, we are interested in modeling competition between input-output mappings in which some mappings require velar palatalization. The input form for these mappings may or may not have the stem extension already specified. If it does, rules specifying that a velar changes into an alveopalatal compete with rules that say that the consonant stays the same in the context of a stem extension that triggers velar palatalization in the lexicon. If not, rules specifying that a velar changes into an alveopalatal also specify the stem extension. Thus a rule like  $k \rightarrow tʃi$  would compete with  $k \rightarrow ka$  as well as  $C \rightarrow Ci$ .

In addition, the output of the competition can be either a phonetic form, specifying the allophone of /i/ used ([i] or [i]) or a phonemic form, which does not include this specification. Both of these possibilities were examined in modeling but the choice between phonetic and phonemic outputs did not influence the qualitative results. In the case of the diminutive suffixes -ek and -ik, which can be considered allomorphs (or even orthographic variants), it also did not make a difference whether the choice between -ek and -ik followed the stage in which the decision on whether to palatalize the stem was made.

The diminutive suffix *-ek* can attach to words that are already affixed with *-ok*, *-ik*, or *-ek*. Since none of the borrowed English words bear these suffixes, we can consider Russian bases bearing them to be irrelevant and exclude them from the training set or train the model on the full dataset. The qualitative results were not affected by this manipulation, whose effect was only to reduce the predicted use of *-ek*. The graphs below are based on a model that was trained on the full set of Russian diminutives extracted from the dictionary.

The model treats reliability as a value that is conservatively estimated based on the sample of experienced words. It requires a choice of confidence level, which determines the width of the confidence interval around the raw reliability value derived from the lexicon. It then takes the lower bound of the confidence interval as the estimate of reliability. This makes rules that can apply to many words (rules with high scope) more reliable than rules that apply to few words (low scope), other things being equal. The weight of scope relative to raw reliability is determined by width of the confidence interval, which in turn is determined by how confident we want to be that the true reliability of the rule is equal to or greater than the estimated value. For the present data, the three default values (75%, 90%, 95%) produced qualitatively indistinguishable results. The quantitative results reported here are for the 75% confidence interval (which is consistent with Albright and Hayes 2003).

Finally, the model has an optional feature called “impugment”. Impugment is a heuristic designed to punish rules that are internally heterogeneous, generalizing over several internally consistent clusters of forms. Albright and Hayes (2002) give the example of the choice of [t] as the allomorph of the past tense suffix *-ed* in English,

which is selected after voiceless consonants (e.g., [l $\Delta$ f]  $\rightarrow$  [l $\Delta$ ft]) but also (sometimes) after [n] as in [b $\partial$ n]  $\rightarrow$  [b $\partial$ nt]. Albright and Hayes use impugnement to reduce the estimated reliability of the rule stating that [t] is chosen after any consonant, which subsumes the two cases. With impugnement, the model calculates the reliability of the residue (forms covered by the general rule but not by the specific rule), which is taken to be the difference in the numbers of hits between the more general rule and the more specific rule divided by the difference in the scopes of the rules. If the upper confidence limit of the estimated reliability of the residue is lower than the lower confidence limit of the estimated reliability of the more general rule, the reliability value of the general rule is replaced by the reliability value of the residue. Versions of the model with and without impugnement were applied to the data on both diminutive formation and stem extension.

#### 5.2.2.2. Testing the model

The model is presented with the set of English verbs found to be borrowed into Russian in the corpus study. To estimate the probability of a given verb undergoing velar palatalization given that a particular suffix is chosen we can divide the reliability of the most reliable rule that requires palatalization by the sum of its reliability and the reliability of the rule that does not require palatalization but still attaches the same suffix. For instance, suppose the verb is /dig/ and the model has extracted the rules in (3) with reliability estimates shown in parentheses. The only rules that can apply to /dig/ are (a), (d), (e), (h), (i), and (j). Of these, the only rules that require velar palatalization are rules h and i. Rule h is more reliable than rule i, so it would get to apply. Its reliability is .272. The rule that attaches –i without palatalizing the stem-final /g/ is rule j. Its reliability is

.232. Therefore, the predicted probability that the final consonant of /dig/ will be palatalized, given that –i is selected as the stem extension, is  $.272 / (.272 + .232) = 54\%$  (cf. Albright and Hayes 2003:128).

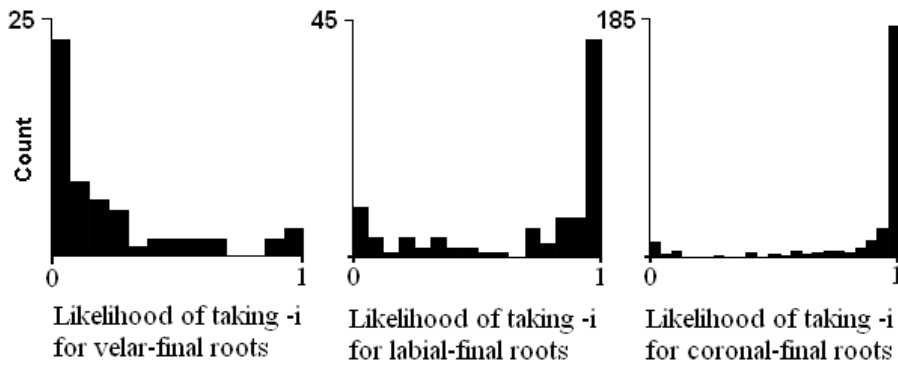
- (3)
- a.  $[\_] \rightarrow a/\{i;l\}g\_ (.723)$
  - b.  $[\_] \rightarrow a/Cag\_ (.718)$
  - c.  $[\_] \rightarrow a/\{l;r\}eg\_ (.718)$
  - d.  $[\_] \rightarrow a/\{i;l;n;r\}g\_ (.670)$
  - e.  $[\_] \rightarrow a/[velar]\_ (.641)$
  - f.  $g \rightarrow \text{ʒ}i/V_{[+back;-high]}\_ (.475)$
  - g.  $g \rightarrow \text{ʒ}i/V_{[-high]}\_ (.350)$
  - h.  $g \rightarrow \text{ʒ}i/V\_ (.272)$
  - i.  $g \rightarrow \text{ʒ}i/[+voice]\_ (.195)$
  - j.  $[\_] \rightarrow i/C_{[+voiced]}\_ (.232)$

### 5.3. Results

Figure 1 shows that most velar-final verbs are highly unlikely to take –i while most labial-final and coronal-final verbs are very likely to take –i. Thus, the stem extension that triggers a stem change in the lexicon is disfavored by the stems that can undergo the change.



Figure 5.1: Histograms showing that most velar-final stems are unlikely to take –i while most labial-final and coronal-final stems tend to take –i.



Since the population distribution is skewed and bimodal, there is no monotonic transformation that will restore normality, which makes standard statistical tests inapplicable, which means that bootstrapping should be done. For this test, I treated the labial-final roots and coronal-final roots as the null population and generated 2000 samples of 56 verbs from this population, calculating mean rate of taking –i in each sample. The mean rate of taking –i in the sample of velar-final stems (33%) falls very far outside the distribution of 2000 samples of 56 verbs from the null population, thus  $p < .0005$  (1/2000). All versions of the model are able to predict that –i is less productive with velar-final stems than with coronal-final and labial-final stems.

Figure 5.2 shows just the velar-final stems that take –i as the stem extension. These are the only stems that undergo velar palatalization in the data, suggesting that the speakers are using a source-oriented generalization mapping velars onto alveopalatals, rather than a purely product-oriented generalization requiring alveopalatals before –i (Pierrehumbert 2006).<sup>60</sup> A product-oriented generalization specifies only the shape of the

<sup>60</sup> The source-oriented generalization could be a rule or a paradigm uniformity constraint.

output, thus imposing no restrictions on what changes can be done to the input to produce the output (for examples of such product-oriented behavior, see Bybee 2001:126-129).

The white bars show the observed likelihood of failure of velar palatalization before –i in various contexts while the dark bars show probabilities of velar palatalization failure predicted by the model. Figure 5.2 shows that velar palatalization is more likely to fail with /g/ than with /k/ ( $t(26)=4.803$ ,  $p<.0005$ ), and when the verb ends in a consonant cluster as opposed to a VC sequence ( $t(22)=3.415$ ,  $p=.003$ ). There is also a trend for the rule to fail more often after front vowels than after back vowels but it is not statistically significant. In other words, speakers tend to retain the velar if it is /g/ and if it is preceded by a consonant. They tend to replace the velar with an alveopalatal if it is a /k/ preceded by a vowel, especially if the vowel is back.

Despite the fact that the model is trained on a lexicon in which velar palatalization is exceptionless, the model predicts that velar palatalization will not be exceptionless with the borrowed verbs. Mean rate of failure of velar palatalization varies between 43% and 62% depending on parameter settings and approximates the actual mean rate of failure of velar palatalization in the data (56%).

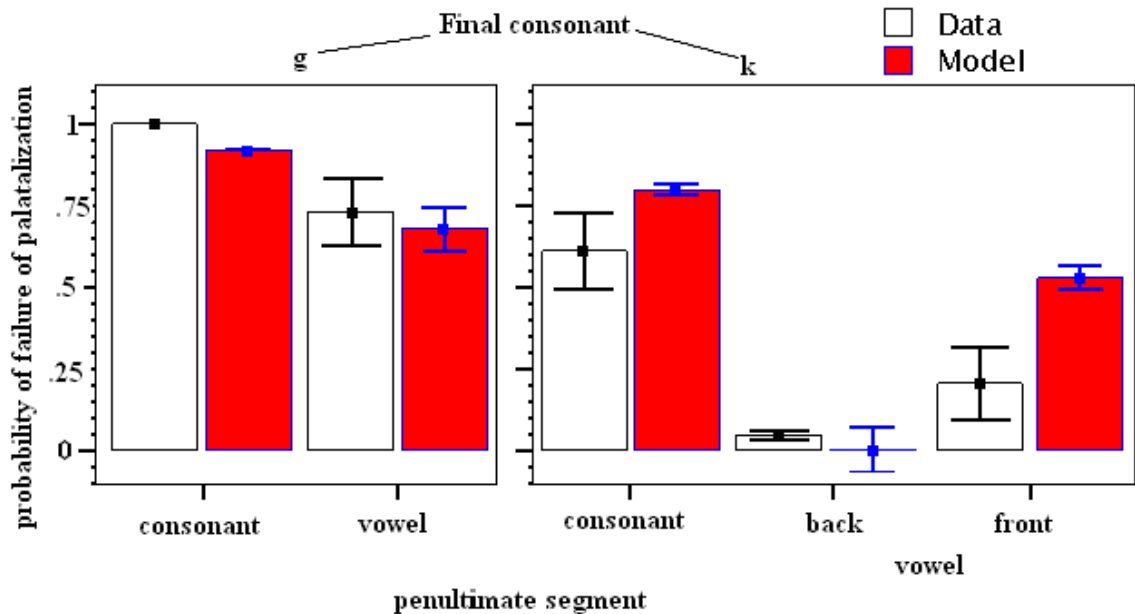
While the mean predicted rate of failure for velar palatalization is similar to the observed rate of failure, the model's predictions are less variable than the data. In order to make them comparable, failure rates predicted by the model were rescaled to have the same standard deviation as the observed failure rates.<sup>61</sup> The qualitative results shown in Figure 5.2 hold for all versions of the model that assume that the stem extension and the stem shape are chosen simultaneously. These versions of the model correctly predict that

---

<sup>61</sup> This is why one of the error bars goes negative.

velar palatalization is more likely to fail when the stem ends in a consonant cluster than when it ends in a single consonant, that penultimate front vowels disfavor palatalization compared to back vowels, and that /k/ is more likely to be palatalized than /g/ (however, all versions of the model underestimate the difference between /k/ and /g/).<sup>62</sup> If the stem extension is chosen first with the decision on whether to change the stem consigned to a subsequent decision stage, the predicted rate of failure of velar palatalization is not significantly affected but the effect of penultimate segment identity disappears.

Figure 5.2. Observed (white bars) vs. predicted (grey bars) probabilities of failure of velar palatalization before the stem extension –i depending on segmental content of the stem.



Observed and predicted rates of failure of velar palatalization in front of

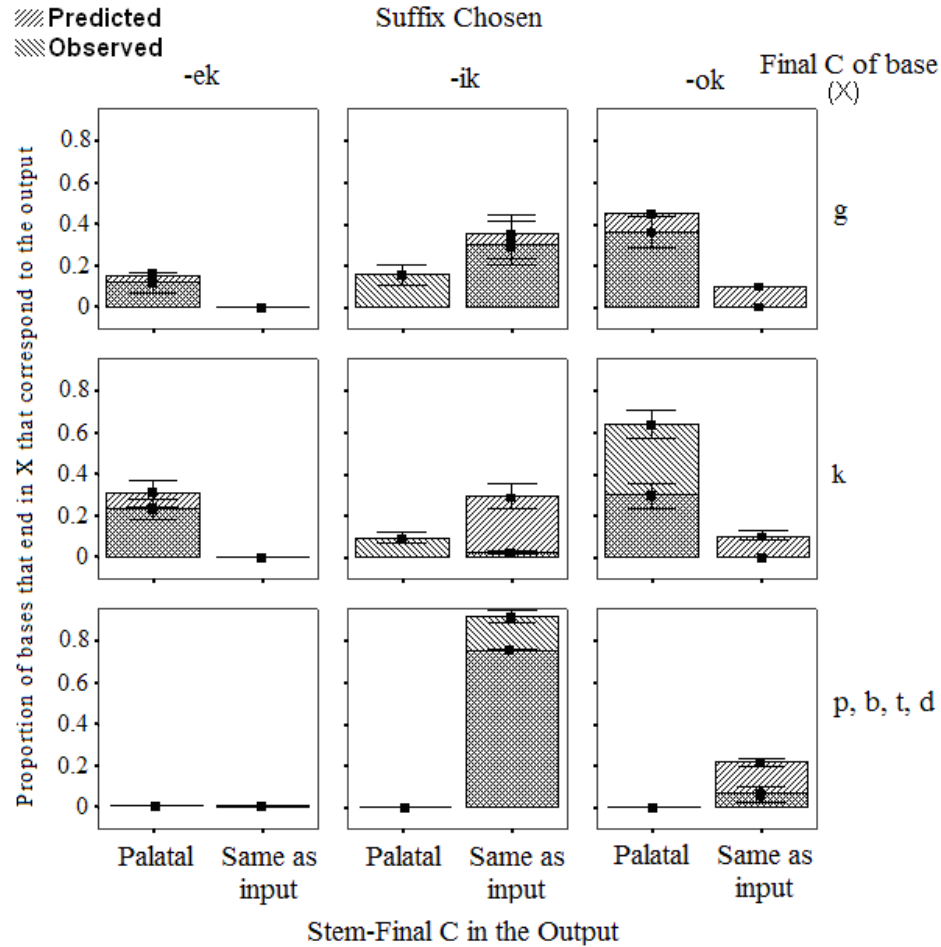
diminutive suffixes are shown in Figure 5.3. As with the stem extensions, velars are the

<sup>62</sup> This problem is exacerbated when impugment is used. While versions of the model without impugment are able to predict that /k/ is more likely to be palatalized than /g/, versions with impugment incorrectly predict the opposite result except for stems with a penultimate back vowel.

only consonants that change into alveopalatals, suggesting a source-oriented generalization. The rate of failure of velar palatalization is significantly higher before the suffix -ik (mean rate of failure = 40%) than before the suffix -ok (mean rate of failure = 0%), according to the paired-samples Wilcoxon signed ranks test ( $Z(16)=3.516$ ,  $p<.0005$ ). Failure of palatalization (which only happens before -ik) is more likely with /g/ (67%) than with /k/ (29%),  $t(15)=2.496$ ,  $p=.025$ . The likelihood of using -ik is lower after /k/ than after /g/ ( $t(17)=5.729$ ,  $p<.0005$ ) and is higher after non-velars than after velars ( $t(45)=12.461$ ,  $p<.0005$ ). Thus, the suffixes -i and -ik tend to attach to non-velar-final inputs and often fail to trigger velar palatalization. The suffixes -ek and -ok tend to attach to velar-final inputs and are strong triggers of velar palatalization. Furthermore, in both the domain of verbal stem extensions and nominal diminutives, the productivity of  $k \rightarrow tʃ$  is greater than the productivity of  $g \rightarrow ʒ$ .

The model successfully learns that -ik is disfavored by velars and that palatalization is likely to fail only if -ik is chosen as the suffix, although the rate of failure of velar palatalization before -ik is overestimated. It predicts that -ek should be more productive with bases ending in /k/ than with bases ending in /g/, a numerical trend in the data. It fails to predict that /k/ is more likely to undergo palatalization and less likely to be followed by -ik than /g/. These predictions are parameter-independent, holding for all versions of the model.

Figure 5.3. Relative likelihoods of various base-diminutive mappings for velar-final and non-velar-final bases of diminutive nouns (likelihoods sum to one across each row of panels).



#### 5.4. Explaining successes and failures of the model

In the present study, the MGL is used as an example of a general class of models that postulate that input-output mappings are involved in a competition that is resolved by the mappings' relative reliability. Therefore it is important to determine the extent to which the successes and failures of the MGL are due to its reliance on this assumption

and to what extent competition between product-oriented generalizations can produce for the same results.

In order to explain why the model performs the way it does let us examine the rules that it abstracts from the lexicon and uses when a velar-final verb is presented. The full list of applicable rules for [g]-final verbs is shown in (3) above. For both [k]-final and [g]-final verbs, there is only one rule that favors adding –i and leaving the final consonant of the stem unchanged. For /g/-final roots, this is the rule  $C_{[+voiced]} \rightarrow C_{[+voiced]}i$  and for /k/-final roots this is the rule  $C \rightarrow Ci$ . Thus, in order for the more specific rules requiring /k/ to change into /tʃ/ or /g/ to change into /ʒ/ to fail, they must lose to an extremely general rule. For this outcome to be likely, 1) a very general rule must be extracted from the lexicon, 2) it should be quite reliable relative to the less general rules requiring stem changes, and 3) it must compete with those rules.

In the Russian lexicon used to train the model, coronal-final and labial-final stems tend to take –i while velar-final stems tend to take –a. Since most stems in the lexicon end up taking –i, the model extracts a very general rule  $C \rightarrow Ci$  and assigns it a moderate reliability. On the other hand, the fact that velar-final stems favor –a drives down the reliabilities of rules that add other stem extensions to velar-final stems. This includes the rules that add –i and change the root-final consonant. As a result, these rules will sometimes lose the competition for application to the more general rule  $C \rightarrow Ci$ . Thus, the model predicts that velar palatalization will often fail before an affix if and only if the affix is more productive after non-velars than after velars. This holds for the stem extension –i and the diminutive suffix –ik but not for the diminutive suffixes –ek and –ok. Therefore, the model correctly predicts that velar palatalization should fail often

before –i and –ik and rarely before –ek and –ok. This prediction follows directly from the hypothesis that input-output mappings compete with the outcome determined by reliability. The prediction also follows from the hypothesis that velar palatalization is driven by positive product-oriented generalizations such as “if the suffix is X, the preceding consonant must be Y”. In the case of –ik and –i, the preceding consonant is rarely an alveopalatal, while in the case of –ek and –ok, it almost always is an alveopalatal (derived from a velar). Thus, velar palatalization is predicted to often fail before –i and –ik but not before –ek and –ok.

The model systematically fails to capture the difference in rate of palatalization between /k/ and /g/, which is observed in both stem extension and diminutive formation. In both cases, the rate of palatalization is underestimated for /k/. Palatalization of /k/ to [tʃ] is much more phonetically natural than palatalization of /g/ to [ʒ]. Bhat (1974:41) notes that velar stops generally become affricates or remain stops as a result of palatalization and if a language palatalizes voiced velars, it also palatalizes voiceless velars but not necessarily vice versa, which suggests that the  $g \rightarrow ʒ$  change is typologically marked. Hock (1991:73-77) proposes that palatalization arises when a fronted velar stop develops a fricative release, with the conversion of the resulting affricate to a fricative being a later development. In addition, the voiceless velar stop [k] is more acoustically similar to [tʃ] than [g] is to [dʒ] in terms of peak spectral frequency and duration of aperiodic noise, leading listeners to misperceive [ki] as [tʃi] much more often than they misperceive [gi] as [dʒi] (Guion 1998). Thus, [g] and [ʒ] can be argued to be more perceptually and articulatorily distinct than [k] and [tʃ] and the  $g \rightarrow ʒ$  alternation can be

argued to be less phonetically natural than the  $k \rightarrow tʃ$  alternation. Phonetic naturalness has been argued to influence learnability of an input-output mapping when the reliability of the mapping is controlled (Finley 2008, Wilson 2003, 2006). The [k]/[g] asymmetry observed in Russian may be another case of this phenomenon. If the palatalization rule for [g] is more difficult to learn than the rule for [k] and the diminutive suffixes –ok and –ek do not permit a velar to precede it without a loss of naturalness, the speaker is driven to choose –ik as the diminutive suffix after [g] more often than after [k], accounting for the relatively high productivity of –ik following [g]. Alternatively, the low productivity of  $g \rightarrow ʒ$  relative to  $k \rightarrow tʃ$  may be explained by the fact that there are fewer examples of products containing [ʒ] than [tʃ] in the lexicon. Thus, if product-oriented generalizations are induced by learners of Russian, as suggested by artificial grammar learning data, the generalization that a suffix Y must be preceded by [ʒ] would be weaker than the generalization that the same suffix must be preceded by [tʃ]. Phonetic naturalness alone cannot account for the data because velar palatalization is much more likely before –ok than before –ik, despite the fact that [o] is a less natural trigger of palatalization than [i].

Another shortcoming of the model is that it underpredicts the rate of velar palatalization before the suffix –ik, especially when –ik attaches to a /k/-final noun. This prediction follows from the fact that –ik never attaches to velar-final inputs in the native lexicon and thus is predicted not to trigger velar palatalization. There are at least two possible explanations for why it should still sometimes trigger velar palatalization. First, the alveopalatal stem-final consonant may be used as a diminutive marker in its own right, especially when the consonant is /tʃ/, which is expected if learners of Russian



induce product-oriented generalizations about typical diminutives, which often end in [tʃik], [tʃek] or [tʃok]. This hypothesis is supported by the fact that some labial-final bases take -tʃik rather than -ik as the diminutive marker, e.g., sup ‘soup’ → suptʃik. The same process of reanalyzing a part of the stem as belonging to the suffix is also observed with other suffixes in Russian as documented in Kapatsinski (2005: 153-156). For instance, Russian has an agentive suffix –nik, similar to English –er, as in *les* ‘forest’ → *lesnik* ‘forester’. One can form verbs from nouns ending in –nik by adding –a-tʃ and palatalizing the [k]. These verbs mean ‘do what an X-nik does’. However, one can also now form verbs with the same meaning by adding –nitʃatʃ to the root (X) above directly. For instance, consider the pair *nerv* ‘nerve’ → *nervnitʃatʃ* ‘to be nervous’. There is no noun \**nervnik* from which the verb could have been formed. Rather, it appears that –nitʃatʃ has fused into a single unit due to its frequent occurrence in verbs with a certain meaning. Similarly, there is a verb *xozʃajnitʃatʃ* ‘to behave like you own the place’ but the noun for ‘owner’ is not \**xozʃajnik* but *xozʃain*. See Kapatsinski (2005: 153-156) for examples with other suffixes.

An alternative explanation for the overuse of [tʃ] before –ik is that –ik and –ek are phonetically identical due to being unstressed. Despite being phonetically identical to –ik, –ek is a much stronger trigger of velar palatalization, thus the two suffixes must constitute different choices in phonology. However, it is possible that some instances of –ik in the (written) data can be cases in which the speaker chose –ek (which triggered velar palatalization) and misspelled it as the more frequent –ik.

## **5.5. The affix and the stem shape are chosen simultaneously**

### **5.5.1. The issue of stages in grammatical processing**

The types of generalizations that can and are likely to enter the grammar when the language provides evidence for them is one of the basic parameters defining the shape of the grammar. Another basic parameter is how the grammar is divided into a series of sets of competing generalizations, which may interact with the types of generalizations that the grammar contains.

The issue of whether something is processed through a single transducer, i.e., a single set of competing input-output mappings, or through a series of transducers in which each subsequent transducer takes in the output of the previous transducer as input is a general issue throughout cognitive science (e.g., Sternberg 1967, 1998) and has been identified as the central issue of phonological theory.

‘We expect [phonology] to answer... in particular questions like ‘what is the sequence of stages traversed... by a speaker in the course of producing utterances... from mnemonic elements merged and structured by the syntax of that speaker’ (Bromberger & Halle 1997: 118).

Johnson (1972) and Karttunen (1993) show that a phonological derivation is equivalent to a sequence of finite-state transducers. The distinctive feature that allows us to tell that some mapping is underlain by two separate transducers is if we can make modifications to one transducer without modifying the other (Sternberg 1998:706). In the present section, I shall consider the issue of whether phonology should be separated from

morphology in the domain of Russian velar palatalization, i.e., if there is a separate stage of affix selection that precedes the stage in which the speaker decides on what phonological consequences (if any) the combining of the affix with a stem would have. However, the general logic can be applied to other domains, linguistic and non-linguistic, as well.

The traditional view of grammar is that phonological changes associated with the concatenation of a pair of morphemes are triggered when the morphemes have already been chosen (e.g., Halle and Marantz 1993). Thus, for instance, the speaker may decide to palatalize or not to palatalize a stem-final consonant after having chosen to affix –i to the stem. The alternative is that the speaker is deciding between complete input-output mappings. Thus, the speaker presented with a new stem ending in [k] would be deciding between producing something ending in [tʃi], something ending in [ka], or something ending in [ki].

The latter alternative is expected under a strict interpretation of Optimality Theory where there are no intermediate stages in the derivation if one way to solve a phonological issue is to choose a different affix (Prince and Smolensky 1993/2004) as well as under Declarative Phonology (Scobbie et al. 1996), but this feature of OT has been abandoned in many recent approaches in an attempt to capture opacity (e.g., Benua 1997, Kiparsky 2000, McCarthy 2007). Opaque alternations require intermediate stages in the derivation in order to be exceptionless. The purported existence of the intermediate stages has been a primary argument in favor of intermediate stages in the derivation and rule-based approaches in general (Chomsky 1995: 223, Kiparsky 1973, 2000, see McCarthy 2002:184 for a comprehensive reference list), although the psychological

reality of the exceptionless serially ordered rules has not been shown.<sup>63</sup> Rule-based approaches to grammar have always used intermediate stages of derivation, with even Natural Generative Phonology, an approach geared towards reducing abstractness and increasing restrictiveness on the range of possible patterns, allowing for a rule to take an output of a preceding rule as input (Hooper 1976b:18). There has been some (theoretical) disagreement about whether rules that require an intermediate stage in the derivation should be harder to learn (Kiparsky 1973 vs. Kaye 1974, Kisseberth 1973, Thomason 1976).

On the other hand, analogical and connectionist approaches have favored a single-stage approach to morphophonology where there is a single transducer converting a form X into a form Y without some intermediate form that is not attested in actual speech (Bybee 1985, 2001, Rumelhart and McClelland 1986, Skousen 1989, 2002). For instance, the Rumelhart and McClelland (1986) network does not first decide to use the regular past tense affix –ed and then decide on the phonetic shape of the suffix depending on whether the preceding consonant of the stem is voiceless. Rather, the allomorphs of the regular suffix compete with each other and the irregular stem changes in a single decision stage (a feature of the model Pinker and Prince 1988 criticize). Of course, there is nothing in the connectionist approach that precludes stacking two networks so that the first learns pure morphology, while the second takes the output of the first network and handles phonological alternations (e.g., Gasser 1997).

---

<sup>63</sup> While Ettlinger (2009) has recently convincingly demonstrated that opaque rule interactions are learnable, he notes that the learned patterns can be handled by single-stage mechanisms, including four-part analogy.

### 5.5.2. Russian velar palatalization: A single-stage process

Interestingly, the penultimate segment effect on palatalization rate for stem changes is only obtained if a particular assumption is made about the sequence of processing stages, allowing us to distinguish between the two models in (4). Each stage in (4) is modeled by a separate Rule-Based Learner trained on the relevant input-output mappings.

(4)

Two-stage Model:

Stage I: Choose the suffix based on the borrowed base:

$[\ ] \rightarrow \text{suffix} / \text{Base\_}$

Stage II: Modify the base to fit the suffix:

$/\text{Base}/ \rightarrow [\text{Base}] / \_ \text{suffix}$

One-stage Model:

Stage I: Choose the suffix based on the borrowed base and modify the base to fit the suffix:

$/\text{Base}/ \rightarrow [\text{Base}] + \text{suffix}$

The effects of the penultimate segment shown in Figure 7.2 (and only those effects) are not predicted if we assume that the stem extension (-i vs. -a) is chosen first, followed by the decision on whether to change the stem (the two-stage model). Let us now examine why this is the case.

In the one-stage model, the palatalizing rules that are applicable to a given stem differ in their reliability, with some rules being more likely to outcompete the general non-palatalizing rule than others. For instance, the stem /overlok/ is likely to undergo velar palatalization because the most reliable palatalizing rule that can apply to it ( $k \rightarrow tʃi / [+cons; +son] o \_$ ) is very reliable (.805) and can easily outcompete the applicable general rule ( $C \rightarrow Ci$ ) with its .2 reliability. By contrast, the most reliable palatalizing rule that can apply to the stem /drink/ ( $[+son]k \rightarrow [+son]tʃi$ ) has a reliability of only .125, which means that it is likely to lose to the more general rule  $C \rightarrow Ci$  whose reliability is .2, resulting in failure of palatalization.

Suppose instead that the suffix has already been chosen and it is -i. The model now needs to decide whether to palatalize the stem. Interestingly, although the rules changing  $k \rightarrow tʃ$  and  $g \rightarrow ʒ$  are exceptionless and thus have a reliability value of 1, they can still sometimes lose to the more general rule “do nothing” because the reliability of “do nothing” is also quite high (86%). This is because most stems in the lexicon take -i and remain the same after the addition of -i.

However, with the stem change choice following affix choice raw reliability predicts no effect of penultimate segment identity. In this model, the reliabilities of all stem-changing rules are at 1, regardless of penultimate segment identity because velar palatalization never fails before -i in the lexicon on which the model is trained. Therefore, the model can capture segmental context effects only if they correspond to differences in rule type frequency (i.e., the number of word pairs supporting the rule), which in this case they do not. Thus, the effect of the penultimate segment is accounted for by the model only if the stem change and the affix are chosen during a single decision

stage in which the palatalizing rules compete with rules adding other stem extensions, such as –a (the one-stage model).

The finding that morphophonological processing appears to happen in a single processing stage is predicted by theories in which generalizations induced by learners are product-oriented, since the product is typically claimed to be a surface form rather than a form at an intermediate stage in the derivation (Bybee 2001, Burzio 2002). It is possible to hypothesize that there are multiple serially arranged transducers, each of which contains product-oriented generalizations, underlying grammatical processing and, as we shall see in the following chapter, this model would make the same incorrect predictions for Russian velar palatalization as a multistage source-oriented model. However, it is much easier to imagine why a learner would make generalizations about typical forms s/he experiences than why s/he makes generalizations about typical forms that s/he produces internally as an intermediate stage in the derivation. Knowledge about typical properties of experienced forms of a certain grammatical category, such as plurals, would help with identifying the grammatical category of a form in perception, while typical properties of internally generated intermediate forms that are never articulated have no such useful function.

## **5.6. Conclusion**

As we have seen in Chapter 2, the hypothesis of reliability- or type-frequency-based competition between rules or product-oriented generalizations predicts that a morphophonemic alternation will lose productivity if the triggering affix comes to increasingly attach to forms that cannot alternate due to not having the alternating

segment. This hypothesis is supported by experimental data from artificial grammar learning, which establish the direction of causation, and correlational data from spontaneous loanword adaptation in a natural language, which lend the experimental data some ecological validity. The present data place three restrictions on the theory of morphophonology. First, the affix and the ‘triggered’ stem change are actually chosen at the same time, rather than the affix being chosen first and then triggering or failing to trigger a stem change. Second, like in artificial grammar learning, the choice between rules must be probabilistic in nature, rather than the subjects always applying the most reliable applicable rule because the most reliable applicable rule is the palatalizing rule. Finally, existing morphologically complex words are stored in memory and retrieved for production (e.g., Bybee 1985, 2001, Halle 1973, Hooper 1976a, Vennemann 1974). Stochastic choice between generalizations does not optimize accuracy (Hudson Kam & Newport 2005, Norris & McQueen 2008), yet derived forms of existing words are always or almost always produced correctly.

Like artificial languages studied in Chapter 3 and Chapter 4, Russian provides evidence for both source-oriented and product-oriented generalizations. Source-oriented generalizations are supported by the fact that only velars are changed into alveopalatals when a diminutive suffix or a verbal stem extension is added. Thus Russian speakers must know that only velars change (into alveopalatals). Product-oriented generalizations are supported in particular by the phenomenon of affix fusion (see also Kapatsinski 2005: 153-156) where parts of the stem that often co-occur with the following suffix fuse to the suffix forming a larger suffix that can be used to derive new words with the same meaning as the words in which reanalysis has occurred. Thus Russian speakers are



sensitive to typical shapes of words from particular grammatical and semantic categories, such as diminutives. Product-oriented generalizations are also able to explain overuse of [tʃ] before the diminutive suffix –ik, higher productivity of  $k \rightarrow tʃ$  relative to  $g \rightarrow ʒ$  and the single-stage nature of Russian morphophonology, which a purely source-oriented account leaves unexplained.

## APPENDIX 1: FEATURE SETS

### Compact set

Segment	cons	son	lab	cor	pal	vel	voice	cont	long
b	1	0	1	0	0	0	1	0	1
d	1	0	0	1	0	0	1	0	1
g	1	0	0	0	0	1	1	0	1
p	1	0	1	0	0	0	0	0	1
t	1	0	0	1	0	0	0	0	1
k	1	0	0	0	0	1	0	0	1
tʃ	1	0	0	0	1	1	0	-1	1
ts	1	0	0	1	0	0	0	-1	1
ʃ	1	0	0	0	1	1	0	1	1
ʃtʃ	1	0	0	0	1	1	0	1	2
s	1	0	0	1	0	0	0	1	1
f	1	0	1	0	0	0	0	1	1
x	1	0	0	0	0	1	0	1	1
v	1	0	1	0	0	0	1	1	1
z	1	0	0	1	0	0	1	1	1
ʒ	1	0	0	0	1	1	1	1	1
r	1	2	0	1	0	0	1	0	1
l	1	2	0	1	0	0	1	1	1
m	1	1	1	0	0	0	1	0	1
n	1	1	0	1	0	0	1	0	1

j	1	3	0	0	1	0	1	1	1
a	0	4	0	0	0	1	1	1	1
e	0	4	0	1	1	0	1	1	1
i	0	3	0	1	1	0	1	1	1
o	0	4	1	0	0	1	1	1	1
u	0	3	1	0	0	1	1	1	1
ɪ	0	3	0	0	0	1	1	1	1
<sup>j</sup> a	0	4	0	0	1	1	1	1	1
<sup>j</sup> u	0	3	1	0	1	1	1	1	1
<sup>j</sup> o	0	4	1	0	1	1	1	1	1

#### Redundant set

Seg.	cons	nas	lat	son	cor	pal	vel	voice	cont	long	front
	high	round									
b	1	0	0	0	0	0	0	1	0	1	-1
	-1	-1									
d	1	0	0	0	1	0	0	1	0	1	-1
	-1	-1									
g	1	0	0	0	0	0	1	1	0	1	-1
	-1	-1									
p	1	0	0	0	0	0	0	0	0	1	-1
	-1	-1									

t	1	0	0	0	1	0	0	0	0	1	-1
	-1	-1									
k	1	0	0	0	0	0	1	0	0	1	-1
	-1	-1									
tʃ	1	0	0	0	0	1	1	0	-1	1	-1
	-1	-1									
ts	1	0	0	0	1	0	0	0	-1	1	-1
	-1	-1									
ʃ	1	0	0	0	0	1	1	0	1	1	-1
	-1	-1									
ʃtʃ	1	0	0	0	0	1	1	0	1	2	-1
	-1	-1									
s	1	0	0	0	1	0	0	0	1	1	-1
	-1	-1									
f	1	0	0	0	0	0	0	0	1	1	-1
	-1	-1									
x	1	0	0	0	0	0	1	0	1	1	-1
	-1	-1									
v	1	0	0	0	0	0	0	1	1	1	-1
	-1	-1									
z	1	0	0	0	1	0	0	1	1	1	-1
	-1	-1									

3	1	0	0	0	0	1	1	1	1	1	-1
	-1	-1									
r	1	0	0	1	1	0	0	1	1	1	-1
	-1	-1									
l	1	0	1	1	1	0	0	1	1	1	-1
	-1	-1									
m	1	1	0	1	0	0	0	1	0	1	-1
	-1	-1									
n	1	1	0	1	1	0	0	1	0	1	-1
	-1	-1									
j	1	0	0	1	0	1	0	1	1	1	-1
	-1	-1									
a	0	-1	-1	1	-1	0	-1	-1	-1	1	0
	0	0									
e	0	-1	-1	1	-1	1	-1	-1	-1	1	1
	0	0									
i	0	-1	-1	1	-1	1	-1	-1	-1	1	1
	1	0									
o	0	-1	-1	1	-1	0	-1	-1	-1	1	0
	0	1									
u	0	-1	-1	1	-1	0	-1	-1	-1	1	0
	1	1									

i	0	-1	-1	1	-1	0	-1	-1	-1	1	0
	1	0									
j <sub>a</sub>	0	-1	-1	1	-1	1	-1	-1	-1	1	-1
	0	0									
j <sub>u</sub>	0	-1	-1	1	-1	1	-1	-1	-1	1	-1
	1	1									
j <sub>o</sub>	0	-1	-1	1	-1	1	-1	-1	-1	1	-1
	0	1									

## APPENDIX 2: THE FULL SET OF RULES EXTRACTED AND USED BY ONE VERSION OF THE MODEL

Note: the program outputs segment lists, rather than featural descriptions as contexts, thus the descriptions are induced by me and are at the moment not necessarily the ones generated internally by the model. For instance, I simplified the descriptions of vowels in terms of the features ‘high’ and ‘back’ instead of the sonority-by-place classification system used by the model.

a/{i;l}g_ .723	tfi/{d;n;r}o_ .765	a/[-cont;-son]_ .474
a/Cag_ .718	tfi/C[+Labial;-son]a_ .610	i/[+round]b_ .878
a/{l;r}eg_ .718	tfi/{o;a}_ .537	i/[+back]b_ .804
a/{i;l;n;r}g_ .670	tfi/V[-high]_ .187	i/[+Labial;+voice;-son]_
a/[velar]_ .641	tfi/[+son]_ .125	.78
fi/V[+back;-high]_ .475	tfi/[+cont]_ .092	a/[son<max,+vce,+Cor]
fi/V[-high]_ .350	i/C_ .200	d_ .4
fi/V_ .272	a/[+Dorsal]_ .412	a/[-son]end_ .718
fi/[+voice]_ .195	i/[+Coronal;+cont]_	a/[-Labial;-cont;-son]_
i/C[+voiced]_ .232	.586	.368
a/C[-velar]ek_ .916	a/[+Dorsal;-voice]_ .296	i/[+Labial]d_ .797
a/C[-ant]k_ .908	a/Vrtf_ .68	i/[+Lab;son<max]d_
a/V[+back;-high]rk_ .870	a/{i;l;r}tf_ .641	.877
a/Vrk_ .848	a/{m;n;r}tf_ .615	i/{d;n;r}_ .61
a/Cek_ .845	i/[+Cor;+Dor;-voice]_	i/{l;r}d_ .852
a/{m;n;r}k_ .833	.704	a/{f;k;p;x}_ .604
a/V[-back]k_ .831	i/[-back]tf_ .805	i/{V[-round];j}f_ .872
a/[-voice;-son]ak_ .825	i/[-back;-high]tf_ .872	i/{f;v}_ .828
a/[+nasal]ok_ .786	i/[+son]f_ .768	a/[-Labial]_ .224
a/[-velar]k_ .733	i/V[-high]f_ .813	a/{f;v;x}il_ .57
a/[+velar;-voice]_ .641	i/V[-high;+back]f_ .845	i/[-Labial]ul_ .898
tfi/{l;r}o_ .805		i/C[-Labial]ul_ .909

i/{a;e}l_ .896	a/{a;y}p_ .803	i/[+voice;-cont;-Vel]as_
i/[+voice;-Coral]l_ .814	a/C <sub>[-Labial]</sub> ap_ .887	.825
i/[+Labial;-son]l_ .872	a/{i;y}p_ .82	i/[+Labial]s_ .765
i/{l;n;r}_ .795	a/{l;r}ip_ .825	i/[+cont;-Velar;son<j]_
a/[-cont;-Coronal]_ .335	i/[-Cor;-son;+cont]op_	.724
a/{e;i;j}m_ .604	.718	i/{l;n;r}os_ .805
a/im_ .825	i/[-Velar]_ .546	i/{u;i}s_ .852
i/[+round]m_ .823	i/[+Cor;-cont]op_ .786	a/[+cont;-son]ast_ .57
i/[+back;-high]m_ .804	a/[-son]ir_ .887	a/[-son;-Velar]it_ .607
i/C <sub>[+Coronal;-Dorsal]</sub> m_ .852	a/{i;j}r_ .746	a/[+voice;-son;-Vel]ist_
i/{d;g;n;r}am_ .825	i/[-Labia l ;-son]or_ .953	.786
i/{m;n;r}_ .75	i/Vr_ .699	a/[son<=nasal]ot_ .746
a/{f;j}n_ .57	i/{u;a;e}r_ .95	a/[son<=nasal;-cont]ot_
a{-cont;-Coronal}in_	i/{a;o}r_ .944	.776
.786	i/[+Velar;son>min]r_	a/{b;p}et_ .718
i/[-Coronal]an_ .97	.474	a/{f;k;p;x}et_ .526
i/[-high]n_ .817	a/[-Cor;-son]is_ .718	a/{i;i}st_ .61
i/{[-round];j;l}n_ .674	a/[-voice;-son]_ .362	a/{k;p;t}_ .375
i/[-round]n_ .687	i/[-Coronal]s_ .743	a/let_ .57
i/{n;r}_ .79	i/[+back]s_ .762	i/Vnt_ .887
a/[-V <sub>[-high]</sub> ]p_ .649	i/[+voice;son<j;-Vel]es_	i/Cat_ .719
a/[-V <sub>{-high;+round}</sub> ]p_ .792	.765	i/[-cont]at_ .916
a/[-round]p_ .73		i/[-cont;-Velar]t_ .604



i/[C[-Coronal]ost\_ .852  
i/[-cont;-Labial]t\_ .58  
i/{d;n;r;t}\_ .563  
i/{j;n;l;r}t\_ .845  
i/{l;v;z}ot\_ .706  
i/{o;u}st\_ .739  
a/[-son;-Coronal]\_ .347  
i/[+Velar]v\_ .846  
i/{i;i;j;l}v\_ .872  
i/{l;r}ov\_ .872  
a/{e;i;l}z\_ .581  
a/[-son;-Labial;-voice]\_  
.359  
i/Coz\_ .845  
i/[son<=nasal]oz .916  
i/[-Velar;+voice]\_ .671  
i/{o;u}z\_ .795

## CHAPTER 6

### GENERAL DISCUSSION AND CONCLUSION

#### **6.1. Formal and substantive biases in learning morphophonology**

In this thesis, I examined velar palatalization in artificial languages acquired in a learning situation that is expected to favor source-oriented generalizations (i.e., directional generalizations over pairs of words from the same morphological family), artificial languages acquired in a learning situation expected to favor product-oriented generalizations (i.e., generalizations about typical shapes of wordforms belonging to the same cell in a morphological paradigm), and in a natural language (Russian). In all cases, the choice between competing generalizations was shown to be stochastic in nature, with the competing generalizations weighted by the lexicostatistical evidence for them. This finding supports probabilistic and usage-based approaches to grammar (e.g., Albright & Hayes 2003, Bybee 1985, 2001, Hayes & Wilson 2008, Pierrehumbert 2006) and is contrary to the traditional view in morphology and phonology that knowledge of grammar is knowledge of what is possible and not what is probable (e.g., Chomsky & Halle 1968, Halle & Marantz 1993, Plag 2003). Furthermore, the learners were shown to obey generalizations at a rate that is proportional to the relative amount of evidence supporting them, rather than choosing the best-supported generalization 100% of the time (cf. Albright & Hayes 2003).

Generalization is not minimal in the present study. This is a violation of the popular Subset Principle (Albright & Hayes 2003, Berwick 1986, Dell 1981, Hale & Reiss 2003, 2008, Langacker 1987), and is specifically inconsistent with Maye et al.'s (2008: 31) hypothesis that "Infants appear to extract the featural properties of the input

speech, while *adult learning may be restricted to the segmental level*" (emphasis mine).<sup>35</sup> In Chapter 3, subjects are shown to generalize from singulars ending in velars to singulars ending in alveopalatals and from singulars ending in alveopalatals to singulars ending in velars or alveolars. Thus they observe that –i attaches to velars and conclude that it would attach to alveopalatals at least as often. They observe examples of  $tʃ \rightarrow tʃi$  and take those examples to support adding –i to velars and mapping of [t] onto [tʃi], unlike the Minimal Generalization Learner (Albright & Hayes 2003), which comes up with a much more constrained rule that restricts itself to alveopalatals. In Chapter 4, velar palatalization is extended to alveolar and sometimes even labial inputs. In Chapter 5, the most common diminutive suffix –ik is extended by Russian speakers to velar-final stems, even though such stems are expected to take –ok or –ek based on the dictionary. Some of the apparent overgeneralizations, including the generalization from  $tʃ \rightarrow tʃi$  or  $k \rightarrow tʃi$  to  $t \rightarrow tʃi$  are perhaps expected if it is assumed that the output of perception is not the identity of the most probable percept but a probability distribution over possible percepts, as proposed by Bayesian approaches (Kruschke 2008, Levy 2009), such that a perceiver hearing [tʃi] assigns some probability to having heard [tʃi]. Others, including the generalization from  $k \rightarrow tʃi$  to  $p \rightarrow tʃi$  appear to be “genuine” overgeneralization due solely to product-oriented schemas like “plurals must end in [tʃi]”, and are much less frequent.

The present data provide very little support for the idea that generalization is constrained by phonetic naturalness (Wilson 2003, 2006). In artificial languages,

---

<sup>35</sup> Maye et al. (2008) cite Maye & Gerken’s (2001) finding that VOT discrimination learning based on a bimodal VOT distribution does not appear to generalize across places of articulation. It may be the case that the type of learning examined by Maye and colleagues is less prone to overgeneralization by adults, although more work needs to be done.

palatalization is often extended to /t/ but almost never to /p/. While this might be explained by knowledge that /p/ is unlikely to palatalize in human languages, a plausible alternative explanation for this finding may lie in the greater similarity between [tʲi] and [tʃi] than between [pʲi] and [tʃi]. Thus, [tʃi] is likely to be assigned a higher probability increment during the perception of [tʲi] than during the perception of [pʲi]. In Russian, k/tʃ alternation is much more productive than the g/ʒ alternation, which is not predicted by the Minimal Generalization Learner. This may be explained by knowledge of the unnaturalness of the g/ʒ alternation. A possible alternative explanation may be provided by product-oriented generalizations, which are not extracted by the Minimal Generalization Learner, since [ʒi] is much less frequent than [tʃi]. In any case, the influence of knowledge about phonetic naturalness appears to be quite limited, since velar palatalization in Russian is less productive before the natural triggers, -i and -ik, than before the unnatural trigger -ok, a pattern that is explained by differences in the lexical distribution of the suffixes.

Elicited production was more product-oriented when a relatively small number of different wordforms was presented in random order than when a large number of different wordforms was presented in singular-plural pairs. Thus, the relative weighting of source-oriented and product-oriented generalizations was shown to be influenced by the way in which the learners experienced language. Thus, the grammar extracted by a learner exposed to a lexicon is influenced not only by the learner's predispositions towards certain types of generalizations (e.g., Goldrick & Larson in press, Moreton 2008, Wilson 2003, 2006) and the amount of evidence the lexicon provides for each generalization (e.g., Hayes & Wilson 2008) but also by how easy the evidence is to

“absorb” for the human learner given the characteristics (and particularly the temporal structure) of the learning task.

Learners exposed to either kind of training used product-oriented generalizations to judge how likely a given plural form is to be the right plural form for a given singular. This finding suggests that judgments are made by taking into account the typicality of the product, and not just the typicality of the source-product mapping. There is some evidence for perception being more product-oriented than production in that examples of  $tʃ \rightarrow tʃi$  favored  $t \rightarrow tʃi$  and showed a trend towards favoring  $k \rightarrow tʃi$  in perception while disfavoring  $k \rightarrow tʃi$  in production following source-oriented training. This suggests that even production likelihood judgment does not proceed by simulation of elicited production (contra Albright & Hayes 2003). I would expect that judgments would match production even less closely when ‘acceptability’ or ‘grammaticality’ are rated. Thus product-oriented generalizations are suggested to play a greater role when a speaker/hearer judges the grammaticality or acceptability of a perceived form than when the same speaker/hearer produces a novel wordform from a known form of the same word.

It is important to note that neither experimental nor natural language data provide support for a purely product-oriented model (Bybee 2001) because learners always place restrictions on source forms that can be mapped onto a product. Thus, even in the product-oriented training paradigm, learners tend not to map labial-final sources onto  $[tʃi]$ -final products. At a minimum, learners must realize which source consonants are changeable and which product consonants result from changes to the source, both source-oriented generalizations, since they are based on generalizing over source-product

mappings. Thus, the present data support the hypothesis that learners extract both source-oriented and product-oriented generalizations (e.g., Pierrehumbert 2006).

## **6.2. Negative product-oriented generalizations and Optimality Theory**

The dominant current approach to phonology is Optimality Theory (Prince & Smolensky 1993/2004). In Optimality Theory, the grammar consists of negative product-oriented generalizations (markedness constraints) and source-oriented generalizations militating against non-identity between morphologically related forms (paradigm uniformity constraints).<sup>36</sup> Standard Optimality Theory cannot account for the present data because competition between generalizations is resolved by strict ranking (i.e., the stronger generalization is obeyed all the time) while in the present data competition between generalizations is resolved stochastically. However, Smolensky & Legendre (2006) propose a stochastic version of Optimality Theory, called Harmonic Grammar, which is consistent with this aspect of the present data.

The artificial languages with the largest difference between the frequency with which [ki] is observed (zero) and the frequency with which it is expected to be observed (languages 2 and 4) are the languages in which [ki] is produced relatively often. This is inconsistent with the hypothesis that the learner's avoidance of [ki] is a function of the difference between how often [ki] is observed and how often it is expected despite the fact that this difference is a good way to estimate the confidence the learner should have in the fact that the absence of [ki] in the training data is not a statistical accident.

---

<sup>36</sup> The job of paradigm uniformity constraints can also be (and usually is) done by faithfulness constraints, which enforce identity to an underlying form, which is an abstraction over morphologically related forms, rather than directly enforcing identity between observed morphologically related forms. The difference is not essential for the present discussion. Paradigm uniformity constraints are assumed because the task of the learner is less complicated if abstracting an underlying form is not required.

Despite this finding, Harmonic Grammar can account for the observed results if it is coupled with Boersma's Gradual Learning Algorithm (Boersma 1997, Boersma & Hayes 2001). The Gradual Learning Algorithm takes in a lexicon of source-product pairs and attempts to derive each product from the corresponding source given the current constraint weights (by adding a fixed amount of noise to the weights and converting the resulting noisy weights to a ranking). If the derived, i.e., expected, product does not match the observed product, the algorithm reduces the weights of constraints that are violated by the observed product and obeyed by the expected product and increases the weights of constraints that are violated by the expected product and obeyed by the observed product. All weights are adjusted by the same amount, which decreases over the course of training.

This procedure correctly predicts the difference between languages 1 and 3 on the one hand and languages 2 and 4 on the other as long as we assume that the learner has extracted a general constraint like  $*i$  ('Do not produce products ending in  $-i$ ') or  $*C_{[-cont]}i$  ('Do not produce products ending in a stop followed by  $-i$ '). These constraints are in competition with the constraint militating against products ending in  $-a$  ( $*a$ ). In languages 2 and 4,  $*i$  or  $*C_{[-cont]}i$  loses to  $*a$  more often than  $*a$  loses to  $*i$  or  $*C_{[-cont]}i$ . This is why these languages have more products ending in an alveolar or a labial followed by  $-i$  than by  $-a$ . The opposite is true for languages 1 and 3. Whenever  $*i$  (or  $*C_{[-cont]}i$ ) loses to  $*a$ , the weight of  $*i$  (or  $*C_{[-cont]}i$ ) is reduced and products ending in  $-i$ , and specifically, an  $-i$  preceded by a stop (including  $[k]$ ), become more likely to be produced. This increase in  $[ki]$  likelihood happens more often in languages 2 and 4 than in languages 1 and 3, predicting the retention of  $[k]$  before  $-i$  to be more common when –

i tends to often attach to non-velars. Importantly, even if there is a more specific constraint against [ki] (\*ki), it never loses to another constraint in either language, thus its strength is identical in both. Representative constraint weights are shown in Table 6.1-6.4. These weightings were obtained in Praat (Boersma & Weenink 2009) with evaluation noise at 0, plasticity set to 10 with a decrement of 1 and all other settings at default values. The grammars and training sets are shown in the Appendix. As the tables show, the difference in acceptability (shown to the left of the table) between the [ki] form and the [tʃi] form is smaller in Language 2 than in Language 1.

Table 6.1. Language 1 under the grammar containing \*C<sub>[-cont]</sub>i (a.k.a. \*Stopi). The hand shows the best product for each source. The constraints are arranged in decreasing weight order from left to right.


k	*Stopi	Ident-alveolar	Ident-labial	*a	Ident-velar	
ki	*					110.954
ka				*		93.084
⌚ Ci					*	79.661

t	*Stopi	Ident-alveolar	Ident-labial	*a	Ident-velar	
ti	*					110.954
⌚ ta				*		93.084
Ci		*				109.125


p	*Stopi	Ident-alveolar	Ident-labial	*a	Ident-velar	
pi	*					110.954
⌚ pa				*		93.084
Ci			*			107.177



Table 6.2. Language 2 under the grammar containing \*C<sub>[-cont]</sub>i (a.k.a. \*Stopi).

k	Ident-labial	Ident-alveolar	*a	*Stopi	Ident-velar	
ki				*		-96.801
ka			*			-103.358
 Ci					*	-79.743

t	Ident-labial	Ident-alveolar	*a	*Stopi	Ident-velar	
 ti				*		-96.801
ta			*			-103.358
Ci		*				-109.435




p	Ident-labial	Ident-alveolar	*a	*Stopi	Ident-velar	
 pi				*		-96.801
pa			*			-103.358
Ci	*					-110.662

Table 6.3. Language 1 under the grammar containing \*i.

k	*a	*ki	Ident-alveolar	Ident-labial	*i	Ident-velar	
ki		*			*		-199.062
ka	*						-100.938
 Ci					*	*	-63.321

t	*a	*ki	Ident-alveolar	Ident-labial	*i	Ident-velar	
 ti					*		-99.062
ta	*						-100.938
Ci			*				-100.000




p	*a	*ki	Ident-alveolar	Ident-labial	*i	Ident-velar	
 pi					*		-99.062
pa	*						-100.938
Ci				*			-100.000


Table 6.4. Language 2 under the grammar containing \*i.

k	*a	*ki	Ident-alveolar	Ident-labial	*i	Ident-velar	
ki		*			*		198.873
ka	*						101.127
 Ci					*	*	65.808

t	*a	*ki	Ident-alveolar	Ident-labial	*i	Ident-velar	
 ti					*		98.873
ta	*						101.127
Ci			*				100.000

p	*a	*ki	Ident-alveolar	Ident-labial	*i	Ident-velar	
 pi					*		98.873
pa	*						101.127
Ci				*			100.000

Unlike taking the difference between observed and predicted sequence probabilities given the observed frequencies of occurrence of parts of the sequence across contexts, the Gradual Learning Algorithm allows unobserved sequences to profit from the existence of similar sequences, including sequences in which some element(s) are shared with the unobserved sequence. Table 6.5 illustrates this feature of the Gradual Learning Algorithm for a source sequence CD, which is mapped onto a product that either shares one of its elements with AB or shares neither element with it. As can be seen from the table, the harmony of an AB product is increased if AD and CB products occur, provided that AD, CB, and AB are derived from the same source. The reasoning is that if AB consists of subsequences that are avoided, then it will also be likely to be avoided. By contrast, the observed-expected difference punishes sequences that are not observed if there are similar sequences that are observed because the observed sequences count as evidence for the systematicity of the gap. The reasoning is that if AB consists of subsequences that are unobserved, the fact that the entire sequence is unobserved is not good evidence for the avoidance of the sequence. In the Gradual Learning Algorithm

(and Harmonic Grammar / Optimality Theory with paradigm uniformity more generally), avoidance of a sequence is not estimated based on the products but rather based on the source-product mappings, i.e., it is source-oriented. This appears consistent with the present data.

Table 6.5. The Gradual Learning Algorithm rewards sequences for the existence of similar sequences (iff the similar sequences are derived from the same source). The lower the harmony, the worse the sequence is estimated to be.<sup>37</sup>

Probability during training				Harmony of CD→AB after training
CD→CB	CD→AD	CD→CF	CD→ED	
50	50	0	0	-366
0	0	50	50	-433

The difference between Harmonic Grammar (Smolensky & Legendre 2006) learned using the Gradual Learning Algorithm (Boersma 1997) and the Minimal Generalization Learner (Albright & Hayes 2003) is that Harmonic Grammar predicts that the existence of a competitor suffix (-a) is necessary for the difference between languages 1 and 3 on the one hand and languages 2 and 4 on the other to be obtained, since otherwise the weight of \*i is changed only when it is in competition with another constraint (\*a). Minimal Generalization Learner predicts that the removal of examples of –a from training should not eliminate the difference between the two types of languages,

<sup>37</sup> The table is created by a grammar containing \*A, \*B, \*C, \*D, \*E, \*F, \*AB, \*EF, Ident-C, and Ident-D. The relative ordering of the harmony values and the importance of attestation of similar strings rather than frequency are constant across parameter settings. For this simulation, evaluation noise was set to 0, initial weights were at 100, with default settings for all parameters.

since the learner has more confidence in the productivity of ‘just add –i’ when it is supported by many examples. The same prediction is also obtained for positive product-oriented generalizations, which are weighted by type frequency.

### **6.3. Influences on word learning**

In the delayed cued recall tested in the product-oriented training paradigm, learners tend to remember the forms of a word that are most helpful for unambiguously deriving other forms of the same word. In immediate recall examined in the source-oriented training paradigm, the plural, which always follows the singular during training, is recalled better than the singular and is able to influence the form of the singular’s stem. While the present study was not designed to address the question of what makes a wordform memorable, several hypotheses that can be tested in the present experimental paradigm come to mind. For instance, are wordforms that are existing words in the learner’s lexicon remembered better than wordforms that are not existing words? That is, does the learner benefit from having some familiarity with the wordform and/or is hindered by the inappropriate connection to semantics that s/he has learned in English? What is the influence of similarity between different wordforms in the artificial lexicon? Is it easier to remember words that are similar to many other words in the lexicon, or those that are relatively distinctive (cf. Storkel 2004 vs. Swingley & Aslin 2007)? Do learners zero in on and memorize exceptions (Pinker 1999), e.g., plurals exemplifying  $p \rightarrow pa$  in languages where the dominant mapping is  $p \rightarrow pi$ ? What sorts of acoustic and contextual variability are helpful for word recall? What is the influence of the connection to semantics? Are polysemy and homonymy helpful or harmful?

Another interesting question that can be addressed with the present paradigm is how learners acquire the meanings of affixes and other constructions. For instance, how do the learners associate the presented affixes with plural semantics? Is the meaning ‘plural’ acquired equally easily from examples of plural wordforms referring to two referents and multiple referents (Zapf & Smith 2008)? Are learners misled by variation in the exact number of objects presented together with a plural wordform into thinking that the suffix means ‘five’, ‘three’, or ‘two’, only the last possibility being attested in real languages? Artificial lexicon studies are a promising way to address questions like these because all characteristics of the lexicon and the training are under direct control of the experimenter.

#### **6.4. The influence of the learning task on the architecture of grammar**

A primary aim of the present thesis has been to determine whether typical characteristics of the learning task faced by a person acquiring language naturally may lead to a preference for product-oriented generalizations over source-oriented generalizations. To address this issue we may compare the results of Chapter 3, in which source-oriented generalizations dominated elicited production, to the results of Chapter 4, in which elicited production was product-oriented. The difference between the experiments reported in Chapter 3 and Chapter 4 is the training paradigm. The training paradigm used in Chapter 4 is closer to the situation faced by a language learner ‘in the wild’ in several respects:

1) the primary aim of the language learner in Chapter 4 is to learn names for objects, just as (arguably) in real language learning, whereas the primary aim of the learner in Chapter 3 is to extract grammatical regularities;

2) the learner in Chapter 4 is presented with one wordform from a paradigm at a time, just like a language learner in the wild, while the learner in Chapter 3 is presented with pairs of words that share the stem; and

3) the learner in Chapter 4 is able to acquire the lexicon of the presented language and, just like the speaker of a natural language, has a choice between generating wordforms on the fly or retrieving them from memory, while the learner in Chapter 3 faces an overabundance of confusable stems, which leads him/her to induce a grammar without memorizing a lexicon.

Thus, the more natural training paradigm leads to an increased reliance on product-oriented generalizations relative to source-oriented generalizations. This result supports the idea that the types of generalizations that are relied upon by a speaker/hearer in extending his/her lexicon are influenced by the way the speaker/hearer experiences language, and not just by an innate Universal Grammar, suggesting that even formal properties of the grammar may be emergent from patterns of language usage (Bybee 2008). While natural languages seem to prefer product-oriented generalizations (Becker & Fanleib 2009, Bybee 2001), this may be due to the way those languages are experienced by their learners. Particularly, learners don't hear multiple forms of the same lexeme one after another. If there were a language that was acquired in this manner, its speakers would presumably rely heavily on source-oriented generalizations. To conclude, formal properties of the grammar, including the types of rules, schemas or constraints

that the grammar consists of, are influenced not just by 'the structural properties of the cognitive system' (Goldrick & Larson in press) but also by the structure and affordances of the task faced by the learners as s/he acquires the language.

At least three predictions for natural languages follow from the observed effect of the learning task. First, reliance on source-oriented generalizations may be more expected in non-native speakers of a language, who experience language through textbooks that explicitly teach the reader to conjugate verbs and decline nouns, than in native speakers who experience language one wordform of a time. Second, source-oriented generalizations should form when wordforms sharing a stem tend to appear in close temporal proximity. This is, perhaps, the case for noun-adjective pairs of the type 'electric-electricity' in English, which are shown to be formed using source-oriented generalizations by Pierrehumbert (2006). Some previous support for this hypothesis is provided by Morgan et al. (1989) who found that the acquisition of a phrase structure grammar was facilitated when learners were provided with pairs of sentences that could be related by pronominalization or movement rules but were unable to replicate the effect with related pairs of sentences being randomly interspersed with other, unrelated sentences.<sup>38</sup> Finally, product-oriented generalizations may be favored over source-oriented generalizations especially strongly if both have to be acquired over a small set of word types where the inherently lower type frequency of source-oriented generalizations is of particular importance.

---

<sup>38</sup> Cf. also Valian & Coulson (1988: 78): "Our actual linguistic competence, and our acquisition of competence is mediated by the performance system. That performance system is a composite of representational, acquisitional, analytic, and memorial abilities. As such, ... it even limits us to acquiring a language only under presentation conditions which are cognitively favorable."

### **6.5. The number of stages underlying grammatical processing**

A grammar is defined not only by the types of generalizations that compete with each other during grammatical processing but also by the set of stages or serially ordered sets of competing generalizations that underlie the observed stimulus-response pairings. In the present thesis, I have shown that velar palatalization in a natural language (Russian) is (surprisingly) a single-stage process: the affix choice and stem shape are chosen simultaneously, rather than the affix being chosen first and then triggering or failing to trigger a stem change.

The traditional view of grammar is that phonological changes associated with the concatenation of a pair of morphemes are triggered when the morphemes have already been chosen (e.g., Halle & Marantz, 1993). Thus, for instance, the speaker may decide to palatalize or not to palatalize a stem-final consonant after having chosen to affix –i to the stem. The alternative is that the speaker is deciding between complete input-output mappings. Thus, the speaker presented with a new stem ending in [k] would be deciding between producing something ending in [tʃi], something ending in [ka], or something ending in [ki] (Bybee, 1985, 2001; Rumelhart & McClelland, 1986; Skousen, 1989).

It appears possible to examine the issue of the number of stages underlying grammatical processing in an artificial language. Let us consider two more languages in which there are two plural suffixes, –i and –a, where –i always triggers velar palatalization while –a never does. One of the languages has an additional rule that says that –a cannot be attached to stems ending in [ik]. The other language has a rule that says that –a should not attach to stems ending in [ak]. Vowel harmony and disharmony are



equally learnable (Pycha et al., 2003), and both are attested in languages (e.g., Walter, 2008). In both languages,  $ik \rightarrow itʃi$  is exemplified by 10 noun pairs as is  $ak \rightarrow atʃi$  and the selection of  $-a$  vs.  $-i$  does not depend on the preceding vowel if the final consonant of the singular is not a velar. The two languages are shown in Table 6.6.

Table 6.6. Two languages that can be used to test the number of stages underlying morphophonological processing.

	Language V	Language VI
$ik \rightarrow itʃi$	N	
$ak \rightarrow atʃi$	N	
$ik \rightarrow ika$	N	0
$ak \rightarrow aka$	0	N
$\{i;a\}\{t;p;tʃ\} \rightarrow \{i;a\}\{t;p;tʃ\}i$	M	

Now let us consider whether the productivity of velar palatalization before  $-i$  depends on the vowel of the stem in each of the two languages under the one-stage and two-stage models. According to the one-stage model, there is competition between the complete input-output mappings  $k \rightarrow tʃi$  /  $i$ \_\_,  $k \rightarrow ka$  /  $i$ \_\_,  $k \rightarrow tʃi$  /  $a$ \_\_,  $k \rightarrow ka$  /  $a$ \_\_,  $C \rightarrow Ci$ , and  $C \rightarrow Ca$ . Velar palatalization happens in a particular context (e.g.,  $i$ \_\_) if the palatalizing mapping applicable to that context wins over the  $-a$ -adding mapping and the mapping that simply adds  $-i$ . Palatalization is predicted to fail before  $-i$  if the mapping that just adds  $-i$  wins. The palatalizing mappings are exemplified by the same number of

nouns regardless of context in both languages. However, in Language I, the –a-adding mapping is more productive after [a] than after [i] while in Language II it is more productive after [i] than after [a]. So the palatalizing mapping is less reliable (and more likely to lose to ‘just add –i’, which is equally reliable in all contexts) after [i] than after [a] in Language V while the opposite is true in Language VI.

Now let us consider the two-stage model. The first stage involves competition between –i and –a and does not influence the probability of failure of velar palatalization before –i. If –i is chosen, the second stage involves competition between  $k \rightarrow tʃ / i\_i$ ,  $k \rightarrow tʃ / a\_i$ , and ‘do nothing’. The two palatalizing mappings are exemplified by the same number of noun pairs in both languages and have no exceptions, thus being equally reliable. Therefore, velar palatalization rate before –i is expected to be constant across preceding vowel contexts in both languages.

The outlined logic is also applicable to a product-oriented grammar as long as the grammar includes paradigm uniformity constraints and these constraints are learnable, rather than innately specified. In a single-stage product-oriented grammar, ‘end in atʃi’ and ‘end in itʃi’ are equally supported in both languages. However, [k] is often retained in the plural after [i] in Language V and after [a] in Language VI. While a purely product-oriented grammar cannot include any statements that refer to multiple cells in a morphological paradigm, all extant product-oriented models do in fact allow a restricted range of such statements, particularly ones that militate against differences between morphologically related forms (formalized as paradigm uniformity constraints in Optimality Theory, Downing et al., 2005, and association lines in Network Theory, Bybee, 1985, 2001), which, as shown in Chapter 1, are also necessary for restrictions on

the types of input that can correspond to a particular output (documented by Pierrehumbert, 2006, among others). If the reliability of a paradigm uniformity constraint varies by context (and these differences are learnable, unlike in Optimality Theory), “keep the [k]” is more reliable after [i] in Language V and after [a] in Language VI. Therefore, the productivity of velar palatalization is expected to be relatively low after [i] in Language V and after [a] in Language VI.

In the second stage of the two-stage model, the following context is known. Thus, the second stage competition in a product-oriented model would involve competition between product-oriented schemas like ‘plurals must end in [itʃi]’ and paradigm uniformity constraints like ‘a velar in the singular corresponds to a velar in the plural in the context i\_i’. The product-oriented schemas favoring palatalization (‘end in [atʃi]’ and ‘end in [itʃi]’) are supported by the same number of examples in both languages. The competing paradigm uniformity constraint ‘keep the velar’ is also equally reliable when followed by –i regardless of the preceding vowel. Therefore, learners exposed to Language V are not expected to differ from learners exposed to Language VI in how productively they palatalize velars before –i. The same prediction is made by the one-stage model if paradigm uniformity constraints are restricted to a small set of innate constraints that can refer only to a single segment or if the grammar contains exclusively generalizations over products.

Thus the comparison between Languages V and VI pits two-stage grammars, pure product-oriented grammars, and product-oriented grammars with context-independent paradigm uniformity constraints, which predict a null result, against one-stage grammars (see Table 6.7).

Table 6.7. Productivity of velar palatalization in different contexts depending on language (V vs. VI) according to various grammatical architectures.

Grammar	Language V	Language VI
Single-stage rule-based	$i_i < a_i$	$a_i < i_i$
Single-stage product-oriented with context-sensitive paradigm uniformity constraints		
Two-stage	$a_i = i_i$	
Single-stage product-oriented with context-independent paradigm uniformity constraints		

Much of the power of source-oriented, rule-based approaches comes from allowing an unlimited number of processing stages to intervene between the input and the output (e.g., Port & Leary 2005). On the other hand, product-oriented approaches have argued for minimizing the number of processing stages, in the extreme allowing only single-stage surface-to-surface mappings (Bybee 2001, Burzio 2002). Thus an interesting question for future research is whether manipulating the training task in favor of source-oriented generalizations also makes learners more likely to favor multistage processing or if the factors leading to multistage processing are distinct from the factors favoring source-oriented generalization.<sup>39</sup>

---

<sup>39</sup> One possible factor that may influence the choice between two-stage and one-stage processing may be the amount overlap between sets of factors influencing choice in the two stages.

## 6.6. Individual differences in artificial grammar learning

Learners exhibit a very wide degree of variability in the extent to which they are able to learn the artificial languages presented to them (cf. Breitenstein et al. 2005, Johnston et al. 1988, Williams 2003, Williams & Lovatt 2003), in the closeness with which they match the statistics of the input. Previous work (Hudson Kam & Newport 2005) has uncovered differences between children and adults. When children are exposed to an artificial language in which there are two competing rules, and the choice between the rules is random, the children tend to use the dominant rule 100% of the time, whereas adults match the relative frequencies of the two rules in the training data (although see Roberts 1996 for evidence that children eventually learn to match the frequencies of variable forms in their environment, at least for variable {t;d} deletion in English).

While our subjects were all adults, some exhibit ‘childlike’ overgeneralization studied by Hudson Kam and Newport (2005). However, this pattern does not exhaust the space of observed behaviors. While most subjects fall roughly along the diagonal from always attaching –i rather than –a to non-velars and not palatalizing velars before –i to never attaching –i to non-velars and always palatalizing velars before –i, where a given subject falls on that continuum is only roughly predicted by the training data to which s/he is exposed. It appears that the learner first becomes aware of the patterns that exist in the language and only later narrows down the relative frequencies of the patterns (cf. Local 1982). If this is granted, it appears plausible that the learners in the present experiments vary in how narrow their estimates of the relative frequencies of competing patterns in the input are, with output frequencies of subjects whose estimates of input frequencies are relatively uncertain being relatively difficult to predict.

Another parameter along which the learners differ is how much they can escape native language interference. Previously, Magnuson et al. (2003) argued that artificial lexicons were isolated from the native language, on the basis of finding within-lexicon neighborhood density effects on word recognition in an artificial lexicon but no effect of English neighborhood density. On the other hand, Williams (2003) and Williams & Lovatt (2003) find that speakers of languages with grammatical gender learn arbitrary word classes defined by paradigmatic relations between affixes more easily than do native speakers of languages like English, which do not have grammatical gender. Furthermore, the more languages with grammatical gender a subject knew, the better s/he performed. These results suggest that prior experience with other languages may influence artificial grammar learning.

While all subjects in the present experiments used the plural suffixes experienced during training, three subjects attached English plural suffixes to a large proportion (25%-75%) of plural nouns following the vowel indicating the plural in the artificial language. This occurred despite the subjects being warned that the language is not English and that they are NOT to use English plural suffixes. The learners did not attach English suffixes to plural forms they repeated during training.

It appears important to determine whether frequency estimation, frequency matching, and native language interference correlate with success in learning natural second languages (as suggested by Williams 2003, Williams & Lovatt 2003). While Hudson Kam and Newport (2005) suggest that inability to frequency-match makes children great language learners, it appears important to compare language learning abilities of people that differ in frequency-matching abilities (and predispositions to

match input frequency) while controlling for age (cf. Breitenstein et al. 2005, Johnston et al. 1988, Williams 2003, Williams & Lovatt 2003). Conversely, the studies reported in the present thesis have employed adults, and the language learning strategies of adults are not necessarily the same as language-learning strategies of children (cf. Hudson Kam & Newport 2005, Braine et al. 1990), hence it is important to replicate the present studies in children. Finally, it would be important to determine what cognitive variables are responsible for the observed differences between learners, including differences between children and adults. An important step in this direction has been done by Williams and Lovatt (2003) who have documented an effect of working memory on success in learning an artificial language.

## **6.7. Implications for language change**

Both positive product-oriented and competing source-oriented generalizations, which find some support in the artificial grammar experiments supported here, predict that the more an alternation-triggering affix attaches to input that are ineligible to undergo the rule, the less it is able to trigger the alternation. This hypothesis provides an explanation for the diachronic finding that two ‘well-documented paths of change occur in parallel... First, phonetically conditioned sound change creates alternations that gradually acquire morphological or lexical conditioning... Simultaneous[ly]... productive phonetically conditioned alternations are likely to become unproductive’ (Bybee 2008: 114). Unless the suffix conditioning an alternation tends to attach mostly to stems that can alternate, it will eventually lose its ability to trigger the alternation.

This hypothesis not only accounts for the tendency of morphologically conditioned alternations to lose productivity but also generates predictions about which suffixes are likely to be good triggers of an alternation, and which ones should be poor triggers. These predictions are confirmed by the data from velar palatalization in Russian, which is productively triggered by –ok and –ek, which attach mostly to velar-final inputs, but not by –ik and –i, which tend not to attach to velar-final inputs. The data from artificial languages provides support for the hypothesized direction of influence: if one manipulates how often a suffix attaches to velar-final stems in an artificial language, one also influences how productive velar palatalization is to learners who are exposed to the language.

In both of the artificial language experiments reported in the present thesis, there is a tendency for subjects exposed to velar palatalization to infer the existence of alveolar palatalization. This finding makes it tempting to conclude that, insofar as language change is driven by learning biases (see Aitchison 1981:180, Bybee 2001:77-85, and Kerswill 1996 for discussion of alternatives), palatalization should extend from velar source forms to alveolar source forms. However, it is important to realize that the present experiments examine only a single ‘generation’ of learners of the artificial languages. An important question that can not be answered by the present studies is if the direction of change will remain constant when multiple generations are stacked in an iterated learning experiment (Kalish et al. 2007, Kirby et al. 2008) so that the language system induced by generation n-1 generates the data that are used by generation n to induce their language system. While Bybee (2008: 120) writes that “phonetic change in a certain direction tends to continue”, oscillatory patterns are also possible (Wedel 2006). The two alternatives can

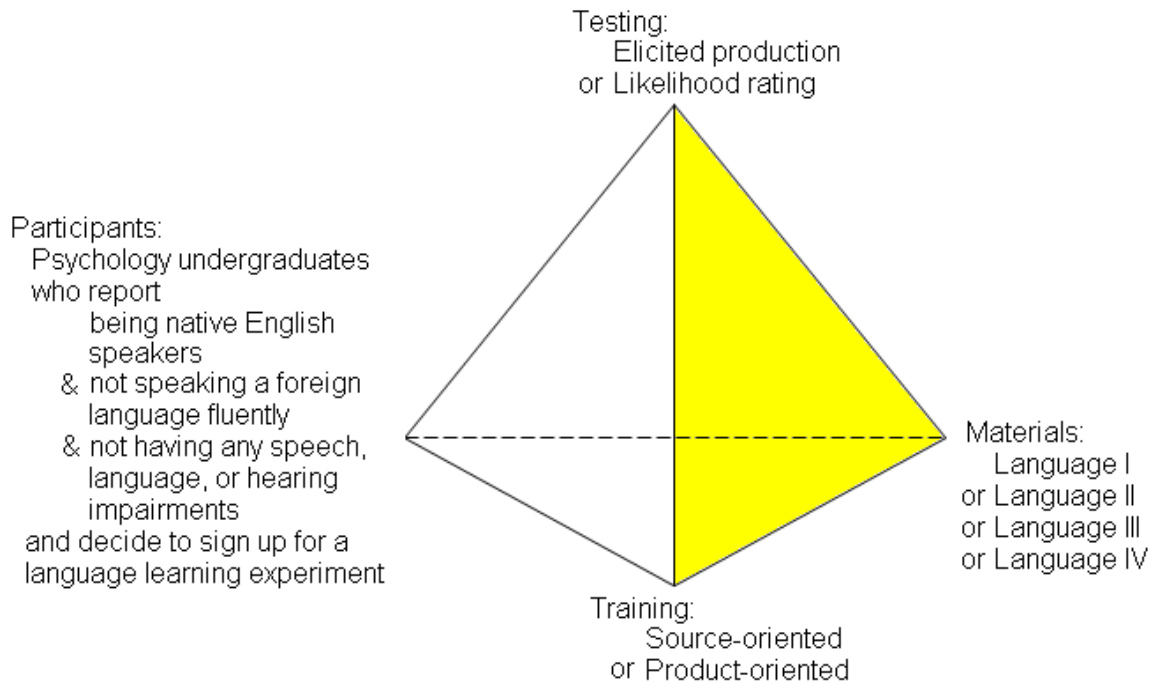


be exemplified by the case in which generation 2 encounters a language in which a singular-final /k/ always corresponds to a plural-final /tʃi/ but a singular-final /t/ only rarely corresponds to a plural-final /tʃi/, most often corresponding to /ti/, i.e., the language implied by the productions of subjects in my experiments (Generation 1). Generation 2 has a choice of whether to extend the  $t \rightarrow tʃi$  pattern (moving the language in the same direction as the previous generation) or regularize unpredictable variation in favor of the dominant  $t \rightarrow ti$  pattern (returning the language to its original state). It is only by examining iterated learning (Kalish et al. 2007, Kirby et al. 2008) that we can delimit possible trajectories of learning-driven language change.

## **6.8. Summary**

Jenkins (1979) and Roediger (2008) write of the “theorist’s tetrahedron”, shown in Figure 6.1, which defines the space of interactions among experimental variables. Thus, type of training may interact with type of testing, stimuli presented, and participant characteristics. In the present study, I examined the interactions between training type (source-oriented vs. product-oriented), characteristics of the language presented to the learners (whether the affix that was shown to turn preceding velars into alveopalatals was also shown to often attached to labials and alveolars, and whether it was also shown to attach to alveopalatals), and testing type (elicited production of a plural from a singular vs. rating the likelihood that the presented plural is the right plural for the presented singular). The possible interactions with participant characteristics remains a matter for future research.

Figure 6.1. The theorist's tetrahedron (based on Jenkins 1979, Roediger 2008) applied to the present thesis. The highlighted face shows the examined interactions.



Characteristics of the training task are shown to influence the extent to which learners rely on source-oriented vs. product-oriented generalizations in both rating and elicited production. Following training in which a small number of word types is presented often and pairs of wordforms sharing the same stem do not co-occur more often than one would expect by chance, the learners extend velar palatalization to alveolar sources and consider examples of alveopalatal sources being mapped onto products ending in [tʃi] to support mapping velar sources onto products ending in [tʃi]. Following training in which singulars and plurals are presented next to each other and a large number of word types is presented, velar palatalization remains restricted to velars in production and examples of alveopalatal sources being mapped onto products ending in [tʃi] are taken to support the generalization that –i should simply be added to an

alveopalatal or a velar to form the plural. These results support the hypothesis that the shape of the grammar extracted by a learner from a lexicon depends on how the lexicon is presented to the learner (e.g., Morgan et al. 1989), i.e., the nature of the learning situation, and not just on lexical statistics and innate biases the learner brings to the task.

Despite the observed interaction of the lexical statistics of an artificial language and the training task, there are observations that hold across training tasks and are also supported by data from natural languages. One such generalization is that learners induce both product-oriented and source-oriented generalizations. Thus, learners exposed to a lexicon of singular and plural form learn at least 1) what typical plurals and singulars are like, 2) which segments of the singular form must be retained in the plural, and 3) which segments of the plural form must be retained in the singular.

Across training paradigms and languages, competing generalizations are weighted relative to each other stochastically. Thus, learners obey competing generalizations in proportion to how much statistical support each competitor receives from the training data, rather than obeying the most strongly supported competitor 100% of the time (cf. Albright & Hayes 2003). This is consistent with prior evidence from English past tense where nonce probes that are similar to regular verbs are more likely to be produced according to a regular pattern compared to nonce probes that are similar to neither regulars nor irregulars, despite the fact that the regular pattern is the strongest of the competing patterns for both classes of probes (Albright & Hayes 2003). It is also consistent with evidence that children learn to match the frequencies with which variable rules are obeyed by their parents (Roberts 1996). This result supports the hypothesis that human language learners attempt to match frequencies of forms and patterns in the

environment (see, e.g., Roberts 1996) and is not consistent with the hypothesis that learners attempt to optimize accuracy of generalization in terms of the number of correct forms produced, a goal best achieved by obeying the most reliable pattern 100% of the time (Hudson Kam & Newport 2005).

Learners do not obey the Subset Principle (e.g., Berwick 1986, Hale & Reiss 2003), which would predict that the learners should induce the most specific generalizations consistent with the training data. The observed overgeneralization patterns are expected if we assume a Bayesian approach to speech perception and word recognition, in which the output of perception is not the identity of the most likely structure but rather a probability distribution over possible structures. For instance, when presented with a  $[t] \rightarrow [t^i]$  pairing the listener is expected to assign some probability to having heard  $[t] \rightarrow [t^f i]$  thus considering the possibility of alveolar palatalization, which is never presented to the learners but is rated as acceptable by them in both training paradigms and is produced relatively frequently after product-oriented training.

## **6.9. Conclusion**

In this thesis, I have proposed that grammars are sets of competing generalizations. The full set of generalizations comprising a grammar is divided into smaller sets, or processing stages, with generalizations in each set weighted relative to each other based on how much evidence for each generalization is provided by the linguistic data in the environment of the learner. I have provided an experimental approach to address the questions of 1) what types of generalizations are extracted from the linguistic data (i.e., both product-oriented and source-oriented positive

generalizations), 2) how the generalizations are divided up into sets of competing generalizations (e.g., the choice of the affix and the stem shape being chosen simultaneously, with production and perception grammars being separate systems, or at least separate sets of weightings), and 3) how the choice between competing generalizations is accomplished (i.e., stochastically).

The learner seems to extract both source-oriented and product-oriented generalizations from the linguistic data, and to weight the extracted generalizations in slightly different ways for the purposes of production and rating. The influence of the training task on the extracted grammar suggests that the learner generalizes over words that s/he experiences in close temporal proximity, which tends to disfavor source-oriented generalizations in natural language learning where one might hear, e.g., a number of plural nouns in rapid succession and only rarely experience a pair of words differing only in the plural inflection next to each other.

While the learner has often been assumed to be extremely conservative, coming up with the most constrained generalizations consistent with the training data (e.g., Berwick 1986, Hale & Reiss 2003), the present data are not consistent with this hypothesis. I would like to suggest that there is a principled reason for this inconsistency, which is that the output of perception is not the identity of the most probable structure but a probability distribution over a set of possible structures. Thus, when a certain stimulus is presented, stimuli that are similar to the one presented become more probable for the learner perceiving the stimulus.

Artificial grammar experiments combined with corpus, typological and wug studies of natural languages provide a promising way to address the architecture of

grammar, the biases (different kinds of) people bring to the task of learning it, and the directions in which these learning biases may drive language change. While I hope only to have scratched the surface of a little corner of the field (the morphophonology of velar palatalization is of course not all there is to linguistics), it is my hope that the experimental methods developed here will be helpful for finding the competing sets (a.k.a. boxes) of generalizations and enumerating their (struggling) contents in other corners of linguistics as well.

**The grammar file for Tables 6.1-6.2:**

```

"ooTextFile"
"OTGrammar 2"
<HarmonicGrammar>
0.0 ! leak
5 ! number of constraints
"*Stopi" 100 100 1
"*a" 100 100 1
"Ident-velar" 1000 1000 1
"Ident-alveolar" 1000 1000 1
"Ident-labial" 1000 1000 1
0 ! number of fixed rankings
4 ! number of accepted inputs
"k" 3 ! input form with number of output candidates
    "ki" 1 0 0 0 0 ! first candidate with violations
    "ka" 0 1 0 0 0 ! second candidate with violations
    "Ci" 0 0 1 0 0
"t" 3
    "ti" 1 0 0 0 0
    "ta" 0 1 0 0 0
    "Ci" 0 0 0 1 0
"p" 3
    "pi" 1 0 0 0 0
    "pa" 0 1 0 0 0
    "Ci" 0 0 0 0 1

```

**The grammar file for Tables 6.3-6.4:**

```

"ooTextFile"
"OTGrammar 2"
<HarmonicGrammar>
0.0 ! leak
6 ! number of constraints
"*a" 100 100 1
"*i" 100 100 1
"*ki" 100 100 1
"Ident-velar" 1000 1000 1
"Ident-alveolar" 1000 1000 1
"Ident-labial" 1000 1000 1
0 ! number of fixed rankings
3 ! number of accepted inputs
"k" 3 ! input form with number of output candidates
    "ki" 0 1 1 0 0 0 ! first candidate with violations
    "ka" 1 0 0 0 0 0 ! second candidate with violations

```

```

      "Ci" 0 1 0 1 0 0
"t" 3
      "ti" 0 1 0 0 0 0
      "ta" 1 0 0 0 0 0
      "Ci" 0 0 0 0 1 0
"p" 3
      "pi" 0 1 0 0 0 0
      "pa" 1 0 0 0 0 0
      "Ci" 0 0 0 0 0 1

```

**The training set for Tables 6.1, 6.3:**

```

"ooTextFile"
"PairDistribution"
5 pairs
"k" "Ci" 100
"t" "ti" 25
"t" "ta" 75
"p" "pi" 25
"p" "pa" 75

```

**The training set for Tables 6.2, 6.4:**

```

"ooTextFile"
"PairDistribution"
5 pairs
"k" "Ci" 100
"t" "ti" 75
"t" "ta" 25
"p" "pi" 75
"p" "pa" 25

```

**The grammar file for Table 6.5:**

```

"ooTextFile"
"OTGrammar 2"
<HarmonicGrammar>
0.0 ! leak
10 ! number of constraints
"*A" 100 100 1
"*B" 100 100 1
"*C" 100 100 1
"*D" 100 100 1
"*E" 100 100 1

```



```

"*F" 100 100 1
"*AB" 100 100 1
"*EF" 100 100 1
"Ident-C" 100 100 1
"Ident-D" 100 100 1
0 ! number of fixed rankings
1 ! number of accepted inputs
"CD" 7 ! input form with number of output candidates
    "AB" 1 1 0 0 0 0 1 0 1 1 ! first candidate with
violations
    "CB" 0 1 1 0 0 0 0 0 0 1 ! second candidate with
violations
    "AD" 1 0 0 1 0 0 0 0 1 0
    "CD" 0 0 1 1 0 0 0 0 0 0
    "CF" 0 0 1 0 0 1 0 0 0 1
    "ED" 0 0 0 1 1 0 0 0 1 0
    "EF" 0 0 0 0 1 1 0 0 1 1

```

### Training sets for Table 6.5:

```

"ooTextFile"
"PairDistribution"
7 pairs
"CD" "AB" 0
"CD" "AD" 50
"CD" "CB" 50
"CD" "CD" 0
"CD" "ED" 0
"CD" "CF" 0
"CD" "EF" 0

```

```

"ooTextFile"
"PairDistribution"
7 pairs
"CD" "AB" 0
"CD" "AD" 0
"CD" "CB" 0
"CD" "CD" 0
"CD" "ED" 50
"CD" "CF" 50
"CD" "EF" 0

```

References:

- Aitchison, J. 1981. *Language change: Progress or decay?* London: Fontana.
- Albright, A. 2005. The morphological basis of paradigm leveling. In Downing et al., 17-43.
- Albright, A. 2008. Explaining universal tendencies and language particulars in analogical change. In J. Good, ed. *Linguistic universals and language change*, 144-84. Oxford, New York: Oxford University Press.
- Albright, A. In prep. Natural classes are not enough: Biased generalization in novel onset clusters. Ms. MIT.
- Albright, A., & B. Hayes. 2002. Modeling English past tense intuitions with minimal generalization. In M. Maxwell, ed. *Proceedings of the sixth meeting of the ACL Special Interest Group in Computational Phonology*. Philadelphia: ACL.
- Albright, A., & B. Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, 90, 119-61.
- Altmann, G. T. M., Z. Dienes, & A. Goode. 1995. Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 899-912.
- Archangeli, D., & D. Pulleyblank. 1994. *Grounded phonology*. Cambridge, MA: MIT Press.
- Bashford, J.A., Jr., R. M. Warren, & P. W. Lenz. 2006. Polling the effective neighborhood of spoken words with the Verbal Transformation Effect. *Journal of the Acoustical Society of America Express Letters*, 119, EL55-EL59.

- Becker, M., & L. Fanleib. 2009. Surface-based generalizations over lexical exceptions. Paper presented at LSA Annual Meeting, San Francisco, CA.
- Becker, M., N. Ketrez, & A. Nevins. 2007. When and why to ignore lexical patterns in Turkish obstruent alternations. Paper presented at LSA Annual Meeting, Anaheim, CA.
- Benua, L. 1997. *Transderivational identity: Phonological relations between words*. Ph.D. Dissertation: University of Massachusetts, Amherst.
- Berent, I., D. Steriade, T. Lennertz, & V. Vaknin. 2007. What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, 104, 591–630.
- Berko, J. 1958. The child's acquisition of English morphology. *Word*, 14, 150–77.
- Berwick, R. C. 1986. *The acquisition of syntactic knowledge*. Cambridge, MA: MIT Press.
- Bhat, D. N. S. 1974. A general study of palatalization. *Working Papers on Language Universals*, 14, 17–58.
- Blevins, J. 1997. Rules in Optimality Theory: Two case studies. In I. Roca, ed. *Derivations and constraints in phonology*, 227–60. Oxford: Clarendon Press.
- Blevins, J. 2004. *Evolutionary Phonology: the emergence of sound patterns*. Cambridge: Cambridge University Press.
- Bock, J. K. 1986. Syntactic persistence in language production. *Cognitive Psychology*, 18, 355–87.
- Boersma, P. 1997. How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, 21, 43–58.

- Boersma, P., & B. Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry*, 32, 45-86.
- Boersma, P., & D. Weenink. 2005. Praat: doing phonetics by computer (Version 4.4.16). [Computer program]. Retrieved August 29, 2005, from <http://www.praat.org/>
- Boersma, P., & D. Weenink. 2009. Praat: doing phonetics by computer (Version 5.1.07). [Computer program]. Retrieved May 19, 2009, from <http://www.praat.org/>
- Braine, M. D. 1987. What is learned in acquiring word classes – A step towards and acquisition theory. In B. MacWhinney, ed. *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum.
- Braine, M. D., R. E. Brody, P. J. Brooks, V. Sudhalter, J. A. Ross, L. Catalano, & S. M. Fisch. 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, 591-610.
- Breitenstein, C., A. Jansen, M. Deppe, A.-F. Foerster, J. Sommer, T. Wolbers, & S. Knecht. 2005. Hippocampus activity differentiates good from poor learners of a novel lexicon. *NeuroImage*, 25, 958-68.
- Bresnan, J. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell.
- Bromberger, S., & M. Halle. 1997. The contents of phonological signs: A comparison between their use in derivational theories and in optimality theories. In I. Roca, ed. *Derivations and constraints in phonology*, 93-122. Oxford: Clarendon Press.
- Brooks, L. R. 1978. Non-analytic concept formation and memory for instances. In E. Rosch & B. Lloyd, eds. *Cognition and concepts*, 169-211. Hillsdale: Erlbaum.

- Brooks, P. J., M. D. S. Braine, L. Catalano, R. E. Brody, & V. Sudhalter. 1993. Acquisition of gender-like noun subclasses in artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language*, 32, 76-92.
- Brooks, R. L., & R. J. Vokey. 1991. Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 316-23.
- Browman, C. P., & L. Goldstein. 1992. Articulatory Phonology: An overview. *Phonetica*, 49, 155-80.
- Burzio, L. 2002. Surface-to-surface morphology: When your representations turn into constraints. In P. Boucher, ed. *Many morphologies*, 142-77. Somerville, MA: Cascadilla.
- Bybee, J. L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.
- Bybee, J. L. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.
- Bybee, J. L. 2008. Formal universals as emergent phenomena: The origins of structure preservation. In J. Good, ed. *Linguistic universals and language change*, 108-24. Oxford, New York: Oxford University Press.
- Bybee, J. L., & C. L. Moder. 1983. Morphological classes as natural categories. *Language*, 59, 251-70.
- Bybee, J. L., & J. E. Newman. 1995. Are stem changes as natural as affixes? *Linguistics*, 33, 633-54.

- Bybee, J. L., & E. Pardo. 1981. On lexical and morphological conditioning of alternations: A nonce-probe experiment with Spanish verbs. *Linguistics*, 19, 937-68.
- Bybee, J. L., & D. I. Slobin. 1982. Rules and schemas in the development and use of the English past. *Language*, 58, 265-89.
- Chang, S. S., M. C. Plauche & J. J. Ohala. 2001. Markedness and consonant confusion asymmetries. In E. Hume & K. Johnson, eds. *The role of speech perception in phonology*, 79-101. San Diego: Academic Press.
- Chomsky, N. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, N., & M. Halle. 1968. *The sound pattern of English*. New York, Evanston, London: Harper and Row.
- Clayards, M., M. K. Tanenhaus, R. N. Aslin, & R. A. Jacobs. 2008. Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, 108, 804-809.
- Coleman, J., & J. Pierrehumbert. 1997. Stochastic phonological grammars and acceptability. *3rd Meeting of the ACL Special Interest Group in Computational Phonology: Proceedings of the Workshop*, 49-56. Somerset, NJ: Association for Computational Linguistics.
- Cristià, A., & A. Seidl. 2008. Is infants' learning of sound patterns constrained by phonological features? *Language Learning and Development*, 4, 203-27.
- Dąbrowska, E. 2008. The effects of frequency and neighbourhood density on adult speakers' productivity with Polish case inflections: An empirical test of usage-based approaches to morphology. *Journal of Memory and Language*, 58, 931-51.
- Daelemans, W., & A. van den Bosch. 2005. *Memory-based language processing*. Cambridge: Cambridge University Press.

- Darcy, I., F. Ramus, A. Christophe, K. Kinzler, & E. Dupoux. 2008. Phonological knowledge in compensation for native and non-native assimilation. In F. Kügler, C. Féry, & R. van de Vijver, eds. *Variation and gradience in phonetics and phonology*. Berlin: Mouton de Gruyter.
- Davidson, L. P. 2006. Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics*, 34, 104-37.
- Dell, F. 1981. On the learnability of optional phonological rules. *Linguistic Inquiry*, 12, 31-37.
- Denton, S. E., J. K. Kruschke, & M. A. Erickson. 2008. Rule-based extrapolation: A continuing challenge for exemplar models. *Psychonomic Bulletin and Review*, 15, 780-86.
- Downing, L. J., T. A. Hall, & R. Raffelsiefen, eds. 2005. *Paradigms in phonological theory*. Oxford: Oxford University Press.
- Eisner, F., & J. McQueen. 2005. The specificity of perceptual learning in speech processing. *Perception and Psychophysics*, 67, 227-34.
- Erickson, M. A., & J. K. Kruschke. 1998. Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, 127, 107-40.
- Erickson, M. A., & J. K. Kruschke. 2002. Rule-based extrapolation in perceptual categorization. *Psychonomic Bulletin & Review*, 9, 160-68.
- Ettlinger, M. 2009. The productivity of opaque interactions. Paper presented at LSA Annual Meeting, San Francisco, CA.
- Finley, S. 2008. *Formal and cognitive restrictions on vowel harmony*. Ph.D. Dissertation: Johns Hopkins University.

- Frigo, L., & J. L. McDonald. 1998. Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language*, 39, 218-45.
- Gasser, M. 1997. Transfer in a connectionist model of the acquisition of morphology. In R. H. Baayen & R. Schreuder, eds. *Yearbook of Morphology, 1996*, 97-116. Dordrecht: Foris.
- Gerken, L. 2006. Decisions, decisions: Infant language learning when multiple generalizations are possible. *Cognition*, 98, B67-B74.
- Gerken, L., & A. Bollt. 2008. Three exemplars allow at least some linguistic generalizations: Implications for generalization mechanisms and constraints. *Language Learning and Development*, 4, 228-48.
- Gerken, L., R. Wilson, & W. Lewis. 2005. Infants can use distributional cues to form syntactic categories. *Journal of Child Language*, 32, 249-68.
- Goldberg, A. F. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Goldberg, A. E., D. Casenhiser, & N. Sethuraman. 2004. Learning argument structure generalizations. *Cognitive Linguistics*, 14, 289-316.
- Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251-279.
- Goldrick, M., & M. Larson. In press. Constraints on the acquisition of variation. In C. Fougerson & M. D'Imperio, eds. *Papers in Laboratory Phonology 10: Variation, Detail and Representation*. Berlin: Mouton de Gruyter.
- Goodman, N. D., J. B. Tenenbaum, J. Feldman, & T. L. Griffiths. In press. A rational analysis of rule-based concept learning. *Cognitive Science*.



- Guest, D., G. Dell & J. Cole. 2000. Violable constraints in language production: Testing the transitivity assumption of Optimality Theory. *Journal of Memory and Language*, 42, 272-99.
- Guion, S. G. 1998. The role of perception in the sound change of velar palatalization. *Phonetica*, 55, 18-52.
- Hale, M., & C. Reiss. 2000. Substance abuse and dysfunctionality: Current trends in phonology. *Linguistic Inquiry*, 31, 157-69.
- Hale, M., & C. Reiss. 2003. The Subset Principle in phonology: Why the *tabula* can't be *rasa*. *Journal of Linguistics*, 39, 219-44.
- Hale, M., & C. Reiss. 2008. *The phonological enterprise*. Oxford: Oxford University Press.
- Halle, M. 1973. Prolegomena to a theory of word formation. *Linguistic Inquiry*, 4, 3-16.
- Halle, M., & A. Marantz. 1993. Distributed Morphology and the pieces of inflection. In K. Hale and S. J. Keiser, eds. *The view from Building 20*, 111-76. Cambridge: MIT Press.
- Hartsuiker, R. J. 2002. The addition bias in Dutch and Spanish phonological errors: The role of lexical context. *Language and Cognitive Processes*, 17, 61-96.
- Hay, J. 2003. *Causes and consequences of word structure*. London: Routledge.
- Hayes, B. 1999. Phonetically-driven phonology: The role of Optimality Theory and inductive grounding. In M. Darnell, E. Moravcsik, M. Noonan, F. Newmeyer, & K. Wheatley, eds. *Functionalism and formalism in linguistics (Vol. I, General papers)*, 243-285. Amsterdam: John Benjamins.

- Hayes, B., & C. Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39, 379-440.
- Hintzman, D. L. 1986. "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, 93, 411-28.
- Hock, H. 1991. *Principles of historical linguistics*. New York: Mouton de Gruyter.
- Hooper, J. 1976a. Word frequency in lexical diffusion and the source of morphophonological change. In W. Christie, ed. *Current progress in historical linguistics*, 96-105. Amsterdam: North Holland.
- Hooper, J. 1976b. *An introduction to Natural Generative Phonology*. New York, San Francisco, London: Academic Press.
- Hudson Kam, C. L., & E. Newport. 2005. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, 1, 151-95.
- Jäkel, F., B. Schölkopf, & F. A. Wichmann. 2008. Generalization and similarity in exemplar models of categorization: Insights from machine learning. *Psychonomic Bulletin & Review*, 15, 256-71.
- Jenkins, J. J. 1979. Four points to remember: A tetrahedral model of memory experiments. In L. S. Cermak & F. I. M. Craik, eds. *Levels of processing in human memory*, 429-46. Hillsdale, NJ: Erlbaum.
- Johnson, C. D. 1972. *Formal aspects of phonological description*. The Hague: Mouton.
- Johnston, J., M. Blatchley, & G. L. Olness. 1988. Miniature language system acquisition by children with different learning proficiencies. *Journal of Speech and Hearing Research*, 33, 335-42.

- Jusczyk, P. W., P. Smolensky, & T. Allocco. 2002. How English-learning infants respond to markedness and faithfulness constraints. *Language Acquisition*, 10, 31-73.
- Jusczyk, P. W., P. Smolensky, K. Arnold, & E. Moreton. 2003. Acquisition of nasal place assimilation by 4.5-month-old infants. In D. Houston, A Seidl, G. Hollich, E. Johnson, & A. Jusczyk, eds. *Jusczyk Lab Final Report*. Online publication.
- Kalish, M. L., T. L. Griffiths, & S. Lewandowsky. 2007. Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin and Review*, 14, 288-94.
- Kapatsinski, V. M. 2005. *Productivity of Russian stem extensions: Evidence for and a formalization of Network Theory*. M.A. Thesis: University of New Mexico.
- Kapatsinski, V. M. 2007. To scheme or to rule: Evidence against the Dual-Mechanism Model. In R. T. Cover and Y. Kim, eds. *Proceedings of the 31st Annual Meeting of the Berkeley Linguistics Society*, 193-204. Berkeley, CA: Berkeley Linguistics Society.
- Kapatsinski, V. M. 2007b. Frequency, neighborhood density, age-of-acquisition, lexicon size, neighborhood density and speed of processing: Towards a domain-general, single-mechanism account. In S. Buescher, K. Holley, E. Ashworth, C. Beckner, B. Jones, & C. Shank, eds. *Proceedings of the 6<sup>th</sup> Annual High Desert Linguistics Society Conference*, 121-40. Albuquerque, NM: High Desert Linguistics Society.
- Kapatsinski, V. M. In press. Testing theories of linguistic constituency with configural learning: The case of the English syllable. *Language*.

- Karttunen, L. 1993. Finite-state constraints. In J. Goldsmith, ed. *The last phonological rule: Reflections on constraints and derivations*, 173-94. Chicago and London: The University of Chicago Press.
- Kavitskaya, D. 2002. *Compensatory lengthening: phonetics, phonology, diachrony*. London & New York: Routledge.
- Kaye, J. 1974. Opacity and recoverability in phonology. *Canadian Journal of Linguistics*, 19, 134-49.
- Kempen, G., & K. Harbusch. 2005. The relationship between grammaticality ratings and corpus frequencies: A case study into word order variability in the midfield of German clauses. In S. Kepser & M. Reis, eds. *Linguistic evidence: Empirical, theoretical, and computational perspectives*, 329-50. Berlin, New York: Mouton de Gruyter.
- Kenstowicz, M. 1996. Base-identity and uniform exponence: alternatives to cyclicity. In J. Durand & B. Laks, eds. *Current trends in phonology: Models and methods*, 363-93. European Studies Research Institute and University of Salford.
- Kerswill, P. 1996. Children, adolescents, and language change. *Language Variation and Change*, 8, 177-202.
- Kiparsky, P. 1973. Abstractness, opacity and global rules. In O. Fujimura, ed. *Three dimensions in phonological theory*, 57-86. Tokyo: TEC Company.
- Kiparsky, P. 2000. Opacity and cyclicity. *The Linguistic Review*, 17, 351-65.
- Kirby, S., H. Cornish, & K. Smith. 2008. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences, USA*, 105, 10681-86.

- Kisseberth, C. 1970. On the functional unity of phonological rules. *Linguistic Inquiry*, 1, 291-306.
- Kisseberth, C. 1973. Is rule ordering necessary in phonology? In B. Kachru, ed. *Issues in linguistics: Papers in honor of Henry and Renee Kahane*, 418-41. Urbana: U of Illinois Press.
- Koo, H., & J. Cole. 2006. On learnability and naturalness as constraints on phonological grammar. In A. Botinis, ed. *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics*, 174-77. Athens: University of Athens.
- Köpcke, K.-M. 1988. Schemas in German plural formation. *Lingua*, 74, 303-35.
- Kruschke, J. K. 1992. ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22-44.
- Kruschke, J. K. 2008. Bayesian approaches to associative learning: From passive to active learning. *Learning and Behavior*, 36, 210-26.
- Langacker, R. W. 1987. *Foundations of Cognitive Grammar: Theoretical prerequisites*. Stanford: CSLI.
- Levikova, S. I. 2003. *Bol'shoy slovar' molodezhnogo slenga*. Moscow: Fair-Press.
- Levy, R. 2009. With uncertain input, rational sentence comprehension is good enough. Paper presented at LSA Annual Meeting, San Francisco, CA.
- Lobben, M. 1991. *Pluralization of Hausa nouns, viewed from psycholinguistic experiments and child language data*. M.Phil. Thesis: University of Oslo.
- Local, J. 1982. Modeling intonational variability in children's speech. In S. Romaine, ed. *Sociolinguistic variation in speech communities*, 85-103. London: Arnold.

- Marantz, A., & M. Halle. To appear. Clarifying “Blur”: Paradigms, defaults and inflectional classes In A. Bachrach & A. Nevins, eds. *The bases of inflectional identity*. Oxford: Oxford University Press.
- Marcus, G. F., S. Vijayan, S. B. Rao, & P. M. Vishton. 1999. Rule learning by seven-month-old infants. *Science*, 283, 77–80.
- Maye, J., & L. Gerken. 2001. Learning phonemes: How far can the input take us? In A. H.-J. Do, L. Dominguez, & A. Johansen, eds. *Proceedings of the 25th Annual Boston University Conference on Language Development*, 480-90. Somerville, MA: Cascadilla.
- Maye, J., D. J. Weiss, & R. N. Aslin. 2008. Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, 11, 122-34.
- McCarthy, J. J. 2002. *A thematic guide to Optimality Theory*. Cambridge: Cambridge University Press.
- McCarthy, J. J. 2005. Optimal paradigms. In Downing et al., 170-210.
- McCarthy, J. J. 2007. *Hidden generalizations: Phonological opacity in Optimality Theory*. Equinox.
- McKinley, S. C., & R. M. Nosofsky. 1995. Investigations of exemplar and decision bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 294-317.
- McLennan, C. T., P. A. Luce, and J. Charles-Luce. 2003. Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 539–553.

- Menn, L., & B. MacWhinney. 1984. The repeated morph constraint: Toward an explanation. *Language*, 60, 519-41.
- Mielke, J. 2008. *The emergence of distinctive features*. Oxford: Oxford University Press.
- Minsky, M. L., & S. A. Papert. 1969. *Perceptrons*. Cambridge, MA: MIT Press.
- Mitchell, T. M. 1997. *Machine learning*. McGraw-Hill.
- Moreton, E. 2008. Analytic bias and phonological typology. *Phonology*, 25, 83-127.
- Morgan, J. L., R. P. Meier, & E. L. Newport. 1989. Facilitating the acquisition of syntax with cross-sentential cues to phrase structure. *Journal of Memory and Language*, 28, 360-74.
- Mowrey, R., & W. Pagliuca. 1995. The reductive character of articulatory evolution. *Rivista di Linguistica*, 7, 37-124.
- Needham, A., G. Dueker, & G. Lockhead. 2005. Infants' formation and use of categories to segregate objects. *Cognition*, 94, 215-40.
- Newmeyer, F. 2005. *Possible and probable languages: A generative perspective on linguistic typology*. Oxford: Oxford University Press.
- Newport, E. L., & R. N. Aslin. 2000. Innately constrained learning: Blending old and new approaches to language acquisition. In S. C. Howell, S. A. Fish, & T. Keith-Lucas, eds. *Proceedings of the 24th Annual Boston University Conference on Language Development*, 1-21. Somerville, MA: Cascadilla.
- Newport, E. L., & R. N. Aslin. 2004. Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48, 127-162.

- Nooteboom, S. G. 1969. The tongue slips into patterns. In A. G. Sciarone, A. J. Van Essen, & A. A. Van Raad, eds. *Nomen. Leyden studies in linguistics and phonetics*, 114-32. The Hague: Mouton.
- Norris, D., & J. M. McQueen. 2008. Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115, 357-95.
- Nosofsky, R. M. 1988. Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 54-65.
- Nosofsky, R. M., & F. B. Bergert. 2007. Limitations of exemplar models of multi-attribute probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 999-1019.
- Nosofsky, R. M., S. E. Clark, & H. J. Shin. 1989. Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 282-304.
- Nosofsky, R. M., T. J. Palmeri, & S. C. McKinley. 1994. Rule-plus-exception model of classification learning. *Psychological Review*, 101, 53-79.
- Ohala, J. J. 1981. The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, & M. Miller, eds. *Papers from the Parasession on Language and Behavior*, 178-203. Chicago: Chicago Linguistic Society.
- Ohala, J. J. 2005. Phonetic explanations for sound patterns: implications for grammars of competence. In W. J. Hardcastle & J. Mackenzie Beck, eds. *A figure of speech: A festschrift for John Laver*, 23-38. Mahwah, NJ: Erlbaum.
- Onishi, K., K. Chambers, & C. Fisher. 2002. Learning phonotactic constraints from brief auditory experience. *Cognition*, 83, B13-B23.



- Palmeri, T. J., S. D. Goldinger, & D. B. Pisoni. 1993. Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 309-28.
- Paradis, C. 1989. On constraints and repair strategies. *The Linguistic Review*, 6, 71-97.
- Pater, J., & A.-M. Tessier. 2003. Phonotactic knowledge and the acquisition of alternations. *Proceedings of the XVth International Congress on Phonetic Sciences*, 1777-1180.
- Peperkamp, S. 2007. Do we have innate knowledge about phonological markedness? Comments on Berent, Steriade, Lennertz, and Vaknin. *Cognition*, 104, 631-7.
- Peperkamp, S., & E. Dupoux. 2007. Learning the mapping from surface to underlying representations in an artificial language. In J. Cole & J. I. Hualde, eds. *Laboratory Phonology 9*, 315-38. Berlin, New York: Mouton de Gruyter.
- Peperkamp, S., K. Skoruppa, & E. Dupoux. 2006. The role of phonetic naturalness in phonological rule acquisition. In D. Bamman, T. Magnitskaia, & C. Zaller, eds. *Proceedings of the 30<sup>th</sup> Annual Boston University Conference on Language Development*, 464-75. Somerville, MA: Cascadilla.
- Perez, E., J. Santiago, A. Palma, & P. O'Seaghda. 2007. Perceptual bias in speech error data collection: Insights from Spanish speech errors. *Journal of Psycholinguistic Research*, 36, 207-35.
- Perfors, A., J. Tenenbaum, & T. Regier. 2006. Poverty of the stimulus? A rational approach. In R. Sun & N. Miyake, eds. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*.

- Pertz, D. L., & T. G. Bever. 1975. Sensitivity to phonological universals in children and adolescents. *Language*, 51, 149–62.
- Pierrehumbert, J. B. 2006. The statistical basis of an unnatural alternation. In L. Goldstein, D.H. Whalen, & C. Best, eds. *Laboratory Phonology VIII, Varieties of phonological competence*, 81-107. Berlin: Mouton de Gruyter.
- Pinker, S. 1999. *Words and rules: The ingredients of language*. New York: Harper Collins.
- Plag, I. 2003. *Word formation in English*. Cambridge: Cambridge University Press.
- Pollard, C., & I. A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago and Stanford: University of Chicago Press and CSLI Publications.
- Port, R. F., & Leary, A. P. 2005. Against formal phonology. *Language*, 81, 927-64.
- Pothos, E. M. 2005. The rules vs. similarity distinction. *Behavioral and Brain Sciences*, 28, 1-49.
- Pothos, E. M. 2007. Theories of artificial grammar learning. *Psychological Bulletin*, 133, 227-44.
- Prince, A. & Smolensky, P. 1993/2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell.
- Pycha, A., P. Nowak, E. Shin, & R. Shosted. 2003. Phonological rule-learning and its implications for a theory of vowel harmony. In M. Tsujimura & G. Garding, eds. *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, 101-14. Somerville, MA: Casacadilla.
- Quinn, P. C., & R. S. Bhatt. 2005. Learning perceptual representations in infancy. *Psychological Science*, 16, 511-15.

- Reber, A. S. 1967. Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6, 317–27.
- Recchia, G., B. T. Johns, & M. N. Jones. 2008. Context repetition benefits are dependent on context redundancy. *Proceedings of the 30th Annual Cognitive Science Society*.
- Roberts, J. 1996. Acquisition of variable rules: A study of (-t, -d) deletion in preschool children. *Journal of Child Language*, 24, 351–72.
- Roca, I. 1997. Derivations of constraints, or derivations and constraints? In I. Roca, ed. *Derivations and constraints in phonology*, 3–41. Oxford: Clarendon Press.
- Roediger, H. L., III. 2008. Relativity of remembering: Why the laws of memory vanished. *Annual Review of Psychology*, 59, 225–54.
- Rumelhart, D. E., & J. L. McClelland. 1986. On learning the past tenses of English verbs. In D. E. Rumelhart & J. L. McClelland, eds. *Parallel distributed processing: Vol.2: Psychological and biological models*, 216–71. Cambridge, MA: MIT Press.
- Saffran, J. R., & E. D. Thiessen. 2003. Pattern induction by infant language learners. *Developmental Psychology*, 39, 484–94.
- Sag, I., & T. Wasow. 1999. *Syntactic Theory: A Formal Introduction*. Stanford: CSLI.
- Schane, S., B. Tranel, & H. Lane. 1974/75. On the psychological reality of a natural rule of syllable structure. *Cognition*, 3/4, 351–58.
- Scobbie, J. M., J. S. Coleman, & S. Bird. 1996. Key aspects of Declarative Phonology. In J. Durand and B. Laks, eds. *Current trends in phonology: Models and methods*, 363–93. European Studies Research Institute and University of Salford.
- Seidl, A., & Buckley, E. 2005. On the learning of arbitrary phonological rules. *Language Learning and Development*, 1, 289–316.

- Shepard, R. N., C. Hovland, & H. M. Jenkins. 1961. *Learning and memorization of classifications*. Psychological Monographs, 75 (13, Whole No. 517).
- Sheveleva, M. S. 1974. *Obratnyj slovar' russkogo jazyka. Okolo 125000 slov*. Moscow: Sovetskaja Enciklopedija.
- Skousen, R. 1989. *Analogical modeling of language*. Dordrecht: Kluwer.
- Skousen, R. 2002. An overview of analogical modeling. In R. Skousen, D. Lonsdale, & D. B. Aprkinson, eds. *Analogical modeling: An exemplar-based approach to language*, 11-26. Amsterdam: John Benjamins.
- Smith, N.V., I.-M. Tsimpli & J. Ouhalla. 1993. Learning the impossible: The acquisition of possible and impossible languages by a polyglot savant. *Lingua*, 91, 279-347.
- Smolensky, P., & G. Legendre. 2006. *The harmonic mind: From neural computation to Optimality-Theoretic Grammar*. Cambridge, MA: MIT Press.
- Stemberger, J. P. 1981. Morphological haplology. *Language*, 57, 791-817.
- Stemberger, J. P. 1990. Wordshape errors in language production. *Cognition*, 35, 123-57.
- Sternberg, S. 1967. Two operations in character-recognition: Some evidence from reaction-time measurements. *Perception and Psychophysics*, 2, 45-53.
- Sternberg, S. 1998. Discovering mental processing stages: The method of additive factors. In D. Scarborough & S. Sternberg, eds. *Methods, models, and conceptual issues. An Invitation to Cognitive Science Vol. 4*, 703-864. Cambridge, MA: MIT press.
- Storkel, H. L. 2004. Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics*, 25, 201-21.

- Sumner, M., and A. G. Samuel. 2005. Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language*, 52, 322-38.
- Swingley, D., & R. N. Aslin. 2007. Lexical competition in young children's word learning. *Cognitive Psychology*, 54, 99-132.
- Thomason, S. G. 1976. What else happens to opaque rules? *Language*, 52, 370-81.
- Tversky, A., and D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-31.
- Valian, V., & S. Coulson. 1988. Anchor points in language learning: The role of marker frequency. *Journal of Memory and Language*, 27, 71-86.
- del Viso, S., J. M. Igoa, & J. E. Garcia-Albea. 1991. On the autonomy of phonological encoding: Evidence from slips of the tongue in Spanish. *Journal of Psycholinguistic Research*, 20, 161-85.
- Wang, H. S., and B. L. Derwing. 1994. Some vowel schemas in three English morphological classes: Experimental evidence. In M. Y. Chen & O. C. L. Tseng, eds. *In honor of Professor William S.-Y. Wang: Interdisciplinary studies on language and language change*, 561-75. Taipei: Pyramid Press.
- Warker, J. A., G. S. Dell, C. A. Whalen, & S. Gereg. 2008. Limits on learning phonotactic constraints from recent production experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1289-95.
- Warren, R. M. 1961. Illusory changes of distinct speech upon repetition—the verbal transformation effect. *British Journal of Psychology*, 52, 249–58.
- Warren, R. M. 1996. Auditory illusions and perceptual processing of speech. In N. J. Lass, ed. *Principles of experimental phonetics*, 435–466. St. Louis: Mosby.

- Wedel, A. B. 2007. Feedback and regularity in the lexicon. *Phonology*, 24, 147-85.
- Weinert, S. 2009. Implicit and explicit modes of learning: Similarities and differences from a developmental perspective. *Linguistics*, 47, 241-71.
- Wexler, K., & P. Culicover. 1980. *Formal principles of language acquisition*. Cambridge, MA: MIT Press.
- Williams, J. N. 2003. Inducing abstract linguistic representations: Human and connectionist learning of noun classes. In R. van Hout, A. Hulk, F. Kuiken, & R. Towell, eds. *The lexicon-syntax interface in second language acquisition*, 151-74. Amsterdam: John Benjamins.
- Williams, J. N., & P. Lovatt. 2003. Phonological memory and rule learning. *Language Learning*, 53, 67-121.
- Wilson, C. 2003. Experimental investigation of phonological naturalness. In M. Tsujimura & G. Garding, eds. *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, 533-46. Somerville, MA: Cascadilla.
- Wilson, C. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*, 30, 945-82.
- Xu, F., & J. B. Tenenbaum. 2007. Word learning as Bayesian inference. *Psychological Review*, 114, 245-72.
- Zaba, A. 2008. *Relative frequency of patterns and learnability: The case of phonological harmony*. Ph.D. Dissertation: University of Utah.
- Zapf, J., & L. B. Smith. 2008. Meaning matters in children's plural productions. *Cognition*, 108, 466-76.

- Zhang, Y., & Y. Lai. 2006. Testing the role of phonetic naturalness in Mandarin tone sandhi. *Kansas Working Papers in Linguistics*, 28, 65-126.
- Zipf, G. K. 1935. *The psychobiology of language*. Houghton-Mifflin.
- Zuraw, K. 2000. *Patterned exceptions in phonology*. Ph.D. Dissertation: UCLA.

## Curriculum vitae

Vsevolod M. Kapatsinski

vkapatsi@umail.iu.edu  
812-855-4893

Speech Research Laboratory  
Psychological and Brain Sciences  
Indiana University  
Bloomington, IN 47405-1301

b. 10/11/1984, Gorky, USSR (now Nizhny Novgorod, Russia)

### Education

Ph.D., Linguistics and Cognitive Science, Indiana University, Bloomington, May 2009.

Thesis: *The architecture of grammar in artificial grammar learning: Formal biases in the acquisition of morphophonology and the nature of the learning task.*

Thesis committee: Kenneth DeJong (co-chair, Linguistics), David Pisoni (co-chair, Cognitive Science), Robert Port, Linda Smith.

M.A., Linguistics, University of New Mexico, 2005.

Thesis: *Productivity of Russian stem extensions: Evidence for and a formalization of Network Theory.* Thesis committee: Joan Bybee (chair), Jill Morford, Caroline Smith

B.A. Linguistics, University of New Mexico, 2003 (4.0 GPA).

A.A. Liberal Arts, Simon's Rock College of Bard, 2002 (3.9 GPA).

Participated in 2003, 2005, 2007 Linguistic Society of America summer Institutes at Michigan State University, M.I.T. and Stanford respectively.

Masterclass on Corpus Methods in Linguistics and Language Pedagogy (co-taught by Drs. Gries, Divjak, Clancy, Goldsmith & Regier), University of Chicago, March 26-30, 2008

### Research Interests

experimental (psycho)linguistics  
corpus linguistics  
computational models of language  
laboratory phonology  
mental lexicon  
explanations for language change  
stochastic grammar  
usage-based linguistics  
morphology  
phonology

### Publications

#### Journals

1. Kapatsinski, V. Under review. *Modularity in the channel: A response to Moreton (2008).* *Phonology*.
2. Kapatsinski, V. Forthcoming a. *Velar palatalization in Russian and artificial grammar: Constraints on models of morphophonology.* *Laboratory Phonology*
- 3.. Kapatsinski, V. Forthcoming b. *Testing theories of linguistic constituency with configural learning: The case of the English syllable.* *Language*.



4. Kapatsinski, V. Forthcoming c. Frequency of use leads to automaticity of production: Evidence from repair in conversation. *Language and Speech*.
5. Kapatsinski, V. Forthcoming d. Adversative conjunction choice in Russian: Semantic and syntactic influences on lexical selection. *Language Variation and Change*.
6. Kapatsinsky, V. 2007. Sex associations of Russian generics. *Journal of Slavic Linguistics*, 14 (1). (written in 2004)

#### **Refereed conference proceedings**

1. Kapatsinski, V., and J. Radicke. Forthcoming. Frequency and the emergence of prefabs: Evidence from monitoring. In R. Corrigan, E. Moravcsik, H. Ouali, & K. Wheatley, eds. *Formulaic Language. Vol. II: Acquisition, loss, psychological reality, functional explanations*. Amsterdam: John Benjamins. (Typological Studies in Language book series).
2. Kapatsinski, V. 2007a. Frequency, neighborhood density, age-of-acquisition, lexicon size, neighborhood density and speed of processing: Towards a domain-general, single-mechanism account. In S. Buescher, K. Holley, E. Ashworth, C. Beckner, B. Jones, and C. Shank, eds. *Proceedings of the 6<sup>th</sup> Annual High Desert Linguistics Society Conference*, 121-40. Albuquerque, NM: High Desert Linguistics Society.
3. Kapatsinski, V. 2006a. Towards a single-mechanism account of frequency effects. In S. J. Hwang, W. J. Sullivan, and A. R. Lommel, eds. *The LACUS Forum XXXII: Networks*, 325-35. The Linguistic Association of Canada and the United States.
4. Kapatsinski, V. 2005a. Characteristics of a rule-based default are dissociable: Evidence against the Dual Mechanism Model. In S. Franks, F. Y. Gladney, and M. Tasseva-Kurkchieva, eds. *[Formal] Approaches to [Slavic] Linguistics: The South Carolina Meeting*, 136-46. Ann Arbor, MI: Michigan Slavic Publications.

#### **Non-refereed conference proceedings**

1. Kapatsinski, V. Forthcoming d. Rethinking rule reliability: Why an exceptionless rule can fail. *CLS* 44.
2. Kapatsinski, V. 2008a. Constituents can exhibit partial overlap: Experimental evidence for an exemplar approach to the mental lexicon. In R. L. Edwards, P. J. Midtlyng, C. L. Sprague, and K. G. Stensrud, eds. *Chicago Linguistic Society 41: The Panels*, 227-242. Chicago, IL: Chicago Linguistic Society.
3. Kapatsinski, V. 2007b. To scheme or to rule: Evidence against the Dual-Mechanism Model. In R. T. Cover and Y. Kim, eds. *Proceedings of the 31<sup>st</sup> Annual Meeting of the Berkeley Linguistics Society*, 193-204. Berkeley, CA: Berkeley Linguistics Society.
4. Kapatsinski, V. 2006b. Phonological similarity relations: Network organization of the mental lexicon. In M. Leyva and M. del Carmen, eds. *Memoria del VIII Encuentro Internacional de Lingüística en el Noroeste, tomo III*. Hermosillo, Mexico: UniSon.
5. Kapatsinski, V. 2005b. Measuring the relationship of structure to use: Determinants of the extent of recycle in repetition repair. In M. Ettlinger, N. Fleischer, and M. Park-Doob, eds. *Proceedings of the 30<sup>th</sup> Annual Meeting of the Berkeley Linguistics Society*, 481-92. Berkeley, CA: Berkeley Linguistics Society.

#### **Working papers and progress reports**

1. Kapatsinski, V. 2008b. Principal components of sound systems: An exercise in multivariate statistical typology. *Indiana University Linguistics Club Working Papers Online* 8.
2. Kapatsinski, V. 2007c. Frequency of use leads to automaticity of production: Evidence from repair in conversation. *Research on Spoken Language Processing Progress Report No. 28*, 161-86. Indiana University: Bloomington, IN.

3. Kapatsinski, V. 2007d. Implementing and testing theories of linguistic constituency I: English syllable structure. *Research on Spoken Language Processing Progress Report No. 28*, 241-76. Indiana University: Bloomington, IN.
4. Kapatsinski, V., & J. Radicke. 2007. Frequency and the emergence of prefabs: Evidence from monitoring. *Research on Spoken Language Processing Progress Report No. 28*, 297-314. Indiana University: Bloomington, IN.
5. Kapatsinski, V. 2005c. Sound similarity relations in the mental lexicon: Modeling the lexicon as a complex network. *Research on Spoken Language Processing Progress Report No. 27*, 133-52. Indiana University: Bloomington, IN.

### **Conference Presentations**

1. Kapatsinski, V. 2009a. Principal components of sound systems. Poster to be presented at Association for Linguistic Typology 8<sup>th</sup> Biennial Meeting, Berkeley, CA, July 23-26.
2. Kapatsinski, V. 2009b. Experimental evidence for product-oriented generalizations. Paper presented at the Linguistic Society of America Annual Meeting, San Francisco, CA, January 8-11.
3. Kapatsinski, V. 2008c. Experimental evidence for product-oriented generalizations: Testing the shape of the grammar. Paper presented at 8<sup>th</sup> High Desert Linguistics Society Conference, Albuquerque, NM, November 6-8.
4. Kapatsinski, V. 2008d. Product-oriented generalization over the (artificial) lexicon. Poster presented at the 6<sup>th</sup> International Conference on the Mental Lexicon, Banff, Alberta, Canada, October 7-10.
5. Kapatsinski, V. 2008e. Learnability of rime-affix vs. body-affix dependencies in English: Introducing a new source of evidence on the nature of constituent structure. Paper presented at the 18<sup>th</sup> International Congress of Linguists, Seoul, Korea, July 21-26.
6. Kapatsinski, V., & D. B. Pisoni. 2008. The role of phonetic detail in associating phonological units. Poster presented at Laboratory Phonology 11, Wellington, New Zealand, June 30 – July 2.
7. Kapatsinski, V. 2008f. Rule reliability and productivity: Velar palatalization in Russian and artificial grammar. Paper presented at Laboratory Phonology 11, Wellington, New Zealand, June 30 – July 2.
8. Kapatsinski, V. 2008g. Why an exceptionless rule can fail: Rethinking rule reliability. Paper presented at the 44<sup>th</sup> annual meeting of the Chicago Linguistic Society, Chicago, IL, April 24-26.
9. Kapatsinski, V. 2008h. Implementing and testing theories of syllable structure. Poster presented at the CUNY Phonology Forum Conference on the Syllable, New York, NY, January 17-19.
10. Kapatsinski, V. 2008i. The influence of syllabic constituency on learning CV-affix vs. VC-affix associations: Constituency is more than dependency. Paper presented at the Annual Meeting of the Linguistic Society of America, Chicago, IL, January 3-6.
11. Kapatsinski, V. 2007e. Does high frequency lead to automaticity? A corpus study. Poster presented at the Workshop on Variation, Gradience, and Frequency in Phonology, Stanford, CA, July 6-8.
12. Kapatsinski, V., & J. Radicke. 2007. Frequency and the emergence of prefabs: Evidence from monitoring. Paper presented at UWM Linguistics Symposium on Formulaic Language, Milwaukee, WI, April 18-21.
13. Kapatsinski, V. 2007f. Rules and analogy in Russian loanword adaptation. Paper presented at the Annual Meeting of the Linguistic Society of America, Anaheim, CA, January 4-7.
14. Kapatsinski, V. 2006c. Frequency and cohesion: Evidence from repair. Paper presented at the 7<sup>th</sup> High Desert Linguistics Society Conference, Albuquerque, NM, November 9-11.

15. Kapatsinski, V. 2006d. Testing psychological reality of complex units with XOR learning. Poster presented at the Fifth International Conference on the Mental Lexicon, Montreal, QC, October 11-13.
16. Kapatsinski, V. 2006e. Syntactic variation in Russian: Adversative conjunctions and sentence production. Paper presented at the 1<sup>st</sup> Conference of the Slavic Linguistics Society, Bloomington, IN, September 8-10.
17. Kapatsinski, V. 2006f. High token frequency leads to greater cohesion: Evidence from repair. Paper presented at the 33<sup>rd</sup> LACUS Forum, Toronto, ON, July 31-August 4.
18. Kapatsinski, V. 2006g. Frequency, salience, and the nature of sound similarity. Poster presented at Laboratory Phonology 10, Paris, France, June 29 – July 1.
19. Kapatsinski, V. 2006h. Ways of testing psychological reality of complex units: XOR learning and unit detection. Invited paper presented at PhonologyFest Workshop: Current Perspectives in Phonology, Bloomington, IN, June 23-24.
20. Kapatsinski, V. 2006i. Popularity does not lead to insensitivity: Evidence contra the Dual Mechanism Model of morphological productivity. Paper presented at Hoosier Mental Life, Bloomington, IN, March 31- April 2.
21. Kapatsinski, V. 2006j. Having something common in common is not the same as sharing something special: Evidence from sound similarity judgments. Paper presented at the Annual Meeting of the Linguistic Society of America, Albuquerque, NM, January 5 – 8.
22. Kapatsinski, V. 2005d. Towards a single-mechanism account of frequency effects. Paper presented at the 32<sup>nd</sup> LACUS Forum, Hanover, NH, August 2 – 6.
23. Kapatsinski, V. 2005e. Words as localist associative networks: Empirical evidence for an exemplar approach to the mental lexicon. Paper presented at the 41<sup>st</sup> Annual Meeting of the Chicago Linguistic Society, Chicago, IL, April 7 - 9.
24. Kapatsinski, V. 2005f. To scheme or to rule: Evidence against the Dual Mechanism Model. Paper presented at the 31<sup>st</sup> Annual Meeting of the Berkeley Linguistics Society, Berkeley, CA, February 8 - 10.
25. Kapatsinski, V. 2004a. Phonological similarity relations: Network organization of the lexicon and phonology. Paper presented at the VIII Encuentro Internacional de Lingüística en el Noroeste, Hermosillo, Sonora, November 17 - 19.
26. Kapatsinski, V. 2004b. Frequency effects in priming and associative learning: Towards a single-mechanism account. Paper presented at the 6<sup>th</sup> High Desert Linguistics Society Conference, Albuquerque, NM, November 4 - 6.
27. Kapatsinski, V. 2004c. Russian *but*s: Variation in adversative constructions. Poster presented at New Ways of Analyzing Variation, 33, Ann Arbor, MI, September 30 – October 3.
28. Kapatsinski, V. 2004d. The role of phonotactics in morphological productivity. Poster presented at the Fourth International Conference on the Mental Lexicon, Windsor, ON, June 30 – July 4.
29. Kapatsinski, V. 2004e. Characteristics of a rule-based default are dissociable: Evidence against the Dual Mechanism Model. Paper presented at Formal Approaches to Slavic Linguistics 13, Columbia, SC, February 27 - 29.
30. Kapatsinski, V. 2004f. Measuring the relationship of structure to use: Determinants of the extent of recycle in repetition repair. Paper presented at the 30<sup>th</sup> Annual Meeting of the Berkeley Linguistics Society, Berkeley, CA, February 13 - 16.
31. Kapatsinski, V. 2003. Names in Russian online discussion: Russian-English contact on the Internet. Paper presented at the 32<sup>nd</sup> Annual Meeting of the Linguistic Association of the Southwest, Edinburg, TX, October 17 – 19.

### **Fellowships and awards**

2008 Best Paper Presented by a Student Award at Laboratory Phonology 11.

2008 NSF Travel Grant for travel to Laboratory Phonology 11 in Wellington NZ, through U of Minnesota

2005-present NIH Pre-Doctoral traineeship awarded to David Pisoni, Indiana University

2003-2005 research assistantship, working with Joan Bybee, University of New Mexico

2003, 2005, 2007 Summer Institute Fellowship, Linguistic Society of America

2004 University of New Mexico GPSA Graduate Research and Development grant

2002 Divisional Prize, Division of Foreign Languages and Literatures, Simon's Rock College

2000-2002 Acceleration to Excellence Scholarship, Simon's Rock College

### **Teaching experience**

2008 Associate Instructor for "Introduction to the Study of Language", Linguistics 103, Dr.

Richard Janda (responsible for three discussion sections per week, grading)

2009 Grader for L210, "Topics in language and society: Language and the Lord of the Rings", Dr. Richard Janda.

### **Professional service**

2006-2008. Associate editor, webmaster for Indiana University Linguistics Club Working Papers Online.

Ad hoc reviewer for *Cognitive Science*, *Laboratory Phonology 11* proceedings, *Origins of Sound Change: Approaches to Phonologization*, *IULC Working Papers Online*, High Desert Linguistics Society 7, International Cognitive Linguistics Conference 11.

### **Professional societies**

Linguistic Society of America

### **Languages**

Natural: English, Russian, Spanish

Computer: Matlab, C, R