Clamoring for Blending: Usage-Based Harmonic Morphophonology

Vsevolod Kapatsinski
University of Oregon

Draft. 2/19/2012

Abstract

The present paper presents Clamoring for Blending (C4B), a formally explicit usage-based theory of morphophonology. Clamoring for Blending proposes that the derivation of an unknown form from a known morphologically-related form is a process of selecting between candidate outputs (Nesset 2008, Prince & Smolensky 1993 / 2004) that are produced and supported by perseverating chunks extracted from the known form and schemas associated with the meaning of the to-be-produced cell in the morphological paradigm (Bybee 1985, 2001). Clamoring for Blending provides an explicit algorithm for extracting product-oriented generalizations / first-order schemas (Bybee 1985, 2001, Bybee & Slobin 1982, Langacker 1987, Nesset 2008) from the lexicon (by means of conditional inference trees) and an explicit mechanism for conflict resolution between competing schemas and chunks (formally equivalent to constraint interaction in Harmonic Grammar, Legendre et al. 1990). We will show how the model accounts for a complex set of data on learning miniature artificial languages in the lab presented in Kapatsinski (2010a, 2010b, in press) and showcase its ability to make interesting predictions for language learning, language change, and language processing. While many of these predictions remain untested, the ones that have been tested are, we suggest, strongly supported by the available data.

1. The goals of the theory

The current theory is largely based on data from language learning in the lab. The advantage of such data is that the human learners and the model can 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 *un*learnable. We can only show that some generalization *is* learnable or that it is learned more easily (from the kind of experience we provided the learners with) than some other generalization.

The theory to be developed is to be held accountable for describing both generalizations that are learnable in the lab and those that are learnable only outside of the lab, for which natural language evidence is provided, and predicting the observed differences in learnability. Given that natural languages feature many "unnatural" productive processes (e.g., Bye 2008, Hayes et al. 2009, Kapatsinski 2010a, 2010b, Mielke 2008, Nevins 2011, Ohala 1978, Paster 2006, Pierrehumbert 2006), it seems prudent to avoid postulating hard biases predicting certain patterns to be unlearnable: a language

exhibiting a pattern that is absent from the typological data at present could always be discovered tomorrow. Rather, I take the ultimate goal of grammatical theory to be predicting 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. In other words, the goal we are pursuing is to define the relationship between usage and grammar (cf. Bybee 2005, Newmeyer 2006).

2. Summary of the data

The learners in Kapatsinski (2010a, 2010b, in press) were adult native English speakers who were exposed to one of a number of miniature artificial languages.[1] All languages featured two plural suffixes –i and –a. In Experiment 1, –i was the only suffix that attached to singulars ending in [k] and always changed the [k] into [tʃ]. There were always more examples of k→tʃi than of any other singular-plural mapping. Both –i and –a could attach to singulars ending in [t] or [p]. There were four languages: Tapa, Tipi, Tapachi, and Tipichi. In Tapa and Tapachi, -i attached to singulars ending in [t] or [p] less often than –a did: 25% -i, 75% -a. The proportions were reversed in Tipichi and Tapachi. Tapachi and Tipichi differed from Tapa and Tipi in featuring examples of –i simply attaching to a [tʃ]-final singular. The languages are shown in (1).

(1) The four types of languages presented to learners (lexicon size varied)

	Tapa	Tipi	Tapachi	Tipichi
{k;g} → {tʃ;dʒ}i	M			
{t;d;p;b} → {t;d;p;b}i	N	3N	N	3N
{t;d;p;b} → {t;d;p;b}a	3N	N	3N	N
{tʃ;dʒ} → {tʃ;dʒ}i	0		K	

The training consisted of hearing singular-plural pairs, as shown in (2), or singulars and plurals in random order paired with pictures of referents appearing on the screen. In either case, subjects were asked to repeat the words after hearing them and went through a cued recall test halfway through training, in which they were asked for the names of the objects presented on screen. The right wordform (either singular or plural) had to be produced.

[1] We take the differences between children and adults with respect to learning morphophonology, given the same kind of experience, to be quantitative in nature and thus capturable by differences in values of free parameters discussed below. All of the proposed mechanisms are assumed to apply to both children and adults.

(2)

Video:						
Audio:		[boʊk]			[boʊtʃi]	
Learner action:	Watch	Watch and listen		Watch	Watch and listen	Repeat aloud, then click
Duration:	300 ms	500-900 ms	500 ms	300 ms	500-900 ms	0-10s

The cued recall stage was followed by more training, which was followed by an elicited production test. The learners were presented with novel singulars and had to say the plural forms (as in (2) but without presentation of the plural form). We analyzed how productive various types of singular-plural mappings (e.g., k→tʃi) were in this elicited production test. This was followed by an acceptability judgment test, in which the learners rated singular-plural pairs, some legal, some illegal, on "how likely this plural is to be the right one for this singular?". We used a seven-point scale with an untimed button-press response, from "impossible" to "very likely". The results were as follows:

(3) Velar palatalization partially overgeneralizes to alveolars but not labials.

(4) –a is overgeneralized to [k]-final and [tʃ]-final singulars (but less to the latter).

(5) In repetition of singular-plural pairs during training, subjects often level stem changes (e.g., erroneously repeating [buk butʃi] as [butʃ butʃi] or (much less commonly) [buk buki].

(6) In recall, erroneous retention of the plural suffix in a singular form is much more common than erroneous omission of a suffix in the plural form

(7) In elicited production, subjects more commonly produce p→ptʃi than p→tʃi. The p→ptʃi errors are very rare occur in repetition during training.

(8) In elicited production, subjects more commonly produce k→tʃi than k→ktʃi.

(9) The k→ktʃi and p→ptʃi errors are very rare in repetition during training.

(10) Examples of tʃ→tʃi support t→tʃi and, to a lesser extent, k→tʃi. Numerically, they also support p→tʃi.

(11) Replacing examples of p→pa and t→ta with examples of p→pi and t→ti (Tapa(chi) vs. Tipi(chi)) helps k→ki over k→tʃi and k→ka.

(12) Temporal adjacency of corresponding singulars and plurals (bup bupi, luk lutʃi, …) hurts stem changes.

(13) Elicited production of a plural given a singular differs from acceptability rating of the singular-plural mapping ("how likely is this plural to be the right one for this singular?") in that stem changes are disfavored by elicited production compared to rating, and

(14) Singular-plural mappings sharing the suffix and differing in the presence/absence of a stem change help each other in rating but compete in elicited production.

(15) Cued recall performance does not correlate with the other tests: one may be unable to recall any examples of k→tʃi but use the mapping productively.

Experiment 2 exposed learners to either velar, alveolar, or labial palatalization. The training was the same as in Experiment 1, except singulars and plurals were only presented in random order: we dropped the "singulars and plurals next to each other" condition. Thus, learners were exposed to either k→tʃi, t→t{i,a}, p→p{i,a} (velar palatalization); t→tʃi, p→p{i,a}, k→k{i,a} (alveolar palatalization); or p→tʃ{i;a}, t→t{i,a}, p→p{i,a} (labial palatalization). The elicited production task was administered but not the rating task. This experiment showed that

(16) Subjects come to the experiment more willing to learn velar palatalization, less willing to learn alveolar palatalization and very unwilling to learn labial palatalization. Velar palatalization overgeneralizes incompletely to alveolars and does not overgeneralize to labials, alveolar palatalization overgeneralizes to velars but not labials, and labial palatalization overgeneralizes to both alveolars and velars.

(17) Subjects produce p→ptʃi when exposed to p→tʃi more often than they produce k→ktʃi after being exposed to k→tʃi.

In Experiment 3, we told subjects that there are two plural suffixes –i and –a and ran them through the elicited production test without training.

(18) Without training, the learners exhibited random (50/50) guessing between –i and –a regardless of the final consonant of the singular.

3. The ingredients of grammar: Chunks and schemas

What kind of model can account for the results? I suggest that when the learner encounters a singular form of a novel word and is asked to produce the plural, s/he internally generates a number of candidate plurals. The candidates are formed by a process of *clamoring for blending* where first-order product-oriented plural schemas (Bybee 1985, 2001) attempt to impose themselves on the product and are sometimes in competition with chunks from the singular form.

Product-oriented schemas are the most natural way of accounting for the finding that tʃ#→tʃi# helps other mappings resulting in tʃi#. Independent support for product-oriented schemas comes from "wig tests", in which subjects are often observed to overuse common output patterns, deriving them in ways unattested in the lexicon (Albright & Hayes 2003, Bybee & Slobin 1982, Köpcke 1988, Lobben 1991, Wang and 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 and MacWhinney 1984, Nesset 2010, Stemberger 1981). For instance, Bybee & 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 a 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', e.g., '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*.

Chunks are relatively independent processing units that can change their positions in speech errors. They thus include non-meaningful 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. We suggest that when a speaker has to produce a novel wordform from a known morphologically-related wordform, chunks of that known wordform are subject to perseveration. This perseveration is largely functional, as *most* of the to-be-produced unknown form *should* come from the known form but can overapply, resulting in the product resembling the source *too much*. Chunks are often schemas but do not have to be: a non-recurrent, rare gesture might be able to persevere even more strongly than a common, well-practiced one (Stemberger 1991) since priming is more effective for rare and especially novel units (Stark & McClelland 2000).

In our data, overapplication of perseveration is common in both repetition during training on singular-plural pairs and elicited production. In repetition of singular-plural pairs during training, perseverations level stem changes (e.g., [bik bitʃi] repeated as [bik biki]). In elicited production, unlike in repetition during training, product-oriented schemas are active and are in competition with persevering chunks. The product-oriented schemas root most strongly for Vtʃi-final plurals, the most common type of plural in the language. However, sometimes an input chunk is perseverated, 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). 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. However, it should not be *as* relevant as for elicited production, making stem changes more productive in rating than in elicited production. This differences 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 (as documented by Kapatsinski in press, 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, see Section 6 for some complications). The strength of an input chunk is the tendency for that chunk to be perseverated on, and is equivalent to a conjunction of chunk-specific output-oriented Max and Ident constraints in OT / HG (Kenstowicz 1996). We shall refer to these constraints with using the notation "[chunk]!", e.g., "[p]!" to indicate that one should output [p] if one is present in the input.[2]

The observed substantive bias against alveolar and especially labial palatalization can be encoded in relative strengths of [p], [t], and [k] prior to training, i.e., "[p]!" >> "[t]!" >> "[k]!" (Howe & Pulleyblank 2004 for Optimality Theory;[3] Saltzman & Munhall 1989 for Articulatory Phonology; Stemberger 1991 for parallel asymmetries in speech errors). As the learner is exposed to tʃi-final plurals (whether tʃ→tʃi or k→tʃi), the strength of PL-tʃi rises, overtaking "[k]!" before "[t]!" before "[p]!". In order for the learner to acquire alveolar palatalization, PL-tʃi must be ranked above "[t]!". Assuming for a moment that the strengths of input chunks do not appreciably change during training, this means that by the time PL-tʃi rises above "[t]!" it will also be ranked above "[k]!". Thus, alveolar palatalization is expected to overgeneralize to velars. By the same logic, labial palatalization is expected to overgeneralize to both velars and alveolars. Since "[chunk]!" constraints are simply tendencies for those chunks to be perseverated upon, we predict a link between perseveration errors involving erroneous retention of a segment and learnability of stem changes involving changing that segment. We do in fact observe the predicted link: perseveration errors are more common with [p] than with [t] and more common with [t] than with [k], and [p], paralleling differences in changeability of the three segments.

"[Chunk]!" constraints are predicted by viewing grammar as competition for production. They are directly grounded in speech errors, being supported by 1) the existence of perseveration errors, and 2)

---

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

[3] Howe & Pulleyblank's (2004) main reason for preferring to encode such biases in faithfulness constraint rankings is to capture "Hooper's generalization" (Hooper 1972), which states that the same segments (e.g., schwa) are most likely to be inserted and deleted. "[Chunk!]" constraints, however, do not directly capture this generalization. The default epenthesis processes in question may be better accounted for by adjustments to gestural timing (Browman & Goldstein 1989, 1991) whereas the deletion processes may be caused by low perceptual salience of the acoustical segments in question (Ohala 1981) or already low gesture magnitude allowing for undershoot to produce deletion (Browman & Goldstein 1991, Bybee 2001).

the Addition Bias, i.e., the finding that speech errors more commonly involve addition than deletion (Goldstein et al. 2007, Hartsuiker 2002, Stemberger 1991). In addition, they help resolve the too-many-solutions problem in Optimality Theory, providing a unified explanation for 1) the Preservation Principle (Paradis & LaCharité 1997), which notes that phonotactic violations in loanword adaptation are overwhelmingly resolved by epenthesis rather than deletion, and 2) 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 also predict 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), i.e., 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, Tessier 2006), since children generally show more motor perseveration than adults (Dell et al. 1997, Smith et al. 1999).[4]

4. Chunk-schema 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 (19). The numbers in the tableaux show the strengths of the various schemas, which is a function of type frequency and redundancy, as shown in (23), 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 times entropy, as discussed in section 7.[5]

[4] 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.

[5] I am assuming the following probabilities of misperception: 5% for misperceiving Labials as Velars, .01% for misperceiving [sonorant], [continuant] or [consonantal]. k is set to 1. No exponentiation is performed.

(19)

bup	[Pal]i# .0006 tʃi# .08	Vtʃi# 10	[-Pal]i# < .0001 [-cont;-Pal]i# < .0001 V[-cont;-Pal]i# .0001 V[-cont;-Pal;-son]i# .10 V[-cont;{Alv;Lab};-son]i# 2	[-cont]a# < .0001 V[-cont]a# .0004 V[-cont;Lab]a# .04 V[-cont;-son;Lab]a# 6	"[p]!"	Total	p(prod)
bupi			2.1			2.1	25%
bupa				6.04		6.04	73%
buptʃi	.08					.08	1%
butʃi	.08	10			-10	.08	1%

The probability of producing a candidate is taken in (19) 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. 1990a, 1990b). 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 3 times more common than [pa]-final plurals in training but are barely more common in the learner's output. This happens because many of the candidates are supported by the same highly general schemas, nullifying the differences between them. There is too much equality between candidates that don't feature stem changes. One solution is to punish forms for violating "chunk!" constraints instead of rewarding forms for opbeying "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 (20), 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 $k << 1$.[6] We will return to the strength of a chunk shortly.

[6] For the present data, $k = .05$ appears to be close to optimal. In other words, competition resolution between schemas is stochastic. Hudson Kam & Newport (2005) report that children faced with unpredictable choice between competing schemas are more likely to apply the more reliable schema 100% of the time rather than matching the probabilities in the input. There are two ways of accounting for this difference in C4B. First, k may be larger in children (which also predicts that children should be faster implicit learners). Alternatively, children might not access as many clamoring schemas for blending, in order to mimize demands on working memory,

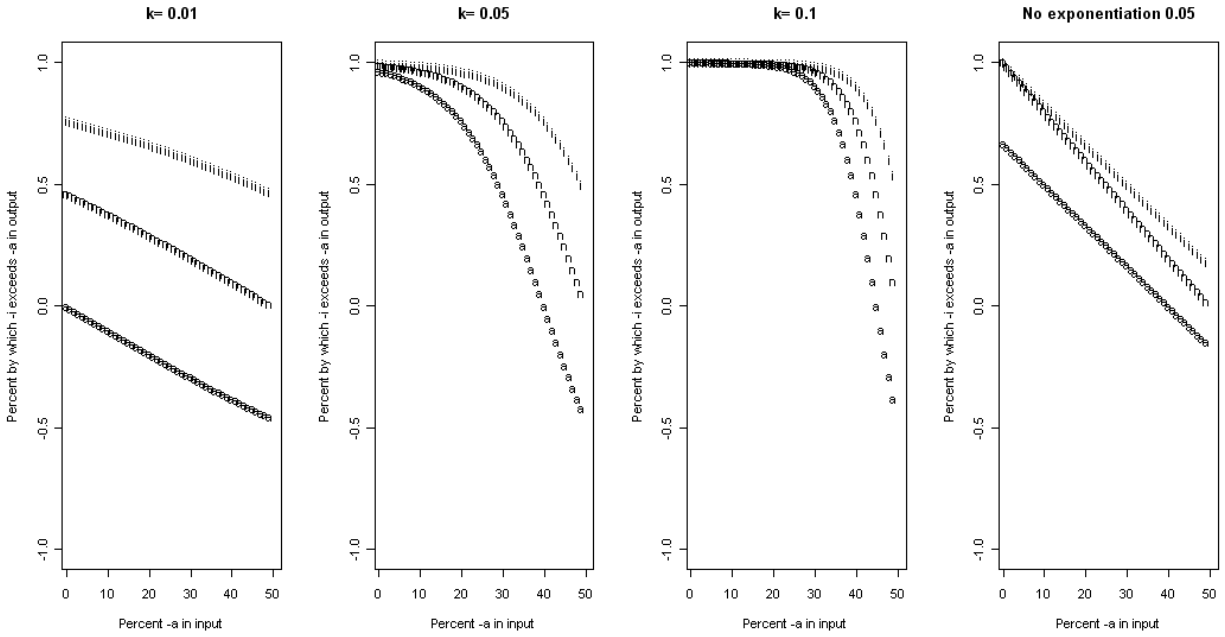
(20) $p_{produceA} = \frac{\exp(support_A)}{\sum \exp(support_i)}$

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→tʃi, t→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 the [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 (20) (in both the numerator and the denominator), this increases the production probability more when the probability of producing A is low (Figure 1). Thus extra 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 but not so weak as to never be produced even with the extra support (as would be the case for butʃi vs. bupi after training on buk--butʃi, butʃ→butʃi, bup→bupi).

---

restricting access to the shorter and/or more reliable schemas (Hudson Kam & Newport 2005, though see Perfors 2011 for evidence that increasing working memory load in adults does not lead to overgeneralization.

Figure 1: The effect of type frequency of a schema on the probability of choosing an output consistent with it, assuming only two competing schemas. The data series labeled as "n" for "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 the type frequency of –i between 100 and 51 (x-axis) for different values of k (which weight type frequency). The "i" line depicts how the differences in production probability between –i and –a changes when we add 1/k extra examples of –i. The "a" line depicts how the differences in production probability between –i and –a changes when we add 1/k extra examples of –a. 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.



Note that other candidates would be worse than the ones shown (harmonically bounded) in incurring extra, entirely superfluous violations of faithfulness. Thus the candidate set appears to be finite, unlike the candidate set in classical OT / HG (Kager 1999, Prince & Smolensky 1993 / 2004). However, one might object that there is currently no theory of schema induction, and thus the set of schemas is potentially infinite in that for every schema that excludes a candidate output that was not observed during training there might be a schema that includes it (Mitchell 1980). We cannot yet provide a complete theory of schema induction, as it is underspecified by our data, but we can make some steps in this direction.

First, meaningful schemas / constructions must contain some morphological boundary (in the Tableaux above, the right edge of the word). There are three reasons for this restriction. First, grammars appear to only count 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), non-past tense is signaled by stem-final alveopalatals; alveopalatals elsewhere do not contribute to non-past 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.

Second, over the course of learning, schemas, at least to a large extent, become more specific, rather than becoming more general. Here we 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, see also Hale & Reiss 2003). In other words, upon exposure to a few words of a novel language, 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. 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 calcified 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. That is, 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 joing probability space defined by meanings and sounds, i.e., which kinds of sequences are unexpectedly frequent in plural forms? Xu & 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". However, if *fep* is presented to you three times, 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. One untested prediction of this way of *specifying* rather than *generalizing* schemas 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#.

General-to-specific schema acquisition results in *automatic overgeneralization* (Vaux 2009), 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 2010 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 re-ranked on the basis of experience. We 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].[7]

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]. For actual languages, automatic overgeneralization also predicts that as an affix triggering a stem change gains in type frequency with stems that end in segments that are not eligible to undergo the change the stem change will lose productivity (as long as the segments eligible to undergo the change form a natural class on their own and a broader natural class together with the segments that are not eligible to undergo the change). This prediction appears to be correct. Kapatsinski (2010a, 2010b) documents that velar palatalization in Russian is fully productive before the diminutive suffixes –ik and –ok but not before the diminutive suffix –ik or the verbal stem extension –i. This puzzle can be explained by the fact that –ik and –i tend

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

not to attach to stems ending in velars, whereas –ok and –ek are largely restricted to velar-final stems. Carlson & Gerfen (2011) show that stem vowel monophthongization in Spanish is more productive before less productive suffixes. Since most Spanish stems do not contain diphthongs, productive suffixes largely attach to stems ineligible to undergo monophthongization, causing monophthongization to become unproductive before these suffixes. Conversely, we predict that a stem change will gain productivity and may expand to other source segments if the type frequency of products that contain the structure that results from the change increases.

Finally, 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, Moreton 2002, e.g., [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 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, Pater et al. 2004, Shvachkin 1948/1973, Stager & Werker 1997, Swingley 2007, Swingley & Aslin 2007 *inter alia*) and adult (Kapatsinski & Johnston 2010) learners tolerate single-feature mismatches despite being able to hear the difference. 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. We 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.

We implement general-to-specific schema extraction as decision tree induction (Daelemans & van den Bosch 2005, Ernestus & Baayen 2011) using the ctree() function in the party package (Hothorn et al. 2006, Strobl et al. 2009) in R (R Development Core Team 2008). We coded each of the words presented to learners in training 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.

The resulting trees are shown in Figure 2. Notice that the tree induction procedure automatically generates both negative and positive schemas, e.g., *[+cont]V#. However, we propose that only the branches with non-zero predicted counts are stored. In other words, schemas describe observed rather than non-existent words (Bybee 1985, 2001, Langacker 1987, Nesset 2008, Taylor 2002, see Section 9 below for justification). Given binary splits, schemas describing a path terminating in a node above a binary split containing one empty branch are fully redundant with the schemas describing the non-empty branch, and thus can be eliminated. Generally, a first-order schema is then defined as in (21), and we propose that all schemas in a decision tree are extracted from the data.

(21) A first-order schema is a path through a binary classification and regression tree in which the predicted variable is type frequency of an ngram and the predictors are semantic and phonological features of words containing that ngram. The path must proceed downwards from the root of the tree terminating in a node that is either 1) a leaf with a non-zero type frequency or 2) an ancestor to at least one leaf with a non-zero type frequency.

For the tree in Figure 2a, the two shortest paths originating at the root can be described as [-cont]X# (left path) and [+cont]X# (right path). However, neither is a schema under (17) because there are no observed words satisfying [+cont]X# and [-cont]X# is sister to [+cont]X#. Continuing down, we obtain [-cont]i# and [-cont]a#, both of which are ancestors to leaves containing existing words, thus both are legitimate first-order schemas describing classes of real plural wordforms. Both are extracted but [-cont]i# receives much greater weight as it is supported by many more words. The schema [-cont]a# dominates one branch that satisfies the criterion for being a schema: [-cont;-son;{Alveolar;Labial}]a#. However, given that this is not a natural class, we split this into two equally-weighted schemas: [-cont;-son;Labial]a# and [-cont;-son;Alveolar]a#. The schema [-cont]i# dominates tʃi#, [-cont;-Palatal]i# and [-cont;-son;{Alveolar;Labial}]i#, which we can again decompose into equally weighted [-cont;-son;Labial]i# and [-cont;-son;Alveolar]i#. Note that we extract both tʃi#, which allows tʃ#→tʃi# to help t#→tʃi# and k#→tʃi#, and [-cont;-Palatal]i#, which allows p#-->pi# and t#-->ti# to help k#→ki#. tʃi# is extracted because it is more common than other [-cont]i# sequences, while [-Palatal;-cont]i# is extracted because of the extraction of tʃi# and the fact that there are [-cont]i# sequences that are not subsumed by tʃi#.

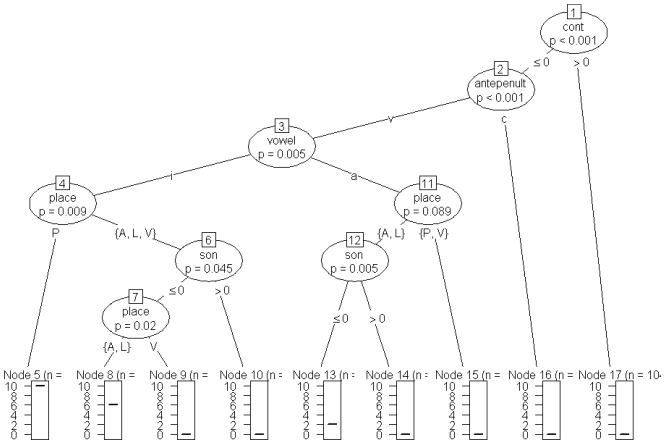
As demonstrated by Daelemans & van den Bosch (2005), abstractionist (grammatical) and analogical (instance-based) models of productivity are both expressible as decision trees of the kind shown in Figure 2 (see also Ernestus & Baayen 2011). The leaves of a decision tree representing an analogical

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, we suggest that the entire tree, complete with exemplars, is stored but time pressure often causes speakers to use only the most reliable schemas in 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 & 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 only accessed when there is time to do so. Otherwise, only the more informative dimensions high in the tree are accessed. We propose that the same should be true for morphological and morphophonological processing tasks like elicited production and acceptability judgment: when done under time pressure, they should only exhibit effects of the most reliable schemas and be more subject to perseveration of input characteristics.

Figure 2. Top-down schema extraction results for Tipi and Tipichi. Observation = word-final trigram, with all English consonants being possible in the stem-final position but trigrams containing most of these consonants having an observed frequency of 0 in the artificial languages.[8]

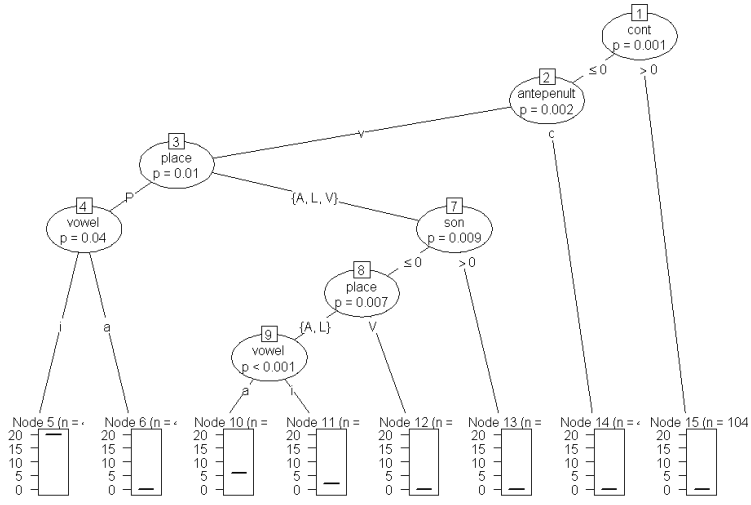
a. Tapa language

Schemas: [-cont]X#, V[-cont]X#, V[-cont]i#, Vtʃi#, V[-cont;-Pal]i# (support for ki#), V[-cont;-Pal;-son]i# (support for ki#), V[-cont;-Pal;-son;{Alv;Lab}]i#, V[-cont]a#, V[-cont;{Alv;Lab}]a#, V[-cont;{Alv;Lab};-son]a#



b. Tapachi language:

Schemas: [-cont]X#, V[-cont]X#, VtʃX#, Vtʃi#, V[-cont;-Pal]X#, V[-cont;-Pal;-son]X#, V[-cont;-son;{Alv;Lab}]X#, V[-cont;-son;{Alv;Lab}]i#, V[-cont;-son;{Alv;Lab}]a#



[8] "cont" = continuant (0 = [-continuant]). "place" = place of articulation, with values "P" = "palatal", "L" = labial, "A" = alveolar, and "V" = "velar". We do not group "P" with "A" or "V" because of the relatively equal typological frequency of velar and alveolar palatalization (Kochetov 2011). We 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 liberal enough to ensure that sequences in each bin have equal token frequency (for present data, micriterion = .8, minbucket = 2, minsplit = 5).

6. Perceptual salience of added segments and the Covered Skeleton Principle

The addition of examples of tʃ#→tʃi# changes the tree in Figure 2a into the tree in Figure 2b. A major difference between the trees is the lower placement of the node corresponding to the identity of the final (suffix) vowel in Figure 2b. This happens because the identity of the word-final vowel is now not as informative as the identity of the stem-final consonant for predicting frequency of occurrence of the word-final segment sequence. Thus where Figure 2a contains the schema [-Pal;-cont]i#, Figure 2b has [-Pal;-cont]V#. In fact, Figure 2b, unlike Figure 2a, contains no schema supporting [ki] over [ka]. This predicts, counter-intuitively, that Ca# should be supported by examples of tʃ#→tʃi#. This prediction appears to be incorrect: Ca# is decreased by examples of tʃ#→tʃi#. Something has to change.

Let us assume that the CV structure of a form is something that is highly salient and focused in early, as suggested by all theories of speech production (Berg 1988, Dell 1986, Levelt 1989, MacKay 1987, Shattuck-Hufnagel 1979, Stemberger 1985, 1990, though cf. Dell et al. 1993). If the CV frame is highly salient, then a plural form must end in a vowel. That vowel cannot be filled in from the singular, since singulars have no corresponding slot in the metrical frame. Therefore, a plural schema that derives a candidate plural must specify the identity of the final vowel. The schema ([-Pal;-cont]V#) then cannot be used to produce a candidate output: it does not specify the vowel. Contrast this with the schema [-Pal;-cont]i#, which *can* be used to derive a candidate output, since it fully specifies all slots in the metrical frame that cannot be filled in from the singular. Thus, if a candidate is only supported by the schemas that give rise to it, the problematic [-Pal;-cont]V# schema is irrelevant for candidate evaluation and therefore cannot boost the acceptability of unattested CV# sequences. Relatively uncontroversial for elicited production, this proposal requires an analysis-by-synthesis account of acceptability rating, which, however, is also required in RBP (Albright & Hayes 2003) and OT / HG (Boersma 2004). That is, schemas that do not play a role in producing an input-output mapping also do not play a role when the acceptability of that mapping is evaluated because this evaluation involves the same candidate generation and evaluation process with a different decision rule. Note, however, that the present account (unlike Albright & Hayes 2003 and Boersma 2004) can accommodate greater tolerance of stem changes in rating compared to elicited production (Kapatsinski in press, Zuraw 2000) insofar as chunk perseveration, being a separate production-internal process, can be ignored in rating, unlike in elicited production.

Figure 2b shows that the language Tapachi does not contain a schema supporting [ki] whereas the language Tapa does. Thus, only the Tapa language contains specific support for [ki]. Thus we expect the addition of –i to {t;p} to be helpful for [ki] when the language does not contain a huge number of tʃi-final (and generally-i final) plurals. To take an extreme case, when a stem alternation is triggered by an affix that has no competitor, the type frequency of that affix with stems that don't have segments eligible to undergo the change is irrelevant for determining the affix's ability to trigger the alternation. The difference in rate of ki# use between Tipi and Tapa is indeed larger than between Tapachi and Tipichi. However, the interaction is nowhere near to being reliable ($p > .1$) whereas the difference

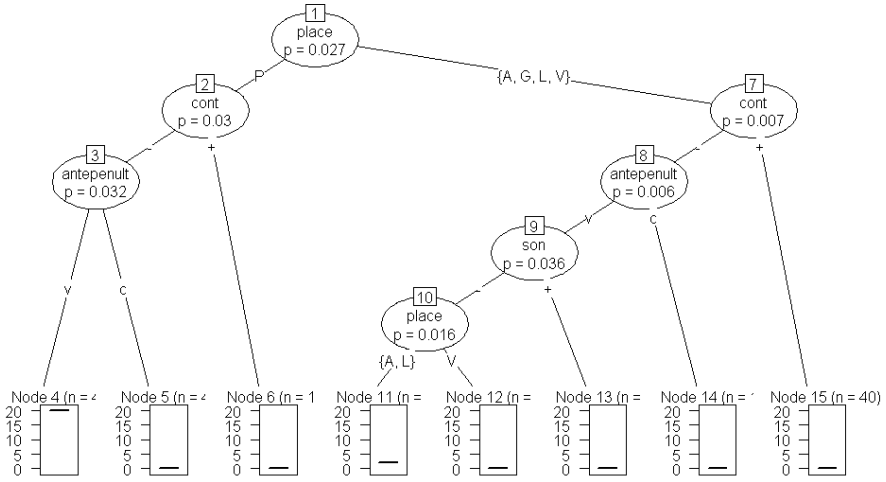
between Tipi and Tipichi is reliable ($p < .01$). Again, the difficulty appears to be that ctree(), knowing nothing about whether a schema will be useful, does not assign any special significance to schemas specifying the final vowel. Thus the identity of the final vowel might not be specified in most schemas in the tree, leading those schemas to be discarded. We thus propose that the parts of the product (slots on the CV skeleton or coupling graph) that have no correspondent in the input are focused on by the learner and become part of every schema (22), and promote final vowel identity to the top of the tree by hand. With this adjustment, the schemas extracted in all of our artificial languages are identical (Figure 3), although their weights vary. Further, differences between languages are appropriately predicted: the addition of examples of tʃ#→tʃi# helps X→tʃi#, and t#→ti# and p#→pi# help k#→ki#.

(22) The Covered Skeleton Principle (CSP): Every schema must fully specify all slots on the output CV skeleton that are not specified by the input.

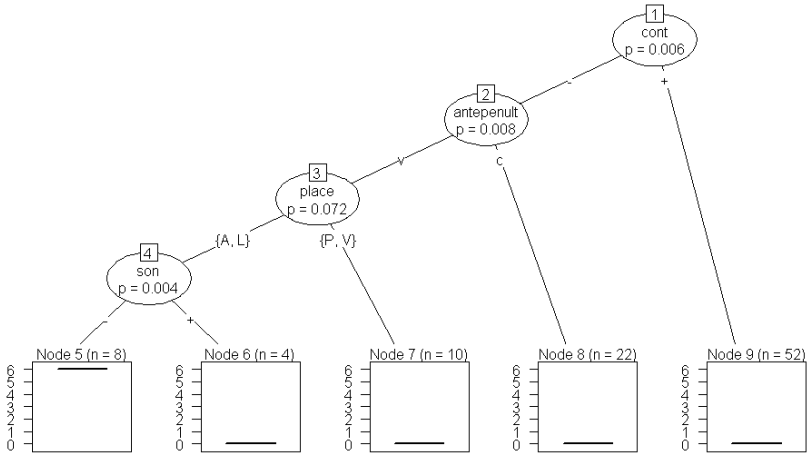
The CSP predicts that mappings involving the same affix help each other, unless they are competing for the same probability mass, whereas mappings involving different affixes do not. This prediction appears to be correct. Kapatsinski (2006, in press) showed that ratings of alternative outputs derived from the same input using the same suffix are positively correlated. For instance, subjects assigninf high ratings to k→tʃi also assign relatively high ratings to k→ki (Kapatsinski in press for artificial language data, Kapatsinski 2006 for Russian). In contrast, ratings of mappings involving different affixes (e.g., k→ki and k→ka) are either uncorrelated or negatively correlated. Further, Kapatsinski (submitted) also shows that providing subjects with examples of tʃ→tʃu does not increase the production probabilities or ratings of t→tʃi relative to t→ti or k→tʃi relative to k→ki.

Figure 3. Schemas extracted if the schema has to specify the final vowel (the part that cannot be filled in from the input). The schemas are the same across the four languages, only weights differ.

Final vowel is –i. Schemas: [Pal]i#, tʃi#, Vtʃi#, [-Pal]i#, [-cont;-Pal]i#, V[-cont;-Pal]i#, V[-cont;-Pal;-son]i#, V[-cont;{Alv;Lab};-son]i#



Final vowel is –a. Schemas: [-cont]a#, V[-cont]a#, V[-cont;{Alv;Lab}]a#, V[-cont;{Alv;Lab};-son]a#.



7. Redundant schema weighting and Bayesian perception

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. We can think of schemas as descriptions of the lexicon. A perfect description of the lexicon would fit all words that are part of the

lexicon and none of the words that are not part of the lexicon. In this sense, if X has only one Y, then Y is a better description of the experienced lexicon than X: X and Y fit the same set of words that are part of the lexicon but X, being more general, also fits some words that are not part of the lexicon. Compare, for instance, the –a-final schemas in Figure 3. Nonetheless, we believe that at least some such redundant schemas are extracted. In other words, the learner assigns some degree of belief to the existence of words that are not quite like the words s/he has experienced so far. This assumption is necessary to account for overgeneralization. If redundant schemas are not extracted, then the only schemas rooting for a stop followed by –i# in Figure 3 are [-cont;Alv;-son]i# and [-cont;Lab;-son]i#, thus there is no reason to overgeneralize simple addition of –i to [k]. However, it seems reasonable to take redundancy into account in schema weighting and to punish non-maximally-specific schemas that have only one corresponding maximally-specific schema for being redundant. This can be accomplished by weighting schemas by entropy (Goldwater & Johnson 2003, Hayes & Wilson 2008) as in (23), where TF is type frequency, and L is the number of leaves subsumed by the schema.

(23) $$wt = k \times (\sum_{i=1}^{L} TF_i) \times \left(-\sum_{i=1}^{L} \left(\frac{TF_i}{\sum TF_i} \log \frac{TF_i}{\sum TF_i}\right)\right)$$

The formula in (23) ensures that specific schemas are generally more highly weighted than general schemas. However, one might notice that fully redundant schemas receive a weight of zero. If we want such schemas to be extracted, something must be changed. Perhaps, the most parsimonious solution is to assume that on some occasions words that fit the more specific schema are erroneously perceived not to fit because some features of the word required by the more specific schema are misperceived or because only the features of the more general schema are accessed in time (McLennan & Luce 2005). Thus entropy then is actually not quite zero. However, as stated, this account faces a problem in accounting for the present data insofar as repetition accuracy during training is very high. How can one repeat a word correctly without correctly perceiving all of its features?

Traditionally, the output of human perception is taken to be a single hypothesis about the identity of the stimulus, thus the only information provided by perception is the identity of the most probable stimulus given the evidence. For instance, Clayards et al. (2008: 804), in a paper arguing for an otherwise Bayesian approach to speech perception, write "the goal of speech perception can be characterized as finding the most likely intended message". On the other hand, from a purely Bauesian perspective, the output of perception should be a probability distribution over possible stimuli (Kruschke 2008). Thus, despite reporting having perceived the stimulus one thinks is the most likely to have been presented, the perceiver assigns other similar stimuli non-zero probabilities of having been presented. For instance, a subject presented with [ti] may report hearing [ti] but also (subconsciously) consider it possible but less likely that [ki] has just been presented. Note that if the learner intends to maximize the probability of being correct, s/he should always report hearing the stimulus s/he considers to be the most probable one (Norris and McQueen 2008) but should update the probability of each possible hypothesis in proportion to how likely s/he believes it to be given the sensory data (Kruschke 2008).

A maximally-specific schema Y differs from its redundant parent X in being specified for additional features. Note that X receives support from a training example whether or not the values of the features that are specified in Y but not X are perceived correctly. Thus, whenever at least one of the N features that are specified in Y but not in X is assigned non-zero probability of having a different value from the one intended by the speaker, X receives support, becoming non-redundant. We can thus replace TF in (23) with the estimated E(TF) defined as in (24). Given Bayesian perception, there are no fully redundant schemas, and non-maximally-specific schemas are expected to be retained in the grammar.

(24) $E(TF_Y) = TF_Y(1 - \bigcup_{i=1}^{N} p(\text{"}\alpha\text{Feature}_i\text{"}|\beta\text{Feature}_i)))$

8. Chunk strength changes as a function of experience

Thus far we have not discussed chunk strength, i.e., structure-specific perseveration tendencies beyond noting that to some extent chunk strength differences are something the subjects bring with them to the experiment. We assumed that learners come to the experiment with the ranking "p!" >> "t!" >> "k!", and then learn that plurals should end in Vtʃi#, with the strength of Vtʃi# gradually rising past "k!", then "t!" and only then "p!" as its experienced frequency grows. This accounts for full overgeneralization of labial palatalization to alveolar and velar sources and alveolar palatalization to only velar sources.

For any given input chunk, potentially relevant training experiences are of two kinds. On the one hand, the learner might notice that the chunk is not retained in the plural form, reducing the weight of the associated "chunk!" constraint. On the other hand, the learner might also notice when the chunk *is* retained in the plural, increasing the weight of the relevant "chunk!" constraint. In either case, weight adjustment requires noticing the relationship between a singular form and a corresponding plural form. It is reasonable to assume that noticing the relationship is easier when the to-be-compared forms are temporally adjacent (Pierrehumbert 1993). Thus, we might expect that strengths adjustment of "chunk!" constraints will be greater in the "source-oriented" training paradigm, in which corresponding singulars and plurals are always temporally adjacent, than in the "product-oriented" training paradigm, in which corresponding singulars and plurals can only meet by chance.

We do in fact observe an effect of training paradigm, such that stem changes are disfavored in the source-oriented training paradigm. This is especially noticeable if one compares changes resulting in "good" products (those favored by first-order schemas). In C4B, the productivities of these changes in the learners' output reflect the weights of "chunk!" constraints relative to the first-order schema Vtʃi#-PL. Since the first-order schema is learned on the basis of plural forms, temporal adjacency of singular and plural forms should not have an effect on its strength. Thus, this effect of training paradigm must be attributed to the effect of temporal adjacency on updating the weights of "chunk!" constraints.

I propose that the detection of the fact that a chunk in the plural is identical to the corresponding chunk in the singular is not automatic and is strongly affected by temporal adjacency or absence of intervening

items (as is now well-accepted for vision, see Mitroff et al. 2004). When the chunks are in adjacent wordforms ([bup bupi]), the fact that they are identical is detected and the weight of the corresponding "chunk!" constraint is increased. When the identical corresponding chunks are *not* in adjacent wordforms ([bup slaik … dwitʃ bupi]), the fact that they are identical is likely to be missed, hence the weight of the corresponding "chunk!" constraint is not increased. This is not an unprecedented claim in phonology: Pierrehumbert (1993) likewise proposes the difficulty of detecting that two segments are identical at a distance as an explanation for why OCP constraints against having multiple nearby segments with the same place of articulation is weaker when the segments are separated.

What about the singular-plural pairs in which a certain chunk is not retained? Presumably, the change should also be easier to detect when the corresponding non-identical chunks are in adjacent wordforms than when the wordforms are far apart ([bup butʃi] vs. [bup slaık … dwitʃ butʃi]). This would then predict that stem changes should be learned *better* when the corresponding singulars and plurals are next to each other, contrary to what we observe. At this point one can propose either that 1) "chunk!" constraints do not weaken even when one detects that a chunk from the singular is not retained in the plural, or 2) temporal adjacency is less important for change detection than for identity detection.

While the data from the present experiment provide no evidence regarding this choice, the second option seems more prudent. First, while we find a bias against labial palatalization, this is a *soft* bias. We know that one can learn to palatalize labials and leave velars unchanged because this (p→tʃ, k→k) is observed as a productive synchronic process in Southern Bantu (Ohala 1978). Such a set of mappings can only be accomplished in C4B by ranking "p!" below "k!". Thus we cannot claim that labial palatalization *must* be completely overgeneralized to velars. Given that labial palatalization is cross-linguistically uncommon and phonetically unmotivated (Bateman 2007, Kochetov 2011), we do not want to abandon the claim of an initial ranking of "p!" above "k!". Therefore, the Southern Bantu ranking of "k!" above "p!" must come about through reweighting "k!" or "p!" or both. It seems unlikely that Southern Bantu contains an exceptionally large number of examples of [k] being retained before the palatalizing suffixes causing the weight of "k!" to soar to unprecedented heights. Rather it seems likely that the weight of "p!" is decreased by examples of p→tʃ. Second, there is (albeit very limited) empirical data directly supporting the second alternative. Massaro (1970) reports on a same-different discrimination task in which subjects judged whether two (non-linguistic) tones were exactly identical. He found that as the delay between the two tones increased, subjects showed a change in response bias: as the delay increased, the rate of erroneously saying "different" when the tones were identical increased faster than the rate of erroneously saying "same" when the tones were physically different. Thus detecting that two stimuli are identical may be more sensitive to memory decay or interference than detecting that two stimuli are *not* identical.

One puzzling finding in the present study (and also Wilson 2006) is that learning biases manifest themselves more in overgeneralization to untrained structures than in accuracy with trained structures. Rates of labial, coronal, and velar palatalization do not differ significantly in Experiment 2. However,

labial palatalization overgeneralizes to alveolars and velars, whereas velar palatalization is not overgeneralized, and alveolar palatalization overgeneralizes only to velars. Likewise, Wilson (2006) found equal accuracy on learning k→tʃi and k→tʃe but found that palatalization generalized from [__e] to [__i] but not from [__i] to [__e]. A simple explanation is that the subjects are at ceiling for the trained mappings: while they are not anywhere near 100% correct in either study, 100% may not be reachable in a group of undermotivated adult undergraduates experiencing the language solely by passive listening. While this can explain the pattern in Wilson's (2006) study, where the testing input is not presented during training, the present study does train the subjects on all input types presented in testing. Something else is necessary to explain why this training on faithful mappings is unsuccessful to the extent that we can still observe the pre-experiment ranking of "chunk!" constraints.

Consider that experience with unfaithful mappings only changes the weights of "chunk!" constraints that are violated in the training data. Thus the weights of "chunk!" constraints for chunks that are not observed to change during training are not decreased during training. The weights of "chunk!" constraints are only increased if the learner notices that two corresponding chunks are identical. Further, our learners appear to only notice that two corresponding chunks are identical if they are in adjacent wordforms whereas changes can be noticed across intervening wordforms. Experiment II, where we observe the learning bias, employs the product-oriented training paradigm, hence corresponding chunks are rarely in adjacent wordforms. Thus the weights of "chunk!" constraints corresponding to chunks that are not observed to change in training do not change from their original weights whereas the weights of "chunk!" constraints that are observed to change are successfully decreased, and learners approach the accuracy ceiling for unfaithful mappings they have experienced. This hypothesis suggests that we should not be as successful at observing pre-experiment "chunk!" constraint weights with source-oriented training.

9. Negative generalizations and the role of production feedback

According to OT / HG, phonology is the interaction of markedness and faithfulness constraints, and any changes to the input are done in order to satisfy some markedness constraint, which is a ban on an articulatorily difficult sound or sound sequence (Kager 1999, Prince & Smolensky 1993 / 2004). Unlike OT / HG, and like CG, C4B proposes that the grammar of an adult native speaker of a language does not contain markedness constraints, or any kinds of explicit generalizations about what forms are not observed in the training data.

If we were to use markedness constraints, velar palatalization could be described as *ki >> Ident-[Dorsal], and the bias for velar palatalization and against labial palatalization could be described as a pre-experimental ranking *ki >> *ti >> *pi. The task facing the learner is to realize that ki# is in fact illegal and that Ident-[Dorsal] can be violated to satisfy *ki. The problem with this account of the data is that examples in which –i simply attaches to a stem that happens to end in [tʃ] have no relevance

for *ki or *ti. Thus, this account fails to capture the fact that examples of tʃ→tʃi help t→tʃi and k→tʃi. One can propose that learners also come to the experiment with a high-ranked *tʃi, whose ranking needs to be decreased to acquire velar palatalization, and that this decrease is helped by tʃ→tʃi. However, the data from Experiment 3 do not support this claim: with no training, learners do not avoid attaching –i to [tʃ]-final stems. Neither is there any evidence for speakers avoiding [ki] compared to [ti] or avoiding [ti] compared to [pi] prior to training, suggesting that the observed substantive bias against alveolar and especially labial palatalization should not be attributed to a ranking of markedness constraints. Rather, we suggest (following Howe & Pulleyblank 2004) that it is accounted for by a ranking of faithfulness ("[chunk]!') constraints.

Another model of phonology that relies heavily on markedness constraints is Maximum Entropy Phonotactics (Hayes & Wilson 2008). Hayes & Wilson (2008) propose that any examples that do not contain a sequence increase one's confidence in a constraint against that sequence. Thus, extra examples of plurals ending in [tʃi] strengthen constraints against all sequences other than [tʃi]. For elicited production, this is equivalent to saying that examples of [tʃi] help [tʃi] because the production probabilities of all sequences must add to 1. For acceptability ratings, however, Hayes & Wilson (2008) predict that examples of tʃ→tʃi should lower ratings of all plurals not ending in [tʃi] whereas C4B predicts that they should raise ratings of plurals ending in [tʃi] and all other sequences ending in [i] (because of the [-cont]i# schema in Figure 3). Further, based on Hayes & Wilson (2008), we would expect extra examples of tʃ→tʃu to increase avoidance of [ki] just as much as examples of tʃ→tʃi increase it. However, we find that such examples do not favor C→tʃi. The predictions of C4B are correct in this case.

Like C4B, CG (with the exception of Frisch et al. 2004) does not use negative generalizations. If grammar is formed by generalizing over lexical items (Bybee 2001, Langacker 1987, Nesset 2008, Taylor 2002), finding phonological characteristics that are shared between wordforms with similar meanings, it is hard to see how negative generalizations could be formed (Nesset 2008, Stemberger & Bernhardt 1999). However, as we have shown above, if schema extraction is done by specification, negative generalizations could be made by looking for "gaps" in the probability distribution: structures that are observed less often than would be expected by randomly sampling structures from a wider natural class. Learners, however, do not seem to be overly concerned with gaps in this sense (contra Frisch et al. 2004). For instance, we observe that increasing the frequency of [Ci] in training while keeping the frequency of [ka] constant and the frequency of [ki] at zero results in increasing acceptability of [ki]. This is unexpected if learners are looking for gaps, since the greater the frequency of [Ci], the greater the difference between how often [ki] is observed (zero times) and how often it would be expected to be observed if it were like other similar –i-final sequences (Frisch et al. 2004,

which could be implemented by comparing differences in type frequencies between nodes with a common parent, indicated by the p values in Figures 2-3).[9]

There are two phenomena that seem to call for negative generalizations, both of which involve *avoidance* of a structure. One is the phenomenon of paradigmatic gaps, where some cells in the morphological paradigms of some words are not filled (e.g., Albright 2003, 2010, Daland et al. 2007, Hetzron 1975, Sims 2007). The gaps appear to be morphophonological rather than morphosyntactic in nature (*pace* Daland et al. 2007) because they tend to happen when there are two competing schemas of similar strength (Albright 2003, 2010) and it is never observed that a form that would normally be homophonous with a form in another cell in the paradigm is arbitrarily outlawed, which would be expected if gaps targeted morphosyntactic feature combinations. However, any account of paradigm gaps requires lexical listing of forms that are "not to be produced". Otherwise, the gaps would have been filled (possibly by two competing forms in free variation), yet they seem to be spreading (Daland et al. 2007, Sims 2007). For instance, Russian speakers know not to say [m^{j}etʃt] as the genitive plural of [m^{j}etʃta] 'dream' yet no-one objects to [matʃt] being the genitive plural of [matʃta] 'dream' (Albright 2010, Daland et al. 2007, Sims 2007). Thus, it is plausible to account for gaps with positive first-order schemas identifying typical properties of wordforms (and meanings of those wordforms) that make listeners cringe, e.g., genitive plurals ending in tʃt# for Russians.

We do believe that negative generalizations are necessary in cases of *production-driven avoidance*. Such avoidance 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 (Schwartz & Leonard 1982, Schwartz et al. 1987, Storkel 2001). Berg (1998) and Martin (2007) also document a tendency for words with universally dispreferred phonotactics to drop out of languages that contain them and be replaced by phonologically unrelated words at a higher rate than words that do not contain hard-to-articulate sequences. 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 competition to easier-to-articulate alternatives with similar meanings. However, we argue that 1) production experience with a difficult-to-produce structure is necessary to learn to avoid that structure (Johnston & Kapatsinski 2011, Redford 2008), one does not learn to avoid a structure simply by not hearing it (see also Boyd & Goldberg 2011, Goldberg 1995 for syntax), and 2) differences in inherent articulatory difficulty are far

[9] Frisch et al. (2004) suggest comparing observed and expected type frequencies using the observed/expected ration, which comes out to be zero regardless of expected frequency if the observed frequency is zero. However, I suggest that a learner searching for non-accidental gaps should be more impressed by some sequence not occurring if its expected frequency of occurrence is 1000 than if it is 1.

more relevant to children than to adults (at least for sounds and sound sequences that occur in the subjects L1): the avoidance behaviors are counterproductive for learning a language and are something one needs to outgrow.

Thus we do not think such feedback plays a major role in the present experiment because the subjects in the present experiments are adults and the structures they learn to associate with (or avoid) in plural nouns are very familiar. Adults are experts at producing sound sequences that occur in their native languages, and motor expertise is known to lead to improved, more flexible control over motor execution (Jordan 1996, Lindblom et al. 1979, Logan 1983, Smith & Goffman 1998). Thus differences in ease of articulation are expected to be of more importance for children than adults. In fact, while children have been shown to avoid hard-to-pronounce words (Schwartz & Leonard 1982, Storkel 2001), Johnston & Kapatsinski (2011) show that adults asked to learn words in a paradigm involving many-to-many word-meaning pairing actually prefer words that begin with phonotactically illegal onset clusters (#lb, #bd, #bn) to ones that begin with legal ones, suggesting that the greater perceptual distinctiveness of the phonotactically unusual words overrides the influence of ease of articulation. The present languages do not involve sequences that are illegal in English: the unobserved sequence [ki] is common in English as is the preferred sequence [tʃi]. There is no reason to expect that either would pose production difficulties for adult native English speakers.

To reiterate, we believe markedness constraints are a kind of generalization language learners make on the basis of production experience but we do not believe they play a major role in the present experiment. The learning task in the present experiments does not involve learning new sounds or sound sequences, where markedness differences are relevant. Rather, it involves learning novel paradigmatic mappings between known sounds and associations between known sounds and particular meanings.

10. Second-order schemas vs. rules

Rule-Based Phonology (Albright & Hayes 2003, Chomsky & Halle 1968, Plag 2003, Reiss 2008 *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 2008). Albright & Hayes (2003) provide an explicit computational model for discovering rules, which is essentially equivalent to what we teach students to do by hand in introductory phonology classes. Consider the singular-plural pairs in (25) exemplifying the Tipi language.

(25)

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, the speakers of the language in (25) are proposed to have made the generalizations in (26), 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.

(26)

a. 0→i/[-cont]__# when plural[10]

b. k→tʃ/___i#

How can the generalizations in (26) be extracted from the data in (25)? First, one splits each pair of forms into change and context, yielding (27), and then generalizes over contexts, yielding (28) (see Albright & Hayes 2003 for a computational model that carries out these two steps for any lexicon). Note that all forms in (25) except those in the left column contribute to the strength of the rule 0→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 on the farthest right in (25).

(27)

k→tʃi/vu__#	0→i/vut__#	0→i/vup__#	0→i/vutʃ__#
k→tʃi/bra_#	0→i/brat__#	0→i/brap__#	0→i/bratʃ__#
k→tʃi/si_#	0→i/sit__#	0→i/sip__#	0→i/sitʃ__#

(28)

a. k→tʃi/V_# b. 0→i/[-cont]_#

The rules in (28) are different from the rules in (26) in only featuring a single processing stage. Further, (28a) and (28b) are in competition for stems that end in /k/: (28a) argues that the plural forms of such stems should end in [tʃi] while (28b) argues that they should end in [ki]. Competition can be minimized

[10] 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 (as we shall see later), {p;t} should not be a natural class, i.e., the feature [Dorsal] is privative. Whether (2a) refers to [-cont], C or nothing at all as the left context is not relevant for the present argument.

by noticing that both rules support the change 0→i but (28a) contains a further change (k→tʃ), which is fed by 0→i (i.e., 0→i produces the context for k→tʃ), resulting in the rules in (26) (modeled computationally by Johnson 1984). The rules in (24) have the advantage over the rules in (26) in correctly predicting that t→ti and p→pi support k→ki. However, they incorrectly predict that examples of tʃ→tʃi should reduce the productivity of velar palatalization by providing support for the competing rule in (28b). On the other hand, these data are easily accounted for by product-oriented, first-order schemas.

While Bybee (2001) considers the possibility of a purely product-oriented grammar,[11] this hypothesis appears to be too restrictive. 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 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*, on the other hand, would yield the observed pattern of [k] being mapped onto [s] and not [l] because adjectives ending in [k] never correspond to nouns ending in [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 its genitive plural will drop it, whereas a novel noun ending in a non-palatalized consonant will gain –of, as in (29). Since the –a is dropped in the genitive plural, the generalizations responsible for the mappings must be source-oriented.[12] This is by no means an isolated example (see Booij 2010 and Nesset 2008 for others).

[11] "[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).

[12] Note that the mappings (a/0 and 0/of) do not pair up affixes based on similarity, thus the mappings cannot be accounted for by minimization of output-output faithfulness violations (as in Kenstowicz 1996).

(29)

Odna	*flarnikrap*	*-a*
One.FEM.SG.NOM	flarnicrop	FEM.SG.NOM
→*Neskol'ko*	*flarnikrap*	
Some	flarnicrop	FEM.PL.GEN

Odin	*flarnikrap*	
One.MASC.SG.NOM	flarnicrop	MASC.SG.NOM
→*Neskol'ko*	*flarnikrap*	*of*
Some	flarnicrop	MASC.PL.GEN

Nesset (2005, 2008) and Booij (2008, 2010) reintroduce source-oriented generalizations into a Cognitive/Constructionist/Usage-based phonology. However, these source-oriented generalizations crucially differ from those of Rule-based Phonology in that they do not involve a change/context split. Nesset (2008) explicitly proposes that they are "second-order schemas", generalizations over first-order product-oriented generalizations. The name "second-order schemas" is designed to imply that they are relatively difficult to acquire (Nesset 2008) because their acquisition relies on the error-prone comparison process. Second-order schemas have in fact been shown to be difficult to acquire 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 & 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 thus unreliable, generalizations over these schemas are not learned. Given this previous data on the difficulty of acquiring second-order schemas, it appears highly unlikely that they play a major role in the present experiments: passive repetition of training items for 15-20 minutes has been shown to be insufficient for acquiring such generalizations, at least by adults. However, given the necessity of such schemas for accounting for patterns in natural language (Booij 2010, Nesset 2008, Pierrehumbert 2006), it is worth showing that tʃ→tʃi continues helping palatalization once second-order schemas are added.

Since second-order schemas are generalizations over product-oriented schemas, they cannot refer to null elements, making 0→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, tʃ→tʃi cannot be an instance of 0→i. Furthermore, they do not involve a split into change and context, thus the 'context' [tʃ] is retained in each side of the schema. Thus the schemas corresponding to the Tipi language are of the format in (30). If you then generalize over these, (30a) and (30b) can be combined together into (31a) whereas (30c) and (30d) can be combined together into (31b). The more specific schemas in (24) are not discarded (Langacker 1982, 1987, Nesset 2008), thus we retain the knowledge that [k] and [tʃ] are

most likely to be mapped onto [tʃi] but also (over)generalize palatalization to [t], which is a stop like [k] but is Coronal like [tʃ]. Importantly, examples of tʃ→tʃi continue contributing to the productivity of palatalization, like in the version of usage-based phonology that lacks source-oriented schemas (Bybee 1985, 2001). Thus, as long as there is no change/context split and there is generalization over forms with similar meanings, we expect common product patterns to be overgeneralized to new sources.

(30)

a. SG-V_itʃ#	/	PL-V_itʃi#
b. SG-V_ik#	/	PL-V_itʃi#
c. SG-V_it#	/	PL-$V_i$ti#
d. SG-V_ip#	/	PL-$V_i$pi#

(31)

a. SG-V_i[-cont;-voice;Lingual]#	/	PL-V_itʃi
b. SG-V_i[-DelRel;-voice]$_j$#	/	PL- V_i[-DelRel;-voice;Ling]$_j$i#

11. Specific > > General or Primacy of the Lexicon?

Langacker (1987) and Nesset (2008) argue that competition between two conflicting schemas that fully match the input is resolved in favor of the more specific schema. Given that the most specific applicable schema is one specific to the individual lexical item, this proposal subsumes the proposal that lexical retrieval takes precedence over grammatical computation (Albright & Hayes 2003, Kapatsinski 2010a, 2010b, Pinker 1991, Zuraw 2000) as a special case. The question is whether this reduction of primacy of lexical retrieval to a more general specific > > general schema ranking is warranted.

The primacy of lexical retrieval is supported by two main findings. First, it is possible for speakers t treat unknown words probabilistically (e.g., choosing the plural suffix –i 30% of the time and –a 70% of the time) while showing complete certainty about the correct plural form of known words (Albright & Hayes 2003, Zuraw 2000). Since novel plurals have to be produced using the grammar, and schema competition in these cases is resolved stochastically, something else must explain the lexically-specific deterministic behavior. Second, it is possible for an alternation to lose productivity, as measured by elicited production or novel loanword adaptation, while not gaining lexical exceptions (Bybee 2001). For instance, Kapatsinski (2010a, 2010b) shows that Russian speakers borrowing English verbs for use on the web may attach the stem extension –i to them (e.g., /blogitʲ/ ‘to blog’) without changing the stem-final [g] into [ʒ] despite the alternation being exceptionless in the native vocabulary (e.g., /drug/ ‘friend’ vs. /druʒitʲ/ ‘be friends with’).

While it is tempting to generalize that a more specific schema always takes precedence over a less specific one (Nesset 2008), this proposal faces a problem: many examples of stochastic competition resolution between schemas involve competition between schemas that fully match the input and differ in generality. For instance, in the present data velar palatalization is less productive in Tipi than in Tapa

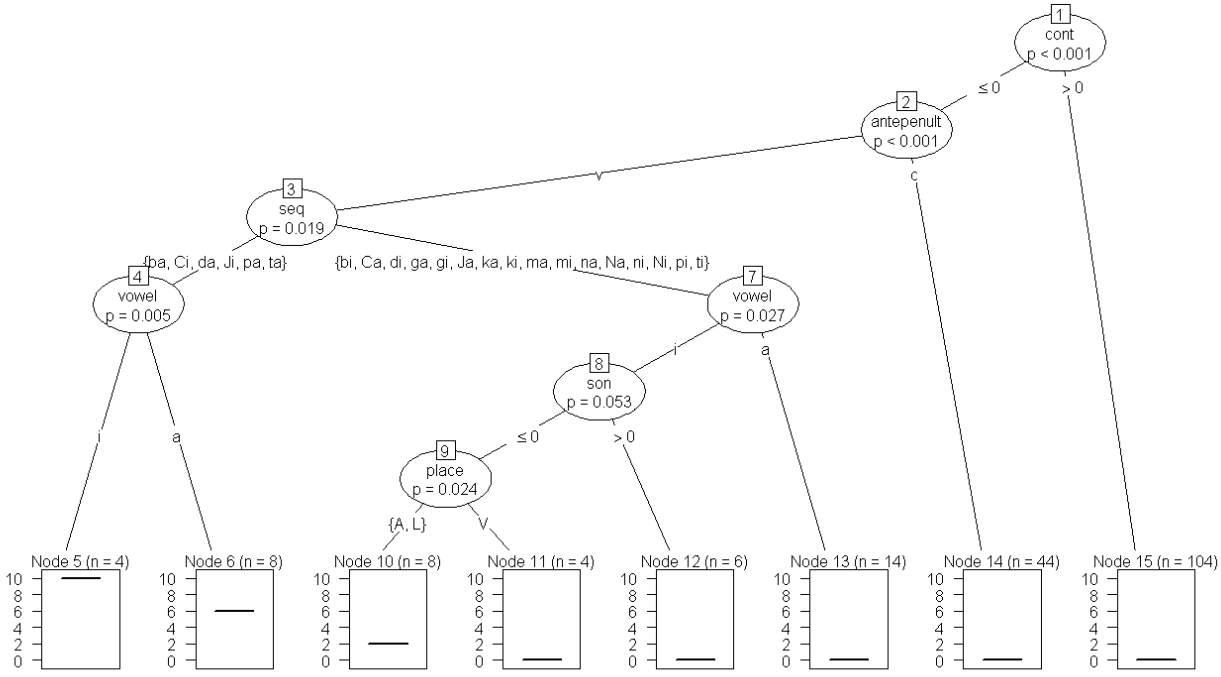
because PL-[-Pal;-cont]i# is stronger in Tipi than in Tapa whereas the more specific PL-tʃi# schema is of equally strong across the two languages. If competition is always resolved in favor of the more specific schema (barring perhaps some "performance errors"), the languages should not differ. Similarly, specific >> general predicts that novel verbs highly similar to specific English irregular verbs (and thus matching the corresponding irregular past tense schema very well) should never be regularized whereas they are in fact regularized most of the time (though the specific extent also depends on the strength of the applicable *regular* schema, Albright & Hayes 2003). It thus seems worthwhile to maintain the special status of lexical retrieval: if one knows the specific wordform that should be produced, that form is produced by lexical retrieval; if there is no form to retrieve, the produced form is the result of a stochastic process of clamoring for blending, in which competition between candidates is resolved stochastically. Assuming that clamoring for blending generally takes longer than lexical retrieval, the two routes can run in parallel (as proposed for word recognition by Baayen et al. 1997) and yet the lexical route will usually win.

12. Unitization and the problem of cross-over interactions

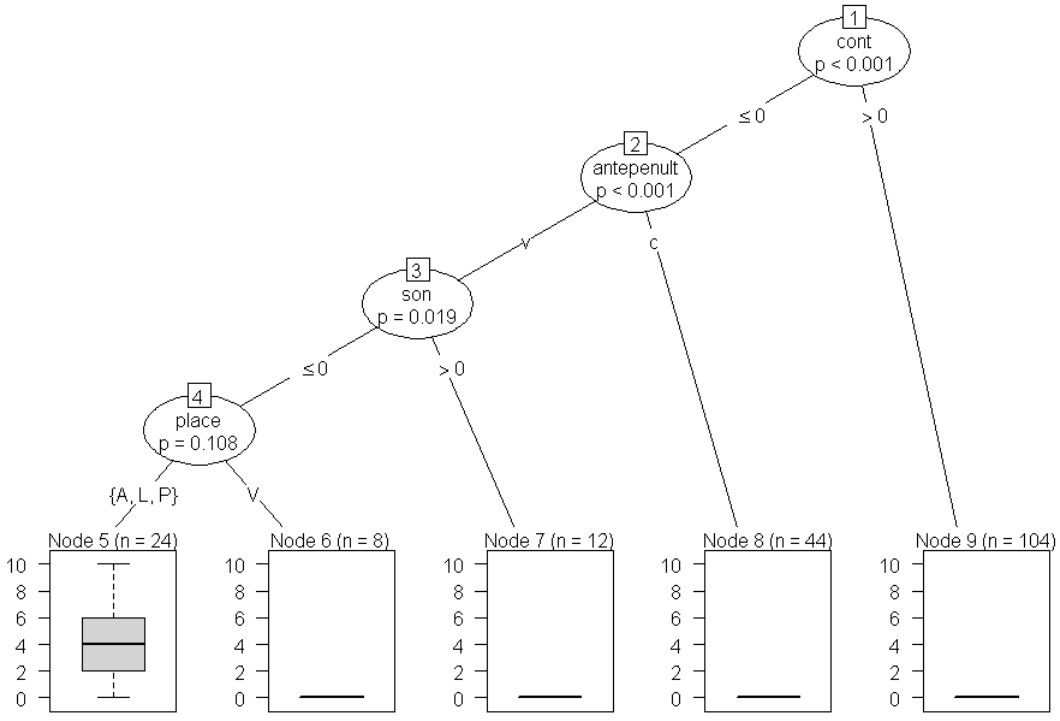
At each split the decision tree is trying to find a feature that, given the features already in the tree, best distinguishes sequences that occur in plurals often from those that occur in plurals rarely (or not at all). However, in some cases this is impossible because the data shows a biconditional or exclusive-or pattern (Strobl et al. 2009). Without the CSP, the Tapa language is one example of this pattern. Figure 4 shows that the common final bigrams [ba], [da], [pa], [ta], [tʃi] and [dʒi] cannot be distinguished from the rare sequences: splits by either vowel identity or consonant identity are uninformative.

Figure 4. Schemas of the Tapa language without the CSP.

a. Final bigrams entered as a possible predictor.



b. Failure of inference without final bigrams as a predictor.



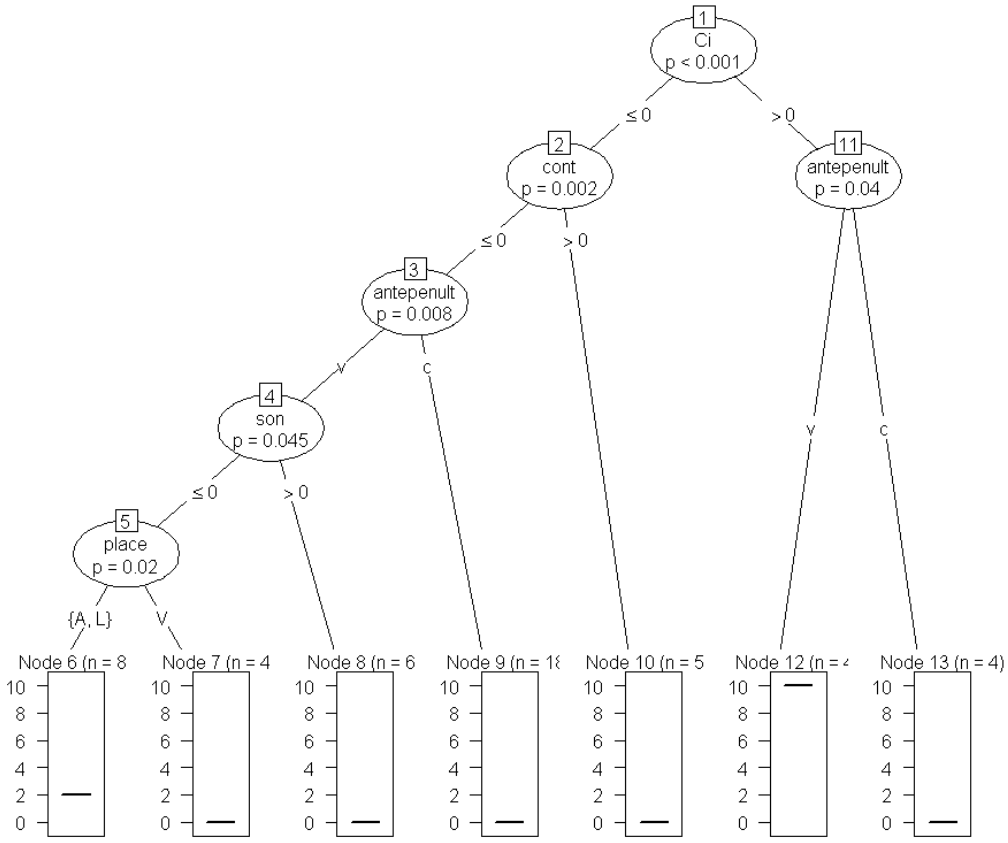
The problem is solved in this case by the CSP but this is not a general solution. It works here because one of the segments forming the to-be-classified bigrams is not part of the input. If both segments were part of the input, the problem would remain. Kapatsinski (2009) presents experimental data regarding the learnability of such patterns. Learners are assigned to either the rime condition or the body condition. In the rime condition, [Cæʃ] or [Cʌg] stems take the affix [mɪn] whereas [Cʌʃ] or [Cæg] stems take the affix [num]. In the body condition, [ʃæC] or [gʌC] stems take the affix [mɪn] whereas [ʃʌC] or [gæC] stems take the affix [num]. In this experiment, neither the vowel nor the consonant alone is predictive of affix choice. Thus, depending on condition, the learner needs to form either rime-affix or body-affix associations for correct prediction. For both prefixes and suffixes, rime-affix associations are shown to be easier to form than body-affix associations, supporting the notion that the rime is a linguistic constituent in English syllables (Fudge 1969, Lee & Goldrick 2009).

Based on these data we suggest that within-stem ngrams that are linguistic constituents should be included as predictors to capture possible cross-over interaction patterns. However, constituency is not all-or-nothing. As suggested by Bybee (2002), "units used together fuse together" (see also Healy 1976). Lee & Goldrick (2008) present experimental data in support of this idea at the subsyllabic level, showing that segmental reordering errors tend to preserve common bigrams more often than rare bigrams, whether those bigrams are rimes or bodies. Following Kapatsinski & Radicke (2009), we take the probability of two units A and B fusing together, and thus the probability that the resulting bigram AB is included as a predictor in the decision tree to be a sigmoidal function of the frequency of the bigram as in (28) where p(AB) is the joint probability of the bigram in the learner's previous experience and LocalityBias is a coefficient indicating a predisposition to consider only short predictors (known to be correlated with age and position on the autism spectrum, e.g., Boyd & Goldberg in press, Johnson 2012).

(32) $$p_{AB\ is\ considered\ for\ tree} \sim \frac{1}{1+LocalityBias^{1-p_{AB}}}$$

If highly-frequent ngrams are added to the tree as predictors, [tʃi], the most common bigram in the language, is very likely to be entered into the tree as a predictor. If this happens, we achieve the tree in Figure 5, which yields the schema PL = tʃi#. This schema is essential to generate the bup→buptʃi error in (15): without it, tʃi-final outputs are produced only by the schema PL = Vtʃi#, which is not satisfied by buptʃi. Thus the existence of errors like [bup→buptʃi] provides support for fusion of [tʃi] into a unit during training on the artificial language.

Figure 5. Top-down schema extraction for –i-final plurals with tʃi# entered in as a binary predictor.



13. Summary and concluding remarks

I have proposed that unknown wordforms are formed from known wordforms by a process of clamoring for blending, where schemas associated with the meaning to be produced and chunks found in the known wordform clamor for candidate outputs that contain them. Competition among candidate outputs is resolved stochastically: a candidate is chosen in proportion to how much support it is getting from the chunks and schemas that have produced it. The process of clamoring for blending is in competition with lexical retrieval, the latter normally winning when the to-be-produced wordform is known.

Schemas are of two kinds: first-order schemas, which describe the forms of words with similar meanings, and second-order schemas, which are generalizations over first-order schemas. First-order schemas gradually become more specific with experience as the learner figures out what the wordforms that share a certain meaning are like in the language s/he is learning. Non-maximally-specific schemas are retained alongside the maximally specific schemas, which provide the tightest possible description of the lexicon. As first-order schemas are acquired, second-order schemas begin forming by a between-schema comparison process.

Like schemas, chunks vary in strength. The strength of a chunk increases when it is repeated in temporally adjacent wordforms, i.e., when it can be thought of as being perseverated (e.g., the strength of [k] increases as a result of encountering and repeating [buk buki]). The strength of a chunk decreases when it is observed not to be retained. Temporal adjacency is not as crucial for this process. A chunk can also be avoided if the speaker has recently experienced difficulty producing it. Chunks also appear to have inherent strength differences, the origin of which is obscure.

In conclusion I would like to emphasize the importance of developing theories that can account for actual linguistic performance rather than pure competence. By considering what *kinds* of experiences lead to a certain kind of linguistic generalization we gain understanding of the roles different kinds of generalizations play in the grammar. For instance, once we realize that rule extraction is dependent on a fallible comparison process, a theory of phonology built on rules becomes much less plausible. The effects of a performance variable can pose major challenges to theories of grammatical competence. For instance, we observe that placing corresponding singulars and plurals next to each other hurts the acquisition of stem changes resulting in good outputs. This fact about performance rules out a purely product-oriented theory of grammatical competence, in which the speaker makes no generalizations about singular-plural mappings and thus temporal relations between singulars and plurals are of no importance. I believe that by pursuing the goal of specifying what learning experiences give rise to a particular kind of linguistic generalization, we should be able to greatly constrain grammatical theory. I hope that C4B will stimulate work on this question by providing a psychologically-plausible and formally explicit theoretical framework linking competence and performance in the domain of morphophonology.

References:

Albright, Adam. 2010. Lexical and morphological conditioning of paradigm gaps. In K. Rice, ed. *When nothing wins: Modeling ungrammaticality in Optimality Theory*. Equinox.

Albright, Adam. 2009. Feature-based generalisation as a source of gradient acceptability. *Phonology*, 26, 9-41.

Albright, Adam. 2003. A quantitative study of Spanish paradigm gaps. *Proceedings of the West Coast Conference on Formal Linguistics*, 22, 1–14.

Albright, Adam, & Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational / experimental study. *Cognition*, 90, 119-161.

Baayen, R. Harald, Tom Dijkstra, & Robert Schreuder. 1997. Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory & Language*, 37, 94-117.

Bateman, Nicoleta. 2007. A crosslinguistic investigation of palatalization. Ph.D. Dissertation, UCSD.

Berent, Iris, Donca Steriade, Tracy Lennertz, & Vered Vaknin. 2007. What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, 104, 591-630.

Berg, Thomas. 1998. *Linguistic structure and change: An explanation from language processing*. Oxford University Press.

Berg, Thomas. 1988. *Die Abbildung des Sprachprodukfionprozess in einem Aktivafionsflussmodel*. Tubingen: Max Niemeyer.

Bergen, Benjamin K. 2004. The psychological reality of phonaesthemes. *Language*, 80, 290-311.

Boersma, Paul. 2004. A stochastic OT account of paralinguistic tasks such as grammaticality and prototypicality judgments. Ms. University of Amsterdam.

Booij, Geert. 2010. *Construction Morphology*. Oxford University Press.

Booij, Geert. 2008. Paradigmatic morphology. In Bernard Fradin, ed. *La raison morphologique. Hommage á la memoire de Danielle Corbin,* 29-38. Amsterdam / Philadelphia: John Benjamins.

Boyd, Jeremy K., & Adele E. Goldberg. In press. Young children fail to fully generalize a novel argument structure construction when exposed to the same input as older learners. *Journal of Child Language*.

Boyd, Jeremy K., & Adele E. Goldberg. 2011. Learning what not to say: Categorization and statistical preemption in "a-adjective" production. *Language*, 87, 55-83.

Braine, M. D. 1987. What is learned in acquiring word classes – A step towards and acquisition theory. In B. MacWhinney, ed. *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum.

Braine, M. D., R. E. Brody, P. J. Brooks, V. Sudhalter, J. A. Ross, L. Catalano, & S. M. Fisch. 1990. Exploring language acquisition in children with a miniature artificial language: Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of Memory and Language* 29: 591-610.

Brooks, P. J., M. D. S. Braine, L. Catalano, R. E. Brody, & V. Sudhalter. 1993. Acquisition of gender-like noun subclasses in artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language* 32: 76-92.

Browman, Catherine P., & Louis Goldstein. 1991. Tiers in articulatory phonology, with some implications for casual speech. In John Kingston & Mary E. Beckman, eds., *Papers in laboratorypPhonology I: Between the grammar and the physics of speech*, 341-76. Cambridge University Press.

Browman, Catherine P., & Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology*, 6, 201-51.

Bybee, Joan L. 2002. Sequentiality as the basis of constituent structure. In Talmy Givón & Bertram F. Malle, eds. *The emergence of language out of pre-language*. John Benjamins.

Bybee, Joan L. 2001. *Phonology and language use*. Cambridge University Press.

Bybee, Joan L. 1985. *Morphology: A study of the relation of meaning and form*. John Benjamins.

Bybee, Joan L., & Dan I. Slobin. 1982. Rules and schemas in the development and use of the English past. *Language*, 58, 265-289.

Bye, Patrik. 2008. Allomorphy: Selection, not optimization. In S. Blaho, P. Bye, & M. Krämer, eds. *Freedom of analysis?*, 63-92. Mouton de Gruyter.

Carlson, Matthew L., & Chip Gerfen. 2011. Productivity is the key: Morphophonology and the riddle of alternating diphthongs in Spanish. *Language*, 87, 510-38.

Charles-Luce, J., & Luce, P. A. 1990. Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language*, 17, 205-15.

Chomsky, Noam, & Morris Halle. 1968. *The sound pattern of English*. Harper & Row.

Clayards, Meghan, Michael K. Tanenhaus, Richard N. Aslin, & Robert A. Jacobs. 2008. Perception of speech reflects optimal use of probabilistic speech cues. *Cognition* 108: 804-809.

Corbin, Danielle. 1989. Form, structure and meaning of constructed words in an associative and stratified lexical component. *Yearbook of Morphology*, 2, 31-54.

Daelemans, Walter, & Antal van den Bosch. 2005. *Memory-based language processing*. Cambridge University Press.

Daland, R., Sims, A. D., and Pierrehumbert, J. B. 2007. Much ado about nothing: A social network model of Russian paradigmatic gaps. *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, 936-943. Prague, Czech Republic.

Dell, Gary S. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283-321.

Dell, Gary S., Lisa K. Burger, & William R. Svec. 1997. Language production and serial order: A functional analysis and a model. *Psychological Review*, 104, 123-47.

Dell, Gary S., Cornell Juliano, & Anita Govindjee. 1993. Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science*, 17, 149-95.

Ernestus,Mirjam, & R. Harald Baayen. 2011. Corpora and exemplars in phonology. In John Goldsmith, Jason Riggle & Alan C. L. Yu, eds. *The handbook of phonological theory, 2nd edition*. Wiley-Blackwell.

Frigo, L., & J. L. McDonald. 1998. Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language*, 39, 218-245.

Frisch, Stefan A., Janet B. Pierrehumbert, & Michael B. Broe. 2004. Similarity avoidance and the OCP. *Natural Language & Linguistic Theory*, 22, 179-228.

Fromkin, Victoria A. 1970. The non-anomalous nature of anomalous utterances. *Language, 46,* 27–52.

Fudge, Erik C. 1969. Syllables. *Journal of Linguistics*, 5, 253-87.

Gerken, L., R. Wilson, & W. Lewis. 2005. Infants can use distributional cues to form syntactic categories. *Journal of Child Language* 32: 249-268.

Goldberg, Adele. 1995. *Constructions: A Construction Grammar approach to argument structure*. University of Chicago Press.

Goldsmith, John. 2001. Unsupervised learning of the morphology of a natural language. *Computational Linguistics*, 27, 153-81.

Goldstein, Louis, Marianne Pouplier, Larissa Chen, Elliot Saltzman, & Dani Bird. 2007. Dynamic action units slip in speech production errors. *Cognition*, 103, 386-412.

Goldwater, Sharon, & Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In Jennifer Spenader, Anders Eriksson, and Östen Dahl, eds. *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, 111–120. Stockholm: Stockholm University, Department of Linguistics.

Hale, Mark, & Charles Reiss. 2003. The Subset Principle in phonology: Why the *tabula* can't be *rasa*. *Journal of Linguistics* 39: 219-244.

Hartsuiker, Robert J. 2002. The addition bias in Dutch and Spanish phonological speech errors: The role of structural context. *Language & Cognitive Processes*, 17, 61-96.

Hayes, Bruce. 2011. Interpreting sonority-projection experiments: the role of phonotactic modeling. *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, China.

Hayes, Bruce. 2008. *Introductory phonology*. Wiley-Blackwell.

Hayes, Bruce. 2004. Phonological acquisition in Optimality Theory: The early stages. In R. Kager, J. Pater & W. Zonneweld, eds. *Fixing priorities: Constraints in phonological acquisition*. Cambridge University Press.

Hayes, Bruce, Petér Siptar, Kie Zuraw & Zsuzsa Londe. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language*, 85, 822-63.

Hayes, Bruce, & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39, 379–440.

Healy, Alice F. 1976. Detection errors on the word *the*: Evidence for reading units larger than letters. *Journal of Experimental Psychology: Human Perception and Performance* , 2(2): 235–42.

Hetzron, Robert. 1975. Where the grammar fails. *Language*, 51, 859-72.

Hooper, Joan Bybee. 1972. A note on inserted and deleted vowels. *Stanford UWP Language Universals*, 10, 141-144.

Hothorn, Torsten, Kurt Hornik & Achim Zeileis. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, 15, 651-74.

Howe, Darin, & Douglas Pulleyblank. 2004. Harmonic scales as faithfulness. *Canadian Journal of Linguistics*, 49, 1-49.

Hudson Kam, Carla, & Elissa Newport. 2005. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning & Development*, 1(2), 151–195.

Johnson, Mark. 1984. A discovery procedure for certain phonological rules. *Proceedings of the 22nd Annual Meeting of the Association for Computational Linguistics*, 344-347. ACL.

Johnson, Matt. 2012. Language generalization in children with autism. Paper presented at the Conference on Sources of Individual Linguistic Differences, Ottawa, March 2-4, 2012.

Johnston, Lamia H., & Vsevolod Kapatsinski. 2011. In the beginning there were the weird: A phonotactic novelty preference in adult word learning. *Proceedings of ICPhS XVII*.

Jordan, M.I., 1996. Computational aspects of motor control and motor learning. In: Heuer, H., Keele, S. (Eds.), *Handbook of Perception and Action, Vol. 2*, 71-120. Academic Press.

Kager, René. 1999. *Optimality Theory*. Cambridge University Press.

Kang, Yoonjung. 2011. Loanword phonology. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice, eds. *The Blackwell Companion to Phonology, vol. IV: Phonological interfaces*, 2258-82. Wiley-Blackwell.

Kapatsinski, Vsevolod. Forthcoming. What statistics do learners track? Rules, constraints or schemas in (artificial) grammar learning. In Stefan Th. Gries & Dagmar Divjak, eds. *Frequency effects in language: Learning and processing*. Berlin: Mouton de Gruyter.

Kapatsinski, Vsevolod. 2010a. Velar palatalization in Russian and artificial grammar: Constraints on models of morphophonology. *Laboratory Phonology*, 1(2), 361-93.

Kapatsinski, Vsevolod. 2010b. Rethinking rule reliability: Why an exceptionless rule can fail. *Chicago Linguistic Society*, 44(2), 277-91.

Kapatsinski, Vsevolod. 2009. Testing theories of linguistic constituency with configural learning: The case of the English syllable. *Language*, 85(2), 248-77.

Kapatsinski, Vsevolod. 2006. To Scheme or to rule: Evidence against the Dual Mechanism Model, In Rebecca T. Cover and Yuni Kim, eds. *Proceedings of the 31st Annual Meeting of the Berkeley Linguistics Society*, 193-204. Berkeley: Berkeley Linguistics Society.

Kapatsinski, Vsevolod. 2005. Productivity of Russian stem extensions: Evidence for and a formalization of Network Theory. M.A. Thesis, University of New Mexico.

Kapatsinski, Vsevolod, & Lamia H. Johnston. 2010. Investigating phonotactics, lexical analogy, and sound symbolism using xenolinguistics: A novel word-picture matching paradigm. In Stellan Ohlsson & Richard Catrambone, eds. *Proceedings of the Annual Conference of the Cognitive Science Society*, 2010-2015. Austin, TX: The Cognitive Science Society.

Kapatsinski, Vsevolod, & Joshua Radicke. 2009. Frequency and the emergence of prefabs: Evidence from monitoring. In R. Corrigan, E. Moravcsik, H. Ouali, & K. Wheatley, eds. *Formulaic Language. Vol. II: Acquisition, loss, psychological reality, functional explanations*. Amsterdam: John Benjamins.

Kenstowicz, M. 1996. Base-identity and uniform exponence: alternatives to cyclicity. In J. Durand & B. Laks, eds. *Current trends in phonology: Models and methods*, 363-93. Manchester, UK: European Studies Research Institute and University of Salford.

Kenstowicz, Michael. 1994. Syllabification in Chukchee: A constraints-based analysis. In A. Davison, N. Maier, G. Silva, & W. S. Yan, eds. *Proceedings of the Formal Linguistics Society of the Midwest 4*, 160-181. Iowa City: University of Iowa.

Kochetov, Alexei. 2011. Palatalization. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice, eds. *The Blackwell Companion to Phonology, vol. III: Phonological processes*, 1666-90. Wiley-Blackwell.

Köpcke, Klaus-Michal. 1988. Schemas in German plural formation. *Lingua* 74: 303-335.

Kruschke, John K. 2008. Bayesian approaches to associative learning: From passive to active learning. *Learning and Behavior* 36: 210-226.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Vol.1: Theoretical prerequisites*. Stanford University Press.

Langacker, Ronald W. 1982. Space Grammar, analyzability, and the English passive. *Language*, 58, 22-80.

Lee, Yongeun, & Matthew Goldrick. 2008. The emergence of sub-syllabic representations. *Journal of Memory & Language* 59.155-68.

Legendre, Géraldine, Yoshiro Miyata, & Paul Smolensky. 1990a. Harmonic Grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, 388–95. Erlbaum.

Legendre, Géraldine, Yoshiro Miyata, & Paul Smolensky. 1990b. Harmonic Grammar: A formal multi-level connectionist theory of linguistic well-formedness: An application. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, 884-91. Erlbaum.

Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Lindblom, B., Lubker, J., Gay, T., 1979. Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *Journal of Phonetics* 7, 147-161.

Lobben, Marit. 1991. *Pluralization of Hausa nouns, viewed from psycholinguistic experiments and child language data*. M.Phil. Thesis: University of Oslo.

Logan, Gordon D. 1983. Time, information, and the various spans in typewriting. In W. E. Cooper, ed., *Cognitive aspects of skilled typewriting*, 197–224. Springer.

Luce, Paul A., & Emily A. Lyons. 1998. Specificity of memory representations for spoken words. *Memory & Cognition*, 26, 708-715.

MacKay, D.G. 1987. *The organization of perception and action: A theory of language and other cognitive skills*. New York: Springer.

Martin, Andrew. 2007. The evolving lexicon. Ph.D. Dissertation, UCLA.

Massaro, Dominic W. 1970. Retroactive interference in short-term recognition memory for pitch. *Journal of Experimental Psychology*, 83, 32-39.

McLennan, Conor T. & 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, 306-321.

Menn, L., & B. MacWhinney. 1984. The repeated morph constraint: Toward an explanation. *Language* 60: 519-541.

Mielke, Jeff. 2008. *The emergence of distinctive features*. Oxford University Press.

Mitchell, Tom M. 1980. The need for biases in learning generalizations. Rutgers University Technical Report CBM-TR-117.

Mitroff. Stephen R., Daniel J. Simons, & Daniel T. Levin. 2004. Nothing compares 2 views: Change blindness can occur despite preserved access to the changed information. *Perception & Psychophysics*, 66, 1268-81.

Mitrovic, Ivana. 2010. Is there a bias for a phonetically natural pattern of velar palatalization? Paper presented at Formal Approaches to Slavic Linguistics 19, College Park, MD.

Moreton, Elliott. 2002. Structural constraints in the perception of English stop-sonorant clusters. *Cognition* 84:55-71.

Nesset, Tore. 2010. Why not? Prototypes and blocking of language change in Russian verbs. In Elzbieta Tabakowska, Michal Choinski, & Lukasz Wiraszka, eds. *Cognitive Linguistics in action: From theory to application and back*, 125-44. Mouton de Gruyter.

Nesset, Tore. 2008. *Abstract phonology in a concrete model: Cognitive Linguistics and the morphology-phonology interface.* Mouton de Gruyter.

Nesset, Tore. 2005. Opaque softening: A usage-based approach. *Poljarnyj Vestnik,* 8, 55-68.

Nevins, Andrew. 2011. Phonologically conditioned allomorph selection. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice, eds. *The Blackwell Companion to Phonology, vol. IV: Phonological interfaces*, 2357-82. Wiley-Blackwell.

Norris, Dennis, & James M. McQueen. 2008. Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review* 115: 357-395.

Ohala, John J. 1981. The listener as a source of sound change. In C. S. Masek, R. A.Hendrick, & M. F. Miller, eds., *Papers from the parasession on language behavior*, 178–203. Chicago Linguistic Society.

Ohala, John J. 1978. Southern Bantu vs. the world: The case of palatalization of labials. *Berkeley Linguistics Society*, 4, 370-86.

Paradis, Carole, & Darlene LaCharité. 1997. Preservation and minimality in loanword adaptation. *Journal of Linguistics*, 33, 379-430.

Paster, Mary. 2005. Subcategorization vs. output optimization in syllable-counting allomorphy. *Proceedings of the West Coast Conference on Formal Linguistics*, 24, 326-333.

Perfors, Amy. 2011. Memory limitations alone do not lead to over-regularization: An experimental and computational investigation. In Laura Carlson, Christoph Hoelscher, & Thomas F. Shipley, eds. *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, 3274-79. Austin, TX Cognitive Science Society.

Pierrehumbert, Janet B. 2006. The statistical basis of an unnatural alternation. In L. Goldstein, D.H. Whalen, & C. Best, eds. *Laboratory Phonology VIII, Varieties of phonological competence*, 81-107. Berlin: Mouton de Gruyter.

Pierrehumbert, Janet B. 1993. Dissimilarity in the Arabic verbal roots. *Papers from the Annual Meeting of the Northeast Linguistics Society*, 23, 367-81.

Pinker, Steven. 1991. Rules of language. *Science*, 253(5019), 530-35.

Plag, Ingo. 2003. *Word formation in English*. Cambridge University Press.

Prince, Alan, & Paul Smolensky. 1993/2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell.

R Development Core Team. 2008. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org.

Redford, Melissa A. 2008. Production constraints on learning novel onset phonotactics. *Cognition*, 107, 785—816.

Reiss, Charles. 2008. Constraining the learning path without constraints, or The OCP and NoBanana. In A. Nevins & B. Vaux, eds. *Rules, constraints and phonological phenomena,* 252-302. Oxford University Press.

Saltzman, Elliot, & Kevin Munhall. 1989. A dynamical approach to gestural patterning in speech production. *Haskins Laboratories Status Report on Speech Research*, 99-100, 38-68.

Schwarz, R. G., Leonard, L. 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, R. G., Leonard, L., Loeb, D., Swanson, L. 1987. Attempted sounds are sometimes not: An expanded view of phonological selection and avoidance. *Journal of Child Language* 14, 411-18.

Shattuck-Hufnagel, Stefanie. 1979. Speech errors as evidence for a serial-order mechanism in sentence production. In W. E. Cooper & E. C. T. Walker, eds., *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, NJ: Erlbaum.

Shattuck-Hufnagel, Stefanie, & Dennis H. Klatt. 1979. The limited use of distinctive features and markedness in phonological speech errors. *Journal of Verbal learning & Verbal Behavior*, 18, 41-55.

Shvachkin, N. K. (1948/1973). The development of phonemic speech perception in early childhood. In C. A. Ferguson, & D. I. Slobin, eds. *Studies of child language development*, 91-127. New York: Holt, Rinehart, and Winston.

Sims, Andrea. 2007. Why defective paradigms are, and aren't, the result of competing morphological patterns. *Chicago Linguistic Society*, 43, 267-81.

Smith, Anne, & Lisa Goffman. 1998. Stability and patterning of speech sequences in children and adults. *Journal of Speech, Language & Hearing Research*, 41, 18-30.

Smith, Linda B., Esther Thelen, Robert Titzer, & Dewey McLin. 1999. Knowing in the context of acting: The task dynamics of the A-not-B error. *Psychological Review*, 106, 235-60.

Stager, Christine L., & Janet F. Werker. 1997. Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, 388, 381-82.

Stanley, Richard. 1967. Redundancy rules in phonology. *Language*, 43, 393-436.

Stark, Craig E. L., & James L. McClelland. 2000. Repetition priming of words, pseudowords, and nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 945-972.

Stemberger, Joseph Paul 1991. Apparent anti-frequency effects in language production: The addition bias and phonological underspecification. *Journal of Memory & Language*, 30, 161-85.

Stemberger, Joseph Paul. 1990. Word shape errors in language production. *Cognition*, 35, 123-57.

Stemberger, Joseph Paul. 1985. An interactive activation model of language production. In A. Ellis, ed. *Progress in the psychology of language*, vol. 1. London: Erlbaum.

Stemberger, Joseph Paul. 1982. The nature of segments in the lexicon: Evidence from speech errors. *Lingua*, 56, 235-59.

Stemberger, Joseph Paul. 1981. Morphological haplology. *Language* 57: 791-817.

Stemberger, Joseph Paul, & Barbara Handford Bernhardt. 1999. The emergence of faithfulness. In B. MacWhinney, ed. *The emergence of language*, 417-446. Mahwah, NJ: Erlbaum.

Storkel, H. L. 2001. Learning new words: Phonotactic probability in language development. *Journal of Speech,Language, & Hearing Research* 44, 1321-37.

Strobl, Carolin, James Malley, & 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, & Richard N. Aslin. 2007. Lexical competition in young children's word learning. *Cognitive Psychology*, 54, 99-132.

Tessier, Anne-Michelle. 2006. Testing for OO-Faithfulness in artificial phonological acquisition. In D. Bamman, T. Magnitskaia & C. Zaller, eds. *Proceedings of the 30th Annual Boston University Conference on Language Development*, 619-39. Somerset, NJ: Cascadilla.

Vaux, Bert. 2009. The Subset Principle vs. Bandwidth Maximization in phonological acquisition. Paper presented at the London Phonology Seminar, 22 April 2009.

Wang, H. S., and B. L. Derwing. 1994. Some vowel schemas in three English morphological classes: Experimental evidence. In M. Y. Chen & O. C. L. Tseng, eds. *In honor of Professor William S.-Y. Wang: Interdisciplinary studies on language and language change*, 561-575. Taipei: Pyramid Press.

Weinert, S. 2009. Implicit and explicit modes of learning: Similarities and differences from a developmental perspective. *Linguistics* 47: 241-271.

Williams, John N. 2003. Inducing abstract linguistic representations: Human and connectionist learning of noun classes. In R. van Hout, A. Hulk, F. Kuiken, & R. Towell, eds. *The lexicon-syntax interface in second language acquisition*, 151-74. Amsterdam: John Benjamins.

Xu, F., & J. B. Tenenbaum. 2007. Word learning as Bayesian inference. *Psychological Review* 114: 245-272.

Zsiga, Elizabeth C. 1995. An acoustic and electropalatographic study of lexical and post-lexical palatalization in American English. In B. Connell & A. Arvaniti, eds. *Phonology and Phonetic Evidence, Papers in Laboratory Phonology IV*, 282-302. Cambridge: Cambridge University Press.

Zuraw, Kie. 2000. Patterned exceptions in phonology. Ph.D. Dissertation, UCLA.