CONSPIRING TO MEAN: EXPERIMENTAL AND COMPUTATIONAL EVIDENCE FOR A USAGE-BASED HARMONIC APPROACH TO MORPHOPHONOLOGY

VSEVOLOD KAPATSINSKI

University of Oregon

This article reports on an experiment with miniature artificial languages that provides support for a synthesis of ideas from USAGE-BASED PHONOLOGY (Bybee 1985, 2001, Nesset 2008) and HARMONIC GRAMMAR (Legendre et al. 1990, Smolensky & Legendre 2006). All miniature artificial languages presented to subjects feature velar palatalization (*k* → *tʃ*) before a plural suffix, *-i*. I show that (i) examples of *-i* simply attaching to a [tʃ]-final stem help palatalization (supporting *t* → *tʃi* over *t* → *ti* and *p* → *tʃi* over *p* → *pi*), a finding that provides specific support for product-oriented schemas like 'plurals should end in [tʃi]'; (ii) learners tend to perseverate on the form they know, leveling stem changes, which provides support for paradigm-uniformity constraints in favor of retaining gestures composing the known form, for example, 'keep labiality'; and (iii) the same plural schema helps untrained singular-plural mappings more than it helps trained ones. This result is accounted for by proposing that schemas and paradigm-uniformity constraints clamor for candidate plural forms that obey them. Given that competition is between candidate outputs, the same schema provides more help to candidates that violate strong paradigm-uniformity constraints and are therefore weak relative to competitor candidates. A computational model of schema extraction is proposed.*

Keywords: morphophonology, miniature artificial-language learning, computational modeling, palatalization, schemas, rules, constraints

1. INTRODUCTION. All currently popular theories of grammar subscribe to the cognitive commitment: we are interested in describing what speakers of human languages know about the languages they speak, and the sound patterns of those languages in particular, not simply in describing the corpus of utterances we observe in the most parsimonious way possible (Albright & Hayes 2003, Bybee 2001, Chomsky & Halle 1968, Daelemans & van den Bosch 2005, Goldberg 1995, Langacker 1987, Nesset 2008, Prince & Smolensky 2004 [1993]). All theories that make a cognitive commitment place restrictions on the types of generalizations language learners make on the basis of the primary linguistic data they are exposed to.

The present article investigates the types of generalizations extracted by learners from a miniature artificial lexicon where the primary linguistic data can be precisely controlled (e.g. Aslin et al. 1998, Peperkamp 2003, Wilson 2006). The advantage of such data is that the human learners and the model can both be exposed to the same training. Thus, the performance of the model is easier to evaluate. The disadvantage is that we can never provide learners with as much experience (and as much experience of the right kind) as in real language acquisition situations. Therefore, we can never prove that some generalization is UNlearnable. We can, however, show that some generalizations ARE learnable and that some generalizations are learned more readily than others (from the kind of experience we provide the learners with). The ultimate goal of this kind of work is what

* Many thanks to Adele Goldberg, Bruce Hayes, Lisa Redford, the anonymous referees, and all students in L407/507 Probabilistic Grammar and L614 Theory of Phonology for comments on previous drafts of this article. I am also grateful to Nina Rinaldi and Matthew Stave for running subjects and to students from the Psychology/Linguistics subject pool for participating in this study. This work was supported by the Junior Professorship Award from the author's university. Parts of this work were presented at the 2010 meeting of the High Desert Linguistics Society, the 2011 and 2012 annual meetings of the Linguistic Society of America, Computational Modeling of Sound Pattern Acquisition, the American International Morphology Meeting, and Laboratory Phonology 12. I thank audience members from these conferences for helpful discussion.

Hayes and Wilson (2008) called LEARNING-THEORETIC PHONOLOGY: ultimately we want to be able to predict which generalizations will be supported by a given perceptual or production experience, given the learner's prior experience and inherent bias, and which other generalizations will lose strength as a result of that experience.

1.1. TYPES OF GENERALIZATIONS IN MORPHOPHONOLOGY. I begin by reviewing the types of morphophonological generalizations that have been proposed and identifying the crucial differences among them that lead to differing predictions with respect to the data from the present study. In particular, it is shown that rules predict that examples like 'singular [sitʃ]'/'plural [sitʃi]' should not help palatalization before *-i* ({*k*;*t*;*p*} → *tʃ*/ __ *i*), whereas schemas predict that they should, even if second-order schemas are allowed. This difference follows from the defining features of rules (changes in context) and schemas (form-meaning pairings).

RULES. RULE-BASED PHONOLOGY (Albright & Hayes 2003, Chomsky & Halle 1968, Plag 2003, Reiss 2004, inter alia) suggests that knowledge of grammar consists largely of knowledge of RULES. A rule describes a change and the context in which that change has been observed to occur and is to be carried out (e.g. Reiss 2004). In its strongest version (Reiss 2004), rule-based phonology proposes that phonology is based ONLY on rules.

In the present study human learners were presented with miniature artificial languages featuring an alternation between [k]-final singulars and [tʃi]-final plurals, as shown in 1. This alternation is known as VELAR PALATALIZATION. The experiment held constant the set of examples of velar palatalization (…k_{SG}/…$tʃi_{PL}$) and varied the numbers of examples of the other types: …{*t*;*p*}$_{SG}$/…{*t*;*p*}i_{PL} and …$tʃ_{SG}$/…$tʃi_{PL}$ or …$tʃ_{SG}$/…$tʃu_{PL}$. As we see below, theories of phonology differ from each other in their predictions about the effects of such examples on the learnability of palatalization. I begin with the predictions of rule-based phonology.

(1)

SG	PL	SG	PL	SG	PL	SG	PL
vuk	vutʃi	vut	vuti	vup	vupi	vutʃ	vutʃi
brak	bratʃi	brat	brati	brap	brapi	bratʃ	bratʃi
sik	sitʃi	sit	siti	sip	sipi	sitʃ	sitʃi
…k	…tʃi	…t	…ti	…p	…pi	…tʃ	…tʃi

In rule-based phonology, learners of the language in 1 are proposed to make the generalizations in 2, where '#' stands for a word boundary: to form the plural one suffixes *-i* to the stem (formally, nothing turns into *-i* in the context of a preceding consonant and a following word boundary), and then [k] changes into [tʃ] before *-i*. Note that all forms in 1 except those in the left-most pair of columns contribute to the strength of the rule ∅ → *i*/*C* __ and say nothing about the strength of the rule *k* → *tʃi*. In other words, they are irrelevant for productivity of palatalization. This includes the *tʃ* → *tʃi* forms in the rightmost pair of columns in 1.

(2) a. ∅ → i/[–cont] __ # when plural[1]
b. k → tʃ/ __ i#

[1] I am adopting the features and notation from Hayes 2009, which I take to be relatively uncontroversial. Nothing hinges on the choice of the feature system, except that, as we see later, negative constraints seem to require binary place features to account for the data (in order to make e.g. [–Pal] stateable), while the minimal generalization learner performs better with privative features. Whether 2a refers to [–cont], C, or nothing at all as the left context is not relevant for the present argument.

How can the generalizations in 2 be extracted from the data in 1? According to the MINIMAL GENERALIZATION LEARNER, a computational model of rule induction proposed in Albright & Hayes 2003, one splits each pair of forms into change and context, yielding 3, and then generalizes over contexts. When the model is applied to the language in 1 it yields the rules in 4.

(3) k → tʃi/vu __ # Ø → i/vut __ # Ø → i/vup __ # Ø → i/vutʃ __ #
k → tʃi/bra __ # Ø → i/brat __ # Ø → i/brap __ # Ø → i/bratʃ __ #
k → tʃi/si __ # Ø → i/sit __ # Ø → i/sip __ # Ø → i/sitʃ __ #

(4) a. k → tʃi/V __ #
b. Ø → i/[–cont] __ #

For the present purposes, there is one crucial difference between the rules in 4 and the rules in 2: the rules in 4 are extracted by an algorithm that does not minimize competition between rules. Rather, Albright and Hayes (2003) reward rules for both reliability and generality. Thus, 4a and 4b are in competition for stems that end in /k/: 4a argues that the plural forms of such stems should end in [tʃi], while 4b argues that they should end in [ki]. This happens because the contexts of the two rules Ø → *i*/[p] __ and Ø → *i*/[t] __ cannot be unified into a single rule while excluding [k]: [p] and [t] are not a natural class, and the more specific rules Ø → *i*/[p] __ and Ø → *i*/[t] __ are punished for their specificity. Empirically, the rules in 4 differ from the rules in 2 in predicting that examples of *t* → *ti* and *p* → *pi* support *k* → *ki* and predicting that examples of *tʃ* → *tʃi* should reduce the productivity of velar palatalization by providing support for the competing rule in 4b.[2]

PHONOTACTICS, CONSTRUCTIONS, AND FIRST-ORDER/PRODUCT-ORIENTED SCHEMAS. By positing rules as the basic type of linguistic generalization, rule-based phonology assumes that human language learners automatically compare pairs of morphologically related word forms and split them into a change and a context (an assumption also shared by some analogical models, e.g. Keuleers 2008). Such a comparison process was also traditionally assumed in the domain of visual scene perception but has famously been shown to be extremely fallible in change-blindness experiments (Simons 1996). A change to a visual display often remains undetected. Further, as shown by Mitroff and colleagues (2004), among others, it is possible for the subject to fail to notice that an object has been replaced in a visual display despite being able to report having seen both the prechange object and the postchange object in follow-up tests. Mitroff and colleagues suggest, in the title of their article, that 'nothing compares two views'. While this might be too strong a statement, the results strongly suggest that the comparison process is fallible.[3] Thus generalizations over comparisons between words (types of changes and contexts for them) seem like a shaky ground on which to build a psychologically realistic theory of phonology. The alternative basic building block, proposed in phonological the-

[2] The rules also differ in whether they assume one or two processing stages. The traditional account claims that the suffix is chosen first, and if the chosen suffix is *-i* then the preceding segment is palatalized. The question of the number of stages is not at issue here: if the minimal generalization learner is asked to learn what to change into [tʃ] by training on only singular-plural pairs involving *-i*, thus modeling the second stage in the traditional analysis, it extracts the competing rules *k* → *tʃ* and Ø → Ø ('do nothing'), resulting in the same prediction: competition for [k] between the palatalizing rule and the 'do nothing' rule, the latter supported by *tʃ* → *tʃi*, *t* → *ti*, and *p* → *pi*.

[3] Massaro (1970) likewise shows the fallibility of sound-comparison processes: whether two stimuli are acoustically identical is difficult to determine if they are not temporally adjacent. Pierrehumbert (1993) uses the proposal that between-segment comparisons are fallible to account for the distance effect on the obligatory contour principle.

ories arising from the functional/cognitive/constructionist tradition, is what Bybee (1985, 2001, Bybee & Moder 1983, Bybee & Slobin 1982) calls PRODUCT-ORIENTED SCHEMAS, also known as CONSTRUCTIONS (Booij 2008, 2010, Goldberg 1995, 2002) and FIRST-ORDER SCHEMAS (Langacker 1987, Nesset 2005, 2008).

Instead of being generalizations about parallel semantic and phonological changes in context, these are generalizations about form-meaning pairings. Thus, one might notice that plurals in a language often end in [tʃi], whether that [tʃi] corresponds to [k] or [tʃ] in the singular. The speaker of such a language might then derive [tʃi]-final plurals from novel singulars ending in [k] (or anything else) because s/he thinks that word-final [tʃi] is a good expression of plural meaning. In this way, examples of *-i* simply attaching to a [tʃ]-final stem are expected to support palatalization (*vutʃ* → *vutʃi* supports *vuk* → *vutʃi*, *vut* → *vutʃi*, and *vup* → *vutʃi*).

Support for product-oriented generalization in phonology comes from several sources. First, a completely rule-based grammar runs into difficulty in accounting for static patterns in the lexicon. For instance, English speakers seem to know that words ending in a stop followed by a nonalveolar stop (e.g. **concetp*) are not legal in the language despite there being no alternations that repair such clusters (Hayes 2009:65). If we want to capture such knowledge, we must allow speakers to make generalizations about characteristics of individual word forms rather than relations between pairs of word forms. Second, product-oriented generalizations are supported by examples of rule conspiracies (Kisseberth 1970), in which a diverse collection of changes results in avoiding or producing the same sound sequence. The importance of capturing such patterns has been recognized in phonology and triggered a paradigm shift from rule-based phonology to OPTIMALITY THEORY (OT; Kager 1999, Prince & Smolensky 2004 [1993]), which can capture some product-oriented generalizations using markedness constraints.

The product-oriented generalizations captured by classical optimality theory are restricted in being phonetically motivated (and therefore arguably universal) and largely negative, for example, *NC (Kager 1999). Product-oriented schemas or constructions may be distinguished from simple phonotactic constraints in being associated with meanings and being therefore clearly nonuniversal and learned by generalizing over the lexicon. Inducing constraints from the lexicon has also recently been proposed within HARMONIC GRAMMAR (HG) by Hayes and Wilson (2008).

The set of product-oriented schemas associated with a particular meaning can then be thought of as a description of the set of word forms associated with that meaning (Bybee 1985). In other words, a good schema is a structure often found in forms with a particular meaning. Construction grammar (CG) schemas or constructions are thus very related to the learned phonotactic constraints of MAXIMUM ENTROPY HARMONIC GRAMMAR (Hayes & Wilson 2008), where the phonotactic grammar is also thought of as a description of the lexicon. The only differences are that (i) the learned phonotactic constraints of Hayes and Wilson (2008) describe what kinds of word forms occur in the lexicon, whereas schemas/constructions describe what kinds of word forms occur in a semantically coherent subpart of the lexicon, say, plural nouns, and (ii) schemas are positive, zeroing in on common properties of word forms, whereas constraints are negative, zeroing in on underattested properties. While some phonotactic constraints may be innate or acquired prior to the acquisition of the lexicon from early production experience, product-oriented schemas are uncontroversially abstractions over words. If there is anything hard about producing a /t/ at the end of a plural noun, this difficulty is due to final /z/ being strongly associated with plural meaning, not to something about the articulation of a final /t/.

Support for product-oriented schemas comes largely from wug tests (Berko 1958), in which subjects are often observed to overuse common output patterns, deriving them in ways unattested in the lexicon (Albright & Hayes 2003, Bybee & Moder 1983, Bybee & Slobin 1982, Köpcke 1988, Lobben 1991, Wang & Derwing 1994). In addition, a morpheme is especially likely to be omitted in forms that sound like they already have it (Bybee 2001:128, Bybee & Slobin 1982, Menn & MacWhinney 1984, Nesset 2010, Stemberger 1981). For instance, Bybee and Slobin (1982) document that children learning the English past tense are more likely to make no-change errors, in which the past-tense form is erroneously identical to the present-tense form, on verbs that happen to already end in [t] or [s] and thus sound like past-tense forms. Stemberger (1981) notes that the progressive form of *lightning* as in *It is thundering and lightning* is at least as likely to be *lightning* as *lightninging*. Another piece of evidence for product-oriented generalizations is affix fusion (Booij 2008, 2010, Corbin 1989, Kapatsinski 2005). For instance, a *de*-N-*ize* verb in English can be formed directly from the noun, skipping the intermediate step of an N-*ize* verb (one can coin *destalinize* in the absence of *Stalinize*; Booij 2008). In Russian, one can form verbs meaning 'act as an X-er' by adding *-nitʃatʲ* (*-nik+ja+tʲ*) to an X that cannot combine with *-nik* '-er'; for example, 'act as an owner' is *xozjajnitʃatʲ*, but 'owner' is *xozjain*, not **xozjajnik* (Kapatsinski 2005). A related phenomenon is hypercharacterization, where a redundant marker is added to make the shape of a word typical for words with that meaning. Booij (2008) mentions the case of *UHD*, the abbreviation for *universitair hoofddocent* 'assistant professor' in Dutch, being often redundantly marked by the agentive *-er* to become *UHD-er*, thus coming to fit the agentive noun schema '…*er*' characterizing most agentive nouns.

Since constructions are generalizations over individual utterances, their boundaries do not have to coincide with morpheme boundaries established by comparing morphologically or syntactically related utterances. Thus if English nouns often end in [ɪsɪti], the whole sequence might become associated with nouniness, whereas generalization over adjective-noun pairs would only associate nouniness with [ɪti]. The constructionist proposal predicts the possibility of attaching common affix sequences to stems and hypercharacterization, for example, the joking English *beauticity* (attested on Google) or the nonjoking Russian *xozjajnitʃatʲ* ('to behave as if you are the *xozjain*', **xozjajnik*).

SECOND-ORDER SCHEMAS. Bybee (2001) considered the possibility of a morphophonology that has only product-oriented schemas and no source-oriented ones: '[R]ules 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' (Bybee 2001:128, emphasis mine; see also Goldberg 2002 vs. Cappelle 2006 in syntax).[4] Booij (2010), Nesset (2008), and Pierrehumbert (2006) point out that this hypothesis is too restrictive, however, reintroducing rule-like source-oriented generalizations into the cognitive/constructionist tradition.

The existence of source-oriented generalizations in morphophonology is suggested by the existence of restrictions on the class of inputs that are productively mapped onto a certain class of outputs. Pierrehumbert (2006) presents a particularly convincing case of a productive restriction of this kind. 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 [k] is changed into an [s] when followed by *-ity*. Pierrehumbert argues that English speakers must be using a

[4] Hayes and Wilson (2008) implemented a purely product-oriented model of grammar but argue that such a grammar is used only prior to acquisition of alternations.

source-oriented generalization like *k* → *s*/ __ *ity* and not a product-oriented one like 'Latinate nouns should end in [sɪti]' or 'Latinate nouns should not end in [kɪti]' because [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, by contrast, 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 [lɪti] but often correspond to nouns ending in [sɪti].

Source-oriented generalizations are also essential for paradigmatic morphology (Booij 2010, Nesset 2008). For instance, a Russian speaker hearing a novel nominative noun ending in *-a* knows that the *-a* will be dropped in the genitive plural, as in 5a, whereas a novel noun ending in a nonpalatalized consonant will gain *-of* in the genitive plural, as in 5b. Since the *-a* is dropped in the genitive plural, the generalizations responsible for the mappings must be source-oriented.[5] This is by no means an isolated example (see Booij 2010 and Nesset 2008).

(5) a. odna flarnikrap-a → neskol' ko flarnikrap-ø
one.F.SG.NOM flarnikrap-F.SG.NOM some flarnikrap-F.PL.GEN
b. odin flarnikrap-ø → neskol' ko flarnikrap-of
one.M.SG.NOM flarnikrap-M.SG.NOM some flarnikrap-M.PL.GEN

In order to capture this kind of knowledge of alternations, source-oriented generalizations have recently been reintroduced into CG (Booij 2008, 2010, Cappelle 2006, Nesset 2005, 2008). Thus no current theory of phonology (except Reiss 2004) denies the existence of either source-oriented or product-oriented generalizations. The differences are rather in the KINDS of source-oriented and product-oriented generalizations various theories propose, the degree to which speakers are proposed to rely on source-oriented vs. product-oriented generalizations in accounting for alternations, and the ways source-oriented and product-oriented generalizations are proposed to interact.

The source-oriented generalizations of CG are not rules in that they do not feature a change-context split: they are not generalizations about possible changes, or transformations. Rather, they are paradigmatic mappings between what Cappelle (2006) calls ALLOstructions. In Nesset's (2008) terms, they are SECOND-ORDER SCHEMAS: generalizations over product-oriented, or first-order, schemas (cf. also the proposal that phonotactic constraints are acquired before, and bootstrap acquisition of alternation patterns in Hayes & Wilson 2008:424). Since second-order schemas are generalizations over product-oriented schemas, they cannot refer to null elements, making ø → *X* an impossible generalization: each of the associated form elements in a schema must contain an element that is a good marker of the associated meaning. Thus, [butʃ]$_{SG}$/[butʃi]$_{PL}$ cannot be an instance of the rule ø → *i*. Furthermore, they do not involve a split into change and context; thus the 'context' [tʃ] is retained in each side of the schema (…*tʃ*$_{SG}$/…*tʃi*$_{PL}$).

Thus, under the CG approach examples of …*tʃ*$_{SG}$/…*tʃi*$_{PL}$ may help palatalization, providing support for the first-order/product-oriented schema 'plurals should end in [tʃi]', which can then serve as input to the process of second-order schema formation, resulting in the source-oriented schema in 7, which can be formed by generalizing over

[5] Note that the mappings (*a*/ø and ø/*of*) do not pair up affixes based on similarity; the mappings cannot therefore be accounted for by minimization of output-output faithfulness violations (as in Kenstowicz 1996).

the schemas in 6. Whether or not second-order schemas are allowed, examples of …*tʃ*$_{SG}$/…*tʃi*$_{PL}$ are expected to lead learners to palatalize stops (…{*k*;*t*;*p*}$_{SG}$/…*tʃi*$_{PL}$) by virtue of making [tʃi] the most common final sound sequence in plurals.

(6) a. SG-V$_i$tʃ# / PL-V$_i$tʃi#
b. SG-V$_i$k# / PL-V$_i$tʃi#

(7) SG-V$_i$[–cont;–voice;Lingual]# / PL-V$_i$tʃi

Examples of …{*t*;*p*}$_{SG}$/…{*t*;*p*}*i*$_{PL}$ might help …*k*$_{SG}$/…*ki*$_{PL}$ by virtue of supporting a schema like …[–Del.Rel]*i*-PL or …[–Del.Rel]-SG/…[–Del.Rel]*i*-PL. Given that CG has no worked-out algorithm for schema extraction, however, it is not clear why the learner would extract schemas of this particular level of generality, rather than more specific or more general ones; that is, why extract [–Del.Rel]*i*-PL and not instead the more specific [Vpi]-PL and [Vti]-PL that provide no support for [ki]-final plurals,[6] or the more general [–cont]*i*-PL, which provides equal support for [ki]-final and [tʃi]-final plurals? (See also Pierrehumbert 2006 for a similar observation.) One contribution of the present article is to work out a schema-extraction algorithm.

The name 'second-order schemas' is designed to imply that they are relatively difficult to acquire (Nesset 2008). The difficulty of acquiring second-order schemas is well documented in 'gender learning' experiments (Braine 1987, Braine et al. 1990, Brooks et al. 1993, Frigo & McDonald 1998, Gerken et al. 2005, Weinert 2009, Williams 2003). In particular, Frigo and MacDonald (1998) find that their subjects do not learn the source-oriented generalizations like *a* → *uk* unless they can extract reliable product-oriented generalizations like 'back vowels are followed by *-uk*' during training. If the product-oriented schemas like PL-*uk* and PL-*im* are in free variation, independent of the stem phonology, and are thus unreliable, generalizations over these schemas are not learned. Thus, unlike rule-based approaches, constructionist approaches propose that source-oriented generalizations over alternations are relatively difficult to form and are formed relatively late, after the learner has already extracted reliable product-oriented generalizations. Further, these source-oriented generalizations are paradigmatic mappings between product-oriented constructions. Given that under these proposals acquisition of first-order schemas is necessary and largely prior to acquisition of second-order schemas, we restrict ourselves to modeling the first half of the process here.

GIVING THE RULES A LEG UP. Whereas rule-based approaches suggest that human language learners automatically compare morphologically related word forms, identifying corresponding phonological and semantic differences, constructionist approaches suggest that this comparison, if it happens, is difficult and error-prone (see also Pierrehumbert 1993). As Valian and Coulson point out,

> 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. (Valian & Coulson 1988:78, emphasis mine)

If learners are predisposed to pick up on changes in context, they should do so when the to-be-compared forms are next to each other and therefore easy to compare. Singular forms are always presented next to the corresponding plural forms, while plural forms that share features are never presented next to each other and occur separated by one word no more often than would be predicted by chance. I argue that even under

[6] Langacker (1987) and Nesset (2008) explicitly assign priority to the most specific constructions that fit the data.

these presentation conditions, learners still pay more attention to typical characteristics of word forms belonging to a particular cell in the paradigm (i.e. constructions) than to how singulars and plurals differ from each other (i.e. the changes that are to be made to a singular to form a plural or vice versa) or the fact that certain sound sequences are underattested. They are also affected by a tendency to perseverate on the stem that is well captured by a specific, novel kind of faithfulness constraint. Drawing on ideas from machine learning (Daelemans & van den Bosch 2005) and harmonic grammar (Legendre et al. 1990, Smolensky & Legendre 2006), I develop a formally explicit model showing how morphophonological constructions can be extracted from the lexicon and how the competition between such constructions and faithfulness constraints is resolved.

2. THE EXPERIMENT.

2.1. THE SIX LANGUAGES. Each participant in the present experiment was presented with one of the six miniature artificial languages shown in Table 1. I call these six languages Tapa, Tipi, Tapatʃi, Tipitʃi, Tapatʃu, and Tipitʃu based on their characteristic plural-final bigrams and use standard phonological notation to refer to sets of these languages; thus Tapa(tʃi) refers to Tapa and Tapatʃi and T{a;i}p{a;i} refers to Tapa and Tipi.[7]

All languages had the two plural suffixes *-i* and *-a*. In all languages [k] obligatorily changed into [tʃ] before *-i*; that is, all languages featured a process of velar palatalization. In all languages, *-i* was the only plural suffix that could attach to [k]. In all languages, both *-i* and *-a* could attach to [t] and [p], which never changed. In Tapa, Tapatʃi, and Tapatʃu, *-i* was unlikely to attach to [t] and [p]: stems ending in [t] or [p] usually took *-a* as the plural suffix. In Tipi, Tipitʃi, and Tipitʃu, *-i* was more likely to attach to [t] and [p] than *-a* was. Participants who were exposed to Tapatʃi or Tipitʃi experienced additional examples in which *-i* simply attached to [tʃ]. These examples were absent from Tapa and Tipi, and were replaced by examples in which *-u* simply attached to [tʃ] in Tapatʃu and Tipitʃu.

	TAPA	TIPI	TAPATʃI	TIPITʃI	TAPATʃU	TIPITʃU
k → tʃi	4					
{t;p} → {t;p}i	2	6	2	6	2	6
{t;p} → {t;p}a	6	2	6	2	6	2
tʃ → tʃi	0		4		0	
tʃ → tʃu	0				4	

TABLE 1. The six languages presented to participants. Numbers show type frequencies of the various mappings. Actual training stimuli are shown in the appendix.

The effect of presenting examples of *tʃ → tʃi* allows us to determine the extent to which learners pay attention to the shape of the product vs. the source-product mapping. Product-oriented schemas and constraints (to be more fully developed in §3) suggest that these mappings should support palatalization by exemplifying the palatalizing product-oriented generalizations 'plurals must end in *-tʃi*' or 'final *-i* must be preceded by [tʃ] in a plural'. Thus, their addition to the training set should increase the productivity of palatalization. On Albright and Hayes's (2003) rule-based account, these singular-

[7] I am indebted to Adele Goldberg for these names.

plural pairings exemplify *C* → *Ci*, which militates against velar palatalization. Thus, their addition should reduce the productivity of velar palatalization. In a more standard rule-based theory, where rule orderings are allowed, *tʃ* → *tʃi* supports ∅ → *i*/ __ # but says nothing about *k* → *tʃ*/ __ *i*. These predictions of rule-based phonology follow directly from the fundamental property of rules as generalizations about changes in context: the change exemplified by *tʃ* → *tʃi* is not the change exemplified by *k* → *tʃi* and *t* → *tʃi*.

2.2. METHODS.

TASKS. The experiment consisted of a training stage followed by a generalization test. Before training, participants were instructed that they are to learn the names of the objects in a made-up language. Training proceeded as follows. Each trial 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 was presented auditorily over headphones. Once the sound finished playing, the picture was removed and replaced with a picture of multiple (five to eight) 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 was expected to facilitate comparisons between corresponding singulars and plurals, which are necessary for rule extraction.[8]

The training was followed by the elicited production test. In this test, the participants were presented with novel singulars (paired with novel objects) that they had not experienced during training. The procedure was exactly the same as in training except no plural form was provided. Instead, the participants were asked to generate the plural form and say it aloud. The participants were not required to repeat the singulars during the test.

STIMULI. Each word pair was presented to learners multiple times during training. One word exemplifying *k* → *tʃi*, one word exemplifying the most frequent *p* → *pV* pattern in each language, one word exemplifying the most frequent *t* → *tV* pattern in each language, and one word exemplifying *tʃ* → *tʃi* or *tʃ* → *tʃu* were presented forty-two times each, while the other words were presented fourteen times each. Several distinct tokens of each word pair were devised using the 'change gender' function in Praat (Boersma & Weenink 2009) as well as by pronouncing the word in four different tones of voice. The same elicited production test was used for all subjects. The stimuli used for elicited production were not presented during training. The trained CV and CCV sequences were reused, however, as shown in the appendix, in order to zero in on the effect of the crucial stem-final consonant.

The auditory stimuli were recorded by me in a soundproof booth onto a computer. 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 learners were asked to repeat the words they heard during training immediately after hearing them. Repetition accuracy, with regard to the crucial stem-final consonants and the preceding and following vowels, was very high (96% of words repeated correctly).

[8] This assumption is justified in Kapatsinski 2012a, which compares this kind of training with a more natural presentation in which singulars and plurals are intermixed. Supporting the idea that this kind of training facilitates source-oriented generalization, we find that the learned restrictions on what kinds of consonants can become [tʃ] are stricter in the present paradigm (in other words, there is less overgeneralization of palatalization to [t] and [p]).

The visual stimuli were a set of made-up creature pictures retrieved from the website sporepedia.com. The number of creatures paired with a plural word form varied between five and eight. All pictures were presented on a black background.

PROCEDURES. Learners were tested one at a time. 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 subject's 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 (a native English speaker) online, as the learner was producing them.[9] The stimuli were presented using E-prime 2.0 Professional. The order of presentation of the stimuli was randomized separately for each learner.

DEPENDENT MEASURES AND ANALYSES. Analyses of the elicited production test presented below are done on production probabilities of singular-plural mappings: for every final consonant of the singular for every participant, we ask: how likely is this participant to produce a certain plural-final two-phoneme sequence given a certain singular-final consonant? For instance, what is the probability of participant 7 producing a plural ending in V_i*tʃi*# given that s/he is presented with a singular ending in V_i*k*#? This is the production probability of *k* → *tʃi* for participant 7, which can also be written as p(PL=…V_itʃi# | SG=…V_ik#). Production probabilities for alternative plural-final sequences that could be produced from a given singular-final consonant sum to 1.

Production probabilities of the various mappings like *k* → *ki*, *k* → *tʃi*, *k* → *ktʃi*, *k* → *ka*, *t* → *ti*, *t* → *tzkmnpta* can then be compared within participants. For instance, one may ask whether, after the training, the participants palatalize velars more than alveolars. This can be determined by comparing the difference between production probabilities of *k* → *tʃi* and *k* → *ki* to the difference between production probabilities of *t* → *tʃi* and *t* → *ti*. Thus, for each participant, the ratio of the number of times that participant produced *k* → *tʃi* to the number of times s/he produced *k* → {*k*;*tʃ*}*i* can be obtained and subtracted from the ratio of the number of times that participant produced *t* → *tʃi* to the number of times s/he produced *t* → {*t*;*tʃ*}*i*. The resulting sample difference scores can then be compared to zero. If the difference scores are significantly above zero, then velar palatalization is significantly more productive than alveolar palatalization for our participants following training. If they are significantly below zero, then velar palatalization is significantly less productive than alveolar palatalization for our participants following training.

The production probability of the same mapping can be compared across languages (and participants). Thus, if we are interested in whether *t* → *tʃi* is more productive in Tipitʃi and Tapatʃi than in Tipi or Tapa, we can determine, for each subject, how often s/he produced a plural ending in [tʃi] when s/he was presented with a singular ending in [t]. We can then compare the scores of subjects exposed to Tipi or Tapa to the scores of subjects exposed to Tipitʃi or Tapatʃi. If we are interested in whether *t* → *tʃi* is supported against *t* → *ti* by examples of *tʃ* → *tʃi*, we can compare the sample of differences in production probabilities between *t* → *tʃi* and *t* → *ti* (p(t → tʃi) – p(t → ti)) obtained from subjects exposed to Tapa or Tipi to the sample of differences in production probabilities obtained from subjects exposed to Tapatʃi and Tipitʃi.

Due to severe nonnormality of production probability distributions, all comparisons are based on nonparametric Wilcoxon tests. All analyses were performed in R (R Development Core Team 2009).

[9] Many thanks to Lucy Gubbins, Nina Rinaldi, and Matthew Stave for data collection and scoring.

PARTICIPANTS. One hundred and twenty participants were recruited from the Linguistics/Psychology undergraduate subject pool at the University of Oregon and participated for course credit. Participants signed up for the experiment blindly to prevent selection bias. Each participant was exposed to only one language. 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. Participants were assigned to languages in the order they came in (subject 3—language 3 (Tapatʃi), subject 6—language 2 (Tipi), etc.). Two subjects were excluded due to using English plurals after the *-i* or *-a* suffixes of the artificial language. Tapa and Tapatʃi were presented to eighteen participants each, while Tipi and Tipitʃi were presented to nineteen participants each. Tapatʃu and Tipitʃu were presented to sixteen and seventeen participants respectively: two participants exposed to Tapatʃu and one exposed to Tipitʃu were excluded for using [eɪ] instead of [i] as a plural suffix.

2.3. RESULTS.

THE EFFECT OF ADDING EXAMPLES OF *tʃ → tʃi*. There is no significant effect of *tʃ → tʃi* examples on velar or labial palatalization production probabilities (*k → tʃi* and *p → tʃi*), although numerically both are favored by examples of *tʃ → tʃi*. As shown in Figure 1, however, the examples of *tʃ → tʃi* given in Tapatʃi and Tipitʃi significantly favor alveolar palatalization (not present in the input); that is, the addition of [tʃ] → [tʃi] examples favors the mapping of [t] onto [tʃi] over mapping it onto [ti] ($p = 0.009$ according to the Wilcoxon test).

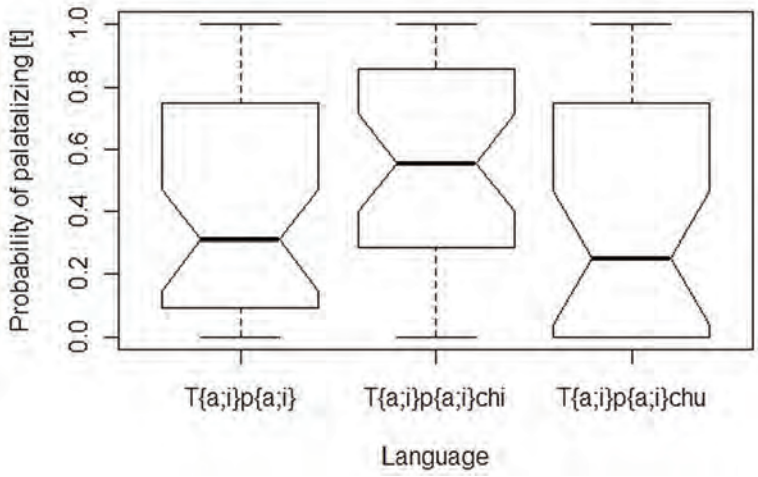


FIGURE 1. The effect of examples of [tʃ] → [tʃi] on the probability of palatalizing [t] before *-i* in elicited production.

THE EFFECT OF ADDING EXAMPLES OF *tʃ → tʃu*. Unsurprisingly, presenting participants with *tʃ → tʃu* examples increased the probability of attaching *-u* to a singular ending in [tʃ] (from 0% to 37%, $p < 0.00001$). Participants presented with *tʃ → tʃu* also occasionally erroneously attached *-u* to [k], [t], and [p] (for learners exposed to *tʃ → tʃu*, p(PL=tʃu# | SG=k#) = 0.14, p(PL=tʃu# | SG=p#) = 0.09, p(PL=tʃu# | SG=t#) = 0.11; these rates are not significantly different from each other in Wilcoxon tests). Most importantly for the present article, the addition of examples of *tʃ → tʃu* did not sig-

nificantly increase rates of alveolar, velar, or labial palatalization before *-i* (comparing T{i;a}p{i;a} vs. T{i;a}p{i;a}tʃu: p(t → tʃ | PL=…i#) = 38% when examples of *tʃ → tʃu* are present and 39% when they are not, p = 0.79 according to the Wilcoxon test; p(k → tʃ | PL=…i#) = 57% when examples of *tʃ → tʃu* are present and 56% when they are not, p = 0.87 according to the Wilcoxon test; rates of labial palatalization actually drop: p(p → tʃ | PL=…i#) = 5% when examples of *tʃ → tʃu* are present and 19% when they are not, p = 0.04). Subjects trained on *tʃ → tʃi* palatalize [t] and [p] more than subjects trained on *tʃ → tʃu* do (58% vs. 38% for [t], p = 0.02; 18% vs. 5% for [p], p = 0.04; there is no significant difference for velar palatalization: 70% vs. 57%, p = 0.22). Thus, as shown in Fig. 1, examples of *tʃ → tʃu* do not help palatalization before *-i*, whereas examples of *tʃ → tʃi* do.

THE EFFECT OF *-i* OFTEN ATTACHING TO {*t*;*p*}. Unsurprisingly, the addition of examples of {*t*;*p*} → {*t*;*p*}*i* increases the probability of {*t*;*p*} → {*t*;*p*}*i* ($p < 0.00001$ according to the Wilcoxon test). More importantly, high probability of attaching *-i* to nonvelars correlates with low productivity of velar palatalization. Comparing Tapa(tʃi)(tʃu) to Tipi(tʃi)(tʃu) as in Figure 2, we observe that participants exposed to the latter languages produce significantly more [ki]-final plurals ($p < 0.00001$) and marginally fewer [tʃi]-final plurals (p = 0.09) from [k]-final singulars.[10]

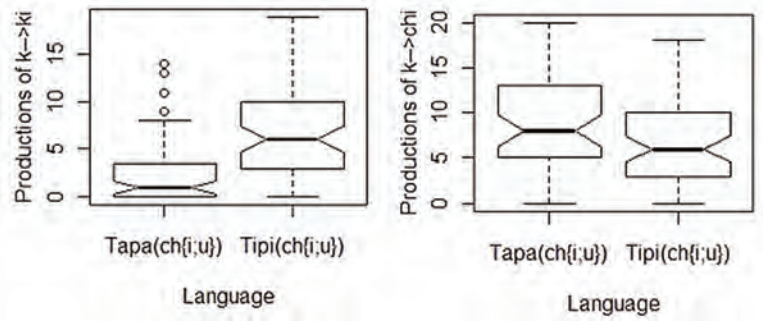


FIGURE 2. Adding *-i* to [k] without changing the stem-final consonant is more productive in Tipi(tʃi)(tʃu) than in Tapa(tʃi)(tʃu).

SOURCE-ORIENTED KNOWLEDGE. Palatalization of alveolars is less likely than palatalization of velars ($p < 0.00001$), as seen in Figure 3. Overgeneralization of velar palatalization to labial sources is much less likely than overgeneralization to alveolar sources (p = 0.0002). Thus the grammars learned at the end of training contain source-oriented generalizations that allow the learners to restrict the types of sources that can give rise to a good product. I do not believe this is due solely to the fact that the training paradigm facilitates formation of source-oriented generalizations, as the same result has been observed when using a training paradigm in which the order of all word forms was randomized (Kapatsinski 2012a).

STEM-FINAL CONSONANT RETENTION. A common, albeit originally unexpected, response type (exhibited by 27/74 participants) during the generalization test was to perseverate on the stem too much, yielding, for example, [buk buktʃi] rather than the

[10] We also exposed forty additional native English speakers to languages that were exactly like T{i;a}p{i;a} but with the addition of two *k → ka* singular-plural pairings. The same difference between the Tipi and Tapa languages was observed ($p < 0.0001$).

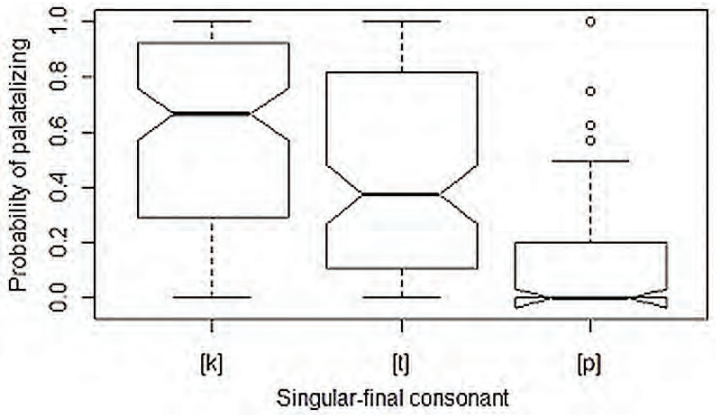


FIGURE 3. Probability of palatalizing [k], [t], and [p].

correct [buk butʃi], or [bup buptʃi] as opposed to the (also incorrect) [bup butʃi]. There were sixty-four *k* → *ktʃi* responses and twenty *p* → *ptʃi* responses, compared to 644 *k* → *tʃi* responses and fifty-one *p* → *tʃi* responses. Thus [k] was erroneously retained in 9% of [tʃi]-final plurals where it could have been retained, whereas [p] was retained in 28% of [tʃi]-final plurals where it could have been retained, suggesting that [p] is more likely to be retained than [k] ($\chi^2(1) = 25.2$, $p < 0.0001$). These stem-final consonant retentions occurred despite the fact that the participants were not presented with stem-final consonant clusters during training. This kind of retention error was relatively uncommon during training. There were fifty-two cases in which a stem-final alternation was neutralized in repetition during training. In forty-seven of these cases, however, the neutralization was in favor of the consonant found in the plural form, not the singular one (e.g. [buk butʃi] would be likely to be erroneously repeated as [butʃ butʃi], not [buk buki] or [buk buktʃi]). Of the remaining cases, one featured retention of both singular and plural stem-final consonants.

2.4. SUMMARY AND DISCUSSION. There are several findings to explain in the present data. First, examples of *tʃ* → *tʃi* support *t* → *tʃi* over *t* → *ti* and *p* → *tʃi* over *p* → *pi*. This effect is mostly driven by *tʃ* → *tʃi* providing support for *C* → *tʃi* rather than reducing the goodness of {*k*;*t*;*p*} → {*k*;*t*;*p*}*i*. Replacing examples of {*t*;*p*} → {*t*;*p*}*a* with examples of {*t*;*p*} → {*t*;*p*}*i* helps *k* → *ki* over *k* → *tʃi* and *k* → *ka*. As discussed below, I attribute these effects to product-oriented generalizations.

Velar palatalization (*k* → *tʃi*) is overgeneralized to alveolars more than to labials. In addition, stem-final consonants of the singular are likely to be erroneously retained when productively making a plural from a known singular. This erroneous retention is most likely for labials. I argue that the erroneous retention of labials suggests perseveration on the labial gesture, which can be captured by means of a constraint militating against nonretention of input chunks.

Examples of *tʃ* → *tʃi* help *k* → *tʃi* (the experienced singular-plural mapping involving a stem change) over *k* → *ki* less than they help the nonexperienced mappings. I attribute this finding to the way competition between chunks and schemas is resolved: the same amount of support from a schema helps candidates that are violating strong constraints clamoring for chunk retention and are thus in danger of losing the competition to other candidate outputs that obey the constraints more than they help candidates that are already well ahead of the competition.

3. THEORETICAL INTERPRETATION.

3.1. RULE-BASED PHONOLOGY: *Ø → ?. As discussed in the introduction, rule-based phonology proposes that generalizations are acquired on the basis of pairs of morphologically related word forms split into change and context (Albright & Hayes 2003). This proposal is inconsistent with the finding that examples like *ʧ → ʧi* help palatalization (*C → ʧi*). Consider again the rules in 8, representing a standard rule-based analysis, and the rules in 9, representing the rules induced by the minimal generalization learner of Albright & Hayes 2003.

(8) a. Ø → i/[–cont] __ # when plural b. k → tʃ/ __ i#

(9) a. k → tʃi/V __ # b. Ø → i/[–cont] __ #

The rules in 9 are different from those in 8 in featuring only a single processing stage. Further, 9a and 9b are in competition for stems that end in /k/: 9a argues that the plural forms of such stems should end in [tʃi], while 9b argues that they should end in [ki]. Competition can be minimized by noticing that both rules support the change Ø → *i*, but 9a contains a further change (*k → ʧ*), which is fed by Ø → *i* (i.e. Ø → *i* produces the context for *k → ʧ*), resulting in the rules in 8 (modeled computationally in Johnson 1984).

The rules in 9 have the advantage over the rules in 8 in correctly predicting that *t → ti* and *p → pi* support *k → ki* over *k → ʧi*. They incorrectly predict, however, that examples of *ʧ → ʧi* should reduce the productivity of velar palatalization by providing support for the competing rule in 9b.

A possible way to maintain the fundamental premise of rule-based phonology that phonological generalizations are made over PAIRS of morphologically related word forms would be to also posit that a rule (extracted by the standard mechanisms above) gains additional strength whenever the structure it produces is encountered. In other words, a rule like *k → ʧi*-PL is supported by [tʃi]-final plurals. Note, however, that a [t]-palatalizing rule (*t → ʧi* or *t → ʧ/ __ i*) should not be extracted from the training data: this mapping is unattested during training. If product-oriented support can only help rules, which are generalizations over pairs of forms split into change and context, it is then puzzling that [tʃi]-final plurals support palatalizing [t] (*t → ʧi*) and, further, that they support palatalizing [t] (an unfamiliar rule) more than they support palatalizing [k] (a familiar rule).

The root of the problem that rule-based phonology has with these data is that rules are formed over pairs of forms SPLIT INTO A CHANGE AND A CONTEXT. Therefore forms involving simple concatenation are analyzed as involving a Ø → *X* change. This assumption of change-context decomposition is what makes it impossible to consider examples of *ʧ → ʧi* and of *k → ʧi* to be examples of the same rule. A simple solution to this issue is to prohibit rules from referring to zeroes, at least in the input, or to give up on the change/context split altogether.

If the split is not made, as in second-order schemas of CG (Booij 2008, 2010, Nesset 2005, 2008), source-oriented generalizations CAN describe the data: surface-to-surface mappings like *k → ʧi* and *ʧ → ʧi* can be generalized over to yield [–cont;–voiced] *→ ʧi*, which also subsumes *t → ʧi* and *p → ʧi*. Note, however, that this generalization is not a simple matter: humans are easily capable of learning that units in specific positions within a symbol string must be identical, as demonstrated by the existence of reduplication and artificial-grammar learning data obtained by Altmann and colleagues (1995) and Marcus and colleagues (1999). Thus, there is nothing about the formalism itself that would prevent considering *ʧ → ʧi* as being more like *t → ti*, *p → pi*, and *k → ki* than like {*k*;*t*;*p*} *→ ʧi*: *ʧ → ʧi*, *t → ti*, and *p → pi* all feature an *X → Xi* map-

ping while {*k*;*t*;*p*} → *ʧi* do not. Therefore it is still necessary to show why *ʧ* → *ʧi* is more like {*k*;*t*;*p*} → *ʧi* than like {*k*;*t*;*p*} → {*k*;*t*;*p*}*i*.

According to CG, second-order schemas are generalizations over pairs of first-order schemas (Booij 2008, 2010, Nesset 2005, 2008), which is supported by the finding that second-order schemas are very difficult to learn, at least in the laboratory (e.g. Braine 1987, Brooks et al. 1993, Frigo & MacDonald 1998, Weinert 2009). In fact, based on this literature, they are unlikely to be discovered with the short training provided to the learners in this experiment. The difficulty learners exhibit with forming second-order schemas is not surprising, given that their formation requires an error-prone and memory-intensive comparison process. A promising hypothesis then is that what makes *ʧ* → *ʧi* like {*k*;*t*;*p*} → *ʧi* is that all of these mappings are supported by a simple first-order schema like 'plurals should end in [tʃi]'. Since second-order schemas are formed by generalizing over first-order schemas, productivities of second-order schemas involving the same first-order schema should correlate positively. Given that examples of *ʧ* → *ʧi* support {*k*;*t*;*p*} → *ʧi*, we want to induce a first-order schema that roots for [tʃi]-final plurals. Given that examples of {*t*;*p*} → {*t*;*p*}*i* support *k* → *ki*, we want to induce a first-order schema that roots for [–Del.Rel]*i* plurals. Thus, the right schemas must be specific enough so that {*k*;*t*;*p*} → {*k*;*t*;*p*;*ʧ*}*i* is not a single unitary schema ([–cont]*i*) but general enough so that [ki] is grouped together with [pi] and [ti] (see Hayes & Wilson 2008, Pierrehumbert 2006 for this issue beyond palatalization). In the rest of this section a computationally explicit model of first-order schema/construction induction that attempts to achieve the right level of generality is developed. The induced schemas are proposed to compete against a specific version of paradigm-uniformity constraints, which we think of as perseveratory tendencies associated with various gestures. The competition is resolved using the candidate generation and evaluation process proposed in harmonic grammar, which is shown to be helpful for explaining why trained unfaithful mappings are helped by experiencing the resulting products less than untrained unfaithful mappings are.

3.2. FIRST-ORDER SCHEMA EXTRACTION.

FOUNDATIONAL CLAIMS. Given that there is currently no explicit theory of schema or construction induction, the first question to address is how constructions/first-order schemas could be learned from the lexicon. Two foundational assumptions are made.

First, over the course of learning, schemas, at least to a large extent, become more specific, rather than becoming more general. Here I depart from the standard approach in CG, where one is supposed to gradually generalize over memorized representations of specific utterances (or words), with schemas growing gradually more general (Bybee 1985, 2001, Goldberg 1995, Nesset 2008).

The crucial prediction of schema specification is PROGRESSIVE DIFFERENTIATION (as claimed in McClelland et al. 1995 and Rogers & McClelland 2004 for semantic development): upon exposure to a few plural word forms, one thinks that plurals in the language can be pretty much anything, rather than thinking they can only be the words one has just experienced. Gradually, plural forms are differentiated from nonplurals through the development of more and more specific knowledge of what plural forms are like. Upon hearing [bupi] paired with multiple novel creatures, one does not think that the plural form of any word is [bupi] and does not overgeneralize [bupi] to suppletively replace plural forms of other words, no matter how often [bupi] is heard. Learners start out thinking that *-i* and *-a* can attach to stems to make plurals, but even at the end of training may not grasp that [ka] and [ki] are illegal. When schemas have not yet calci-

fied into strong preferences for the observed sound sequences, the learner is open to experience and is ready to learn. Once the schemas have calcified, the learner has an idea of what does and does not occur in the language and is not as ready to accept input violating these well-entrenched patterns. Since the to-be-produced form matches the input unless this contradicts a well-entrenched schema, the learner enters the experiment with a bias against stem changes. Thus, the general-to-specific order of schema acquisition captures the developmental decrease in the tendency to perseverate that is captured by a high initial ranking of output-output faithfulness constraints in OT (Hayes 2004).

In the proposed theory, schema specification proceeds by seeking out unexpected bumps in the joint probability space defined by meanings and sounds; that is, which kinds of sequences are unexpectedly frequent in plural forms? Xu and Tenenbaum (2007) document this kind of inference for semantic categories: suppose you are presented with a picture of a Dalmatian paired with the word *fep*. At first you are likely to think that *fep* means 'dog'. If *fep* is presented to you three times, however, each time paired with a picture of a different Dalmatian, you are likely to discard the hypothesis that *fep* means 'dog' as it would be a very suspicious coincidence that a process of randomly sampling dogs would produce three Dalmatians in a row.

Progressive differentiation implies automatic overgeneralization, whereby variation within natural classes is leveled and patterns are extended to segments that are similar to segments known to participate in the pattern. This allows us to account for a data pattern previously claimed to be due to a specific substantive bias against [ki]: Wilson (2006) reports that subjects trained on *k* → *tʃ*/ __ *e* but *k* → *k*/ __ *a* generalize that *k* → *tʃ*/ __ *i*, whereas subjects trained on *k* → *tʃ*/ __ *i* but *k* → *k*/ __ *a* generalize that *k* → {*k*;*tʃ*}/ __ *e* (see also Mitrovic 2012 for the same finding in a natural language). Wilson (2006) interprets this result as supporting an innate ranking of **ki* >> **ke* and an innate difference in susceptibility to reranking for **ki* and **ke* such that **ke* is more easily reranked on the basis of experience. I argue that this is unnecessary: both acoustically and articulatorily, [e] is between [i] and [a]. Given some degree of automatic overgeneralization, [ka] provides some support to [ke] and [tʃi] provides some support for [tʃe], making learners undecided between [tʃe] and [ke] after being trained on [tʃi] and [ka]. By contrast, [tʃe] is much more similar to [tʃi] than [ka] is to [ki]; thus learners trained on [tʃe] and [ka] think that [tʃi] is more likely than [ki].[11] Given automatic overgeneralization, attested segment sequences pull unattested segment sequences similar to them up to (partial) acceptability. This appears to be the right prediction for the present case, where [ki] is pulled up to acceptability by [ti] and [pi].

Automatic overgeneralization appears to be necessary to account for the finding that unattested onset clusters like [bn] and [bd] are judged by English speakers to vary in acceptability and undergo perceptual repair based on their similarity to attested onsets (Berent et al. 2007, Hayes & Wilson 2008, Moreton 2002); for example, [bn], being similar to [bl] and [br], is judged as being better than [bd] and is less likely to be misperceived as containing a schwa. See Albright 2009 and Hayes 2011 for computational modeling showing that the data can be accounted for if one assumes that speakers generalize acceptability from known clusters to similar unattested ones. The fact that similarity to attested clusters affects not only explicit acceptability judgments but also

[11] Of course, the data do not require overgeneralization DURING TRAINING: generalization could also be made on an as-needed basis, with generalizations formed only during test. It seems likely, however, that such generalizations do form spontaneously and do not require an explicit grammaticality-judgment test (as demonstrated for onset clusters in Berent et al. 2007).

probability of misperception (Berent et al. 2007) suggests that generalization beyond experienced clusters is not task-specific.

General-to-specific learning is also supported by apparent underspecification of early lexical representations (Charles-Luce & Luce 1990, Shvachkin 1973 [1948], Stager & Werker 1997, Swingley 2007, Swingley & Aslin 2007, inter alia) and adult learners tolerate single-feature mismatches despite being able to hear the difference (Johnston & Kapatsinski 2011). The general pattern of results is that while correctly pronounced words are recognized more easily than slightly mispronounced ones, mispronunciations of low-frequency familiar words are preferred over unfamiliar words (Swingley 2007). This is expected if learners are gradually strengthening the more specific schema that does not allow for mispronunciations but still retain the more general schema that allows for some featural mismatch. As learners continue hearing a word, the more specific schema strengthens; thus, for highly familiar words featural mismatches are not tolerated even by young children. The phonological representation of a word is thus one instance of a schema where the specification process is relatively uncontroversial. I propose that this extends to all first-order schemas. When the schema is weak, one is willing to accept major deviations from the previously encountered examples, but the tolerance decreases as the distribution of experienced exemplars grows and its believable extent shrinks.

Second, meaningful schemas/constructions must contain some word boundary (in the tableaux below, the right edge of the word). There are three reasons for this restriction. First, grammars appear to count only from edges (e.g. Hayes 2009). Second, assuming extra word boundaries cannot be added, the requirement of including word boundaries prevents schemas that are maximally satisfied by an infinite number of additions of the structure they prefer. Third, a phonological structure does not retain its meaning independently of alignment with morphological boundaries. For instance, in Nesset 2008, nonpast tense is signaled by stem-final alveopalatals; alveopalatals elsewhere do not contribute to nonpast meaning. Similarly, psychologically real phonaesthemes in English (like [gl-] and [sn-] in Bergen 2004) are obligatorily stem-initial. Thus, while schemas can straddle morpheme boundaries, they must have SOME specified alignment with them.

Third, schemas are developed by looking for bumps in the distribution of conditional probabilities of forms given meanings, or joint form-meaning probabilities, not conditional probabilities of meanings given forms. In other words, schemas are not the most reliable cues to meaning; they are the most common realizations of meanings. This can be seen by considering the effect of adding examples like [SG=butʃ PL=butʃi] on the productivity of palatalization. These examples actually reduce the reliability of stem-final [tʃ] as a cue to plurality by introducing cases in which [tʃ] is found in the singular. Nonetheless, such examples help palatalization. This finding makes sense if schemas are fundamentally based on production, rather than perception, experience: they are what is most likely to be produced given that a certain meaning is intended.[12]

IMPLEMENTATION. General-to-specific schema extraction is implemented as decision-tree induction (Daelemans & van den Bosch 2005, Ernestus & Baayen 2011) using the

[12] The disjunction between best cues to meaning and most common realizations of that meaning is also documented at the lexical level (Ellis & Ferreira-Junior 2009, Kapatsinski 2009b). While we find parallel effects of *ɪʃ → ɪʃi* examples on production and acceptability rating of singular-plural pairs, acceptability rating is commonly thought to involve some simulation of production behavior (Kapatsinski 2012a). In purer perception tasks, such as speech in noise, we do not expect them to be helpful for plural meaning identification.

ctree() function in the party package (Hothorn et al. 2006, Strobl et al. 2009) in R (R Development Core Team 2009).[13] Each of the words presented to learners in training was coded in terms of the features of the stem vowel, the stem-final consonant, and the identity of the final vowel. The dependent variable to predict was the type frequency of the resulting word-final trigram. The ctree() recursively partitions the space defined by the predictors into rectangular areas such that at every split entropy reduction is maximized. The predictor producing the best binary split is at the top of the tree, with other predictors entering the tree if they improve predictiveness within the bins defined by the predictors already in the tree. At each step, the predictor that achieves the best split within a branch is entered into that branch.

Each of the words presented to learners in training was coded in terms of the place of articulation, presence/absence of delayed release of the stem-final stop or affricate, the consonantality of the preceding segment (always a vowel in training), and the identity of the final vowel. Other features either were not distinctive in the artificial language (for instance, it contained no nasals in any position), or subjects in the experiment were not tested on generalizing across their values (e.g. height of the stem vowel, or the nasality or continuancy of the stem-final consonant); thus we have no data on which to evaluate the resulting model's predictions for these features. The delayed release feature was included for the stem-final consonant in order to allow schemas specific to stops, and the consonantal feature of the preceding vowel was included to let the model make predictions about whether [pt∫i]- and [kt∫i]-final plurals are likely to be produced.

Each low-frequency word contributed one observation to the training set for the model, whereas each high-frequency word contributed three (based on the fact that for humans these high-frequency words were three times more frequent than the low-frequency words and produced by three artificial speakers rather than one). All words that could be generated by changing one or more of the features of the real words without violating the phonotactics of English were also entered into the training set. The dependent variable to predict was the occurrence ('yes' in the trees below) vs. nonoccurrence ('no' in the trees below) of these words in the lexicon. Thus, each existing word had 'yes' for the dependent variable and each nonexisting word had 'no', and high-frequency existing words were entered three times.

The resulting trees for the Tapa and Tapat∫i languages are shown in Figure 4. Notice that the tree induction procedure automatically generates negative schemas in addition to positive ones, for example, *[+cont]V#. I propose, however, that only the branches with nonzero predicted counts are stored. In other words, schemas describe observed rather than nonexistent words (Bybee 1985, 2001, Langacker 1987, Nesset 2008). Generally, a first-order schema is then defined as in 10, and I propose that all schemas in a decision tree are extracted from the data.

(10) A first-order schema is a path through a conditional inference tree in which the predicted variable is type frequency of an ngram and the predictors are phonological features of words containing that ngram. The path must proceed downward from the root of the tree terminating in a node that is either (i) a leaf with a nonzero type frequency or (ii) an ancestor to at least one leaf with a nonzero type frequency.

Taking as an example the tree in Fig. 4a, which depicts the language Tapa, the two shortest paths originating at the root can be described as CCV# (left path, linking node 1 to node

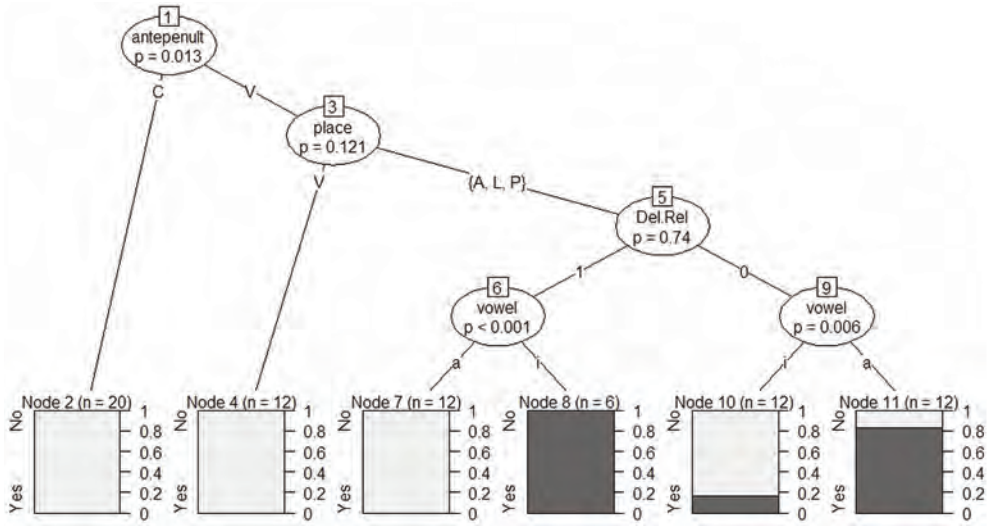
[13] The binary nature of the resulting trees is to be treated as an implementational convenience, rather than a theoretical claim.

2) and VCV# (right path, linking node 1 to node 3). However, there are no observed words satisfying CCV#, and thus CCV# is not a schema under 10. Continuing down, we obtain V[+Velar]V# (linking node 1 to node 4 through node 3) and V[–Velar]V# (linking node 1 to node 5 through node 3), where again only the latter is a schema under the definition. Further down, we obtain V[–Velar;+Del.Rel]V# and V[–Velar;–Del.Rel]V#, both of which are ancestors to leaves containing existing words, and thus both are legitimate first-order schemas describing classes of real plural word forms. Both are extracted. Lower down, V[–Velar;–Del.Rel]*i*# and V[–Velar;–Del.Rel]*a*# are both extracted but the one rooting for *-a* is weighted more highly as it is supported by more existing words. Also extracted is V[–Velar;+Del.Rel]*i*#, which roots for [tʃi]-final outputs. The tree for Tapatʃi in Fig. 4b is interpreted similarly.[14]

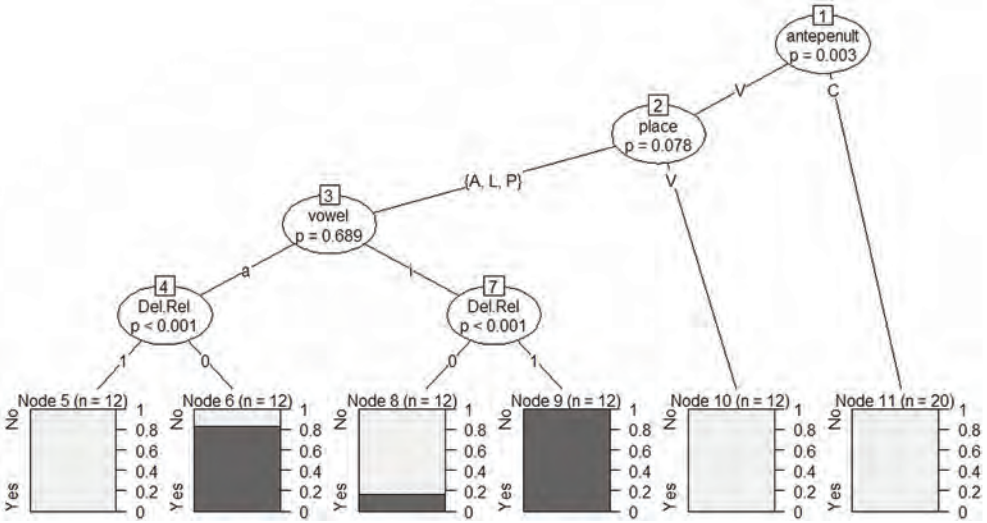
As demonstrated by Daelemans and van den Bosch (2005), abstractionist (grammatical) and analogical (instance-based) models of productivity are both expressible as decision trees of the kind shown in Fig. 4 (see also Ernestus & Baayen 2011). The schemas terminating in the leaves of a decision tree representing an analogical model are individual words. By removing these leaves, one arrives at a grammatical model, in which individual exemplars play no role. While the full tree yields maximum accuracy, the pruned tree yields gains in processing speed (Daelemans & van den Bosch 2005). Thus, the decision-tree implementation of schema induction suggests that, while the entire tree, complete with exemplars, is stored, time pressure may cause speakers to use only the most reliable schemas at the top of the tree (Ernestus & Baayen 2011). The model predicts that processing should become less exemplar-based/less sensitive to the less informative features of the stimulus under time constraints. Evidence for this prediction, at least for word recognition, is provided by McLennan and Luce (2005), who demonstrated that exemplar-specific information is ignored in lexical decision if listeners have to respond under time pressure. Words are easier to recognize if they have been presented previously, whether or not processing is done under time pressure. Without time pressure, the effect of previous presentation is augmented if the word is acoustically identical on both presentations, but this augmentation disappears under time pressure. This suggests that relatively uninformative acoustic details are accessed only when there is time to do so. Otherwise, only the more informative dimensions high in the tree are accessed. We propose that the same may be true for morphological and morphophonological processing tasks like elicited production and acceptability judgment: when done under time pressure, they should exhibit effects only of the most reliable schemas (though reliability may, perhaps, be tempered by a preference for locality, as in Albright & Hayes 2003) and therefore be more subject to perseveration of input characteristics.

Progressive differentiation is also observed in L1 semantic development (Keil 1979, Mandler 2000, Pauen 2002, Warrington 1975). In modeling these data, Rogers and McClelland (2004:84–96) show that progressive differentiation is exhibited by distributed connectionist networks, whose internal representations gradually develop to distinguish related concepts. We can thus think of the present implementation as a symbolic description of the progressive differentiation progress exhibited by the neural network

[14] ‘Del.Rel’ = delayed release (0 = stop). ‘place’ = place of articulation, with values ‘P’ = ‘palatal’, ‘L’ = labial, ‘A’ = alveolar, and ‘V’ = ‘velar’. ‘P’ is not grouped with ‘A’ or ‘V’ because of the relatively equal typological frequency of velar and alveolar palatalization (Kochetov 2011). I assume that subjects take the data as being a noise-free informative sample and strive to perfectly represent the distribution in the training data; thus the criteria for split creation are extremely liberal (for present data, micriterion = .1, minbucket = 1, minsplit = 1).



a. Tapa language. Schemas: VCV#, V[–Velar]V#, V[–Velar;+Del.Rel]V#, V[–Velar;–Del.Rel]V#, V[–Velar;+Del.Rel]*i*#, V[–Velar;–Del.Rel]*i*#, V[–Velar;–Del.Rel]*a*#.



b. Tapatʃi language. Schemas: VCV#, V[–Velar]V#, V[–Velar]*a*#, V[–Velar]*i*#, V[–Velar;–Del.Rel]*a*#, V[–Velar;–Del.Rel]*i*#, V[–Velar;+Del.Rel]*i*#.

FIGURE 4. Top-down schema extraction results for Tapa and Tapatʃi. Observation (counted in *n*) = individual word or phonotactically possible word. Dependent variable is occurrence ('yes', shown in black) or nonoccurrence ('no', shown in light gray). Probability of 'yes' is shown on the right side of the box in each leaf of the tree. The *n*s above the leaves of the tree are numbers of words containing the structure defined by the path from the root of the tree to that particular leaf. Note the high *p*-values for 'vowel', indicating that the identity of the final vowel is not very informative, especially in Tapa. The numbers in square boxes are arbitrary node indices.

underlying form-meaning mapping. As Rogers and McClelland point out, 'there may be good computational reasons … to begin with very similar, undifferentiated internal representations. Specifically, such an internal state would permit very rapid learning about properties common to all things' (2004:96). Similarly, for phonological representations,

starting from a state where forms with different meanings are undifferentiated leaves the learner ready to learn what phonological forms in general, and forms sharing salient semantic features in particular, have in common.

THE PECULIAR SALIENCE OF THE FINAL VOWEL. There is a problem with the trees in Fig. 4. Recall that classification and regression trees enter predictors into the tree when they are informative. For Tapa and, to a lesser extent, Tapatʃi, the identity of the final vowel is not very informative regarding whether the form can occur in plurals: the numbers of *-i* and *-a*-final plurals are quite evenly matched. Thus, suffix vowel identity is not by itself a significant predictor of occurrence. But suffix vowel is extremely informative if the place of articulation of the stem-final consonant is also known: [tʃ]-final stems are always followed by *-i*, whereas [t]- or [p]-final ones are almost always followed by *-a*. Conversely, *-a* is never preceded by [tʃ], whereas *-i* is rarely preceded by [t] or [p]. In other words, there is a cross-over interaction between stem-final consonant identity and suffix vowel identity in these two languages. A perfect cross-over interaction cannot be captured by a classification and regression tree (Strobl et al. 2009), and here we are close to one. Nonetheless, the interaction does seem to be learned, which suggests that the salience of the final vowel to the learners is greater than predicted by its reliability.[15]

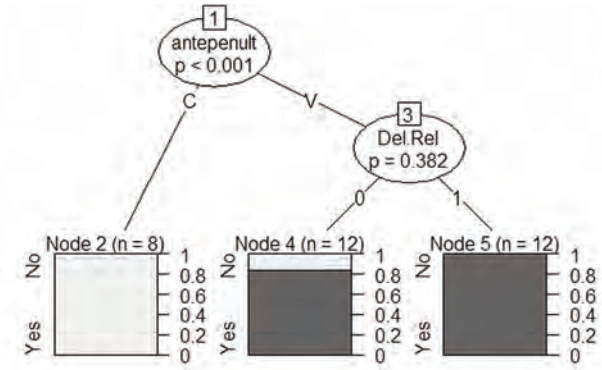
I propose that the final vowel is focused on early and promote final vowel identity to the top of the tree by hand. There are many possible reasons for the special salience of this final vowel: particular salience of the CV structure of a form, the need to fully specify every part of it that cannot be filled in from the input, recency effects, two-stage derivation where the suffix is decided on before the stem allomorph, and so forth.

Promotion of the suffix vowel to the top of the tree predicts that mappings involving the same affix help each other, unless they are competing for the same probability mass,[16] whereas mappings involving different affixes do not. This prediction appears to be correct. In the present experiments, *tʃ* → *tʃi* helps {*t*;*p*} → *tʃi*, whereas *tʃ* → *tʃu* does not. In Kapatsinski 2009a:143–53, ratings of mappings involving different affixes (e.g. *k* → *ki* and *k* → *ka*) are either uncorrelated or negatively correlated across subjects, whereas ratings of mappings involving the same suffix are positively correlated.

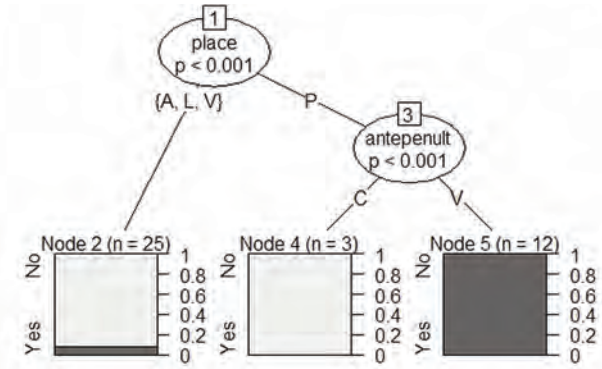
With this adjustment, the schemas extracted in our artificial languages are shown in Figure 5. The differences between languages are appropriately predicted: the addition of examples of *tʃ*# → *tʃi*# helps *X* → *tʃi*# (by driving up the weight of *Vtʃi*# in Tipitʃi compared to Tipi and both *tʃi*# and *Vtʃi*# in Tapatʃi compared to Tapa). The presence of the *tʃi*# schema in Tapatʃi but not Tipitʃi predicts a greater rate of *p* → *ptʃi* and *k* → *ktʃi* errors in the former language. The replacement of examples of *t*# → *ta*# and *p*# → *pa*# with examples of *t*# → *ti*# and *p*# → *pi*# helps *k*# → *ki*# by replacing a weak [–Pal]*i*# schema supported by only two words with a stronger *V*[–Del.Rel]*i*# schema supported by six words. The addition of *tʃ*# → *tʃu*# examples has no effect on palatalization rates before *-i* since the final suffix vowel is not shared.

[15] Alternatively, we could assume that the stem-final consonant and the following vowel are perhaps chunked together in perception, because they are part of the same syllabic unit (Kapatsinski 2009b) or because the vowel carries cues for the consonant's place of articulation. Allowing for such a larger 'configural' feature would also allow the learner to deal with a cross-over interaction by allowing all CVs to behave in ways not predictable from their parts. The claim of the current learner is that such cross-over interactions between features are in fact hard to learn for humans, unless there are large differences in salience (due, for instance, to being close to an edge or prosodic prominence) or the features are chunked together into a larger unit (e.g. the rime in Kapatsinski 2009b).

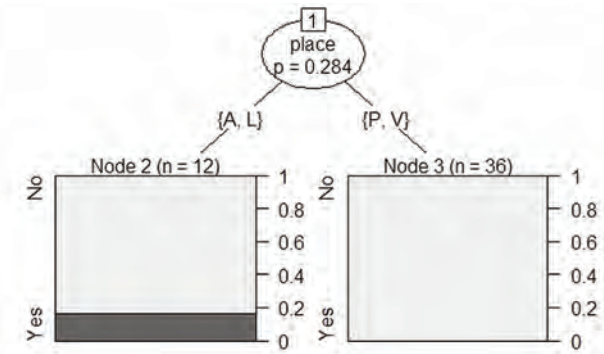
[16] As in elicited production tasks and in rating tasks that pit alternatives against each other (see Kapatsinski 2006 for more details).



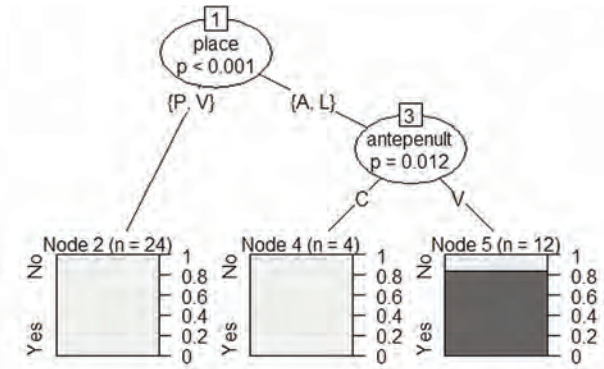
a. Tipi(tʃi), final vowel is -*i*. Schemas: *i*#, VC*i*#, V[+Del.Rel]*i*#, V[–Del.Rel]*i*#.



b. Tapa(tʃi), final vowel is -*i*. Schemas: *i*#, [–Pal]*i*#, V[+Pal]*i*#.



c. Tipi(tʃi), final vowel is -*a*. Schemas: *a*#, {Alv;Lab}*a*#.



d. Tapa(tʃi), final vowel is -*a*. Schemas: *a*#, {Alv;Lab}*a*#, V{Alv;Lab}*a*#.

FIGURE 5. Schemas extracted by models in which the final (suffix) vowel is especially salient.

3.3. POSITIVE VS. NEGATIVE CONSTRAINTS: A COMPARISON WITH MAXENT. I compared the performance of ctree() with the theoretically highly related MAXIMUM ENTROPY (MaxEnt) learner developed by Hayes and Wilson (2008). Both models seek the best description of the lexicon (in this case, of plural forms). The principal difference between MaxEnt as implemented in Hayes & Wilson 2008 and the present model is that Hayes and Wilson use negative constraints while positive ones are used here.[17] Thus the Hayes and Wilson learner focuses on what is underrepresented (what plurals are NOT like), looking for valleys in the probability distribution, while our learner focuses on what plurals typically ARE like, looking for peaks in the distribution.[18]

It is important to note here that MaxEnt was originally developed to account for language-wide phonotactics rather than schemas specific to forms that share semantic similarity. There is a good theoretical reason to prefer positive statements for the latter: schemas are natural descriptions of the result of grammaticalization of specific meaningful word forms in a specific constructional context (e.g. Bybee 2003). While phonetic erosion often accompanies grammaticalization, this erosion can make the schema more and more general but is relatively unlikely to make it split into a set of disjoint forms that are better described by a negative constraint (e.g. *going to* can reduce to [ənə] as it becomes a future marker and could conceivably reduce even further to, say, a floating [+nasal] feature, but it is difficult to think of it generalizing to, for example, **g*, unless some other word grammaticalizes into the nonfuture marker [g]). Thus the lexical patterns facing a learner of schemas are typically better described using positive constraints: the probability distribution of element combinations in a semantically defined area of the lexicon is more likely to have a few isolated peaks than a few isolated holes. This is less obviously true of phonotactics, where feedback from articulation or perception can lead to avoidance of specific sound sequences in speech planning (Martin 2007). Nonetheless, I think it is worthwhile to devote a couple of pages to elucidate the differences in empirical predictions between negative and positive constraints.

I considered two versions of MaxEnt. MaxEnt with only simple negative constraints allowed only constraints against underobserved sound sequences. This version failed to capture the effect on palatalization rates of adding examples of *tʃ* → *tʃi*: predicted palatalization probabilities did not change. The constraints extracted from the data militated against [k], [t], or [p], but had no preference against [tʃi]: all languages gave rise to **k*, **kV/Vk*, **C*{*p*;*k*}, and **Cp*; in addition, languages featuring many examples of

[17] The version of MaxEnt implemented in Praat (Boersma & Weenink 2009) allows for both positive and negative constraint weights; thus **tʃi* can, upon exposure to many [tʃi] examples, gradually shift to have a negative weight, becoming a preference FOR [tʃi] (Goldwater & Johnson 2003, Jäger 2007, Johnson 2002). I do not discuss this further for two reasons. First, like the model used here (C4B), Praat MaxEnt learns a preference for a structure, which grows in strength as the type frequency of the structure in the lexicon increases, making the differences between it and C4B uninteresting for examining the theoretical questions addressed in the present article. Second, unlike Hayes and Wilson's (2008) version and C4B, it does not discover constraints. Rather, the constraint set must be provided a priori. One untested prediction of C4B compared to Praat MaxEnt is that extra examples of *tʃ*# → *tʃi*# should only help [tʃi#]-final plurals if [tʃi#] is (or becomes) an unexpectedly common sequence in plural forms. It does not help to add examples of *tʃ*# → *tʃi*# if they only make *tʃi*# as common as *pi*# and *ti*#. Praat MaxEnt predicts, by contrast, that being unexpected at the beginning of training can cause a sequence to rapidly rise in popularity.

[18] Since MaxEnt as currently implemented requires at least 3,000 training tokens (MaxEnt help), the training data presented to subjects were presented to the model twelve times as often. Thus the difference in amounts of training and lexicon sizes between the artificial languages was maintained. See §3.6 for discussion of this limitation.

[ta] and [pa] and few of [ti] and [pi] gave rise to **XkX* (no medial [k]) and **V{t;k}i*.[19] The weights of these constraints were not significantly affected by the addition of examples of *tʃ → tʃi*: the probability of palatalization was 100% for [k] and 68% for [p] in all languages; probability of palatalizing [t] was 50% in Tipi(tʃi)(tʃu) and 91% in Tapa(tʃi)(tʃu), thus independent of whether examples of *tʃ → tʃi* were presented (it depended instead on the number of *Vti* examples presented). Thus, induction of simple negative constraints failed to account for the effect on probability of palatalization of adding examples of *tʃ → tʃi*: these examples do not greatly change the badness of [ti] or [pi], and the goodness of [tʃi] is not tracked since it is the most common sequence in the data and therefore does not violate any constraints.

This example illustrates the general problem with relying exclusively on negative constraints to describe a small developing lexicon. Given that the lexicon is small and biased in favor of the most easily pronounceable words (Schwarz & Leonard 1982), negative constraints extracted from it are likely to be more restrictive than those extracted from a large lexicon akin to the lexicon of an adult native speaker. Thus the general prediction of theories relying on negative constraints is early conservatism, rather than early liberalism. While this may be true of phonotactics, it is definitely not true of morphophonology: early on, schemas forcing a novel stem to be like other plural stems are relatively weak.

After observing the failure of simple negative constraints to account for the data, I allowed MaxEnt to infer conditional constraints referring to complement natural classes (as in Hayes & Wilson 2008). These constraints take the form of 'only X can {precede;follow} Y' or 'Y can {precede;follow} only X'. This in principle allows MaxEnt to formulate schemas like 'given that the segment in position N is Y, the {preceding;following} segment must be X', where X and Y could be natural classes of segments. These constraints are equivalent to downward paths through a conditional inference tree, but ones that terminate in a leaf with a zero frequency of occurrence. Thus the resulting schemas punish underattested sequences rather than rewarding overattested ones.[20] With this addition, MaxEnt could account for the effect of adding examples of *tʃ → tʃi* to the Tapa language, as shown in Table 2. The addition of these examples led to the extraction of the constraint 'the final segment of the plural cannot be [i] if the preceding segment is not [tʃ]'. This constraint punishes [ki], [ti], and [pi] while not punishing [tʃi], increasing the probability of palatalization.

Unfortunately, the model was unable to account for the effect of adding *tʃ → tʃi* examples to Tipi, since [i] is still the vowel that most commonly follows [t] and [p] in Tipitʃi. A possible solution is to make sure that the negative constraints are conditionalized on the suffix: for example, 'if the final segment is [i], the preceding segment must not be [–Pal]'. This would capture the special significance of the suffix vowel, which is as nonobvious to MaxEnt as to the model used here (C4B). The move would ensure that examples of *tʃ → tʃi* reduce the goodness of [ti] and [pi] by making [t] and [p] less probable given that the following segment is [i].

[19] I used the same feature set used for C4B and varied whether place features are privative or binary. The results reported here hold for both choices. For simplicity, the antepenultimate segment was specified here as simply [+/–syllabic;+antepenult]. Preceding segments were omitted. Using fully specified representations did not improve performance.

[20] The schema/constraint-extraction algorithms also differ in a number of ways. One way in which the current implementation of MaxEnt is superior is that it has a bias in favor of interactions between temporally adjacent or cotemporaneous features. The current implementation of C4B has no knowledge of serial order. This is not a claim of the C4B model but rather a limitation of implementation.

	TAPA		TIPI		TAPATʃI		TIPITʃI		TAPATʃU		TIPITʃU	
	*kV/Ck	3.09[a]	*kV/Ck	3.22	*kV/Ck	4.01	*kV/Ck	5.47	*kV/Ck	4.43	*kV/Ck	4.39
	*k	0.49	*k	0.49	*k	0.33	*k	0.18	*k	0.01	*k	0.07
	*Xk	2.56[b]	*Xk	2.49					*C/ __ {k;p}	5.16	*C/ __ {k;p}	5.16
	*i/V{k;t} __	4.55			***i/[–Pal] __**	**1.67**			*i/V{k;t} __	4.55		
	*Cp	5.15[c]	*Cp	5.16	*Cp	4.46	*Cp	5.38				
[k]	100%		100%		100%		100%		100%		100%	
[t]	**86%**		50%		**97%**		50%		**86%**		50%	
[p]	**68%**		68%		**90%**		68%		**68%**		68%	

TABLE 2. Constraints and their weights and the production probabilities of palatalization for each language, as extracted by MaxEnt. [tʃi] never violates any constraints. Bolded constraints increase the probability of palatalization in Tapatʃi compared to Tapa and Tapatʃu.

[a] Dorsals must be preceded by nondorsals: *[^–Dorsal][+Dorsal]; all vowels are dorsal; the only dorsal consonant in the lexicon is [k].

[b] Penultimate [k] is prohibited (along with vowels): *[+antepenultimate][+Dorsal].

[c] In the test data, 7/11 singular-final labials and 9/11 singular-final coronals are preceded by a vowel.

Once we allow for conditional inference and assume special salience of the final suffix vowel, negative constraints can account for the present data. However, the underlying difference in predictions between positive and negative constraints remains. Positive constraints, as in C4B, predict that similar segment structures with similar meanings help each other gain productivity (see also Abbot-Smith & Behrens 2006 for evidence of this effect in syntax), while unrelated constructions do not interact. As a result, ratings of form-meaning pairings sharing meaning should be either positively correlated or uncorrelated across subjects.[21] In the current version of C4B, form-meaning pairings should help each other if they share the affix and should not interact otherwise (for some supportive rating evidence, see Kapatsinski 2009a:143). The addition of faithfulness constraints introduces a small complication: negative correlations can then be observed but only if one of the mappings involves a stem change while the other does not and both are competing for the same input; for example, ratings of *k* → *tʃi* and *k* → *ki* may be negatively correlated because one involves a violation of the faithfulness constraint clamoring for retaining [k] while the other does not.[22] In theories relying on negative constraints, related form-meaning pairings can hurt each other; for example, exposures to [tʃi] reduce the goodness of [ki] in MaxEnt.[23]

[21] Of course, the ratings must be standardized before this comparison is made (by converting into *z*-scores). Otherwise, ratings of rare structures and frequent structures might be negatively correlated across subjects simply because some subjects will use more of the scale, making rare structures less acceptable and frequent structures more acceptable.

[22] To the extent that faithfulness constraints are perseveratory tendencies internal to production, however, the competition does not need to manifest itself in ratings and positive correlations can still be observed, as in Kapatsinski 2006, 2009a.

[23] This is taken to perhaps the logical extreme in Pierrehumbert 1993 and Frisch et al. 2004, where constraints against unobserved segment combinations are weighted by the difference between how often the sequence is observed to occur and how often it would be expected to occur if the component segments combined randomly. This model then predicts that frequently encountering [k] and [i] on their own without ever encountering [ki] should make one confident that there is a constraint against [ki], and examples of plurals ending in [pi] and [ti] should provide evidence against [ki]. We have seen that this prediction is incorrect: additional examples of [pi] and [ti] help [ki] rather than hurting it. Sound sequences sharing segments can help each other in phonotactic learning (see also Goldrick 2004, Hayes & Wilson 2008). MaxEnt does not share this problematic prediction for the present data because of acquiring general, and partially redundant, constraints against [i] or [a] rather than only more specific constraints against bigrams.

3.4. 'CHUNK!' CONSTRAINTS. There are two effects remaining to be explained: (i) the high proportion of stem-final consonant retentions/no-change errors, especially for labials, and (ii) the finding that learners are more likely to overgeneralize palatalization to [t] than to [p]. Both of these findings are unexplained by the product-oriented schemas extracted via either C4B or MaxEnt. I suggest that these effects are due to the tendency to perseverate on chunks of the singular form, to be implemented as 'CHUNK!' CONSTRAINTS.[24]

A 'chunk!' constraint is equivalent to a conjunction of chunk-specific output-oriented Max and Ident constraints in OT/HG (Kenstowicz 1996). I refer to these constraints using the notation 'chunk!': for example, '[p]!' indicates that one should output [p] if [p] is present in the input.[25]

CHUNKS are relatively independent processing units that can change their positions in speech errors. They thus include nonmeaningful, roughly segment-sized gestural units (e.g. Browman & Goldstein 1989, Dell et al. 1997, Fromkin 1970, Goldstein et al. 2007, Shattuck-Hufnagel & Klatt 1979, Stemberger 1982, 1991) as well as larger meaningful units including morphemes and words. I suggest that when a speaker has to produce a novel word form from a known morphologically related word form, chunks of that known word form are subject to perseveration. This perseveration is largely functional, as MOST of the to-be-produced unknown form SHOULD come from the known form,[26] but can overapply, resulting in the product resembling the source TOO MUCH when a schema that demands a stem change loses the competition to a perseverating chunk. If there is no relevant schema that is incompatible with some features of the stem, which is particularly likely early during lexical acquisition when the schemas still accept pretty much anything, 'chunk!' constraints are free to drive production, leveling stem changes. As schemas become more specific in the course of lexical acquisition, stem changes become more productive (this is what is captured with high initial ranking of output-output faithfulness constraints in OT; Hayes 2004).

In our data, overapplication of perseveration is particularly common in elicited production. In repetition of singular-plural pairs during training, errors sometimes level stem changes, but the leveling usually proceeds from the more recent plural form to the singular (e.g. [bik bitʃi] repeated as [bitʃ bitʃi], not [bik biki]). In elicited production, unlike in repetition during training, the only presented form is the singular; thus perseveration is the only direction for speech errors. The perseverating chunks are sometimes in competition with product-oriented schemas rooting for common plural patterns. Namely, the product-oriented schemas root most strongly for [Vtʃi]-final plurals, the most common type of plural in the language. Sometimes an input chunk is perseverated, however, resulting in errors like [buptʃi] from [bup] or [buktʃi] from [buk]. Such outputs happen more often for [p] than for [k], likely because [p] and [tʃ] do not share articulators and thus can both surface without interfering with each other, whereas [k] and [tʃ] are more likely to be blended together (Browman & Goldstein 1991).

[24] The remaining effect to be explained—that examples of *tʃ* → *tʃi* help untrained *C* → *tʃi* mappings—requires an interaction between 'chunk!' constraints and schemas, and is therefore left until §3.5.

[25] While the current statements of the 'chunk!' constraints use segments, gestures are likely to work better (Browman & Goldstein 1989, 1991, Gafos 2002): chunks are production units and a segment can contain more than one independently controllable production unit. I stick to segments for the present article for reasons of their greater familiarity.

[26] Lack of any tendency to perseverate on aspects of the input may be one reason Rumelhart and McClelland's (1986) connectionist model of the English past tense famously produced 'bizarre' stem changes like *mail-membled* (Pinker & Prince 1988).

To the extent that acceptability rating models the production process (see Albright & Hayes 2003, Boersma 2004, Kapatsinski 2006, Zuraw 2000), the tendency for perseveration should also be relevant for acceptability rating. It should not be AS relevant as for elicited production, however, making stem changes more productive in rating than in elicited production. This difference can lead to a paradoxical situation in which a form featuring a stem change can be rated as being more acceptable than an alternative that preserves the stem faithfully (due to being a better match to the schemas of the language) and yet be less likely to be produced when given the stem (Zuraw 2000).

Each schema and each chunk clamors for expression, but some chunks and schemas have a stronger voice and can thus clamor more effectively. The strength of a schema is determined largely by its type frequency (Bybee 1985, 2001). The strength of an input chunk is the tendency for that chunk to be perseverated on. I propose that chunk strength increases whenever one perseverates on a chunk (e.g. saying [buk buktʃi] or [buk buki] increases the weight of '[k]!' among other chunks, and saying [bup bupi] increases the weight of '[p]!') and decreases whenever one has to replace the chunk with something else in production (e.g. saying [buk butʃi] reduces the weight of '[k]!'). Thus training on velar palatalization reduces the weight of the constraint clamoring for perseverating on [k]. Reducing the weights of violated 'chunk!' constraints appears necessary to account for phonetically unnatural patterns of palatalization. For instance, Ohala (1978) discusses productive pre-[w] palatalization in Southern Bantu, which affects [p] but not [k] or [t], despite labial palatalization being crosslinguistically rare compared to velar or alveolar palatalization (Kochetov 2011). How do speakers of these languages learn this pattern? While it is possible that the Southern Bantu lexicon contains an abnormally large number of [k]-final and [t]-final stems that never change, thus raising the strengths of '[k]!' and '[t]!' so much that the 'tʃw' schema cannot overcome them, it seems more likely that changing labials and not velars is learned in part by lowering the weight of '[p]!'.

'Chunk!' constraints are directly grounded in speech errors, being supported by (i) the existence of perseveration errors, and (ii) the ADDITION BIAS, that is, the finding that speech errors more commonly involve addition than deletion (Goldstein et al. 2007, Hartsuiker 2002, Stemberger 1991).

'Chunk!' constraints help resolve the too-many-solutions problem in optimality theory, providing a unified explanation for (i) the PRESERVATION PRINCIPLE (Paradis & LaCharité 1997), which notes that phonotactic violations in loanword adaptation are overwhelmingly resolved by epenthesis rather than deletion, and (ii) the CONTIGUITY CONSTRAINT (Kenstowicz 1994), which captures the tendency of epenthesis to happen at morpheme edges rather than morpheme-internally. Both the preservation principle and the contiguity constraint ensure preservation of input chunks. As Kang (2011:2272) points out, 'all languages … that choose deletion repair in coda position have a strong preference for monosyllabic morphemes. But … even these languages do not systematically prefer deletion for onset clusters'. In other words, speakers of a language tend to give up an input chunk only if forced to do so by a strong product-oriented schema. Furthermore, onset chunks, which we know to be more strongly activated (Dell 1986) and more tightly fused than coda chunks (Browman & Goldstein 1989), are also more likely to persevere.

'Chunk!' constraints are equivalent to stating that Max-B{R;A} and Ident-B{R;A} should generally be ranked at least as high as Dep-B{R;A} constraints (Kenstowicz 1996); that is, paradigm uniformity is enforced by blocking deletion or stem changes but not by blocking insertion, unless that insertion breaks up a sequence of segments

that would otherwise be identical to an input chunk. Treating 'chunk!' constraints as accidental conjunctions of Ident-[] and Max-[] constraints fails to predict that phonotactically illegal sequences are more commonly repaired by insertion, and especially insertion at the boundary, than deletion or change (Kang 2011). Finally, if paradigm-uniformity constraints are perseveratory tendencies, we expect these constraints to be high-ranked in childhood (as suggested in the literature, e.g. Hayes 2004), since children generally show more motor perseveration than adults (Dell et al. 1997, Smith et al. 1999) and in fact are observed to often perseverate on inflections recently produced by them or their caregivers (which, more often than not, increases accuracy; Ambridge & Lieven 2011:165–66, Farrar 1992, Rubino & Pine 1998).[27]

3.5. SCHEMA-CHUNK COMPETITION. The grammar generates all candidates that instantiate at least one plural schema and perseverate on all segments of the stem that are not in contradiction with the schema being instantiated. For instance, given [bup], one might generate candidate plurals shown in the tableaux in Table 3. The numbers in the tableaux show the strengths of the various schemas, which is a function of type frequency and redundancy, and the strength of '[p]!', which is arbitrarily set to 10. Each column contains all schemas that support the same candidate output or set of candidate outputs (from among those shown in the tableaux). The bottom-most schema is the most specific schema supporting that candidate output. Its weight is type frequency. Schemas above it are more general versions of the same schema. Their weights are given as type frequency × entropy.[28]

bup	[Pal]i#	1.7	[Pal]i#	1.7	[–Pal]i#	0	[–cont]a#	1.9				
	tʃi#	2.7	tʃi#	2.7	[–cont;–Pal]i#	0	V[–cont]a#	3				
			Vtʃi#	4	V[–cont;–Pal]i#	0.1	V[–cont;Lab]a#	4.1				
					V[–cont;–Pal;–son]i#	1.2	V[–cont;–son;Lab]a#	6				
					V[–cont;{Alv;Lab};–son]i#	2			'[p]!'	TOTAL	*p*(prod)	
bupi						3.3				3.3	11%	
bupa								15.0		15.0	52%	
buptʃi		4.4								4.4	15%	
butʃi				8.4					–2	6.4	21%	

TABLE 3. The candidate plural forms for the singular form [bup] and schemas supporting each candidate in Tapatʃi.[29]

[27] An untested prediction of the faithfulness-based account of the bias is that *p* → *tʃ* should be dispreferred over *k* → *tʃ* regardless of the trigger of the change.

[28] Raw type frequency results in too much overgeneralization in that it favors very general schemas like PL=…i# over specific schemas like PL=…aki#. The proper way to penalize general schemas that overgenerate is beyond the scope of this article.

[29] A column contains all schemas supporting a single candidate. Total shows total amount of support for a candidate. *p*(prod) is the corresponding production probability for a candidate. We can distinguish between first-order schemas that are paths terminating in a leaf (let us call them MAXIMALLY SPECIFIC SCHEMAS) and schemas that are more general. Maximally specific schemas can be weighted by type frequency. Given the proposed process of schema extraction, for every schema X that is not maximally specific there is at least one corresponding schema Y that is maximally specific, such that the features constituting Y are a superset of features constituting X. Thus the non-maximally-specific schemas are redundant. Redundant schemas are punished by entropy using the formula in (i). TF is type frequency, and L is the number of leaves subsumed by the schema. R is a coefficient weighting the importance of nonredundancy and varies between zero and one. When R is one, redundancy does not matter. When R is zero, redundancy is punished the most. In fact, fully redundant schemas receive a weight of zero. However, there may not be any fully redundant schemas: on at least some occasions words that fit the more specific schema may be erroneously perceived not to fit because some features of the word required by the more specific schema are misperceived or because only the features

The probability of producing a candidate is taken in 11 to be the sum of the strengths of schemas and chunks supporting that candidate divided by the sum of strengths of all relevant chunks and schemas (Legendre et al. 1990, Smolensky & Legendre 2006). The relevant 'chunk!' constraint punishes outputs that violate it rather than supporting the outputs that do not violate it. The reason for this decision is that the alternative solution leads to shrinking differences between alternative outputs that are due to schema strength. For instance, if 10 were added to the strengths of [bupi], [bupa], and [buptʃi], then their production probabilities would become 27%, 31%, and 20% respectively, despite training probabilities of 25%, 75%, and 0% respectively. Thus the system would fail to match the probabilities in the training data: [pi]-final plurals are three times more common than [pa]-final plurals in training but are barely more common in the learner's output. One solution is to punish forms for violating 'chunk!' constraints instead of rewarding forms for obeying 'chunk!' constraints. Alternatively, we can stretch the differences in predicted acceptability between well-supported forms. This can be accomplished by exponentiating the support values (Goldwater & Johnson 2003, Hayes & Wilson 2008), so the probability of producing candidate A given a set of *i* candidates is as given in 11, where support_A is the sum of strengths of the schemas and chunks supporting A and the strength of a schema is k * type_frequency, with k << 1.[30]

$$(11)\quad p_{produceA} = \frac{\exp(support_A)}{\Sigma \exp(support_i)}$$

The two-level generation/evaluation scheme with stochastic choice naturally produces the finding that additional support for a product helps unfamiliar mappings resulting in that product more than it helps familiar ones. Consider [butʃi] derived from [buk] vs. [butʃi] derived from [but] after training on $k \rightarrow tʃi$, $t \rightarrow ti$. When [butʃi] is derived from [buk], it competed with [buki] and is well ahead of [buki] based on the training data and a weak '[k]!'. When it is derived from [but], it competes with [buti] and is likely to lose, since [buti] is specifically supported by training examples and the stronger '[t]!'. If you add 1 to support_A in 11 (in both the numerator and the denominator), this increases the production probability more when the probability of producing A is low (Figure 6). Thus infusing extra strength into a product-oriented schema that supports a candidate is most likely to appreciably help that candidate when the candidate is weak relative to the competition.

3.6. LIMITATIONS AND FUTURE WORK. There are a number of limitations of this model. First, like HG and CG, the model does not include any kinds of source-oriented generalizations beyond identity relations. Yet such generalizations appear to be required for phonologically unmotivated alternations found in natural languages (Booij 2010, Nesset 2008, inter alia). The CG solution is to propose the existence of generalization across first-order schemas/constructions. But how exactly this generalization process works and how the extracted second-order schemas interact with first-order schemas in processing have not yet been worked out.

---

of the more general schema are accessed in time (McLennan & Luce 2005). I am assuming the following probabilities of misperception: 5% for misperceiving place of articulation, 0.01% for misperceiving [sonorant], [continuant], or [consonantal]. Here R is set arbitrarily to .5. The weight of '[p]!' is set to 2 by hand.

$$\text{(i)}\quad wt = \left(\sum_{i=1}^{L} TF_i\right) \times \left(R + (1-R) \times \left(-\sum_{i=1}^{L}\left(\frac{TF_i}{\Sigma TF_i}\log\frac{TF_i}{\Sigma TF_i}\right)\right)\right)$$

[30] For the present data, k = 0.05 appears to be close to optimal. In other words, competition resolution between schemas is stochastic (cf. Hayes et al. 2009's 'law of probability matching').

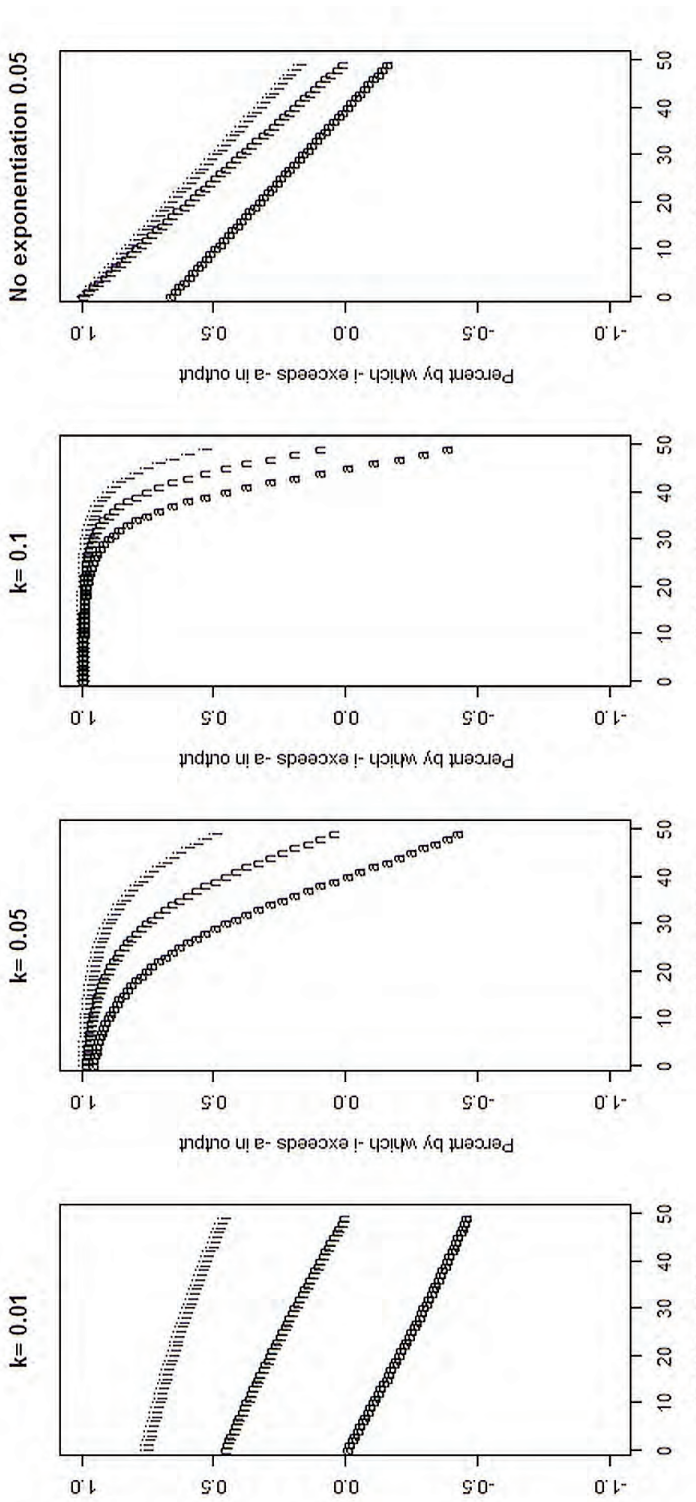


FIGURE 6. The effect of a schema's type frequency on the probability of choosing an output consistent with it, assuming only two competing schemas. Data series 'n' ('neutral') depicts what happens to the likelihood of choosing *-i* as the suffix as we vary the type frequency of *-a* between 0 and 49 and of *-i* between 100 and 51 (x-axis) for different values of k (which weight type frequency). The other lines depict how the differences in production probability between *-i* and *-a* change: 'i' when 1/k extra examples of *-i* added, 'a' when 1/k extra examples of *-a* added. The 'n' line is closer to the 'i' line than to the 'a' line, especially when *-i* is quite a bit more frequent than *-a* but not so frequent as to be chosen almost 100% of the time. This holds for both solutions to the too-much-equality problem above.

Second, like CG, the present model assumes generalization over forms within semantically coherent sets of word forms. I have not, however, proposed any algorithm for identifying these semantically coherent sets. Generally, it appears implausible that the necessary semantic clustering happens before the phonological-construction extraction. Rather, both must happen at the same time, so that meanings and forms coevolve. This might reduce some paradigm-uniformity constraints to product-oriented schemas learned across forms sharing the same stem and thus the associated semantics. This is an important direction for future research.

Third, there are important differences in predictions between negative constraints and positive schemas as well as differences in how the schema/constraint-induction process is implemented in C4B and MaxEnt. MaxEnt can be thought of as a logistic regression model (Hayes & Wilson 2008, Johnson 2002). C4B constraint induction uses conditional inference trees. Logistic regression looks for main effects of features and all possible interactions, whereas conditional inference trees include a main effect of the most informative feature and then enter the less informative features if they help improve predictiveness within ANY VALUE of the more informative feature one by one. This makes conditional inference trees the tool of choice for cases in which there are more predictors than observations (Strobl et al. 2009). Such a situation arises in language learning when the lexicon is small: as the help files for MaxEnt indicate, MaxEnt cannot be run on lexicons with fewer than 3,000 distinct words. This required us to present the model with each word as if it were a set of homophones with the same number of meanings as the frequency of the word. A language with that many homophones must have extremely restrictive phonotactics, leading MaxEnt to be unduly conservative if trained on realistically detailed representations: every accidental gap looks real. Conditional inference trees could be run even on subsets of the original lexicon with stable predictions.

However, conditional inference trees are also known to miss (near-)perfect crossover interactions (Strobl et al. 2009). Thus, it is possible to come up with a more extreme version of the Tapa language in which common plural forms are equally likely to end in [–Dorsal]*a* and [+Dorsal]*i*. As a result, neither the [Dorsal] feature nor the vowel identity is informative, on its own, about whether the combination will occur in plurals. Thus, neither feature would be entered as a predictor into the tree, making the language unlearnable. This is a special case of the more general CONFIGURAL LEARNING PROBLEM where stimulus features need to be unitized into combinations that can then be associated with distinct responses or predictions to accurately perform the task (Goldstone 2000, Kapatsinski 2009b). What such units are is an open question, although see Kapatsinski 2009b for some evidence that they include syllabic constituents. Given that, as Bybee (2002) describes it, 'units used together fuse together', this unitization process must interact with schema induction, leading the trees to become smaller, bypassing decomposition of common words into their most elementary units.

Fourth, the types and loci of substantive biases brought by learners to the experiment require additional research. There are three kinds of biases directly encodable in C4B. First, learners might attempt to transfer schemas from their native language. For instance, two of our subjects (excluded from the analyses) hypercharacterized the plurals by attaching the English /z/ plural after the *-i* or *-a* suffix from the artificial language. Second, learners might use phonotactic constraints from their native language. Third, learners might perseverate on some chunks more than on others. For instance, in ongoing work (Kapatsinski 2012b), I have observed that subjects exposed to labial palatalization ($p \rightarrow t\int i$, $t \rightarrow ti$, $k \rightarrow ki$) completely overgeneralize it to alveolars and velars, either learning that everything palatalizes or that nothing does. Alveolar palatalization

(*t* → *tʃi*, *p* → *pi*, *k* → *ki*) completely overgeneralizes to velars, whereas velar palatalization (*k* → *tʃi*, *t* → *ti*, *p* → *pi*) generalizes to alveolars, and rarely labials, only incompletely. This can be captured by proposing that learners come to the experiment with a higher weight on '[p]!' than on '[t]!' and a higher weight on '[t]!' than on '[k]!'; thus as PL=…tʃi increases in strength over the course of training, its weight surpasses that of '[k]!' on the way to surpassing the weight of '[t]!' on the way to surpassing the weight of '[p]!' (cf. Howe & Pulleyblank 2004). However, C4B currently has no way of encoding a bias in favor of certain changes in certain contexts; thus, for example, the bias against labial palatalization is expected to be trigger-independent, holding in front of [a] just as well as in front of [i]. Whether this theory of inductive bias is overly constrained remains to be tested.

Finally, different kinds of experience might lead to different types of generalizations. For instance, production experience might be necessary to learn that certain sound combinations are hard to produce, yielding markedness constraints. Such constraints appear necessary in cases of production-driven avoidance, which is well documented in the literature on first language acquisition, where children avoid attempting to produce words containing sounds and sound sequences they have not yet mastered when a semantically similar alternative is available (Schwarz & Leonard 1982, Schwarz et al. 1987, Storkel 2001). Martin (2007) relates these effects to feedback in Dell's (1986) interactive model of speech production: when a word is difficult to produce, this difficulty percolates from articulatory planning and execution back up to the lexical level, lowering resting activation levels of words that are difficult to produce. Thus when those words compete for selection with easier-to-produce alternatives, they are likely to lose the competition. This account is easily expanded to sublexical schemas, where difficult-to-produce structures would be penalized by feedback from articulatory planning and execution and be likely to lose the competition to easier-to-articulate alternatives with similar meanings. Given Martin's (2007) account of these effects, however, production experience should be necessary to learn to avoid that structure. By considering what KINDS of experiences lead to a certain kind of linguistic generalization, we would likely gain a better understanding of the roles different kinds of generalizations play in the grammar and how/when they are acquired.

4. CONCLUSION. I consider this work to be a contribution to developing a learning-theoretic (Hayes & Wilson 2008) or usage-based (Bybee 2001) phonology. The ultimate goal of this enterprise I take to be to describe the learning abilities and biases of human learners so that we would ultimately be able to predict for a given learner which beliefs about the phonology of his/her language will lose strength and which will gain strength as a result of certain perceptual and production experiences. Here I focused on the effect of experiencing examples of simple suffix addition that happen to result in a sound sequence that can also result from suffix addition followed by a stem change. I found that such examples increase the productivity of the stem change and showed that this result is incompatible with the proposal that phonology is acquired by making generalizations about changes in context (rules). Importantly, this result is shown here even under presentation conditions that facilitate comparing morphologically related forms and viewing one of these forms as derived from the other, providing the prerequisites for rule instruction (see Kapatsinski 2012a for even stronger results under more natural presentation conditions).

The results were then shown to be explainable if learners are extracting form-meaning correspondences, noting the common characteristics of forms sharing mean-

ing (in this case, plural forms). I proposed that they learn a set of partially redundant descriptions of typical forms that are semantically similar (schemas/constructions). Schema induction is a conditional inference process: the most salient and common characteristics of forms with a certain meaning are extracted first, and less salient and common characteristics are extracted and used only if they help improve the description once the 'better' characteristics are taken into account. However, as knowledge is built up, these characteristics fuse together, becoming unitary schemas. Schemas are stronger when they are exemplified by many words and when they are maximally specific. Chunks are stronger if they are usually shared by morphologically related forms. These proposals together form a theory of the learning mechanism behind the acquisition of phonology from the kind of experience provided to learners in the present study: namely, auditory and production exposure to words while learning word meanings. Other kinds of experience—in particular, experience in deriving novel forms of known words—might lead to different kinds of generalizations. Other kinds of learners—in particular, ones that do not have a notion of 'plural'—might also behave differently. I believe that future work in learning-theoretic phonology should examine these 'performance' factors to determine what kinds of experience lead to the formation of the various kinds of generalizations we see represented in productive phonologies of natural languages. I believe that taking these performance factors into account constrains a theory of phonological competence. For instance, rule extraction requires a low-error form-comparison process that can compare morphologically related words and analyze them into a change and context. If we can then show that the required comparison process places high demands on working memory and as a result is prone to error, a rule-based theory of phonological competence loses its appeal, since the proposed form of phonological knowledge cannot be obtained, given human learners' limitations in tracking the requisite environmental contingencies.

The proposed theory of phonological grammar is also connected to the observed elicited production data. An explicit connection of this kind is necessary for evaluating the descriptive adequacy of a phonological theory. Here the goal is to provide an explicit account of linguistic, and specifically morphophonological, creativity: we want to know what speakers do when they produce a novel form of a word from a form they know. I have argued that learners attempt to (i) reuse as much of the known form as they can (perseverating on the input chunks) and (ii) produce the most common characteristics of forms with the to-be-expressed meaning (fitting the schemas associated with the meaning being produced). When chunks and schemas are in competition, they are used to derive candidate output forms. The actual output is selected from the set of competing candidate outputs stochastically, with the probability of selecting an output proportional to the summed strength of chunks and schemas that have produced that candidate. Thus speakers have a tendency to level stem changes, reusing too much of the known word form, and this tendency is stronger early in the learning process when the schemas are still weak and thus the learner is likely to accept a wider variety of output forms. Perseveration is fundamentally a 'performance' factor; thus the perseveratory tendencies observed are likely to be task-specific (dominating production, and especially rapid on-the-fly production, but not perception or selection among alternatives, as shown by, for example, Kapatsinski 2012a, Mitrovic 2012, and Zuraw 2000) and subject to individual variation: some learners will be more likely to perseverate on the input than others.

In conclusion, I would like to call for working toward a truly learning-theoretic or usage-based phonology, in which the connection between experience, prior bias, induc-

tion of generalizations, and the use of these generalizations in production and perception is made fully explicit. I would also like to caution against overinterpreting the results of artificial-grammar learning experiments as being about learnability or nonlearnability of various kinds of generalizations. What has been shown is that certain types of generalizations are EASIER to learn than others GIVEN A CERTAIN TYPE OF EXPERIENCE. We can never give learners in experiments enough experience, and enough experience of the right kind, to approach the complexity of natural language phonology. Thus, to paraphrase Shakespeare, there will always be more patterns out there than are dreamt of in your favorite theory of phonology (e.g. as Ohala (1978) points out, Southern Bantu does have labial palatalization without velar palatalization, before, of all things, [w]). A better way to think of this kind of experimental research is as a way of determining what kinds of experience are necessary for the types of generalizations we see active in natural language phonologies to be formed. I believe that great progress toward understanding of the phonological grammar can be made by a systematic examination of the types of experience that lead to formation of various kinds of phonological generalizations and the integration of these generalizations into an explicit theory of phonology whose validity can be tested by confronting it with data about the time course of learning and wug tests of pattern productivity from both artificial and natural language.

APPENDIX: STIMULI

Stimuli presented during training by language. Bolded stimuli were frequent in training.

ALL	TAPA, TAPATʃI	TIPI, TIPITʃI	TAPATʃI, TIPITʃI	TAPATʃU, TIPITʃU
blaɪk → blaɪtʃi	blort → blorti	blort → blorta	bortʃ → bortʃi	bortʃ → bortʃu
truk → trutʃi	hit → hita	hit → hiti	dwitʃ → dwitʃi	dwitʃ → dwitʃu
swik → switʃi	**ðaɪt → ðaɪta**	**ðaɪt → ðaɪti**	frutʃ → frutʃi	frutʃ → frutʃu
vork → vortʃi	flort → florta	flort → florti	**slaɪtʃ → slaɪtʃi**	**slaɪtʃ → slaɪtʃu**
	bup → bupa	**bup → bupi**		
	floʊp → floʊpa	floʊp → floʊpi		
	gwip → gwipa	gwip → gwipi		
	klup → klupi	klup → klupa		

Elicited production stimuli. Minimal pairs are introduced in this stage to ensure that subjects focus on the crucial stem-final consonant.

{blaɪ;swi;tru;vor} {t;p;tʃ}
{blor;flor;hi;ðaɪ} {k;p;tʃ}
{bu;floʊ;gwi;klu} {k;t;tʃ}
{bor;dwi;slaɪ} {k;t;p}
{fl;kr;w;skl} {a;i}k

REFERENCES

ABBOT-SMITH, KIRSTEN, and HEIKE BEHRENS. 2006. How known constructions influence the acquisition of other constructions: The German passive and future constructions. *Cognitive Science* 30.995–1026.

ALBRIGHT, ADAM. 2009. Feature-based generalisation as a source of gradient acceptability. *Phonology* 26.9–41.

ALBRIGHT, ADAM, and BRUCE HAYES. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90.119–61.

ALTMANN, GERRY T. M.; ZOLTÁN DIENES; and ALASTAIR GOODE. 1995. Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21.899–912.

AMBRIDGE, BEN, and ELENA V. M. LIEVEN. 2011. *Child language acquisition: Contrasting theoretical viewpoints*. Cambridge: Cambridge University Press.

ASLIN, RICHARD N.; JENNY R. SAFFRAN; and ELISSA N. NEWPORT. 1998. Computation of conditional probability statistics by 8-month-old infants. *Psychological Science* 9.321–24.

BERENT, IRIS; DONCA STERIADE; TRACY LENNERTZ; and VERED VAKNIN. 2007. What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* 104. 591–630.

BERGEN, BENJAMIN K. 2004. The psychological reality of phonaesthemes. *Language* 80. 290–311.

BERKO, JEAN. 1958. The child's learning of English morphology. *Word* 14.150–77.

BOERSMA, PAUL. 2004. A stochastic OT account of paralinguistic tasks such as grammaticality and prototypicality judgments. Amsterdam: University of Amsterdam, MS. Online: http://www.fon.hum.uva.nl/paul/papers/ParalinguisticTasks.pdf.

BOERSMA, PAUL, and DAVID WEENINK. 2009. Praat: Doing phonetics by computer (version 5.1.07). Online: http://www.praat.org/.

BOOIJ, GEERT. 2008. Paradigmatic morphology. *La raison morphologique: Hommage á la memoire de Danielle Corbin*, ed. by Bernard Fradin, 29–38. Amsterdam: John Benjamins.

BOOIJ, GEERT. 2010. *Construction morphology*. Oxford: Oxford University Press.

BRAINE, MARTIN D. 1987. What is learned in acquiring word classes—A step toward an acquisition theory. *Mechanisms of language acquisition*, ed. by Brian MacWhinney, 65–87. Hillsdale, NJ: Lawrence Erlbaum.

BRAINE, MARTIN D.; RUTH E. BRODY; PATRICIA J. BROOKS; VICKI SUDHALTER; JULIE A. ROSS; LISA CATALANO; and SHALOM 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.

BROOKS, PATRICIA J.; MARTIN D. S. BRAINE; LISA CATALANO; RUTH E. BRODY; and VICKI 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.

BROWMAN, CATHERINE P., and LOUIS GOLDSTEIN. 1989. Articulatory gestures as phonological units. *Phonology* 6.201–51.

BROWMAN, CATHERINE P., and LOUIS GOLDSTEIN. 1991. Tiers in articulatory phonology, with some implications for casual speech. *Papers in laboratory phonology 1: Between the grammar and the physics of speech*, ed. by John Kingston and Mary E. Beckman, 341–76. Cambridge: Cambridge University Press.

BYBEE, JOAN L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.

BYBEE, JOAN L. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

BYBEE, JOAN L. 2002. Sequentiality as the basis of constituent structure. *The evolution of language out of pre-language*, ed. by Talmy Givón and Bertram F. Malle, 109–34. Amsterdam: John Benjamins.

BYBEE, JOAN L. 2003. Cognitive processes in grammaticalization. *The new psychology of language: Cognitive and functional approaches to language structure*, vol. 2, ed. by Michael Tomasello, 145–68. Mahwah, NJ: Lawrence Erlbaum.

BYBEE, JOAN L., and CAROL LYNN MODER. 1983. Morphological classes as natural categories. *Language* 59.251–70.

BYBEE, JOAN L., and DAN I. SLOBIN. 1982. Rules and schemas in the development and use of the English past. *Language* 58.265–89.

CAPPELLE, BERT. 2006. Particle placement and the case for 'allostructions'. *Constructions*, special volume 1. Online: http://elanguage.net/journals/constructions/.

CHARLES-LUCE, JAN, and PAUL A. LUCE. 1990. Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language* 17.205–15.

CHOMSKY, NOAM, and MORRIS HALLE. 1968. *The sound pattern of English*. New York: Harper and Row.

CORBIN, DANIELLE. 1989. Form, structure and meaning of constructed words in an associative and stratified lexical component. *Yearbook of Morphology* 1989.31–54.

DAELEMANS, WALTER, and ANTAL VAN DEN BOSCH. 2005. *Memory-based language processing*. Cambridge: Cambridge University Press.

DELL, GARY S. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93.283–321.

DELL, GARY S.; LISA K. BURGER; and WILLIAM R. SVEC. 1997. Language production and serial order: A functional analysis and a model. *Psychological Review* 104.123–47.

ELLIS, NICK C., and FERNANDO FERREIRA-JUNIOR. 2009. Constructions and their acquisition: Islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics* 7.187–220.

ERNESTUS, MIRJAM, and R. HARALD BAAYEN. 2011. Corpora and exemplars in phonology. *The handbook of phonological theory*, 2nd edn., ed. by John Goldsmith, Jason Riggle, and Alan C. L. Yu, 374–400. Malden, MA: Wiley-Blackwell.

FARRAR, MICHAEL J. 1992. Negative evidence and grammatical morpheme acquisition. *Developmental Psychology* 28.90–98.

FRIGO, LENORE, and JANET L. MCDONALD. 1998. Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language* 39.218–45.

FRISCH, STEFAN A.; JANET B. PIERREHUMBERT; and MICHAEL B. BROE. 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22.179–228.

FROMKIN, VICTORIA A. 1970. The non-anomalous nature of anomalous utterances. *Language* 46.27–52.

GAFOS, ADAMANTIOS I. 2002. A grammar of gestural coordination. *Natural Language and Linguistic Theory* 20.269–337.

GERKEN, LOUANN; RACHEL WILSON; and WILLIAM LEWIS. 2005. Infants can use distributional cues to form syntactic categories. *Journal of Child Language* 32.249–68.

GOLDBERG, ADELE E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

GOLDBERG, ADELE E. 2002. Surface generalizations: An alternative to alternations. *Cognitive Linguistics* 13.327–56.

GOLDRICK, MATTHEW. 2004. Phonological features and phonotactic constraints in speech production. *Journal of Memory and Language* 51.586–603.

GOLDSTEIN, LOUIS; MARIANNE POUPLIER; LARISSA CHEN; ELLIOT SALTZMAN; and DANI BIRD. 2007. Dynamic action units slip in speech production errors. *Cognition* 103.386–412.

GOLDSTONE, ROBERT L. 2000. Unitization during category learning. *Journal of Experimental Psychology: Human Perception and Performance* 26.86–112.

GOLDWATER, SHARON, and MARK JOHNSON. 2003. Learning OT constraint rankings using a maximum entropy model. *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, ed. by Jennifer Spenader, Anders Eriksson, and Östen Dahl, 111–20. Stockholm: Stockholm University, Department of Linguistics.

HARTSUIKER, ROBERT J. 2002. The addition bias in Dutch and Spanish phonological speech errors: The role of structural context. *Language and Cognitive Processes* 17.61–96.

HAYES, BRUCE. 2004. Phonological acquisition in optimality theory: The early stages. *Fixing priorities: Constraints in phonological acquisition*, ed. by René Kager, Joe Pater, and Wim Zonneveld, 158–203. Cambridge: Cambridge University Press.

HAYES, BRUCE. 2009. *Introductory phonology*. Malden, MA: Wiley-Blackwell.

HAYES, BRUCE. 2011. Interpreting sonority-projection experiments: The role of phonotactic modeling. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)*, Hong Kong, 835–38. Online: http://www.icphs2011.hk/ICPHS_CongressProceedings.htm.

HAYES, BRUCE, and COLIN WILSON. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39.379–440.

HAYES, BRUCE; KIE ZURAW; PÉTER SIPTÁR; and ZSUZSA LONDE. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85.822–63.

HOTHORN, TORSTEN; KURT HORNIK; and ACHIM ZEILEIS. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* 15.651–74.

HOWE, DARIN, and DOUGLAS PULLEYBLANK. 2004. Harmonic scales as faithfulness. *Canadian Journal of Linguistics* 49.1–49.

JÄGER, GERHARD. 2007. Maximum entropy models and stochastic optimality theory. *Architectures, rules, and preferences: A festschrift for Joan Bresnan*, ed. by Jane Grimshaw, Joan Maling, Chris Manning, Jane Simpson, and Annie Zaenen, 467–79. Stanford, CA: CSLI Publications.

JOHNSON, MARK. 1984. A discovery procedure for certain phonological rules. *Proceedings of the International Conference on Computational Linguistics (COLING '84)* 10.344–47.

JOHNSON, MARK. 2002. Optimality-theoretic lexical functional grammar. *The lexical basis of sentence processing: Formal, computational and experimental issues*, ed. by Suzanne Stevenson and Paolo Merlo, 59–73. Amsterdam: John Benjamins.

JOHNSTON, LAMIA HADDAD, and VSEVOLOD KAPATSINSKI. 2011. In the beginning there were the weird: A phonotactic novelty preference in adult word learning. *Proceedings of the*

17th International Congress of Phonetic Sciences (ICPhS), Hong Kong, 978–81. Online: http://www.icphs2011.hk/ICPHS_CongressProceedings.htm.

KAGER, RENÉ. 1999. *Optimality theory*. Cambridge: Cambridge University Press.

KANG, YOONJUNG. 2011. Loanword phonology. *The Blackwell companion to phonology, vol. 4: Phonological interfaces*, ed. by Mark van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice, 2258–82. Malden, MA: Wiley-Blackwell.

KAPATSINSKI, VSEVOLOD. 2005. Productivity of Russian stem extensions: Evidence for and a formalization of network theory. Albuquerque: University of New Mexico master's thesis.

KAPATSINSKI, VSEVOLOD. 2006. To scheme or to rule: Evidence against the dual-mechanism model. *Berkeley Linguistics Society* 31.193–204.

KAPATSINSKI, VSEVOLOD. 2009a. *The architecture of grammar in artificial grammar learning: Formal biases in the acquisition of morphophonology and the nature of the learning task*. Bloomington: Indiana University dissertation.

KAPATSINSKI, VSEVOLOD. 2009b. Testing theories of linguistic constituency with configural learning: The case of the English syllable. *Language* 85.248–77.

KAPATSINSKI, VSEVOLOD. 2012a. What statistics do learners track? Rules, constraints and schemas in (artificial) language learning. *Frequency effects in language: Learning and processing*, ed. by Stefan Th. Gries and Dagmar Divjak, 53–82. Berlin: Mouton de Gruyter.

KAPATSINSKI, VSEVOLOD. 2012b. Chunk-schema competition in deriving new forms of known words. Paper presented at the American International Morphology Meeting, Amherst, MA. Online: http://pages.uoregon.edu/vkapatsi/AIMM_C4B_Sept.pdf.

KEIL, FRANK C. 1979. *Semantic and conceptual development: An ontological perspective*. Cambridge, MA: Harvard University Press.

KENSTOWICZ, MICHAEL. 1994. Syllabification in Chukchee: A constraints-based analysis. *Proceedings of the Formal Linguistics Society of the Midwest* 4.160–81.

KENSTOWICZ, MICHAEL. 1996. Base-identity and uniform exponence: Alternatives to cyclicity. *Current trends in phonology: Models and methods*, ed. by Jacques Durand and Bernard Laks, 363–93. Salford: European Studies Research Institute and University of Salford.

KEULEERS, EMMANUEL. 2008. *Memory-based learning of inflectional morphology*. Antwerp: University of Antwerp dissertation.

KISSEBERTH, CHARLES. 1970. On the functional unity of phonological rules. *Linguistic Inquiry* 1.291–306.

KOCHETOV, ALEXEI. 2011. Palatalization. *The Blackwell companion to phonology, vol. 3: Phonological processes*, ed. by Mark van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice, 1666–90. Malden, MA: Wiley-Blackwell.

KÖPCKE, KLAUS-MICHAEL. 1988. Schemas in German plural formation. *Lingua* 74.303–35.

LANGACKER, RONALD W. 1987. *Foundations of cognitive grammar, vol. 1: Theoretical prerequisites*. Stanford, CA: Stanford University Press.

LEGENDRE, GERALDINE; YOSHIRO MIYATA; and PAUL SMOLENSKY. 1990. Harmonic grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. *Proceedings of the 12th annual conference of the Cognitive Science Society*, 388–95.

LOBBEN, MARIT. 1991. Pluralization of Hausa nouns, viewed from psycholinguistic experiments and child language data. Oslo: University of Oslo master's thesis.

MANDLER, JEAN M. 2000. Perceptual and conceptual processes in infancy. *Journal of Cognition and Development* 1.3–36.

MARCUS, G. F.; S. VIJAYAN; S. BANDI RAO; and P. M. VISHTON. 1999. Rule learning by seven-month-old infants. *Science* 283.77–80.

MARTIN, ANDREW. 2007. *The evolving lexicon*. Los Angeles: University of California, Los Angeles dissertation.

MASSARO, DOMINIC W. 1970. Retroactive interference in short-term recognition memory for pitch. *Journal of Experimental Psychology* 83.32–39.

MCCLELLAND, JAMES L.; BRUCE L. MCNAUGHTON; and RANDALL C. O'REILLY. 1995. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review* 102.419–57.

McLennan, Conor T., and Paul A. Luce. 2005. Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31.306–21.
Menn, Lise, and Brian MacWhinney. 1984. The repeated morph constraint: Toward an explanation. *Language* 60.519–41.
Mitroff, Stephen R.; Daniel J. Simons; and Daniel T. Levin. 2004. Nothing compares 2 views: Change blindness can occur despite preserved access to the changed information. *Perception and Psychophysics* 66.1268–81.
Mitrovic, Ivana. 2012. A phonetically natural vs. native language pattern: An experimental study of velar palatalization in Serbian. *Journal of Slavic Linguistics* 20.229–68.
Moreton, Elliott. 2002. Structural constraints in the perception of English stop-sonorant clusters. *Cognition* 84.55–71.
Nesset, Tore. 2005. Opaque softening: A usage-based approach. *Poljarnyj Vestnik* 8.55–68.
Nesset, Tore. 2008. *Abstract phonology in a concrete model: Cognitive linguistics and the morphology-phonology interface*. Berlin: Mouton de Gruyter.
Nesset, Tore. 2010. Why not? Prototypes and blocking of language change in Russian verbs. *Cognitive linguistics in action: From theory to application and back*, ed. by Elżbieta Tabakowska, Michal Choińsky, and Ukasz Wiraszka, 125–44. Berlin: Mouton de Gruyter.
Ohala, John J. 1978. Southern Bantu vs. the world: The case of palatalization of labials. *Berkeley Linguistics Society* 4.370–86.
Paradis, Carole, and Darlene LaCharité. 1997. Preservation and minimality in loanword adaptation. *Journal of Linguistics* 33.379–430.
Pauen, Sabina. 2002. The global-to-basic shift in infants' categorical thinking: First evidence from a longitudinal study. *International Journal of Behavioral Development* 26. 492–99.
Peperkamp, Sharon. 2003. Phonological acquisition: Recent attainments and new challenges. *Language and Speech* 46.97–113.
Pierrehumbert, Janet B. 1993. Dissimilarity in the Arabic verbal roots. *North East Linguistic Society (NELS)* 23.367–81.
Pierrehumbert, Janet B. 2006. The statistical basis of an unnatural alternation. *Laboratory phonology 8: Varieties of phonological competence*, ed. by Louis Goldstein, D. H. Whalen, and Catherine Best, 81–107. Berlin: Mouton de Gruyter.
Pinker, Steven, and Alan Prince. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28.73–193.
Plag, Ingo. 2003. *Word formation in English*. Cambridge: Cambridge University Press.
Prince, Alan, and Paul Smolensky. 2004 [1993]. *Optimality theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell.
R Development Core Team. 2009. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Online: http://www.R-project.org.
Reiss, Charles. 2004. Constraining the learning path without constraints, or the OCP and NoBanana. *Rules, constraints, and phonological phenomena*, ed. by Bert Vaux and Andrew Nevins, 252–301. Oxford: Oxford University Press.
Rogers, Timothy T., and James L. McClelland. 2004. *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.
Rubino, Rejane B., and Julian M. Pine. 1998. Subject-verb agreement in Brazilian Portuguese: What low error rates hide. *Journal of Child Language* 25.35–59.
Rumelhart, David E., and James L. McClelland. 1986. On learning the past tenses of English verbs. *Parallel distributed processing, vol. 2: Psychological and biological models*, ed. by David E. Rumelhart and James L. McClelland, 216–71. Cambridge, MA: MIT Press.
Schwarz, Richard G., and Lawrence Leonard. 1982. Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language* 9.319–36.
Schwarz, Richard G.; Lawrence Leonard; Diane M. Frome Loeb; and Lori A. Swanson. 1987. Attempted sounds are sometimes not: An expanded view of phonological selection and avoidance. *Journal of Child Language* 14.411–18.
Shattuck-Hufnagel, Stefanie, and Dennis H. Klatt. 1979. The limited use of distinctive features and markedness in phonological speech errors. *Journal of Verbal Learning and Verbal Behavior* 18.41–55.

SHVACHKIN, N. KH. 1973 [1948]. The development of phonemic speech perception in early childhood. *Studies of child language development*, ed. by Charles A. Ferguson and Dan I. Slobin, 91–127. New York: Holt, Rinehart, and Winston.
SIMONS, DANIEL J. 1996. In sight, out of mind: When object representations fail. *Psychological Science* 7.301–5.
SMITH, LINDA B.; ESTHER THELEN; ROBERT TITZER; and DEWEY MCLIN. 1999. Knowing in the context of acting: The task dynamics of the A-not-B error. *Psychological Review* 106.235–60.
SMOLENSKY, PAUL, and GERALDINE LEGENDRE. 2006. *The harmonic mind: From neural computation to optimality-theoretic grammar*. Cambridge, MA: MIT Press.
STAGER, CHRISTINE L., and JANET F. WERKER. 1997. Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature* 388.381–82.
STEMBERGER, JOSEPH PAUL. 1981. Morphological haplology. *Language* 57.791–817.
STEMBERGER, JOSEPH PAUL. 1982. The nature of segments in the lexicon: Evidence from speech errors. *Lingua* 56.235–59.
STEMBERGER, JOSEPH PAUL. 1991. Apparent anti-frequency effects in language production: The addition bias and phonological underspecification. *Journal of Memory and Language* 30.161–85.
STORKEL, HOLLY L. 2001. Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research* 44.1321–37.
STROBL, CAROLIN; JAMES MALLEY; and GERHARD TUTZ. 2009. An introduction to recursive partitioning: Rationale, application and characteristics of classification and regression trees, bagging and random forests. *Psychological Methods* 14.323–48.
SWINGLEY, DANIEL. 2007. Lexical exposure and word-form encoding in 1.5-year-olds. *Developmental Science* 43.454–64.
SWINGLEY, DANIEL, and RICHARD N. ASLIN. 2007. Lexical competition in young children's word learning. *Cognitive Psychology* 54.99–132.
VALIAN, VIRGINIA, and SHAWNA COULSON. 1988. Anchor points in language learning: The role of marker frequency. *Journal of Memory and Language* 27.71–86.
WANG, H. SAMUEL, and BRUCE L. DERWING. 1994. Some vowel schemas in three English morphological classes: Experimental evidence. *In honor of Professor William S.-Y. Wang: Interdisciplinary studies on language and language change*, ed. by Matthew Y. Chen and Ovid C. L. Tseng, 561–75. Taipei: Pyramid.
WARRINGTON, ELIZABETH K. 1975. Selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology* 27.635–57.
WEINERT, SABINE. 2009. Implicit and explicit modes of learning: Similarities and differences from a developmental perspective. *Linguistics* 47.241–71.
WILLIAMS, JOHN N. 2003. Inducing abstract linguistic representations: Human and connectionist learning of noun classes. *The lexicon-syntax interface in second language acquisition*, ed. by Roeland van Hout, Aafke Hulk, Folkert Kuiken, and Richard J. Towell, 151–74. Amsterdam: John Benjamins.
WILSON, COLIN. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30.945–82.
XU, FEI, and JOSHUA B. TENENBAUM. 2007. Word learning as Bayesian inference. *Psychological Review* 114.245–72.
ZURAW, KIE. 2000. *Patterned exceptions in phonology*. Los Angeles: University of California, Los Angeles dissertation.

Department of Linguistics
1290 University of Oregon
Eugene, OR 97403
[vkapatsi@uoregon.edu]

[Received 25 September 2009;
revision invited 20 July 2010;
revision received 23 March 2011;
revision invited 8 October 2011;
revision received 14 June 2012;
accepted 7 December 2012]