

Grammars leak: Modeling how phonotactic generalizations interact within the grammar

Abstract

In this paper, I present evidence from two unrelated languages—Navajo and English—that weaker, gradient versions of morpheme-internal phonotactic constraints hold even across prosodic word boundaries. In English, for example, compounds like *bookcase*, which contains an otherwise illegal geminate consonant, are statistically underrepresented when compared to compounds with legal clusters, such as *swan dive*. I argue that this is the result of a learning bias which is a consequence of the trade-off during language acquisition between a need to accurately model the data the learner is exposed to and a pressure to construct a grammar that generalizes beyond the data. I formalize this trade-off with a Maximum Entropy phonotactic learning algorithm that assigns weights to a given set of constraints when exposed to training data. The algorithm attempts to maximize the probability of the learning data, but also contains a smoothing term that penalizes complex grammars. When this learner attempts to construct a grammar in which some constraints are blind to morphological structure, and thus cannot distinguish between tautomorphemic and heteromorphemic geminates, the smoothing term forces the learner to assign a small but nonzero weight to a structure-blind constraint against geminates, even when the training data contains equal numbers of compounds with and without geminates. The algorithm, when equipped with structure-blind constraints, is thus inherently biased—it underpredicts the frequency of compounds that violate a morpheme-internal phonotactic. I further show how, over time, this learning bias could plausibly lead to the lexical biases seen in Navajo and English.

1. Introduction

A popular explanation for language change attributes linguistic differences across generations to *mislearning*, that is, cases in which one generation fails to learn exactly the same grammar as that used by the previous generation (e.g., Kiparsky 1968, Lightfoot 1979, 1991, 1999, Clark and Roberts 1993, Hale 1998, Blevins 2004, Hudson Kam and Newport 2005, 2009). In this paper I apply this idea to explain, not diachronic change *per se*, but statistical biases present in the lexicons of Navajo and English. In these languages, compounds license violations of phonotactic constraints that hold within morphemes—for example, in Navajo, multiple sibilants within a root must agree in anteriority, but compounds that combine two roots with disagreeing sibilants are permitted. Likewise, in English,

compounds may contain geminate consonants, which are illegal within morphemes. I show that in both languages, compounds that create violations of morpheme-internal generalizations, although legal, are statistically underrepresented.

I argue that these lexical biases are the result of a bias in the phonotactic learning mechanism. Learners of Navajo and English construct a probabilistic grammar on the basis of the input they are exposed to, but the absence of monomorphemes that violate a given phonotactic constraint causes them to acquire grammars which underpredict the frequency of compounds that violate the same constraint, in what might be described as a “leakage”¹ of a phonotactic generalization from the tautomorphemic domain to the heteromorphemic. Such compounds effectively sound worse to learners than they should, given their frequency. When these same learners go on to form compounds which may in turn become part of the language, this bias will cause them to prefer compounds that obey the constraint, resulting eventually in a lexicon in which compounds that violate the constraint are underrepresented.

My model of this mislearning process consists of two main components. The first is a distinction between two types of phonotactic constraint: *structure-sensitive* constraints, which take into account morphological structure, and *structure-blind* constraints, which ignore morphological structure. Structure-sensitive constraints, which encode generalizations such as “geminate consonants are only permitted across morpheme boundaries,” are necessary in order to correctly model the phonotactic differences between monomorphemes and compounds. Structure-blind constraints, on the other hand, encode generalizations such as “geminate consonants are rare,” which although strictly true in English, fails to capture the reason that geminate consonants are rare, namely their restriction to a specific context. If in addition to structure-sensitive

¹ Although I allude to Sapir’s (1921) famous comment in my choice of terminology (and title), I do not use the word “leak” in the same sense that Sapir did—he was referring to the tendency for linguistic generalizations to have exceptions (“no language is tyrannically consistent”), while I use the word to mean a process by which a generalization in one domain influences a generalization in another domain.

constraints, learners are also equipped with structure-blind constraints (possibly representing a holdover from very early phonotactic learning), the resulting grammar may be biased against compounds that violate stem-internal generalizations, even when there is no bias in the input data.

The conditions under which this leakage can occur are formally described by the second component of the model, a Maximum Entropy (*MaxEnt*) learning algorithm which assigns probabilities to the space of possible words by constructing a grammar of weighted phonotactic constraints. The algorithm incorporates the idea, prevalent in the machine learning literature, that learning probabilistic generalizations involves a trade-off between accuracy and generality. One consequence of this tradeoff is that under certain conditions learners sacrifice accuracy for generality, choosing a grammar that does not model the training data perfectly, but is more general than grammars that are more consistent with the data. Combined with the structure-blind constraints described above, this learning algorithm can account for the mislearning in Navajo and English.

In addition to the learning model, I also propose a model of how newly formed words compete with existing words to be used by speakers, and thereby become part of the language. Assuming that the phonotactic grammar constructed by the learning algorithm influences the creation or adoption of novel words, the model predicts that over several generations, the lexicon used by a speech community will acquire a bias against compounds violating a stem-internal phonotactic constraint, even if the initial lexicon exhibits no such bias. I also show that, assuming that compound formation is influenced by both semantic and phonological factors, the lexical underrepresentation will remain stable across generations—despite certain compound types being dispreferred by speakers, they never completely die out.

The remainder of the paper can be broadly divided into three parts. First, in Sections 2 and 3, I describe the Navajo and English data respectively. Second, in Section 4, I explicate the Maximum Entropy learning algorithm and show that, when given certain constraints, the algorithm consistently underpredicts the frequency of compounds that violate a tautomorphic constraint. Third, in Section 5, I use a simulation of multi-generational learning to demonstrate that the bias introduced by the learning model can result in a stable lexical pattern in which the dispreferred compounds are underrepresented.

2. Navajo

2.1. Navajo sibilant harmony

All sibilants in a Navajo root must agree in their specification for the [anterior] feature; thus, a single root can only contain sibilants that are either all anterior or all posterior (Sapir and Hoiyer 1967, Kari 1976, McDonough 1991, 2003, Fountain 1998). The two sets of consonants are summarized in the chart below.

(1) Navajo sibilant classes

[+anterior]	[-anterior]
s	ʃ
z	ʒ
ts ^h	tʃ ^h
ts	tʃ
ts'	tʃ'

Thus, for example, roots like /tʃ'óʒ/ 'worm' or /ts'ózi/ 'slender' are attested, but */soʃ/ is not a possible Navajo root.

This is not only a cooccurrence restriction on roots—sibilants in affixes must also agree in anteriority with sibilants in the root, resulting in alternations in sibilant-bearing

affixes (Sapir and Hoijer 1967). The examples in (2) demonstrate the alternations in prefixed forms (sibilants are in bold).

(2) Examples of sibilant harmony (Fountain 1998)²

- (a) /ji-**s**-lééʒ/ → [ji-**ʃ**-tlééʒ] ‘it was painted’
 (b) /ji-**s**-tiz/ → [ji-**s**-tiz] ‘it was spun’

Typically, assimilation proceeds from the root to the prefixes.

In compounds, however, which contain multiple roots, sibilant harmony does not necessarily apply, meaning that such words can contain disagreeing sibilants.³

(3) Exceptions to sibilant harmony in compounds (Young and Morgan 1987)

- (a) **tʃ**éí- **ts**’iin ‘rib cage’
 heart bone
 (b) **ts**^hé- **tʃ**ééʔ ‘amber’
 stone resin

In the next section, I will show that compounds in Navajo, although they may violate sibilant harmony, tend to combine roots whose sibilants already agree.

2.2. Navajo compounds

The data described here are taken from the Young and Morgan (1987) dictionary of Navajo. From this dictionary a list of all compounds containing exactly two sibilants, each sibilant in a different root, was compiled, a total of 140 words—this represents all the words that could potentially violate sibilant harmony. Because sibilant cooccurrence in compounds is sensitive to distance, the data discussed here are limited to the subset of these words in

² A note on transcriptions: Navajo examples are given in IPA, with acute accents marking high tones (low tones are unmarked).

³ A handful of compounds do undergo sibilant harmony, such as *tsaa-nééʒ* ‘mule’, from /tʃaa/ ‘ear’ + /nééʒ/ ‘long’ (Sapir and Hoijer 1967). I suspect that these words undergo harmony because they have been stored as single units by speakers due to their semantic opacity, but I have included them in the analysis in their underlying (i.e., disagreeing) form, on the assumption that the sibilants disagreed when the compound was originally formed.

which the sibilants are in adjacent syllables (there were no cases in which sibilants were in the same syllable, but different roots), a total of 97 words. Representative examples are given in (4).

(4) Examples of compounds with two sibilants in adjacent syllables (one per root)

- | | | | | |
|-----|--------------------------|---------------|-----------|--------------|
| (a) | ts^hee- | ts'iin | | 'tailbone' |
| | tail | bone | | |
| (b) | k'iif- | ʒin- | ii | 'blue beech' |
| | alder | black | one | |
| (c) | ts^hé- | zéi | | 'gravel' |
| | rock | crumbs | | |

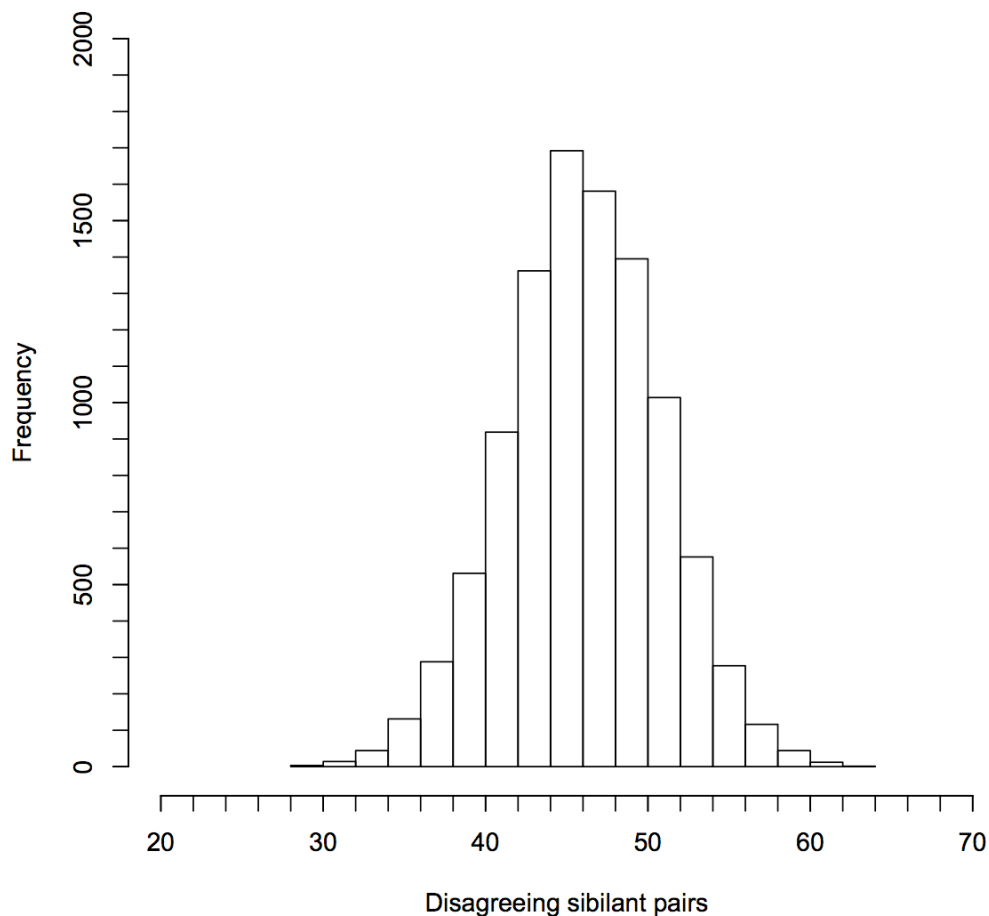
Of these 97 words, 29 (29.9%) contain disagreeing sibilants, violating the stem-internal phonotactic. In order to determine whether this agreement rate significantly differs from chance, I used a Monte Carlo procedure (Kessler 2001) to approximate the distribution of the expected rate. The procedure is described in detail in the following section.

2.3. *The Monte Carlo test for significance*

The Monte Carlo test is performed by randomly recombining the roots that make up the set of compounds in question. The initial root in each compound is combined with another root, pseudo-randomly selected from the same list of compounds (position in the compound is fixed, so that initial roots always remain initial, and final roots remain final). After each such shuffling, the number of disagreeing sibilant pairs under the new permutation is calculated, and the entire process may be repeated as many times as necessary. If the process is repeated sufficiently many times, the result will be a reliable estimate not only of the average expected number of disagreeing sibilants that would occur by chance, but of the entire distribution of this expected value. With this information, we can determine how likely the actual, attested value is.

The histogram in (5) presents the results of the Monte Carlo procedure on the entire list of Navajo compounds which contain exactly two sibilants in different roots. The x -axis represents the number of sibilant pairs (out of a total of 97) in which the sibilants disagreed in anteriority, and the y -axis represents the number of iterations (out of 10,000 total⁴) in which a given disagreement rate occurred.

(5) Results of Monte Carlo test on Navajo two-sibilant compounds



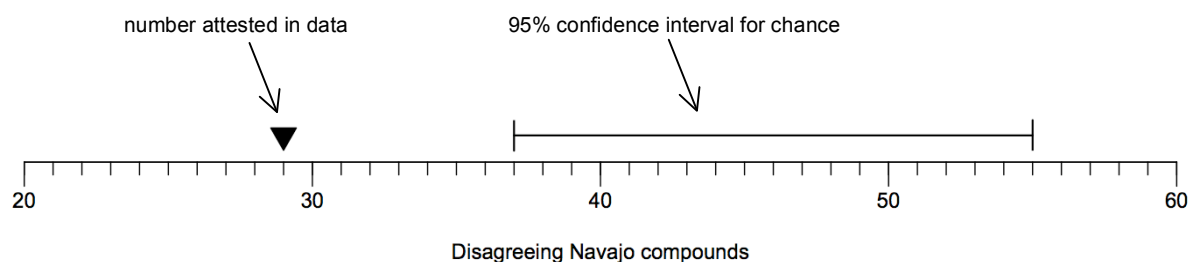
The histogram shows that the values generated in the Monte Carlo test are approximately normally distributed around the mean value of 46.0. The actual number of disagreeing sibilants in the Young and Morgan data, 29 pairs, is extremely unlikely to have arisen by chance—a value this low only occurred three times out of the 10,000 iterations of

⁴ All of the Monte Carlo tests reported in this paper were performed using 10,000 iterations each.

the Monte Carlo test. From this we can conclude that the actual disagreement rate is significantly ($p < 0.001$) below chance.

In the remainder of this paper, Monte Carlo results will be summarized by omitting the histogram and simply reporting the 95% confidence interval and the actual value, as in the chart in (6) (which represents the same test reported in (5)). In this and succeeding charts, the triangle indicates the actual value found in the data (in this case the actual number of Navajo compounds with disagreeing sibilants), while the horizontal bar represents the 95% confidence interval derived from the Monte Carlo test.

(6) Comparing attested Navajo sibilant pairs to Monte Carlo results



The results of this test show that Navajo compounds tend to obey the sibilant harmony constraint, even though violations of the constraint are permitted, and that this tendency is unlikely to be due to chance. In the next section I present a parallel case from English.

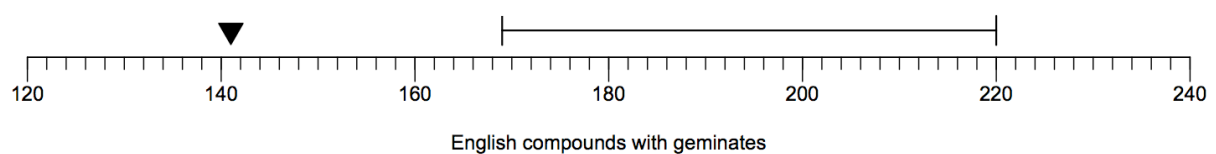
3. English

Geminate consonants in English are permitted only across morpheme boundaries (Hammond 1999, Ladefoged 2001, Kaye 2005). Words like *unknown*, *solely*, and *bookcase* are typically pronounced with geminates that have been created by combining morphemes that end and begin with the same consonant. These morphologically-created geminates are often called “fake geminates” (Hayes 1986) to differentiate them from morpheme-internal

long consonants—the two types of geminate frequently exhibit different phonological behavior (Payne 2005, Ridouane 2007, Oh and Redford 2009, Pycha 2010). Minimal pairs differing only in consonant length, as in the compounds *carpool* and *carp pool*, may be found in multimorphemic words; in monomorphemic words, however, no such minimal pairs exist—the hypothetical word [hæppi], which would form a minimal pair with existing *happy* [hæpi], is not a possible monomorpheme of English. In the following sections, I show that geminate consonants created by compounding are statistically underrepresented in the lexicon of English.

In order to determine the number of compounds in English that contain geminates, I extracted all of the words marked as noun-noun compounds from the lemmatized version of the CELEX database (Baayen et al. 1993), a total of 4,758 words. Of these, 141 words (3.0%) contain fake geminates—e.g., *bus stop*, *hat trick*, *penknife*, *bookkeeper*. The results of a Monte Carlo test on the CELEX compounds are shown in (6).

(7) Geminates are underrepresented in English compounds

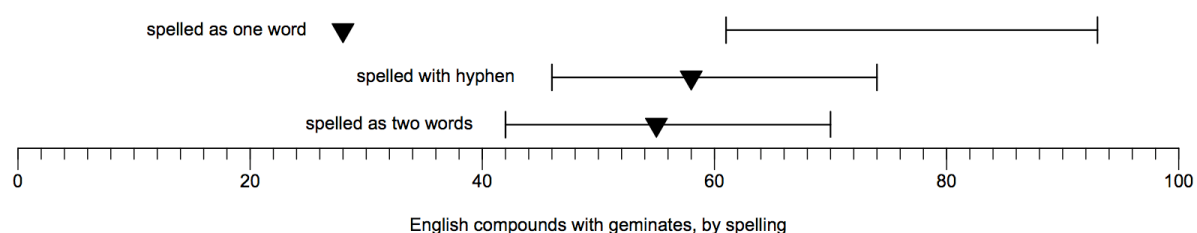


As the chart makes clear, the number of geminates found in the actual compounds, 141 (3.0%), is significantly lower than expected ($p < .001$).

Before accepting that this result tells us something about the compound formation behavior of English speakers, however, a potential confound must be dealt with. The compounds listed in CELEX were collected by applying an automatic parser to a large text corpus. This raises the possibility that some compounds spelled as separate words (e.g., *sand dune*), might have been misidentified by the parser as separate words. This could bias the

results of the Monte Carlo test because of the fact that compounds with geminates are more likely to be spelled with a hyphen or space between the members than as a single word (Sepp 2006). The underrepresentation of geminates could thus be an artifact of the parsing process, combined with people’s tendency to spell compounds according to their junctural phonotactics. The effects of the spelling bias are shown in (8), which makes it clear that the underrepresentation is limited to those compounds that are spelled as a single word (this chart depicts the same set of words from CELEX described in (6), divided according to how they are spelled in the CELEX entry).⁵

(8) Compound spelling is biased by presence of geminate

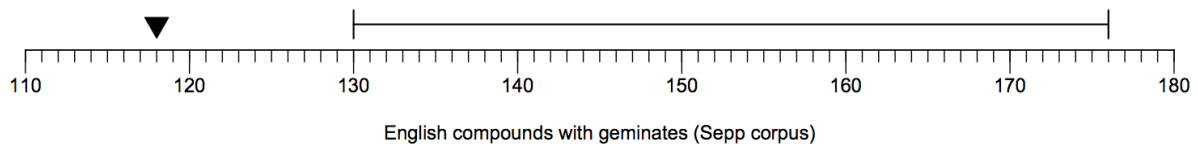


To show that the geminate underrepresentation in English is not solely an artifact of the spelling bias, I ran the same Monte Carlo test on the list of compounds compiled by Sepp (2006) from a 14-million-word corpus of written American English (see Sepp 2006 for details of the construction of this corpus). Sepp used a part-of-speech tagger and computational parser to extract all potential noun-noun compounds (including any sequence of nouns separated by a space), and then further filtered the list by hand, removing all non-compounds. Because every compound was checked by hand, the likelihood of undercounting open compounds is lower than it would be if all parsing were done by algorithm.

⁵ Although the same compound can be spelled different ways by different writers, each compound is listed with a single spelling in CELEX. It is unclear how this spelling was determined. My intuition is that nearly all of the words spelled with hyphens in CELEX would be most often spelled with a space by native speakers (e.g., *space-vehicle*, *rabbit-hutch*, *slot-machine*), a suspicion that is strengthened by the nearly indistinguishable behavior of hyphenated and spaced compounds in (8). This accords with Sepp’s findings that fewer than 5% of the noun-noun compounds in her corpus are spelled with a hyphen more often than either with a space or as one word (many of those are either dvandva compounds (*Clinton-Gore*, *hip-hop*), or involve abbreviations (*op-ed*)).

Of the 3,222 noun-noun compounds that occur in Sepp’s corpus (including both those spelled as single and separate words), 118 (3.6%) contain false geminates. The results of a Monte Carlo test on these compounds, shown in (9), demonstrate that, just as with the CELEX compounds, the actual number of geminates is significantly lower ($p < .01$) than the mean expected number of 152.1 (4.7%).

(9) Geminates are underrepresented in compounds in Sepp corpus



Thus, even when the risk of a counting bias is minimized by careful hand-checking, geminates are still underrepresented in compounds overall. This suggests that any orthographic bias that may exist is in addition to a general bias against forming compounds that create geminates.

4. Discussion

The phonotactics in English and Navajo discussed above obey the same generalization: some phonotactic constraint holds within morphemes, and a weaker version of the same constraint holds across morpheme boundaries. Morpheme boundaries, in other words, license violations of phonotactic constraints, but only up to a point. Tautomorphemic and heteromorphemic phonotactic generalizations are thus in some sense entangled rather than computed independently. This property of the phonotactic grammar could simply be stipulated to be part of Universal Grammar—I will argue, however, that this entanglement follows from more general constraints on the human phonotactic learning mechanism. More specifically, I will describe a learning algorithm that when trained on data which exhibits a

tautomorphemic phonotactic restriction, cannot help but learn a grammar that encodes a weaker version of the same restriction across morpheme boundaries.

The account that I will present of the facts in Navajo and English is thus a theory of linguistic competence. Another possible account of the same facts might appeal instead to universal performance factors. On this view, the tautomorphemic and heteromorphemic phonotactics could both result from the same phonetic pressure. Imagine, for example, that sequences of agreeing sibilants are easier to articulate (or process) than disagreeing sequences. This fact of human physiology could have become phonologized within Navajo roots, resulting in a categorical sibilant harmony constraint, and could also exert a pressure on compound formation, resulting in the underrepresentation of disagreeing compounds. Under this hypothesis, the tautomorphemic constraint does not cause the heteromorphemic constraint; rather, they are both caused by the same underlying factor. A learning bias would thus not need to be invoked to explain the Navajo facts.

Such a performance-only theory, however, runs into empirical problems regarding the universality of these performance factors. If it is true that disagreeing sibilant sequences are universally more difficult than agreeing sequences, and this difficulty exerts a pressure on the contents of the lexicon, then it would be surprising to find many languages in which disagreeing sibilants are *over*represented. English, however, is such a language: /s..ʃ/ sequences are overattested compared to /s..s/ sequences (Berkley 1994, 2000). Gradient similarity *avoidance* in consonants, in fact, appears to a robust cross-linguistic phenomenon, and has been shown in a number of languages: Arabic (Frisch et al. 2004), Maltese (Frisch et al. 2004), Italian (Frisch et al. 2004), Muna (Coetzee and Pater 2008), Japanese (Kawahara et al. 2005), and Russian (Padgett 1995). In the most complete study of the phenomenon to date, Pozdniakov and Seeger (2007) find evidence for gradient similarity avoidance in 31

different languages, from which they conclude that it “is a likely universal property of human language” (308). If anything, the evidence for universal performance factors appears to point in the opposite direction of the Navajo pattern.

Likewise the preference for consonant clusters over geminates seen in English is not universal. In Japanese, for example, geminate consonants are permitted within morphemes while non-geminate clusters (with the exception of homorganic nasal+stop clusters) are not (e.g., /happa/ ‘leaf’, but */hapta/). Furthermore, non-geminate clusters are repaired when created by the morphology (e.g., /kak/ ‘write’ + /ta/ ‘PAST’ → [kaita] ‘wrote’; c.f. /kat/ ‘win’ + /ta/ ‘PAST’ → [katta] ‘won’). Luganda is similar to Japanese in that it allows geminates but prohibits other cluster types (Wiltshire 1999). In some languages, like English, geminates are the worst type of cluster; in others, like Japanese, they are the best.

This is not to say that functional, performance-based factors play no role in these phonotactic generalizations. Rather, it is likely that there are multiple functional factors, some of which are at odds with others. Agreeing sibilant pairs and disagreeing pairs may each be considered “better” along different dimensions, and each language represents a separate compromise among these competing forces. My claim is simply that the correlations between tautomorphemic and heteromorphemic phonotactics in Navajo and English are unlikely to be the result of performance *alone*—the speakers of these languages must learn the language-particular phonotactic generalizations present in morphemes, which they then extend to the formation of complex words. In the remainder of this section I propose a mechanism by which this overgeneralization can occur.

4.1. *The phonotactic learner*

In order to model the gradient constraints observed in the compound data presented above, I will assume that speakers make use of a grammar that assigns probabilities to possible words (Hayes and Wilson 2008). On this view, the actual lexicon can be thought of as a finite sample drawn from this probability distribution—the learner’s task is to reconstruct the grammar, and therefore the distribution, based on that sample. In the case of English, which I will use throughout the rest of the paper to illustrate how the learner works, the final grammar should assign very low probabilities to words containing stem-internal geminates, high probabilities to words containing only legal clusters, and intermediate probabilities to words containing geminates across morpheme boundaries.

It would be trivial to construct a probabilistic learner that, fed the lexicon of English, could learn that compounds with geminates are underrepresented. Any algorithm that is capable of simply counting the number of compounds with and without geminates could succeed at this task. My goal, however, is not to produce a learner that correctly learns the patterns in Navajo and English, but to explain why Navajo and English are the way they are. My strategy will therefore be to describe a learner which systematically mislearns generalizations across morpheme boundaries—when given input in which the two compound types are equally frequent, it will nonetheless construct a grammar that assigns a lower well-formedness value to compounds which violate the stem-internal phonotactic.

4.2. *Maximum Entropy grammars*

The Maximum Entropy formalism has long been a staple of the machine learning literature (Jaynes 1957, Berger et al. 1996, Abney 1997, Della Pietra et al. 1997), and has recently been successfully applied to problems of phonological learning (Goldwater and

Johnson 2003, Jäger 2007, Fischer 2005, Wilson 2006, Hayes and Wilson 2008, Hayes et al. 2009). A MaxEnt learning algorithm learns a probability distribution over the members of some set given a sample drawn from that distribution.

A MaxEnt grammar consists of a set of numerically weighted constraints. The constraints that will be used in this paper ban structures in the output (e.g., “no geminates within a morpheme”), and are equivalent to the markedness constraints used in Optimality Theory (Prince and Smolensky 2004). Unlike classical Optimality Theory, in which constraints are strictly ranked, each constraint in a MaxEnt grammar has a weight, represented by a real number, which represents the “strength” of that constraint. The set of constraints and their weights (which together constitute the grammar) determine a probability for every possible surface form, which is a function of the set of constraints violated by the form and their weights. Specifically, a word’s probability is a function of what Hayes and Wilson (2008) call its *score* $h(x)$, which is calculated by simply summing the (weight \times number of violations) for every constraint in the grammar, as shown in (10).

(10) Definition of score

$$h(x) = \sum_{i=1}^M w_i C_i(x)$$

where

M : number of constraints

w_1, w_2, \dots, w_M : constraint weights

x : representation of candidate

$C_i(x)$: number of violations assigned to x by constraint C_i

For a grammar with three constraints, for example, C_1 (weight 1.0), C_2 (weight 2.0), and C_3 (weight 3.0), an output form x violating C_1 twice and C_3 once would be assigned a score of $h(x) = (1.0 \times 2) + (2.0 \times 0) + (3.0 \times 1) = 2.0 + 0 + 3.0 = 5.0$.

A word’s score is identical to its *harmony* in the Harmonic Grammar framework (Legendre et al. 1990, Smolensky and Legendre 2006); unlike the Harmonic Grammar

approach, however, in a MaxEnt grammar a word's probability is directly related to its score. The equation in (11) describes how scores are mapped to probabilities (Ω represents the set of all possible words).

(11) Determining candidate probability

$$P(x) = \frac{e^{-h(x)}}{\sum_{y \in \Omega} e^{-h(y)}}$$

The goal of the learning algorithm is to reproduce the probability distribution over constraint violations in the learning data. It does this by adjusting the constraint weights so as to maximize the probability of the data—the algorithm thus represents an example of *maximum likelihood* learning. The probability of the data is calculated by simply multiplying the probabilities of all of the words in the data to arrive at their joint probability, equivalently stated as the sum of the log probabilities of each word, or $\sum_{i=1}^N \log P(x_i)$, where N is the number of words encountered during learning.

An algorithm which simply maximizes the probability of the data, however, is prone to *overfitting*. Because the learner is only given a finite sample of data, a pure maximum likelihood learner will tend to overestimate the probability of items that are in the sample, and underestimate the probability of items that didn't happen to occur in the sample. In other words, a probability distribution learned from a finite sample will tend to be skewed in the direction of the observed data.

The standard way to avoid overfitting is to introduce a smoothing term into the learning function (Martin et al. 1999). The smoothing term penalizes skewed distributions and causes the learner to favor more uniform distributions, which ameliorates the tendency to overfit. I use a Gaussian prior over the constraint weights, which prefers that the constraint

weights be as uniform as possible. The prior term is subtracted from the likelihood term, resulting in the learning function in (12).

(12) MaxEnt learning function

$$\sum_{i=1}^N \log P(x_i) - \sum_{j=1}^M \frac{(w_j - \mu_j)^2}{2\sigma_j^2}$$

The Gaussian prior assesses a penalty for constraint weights that deviate from their ideal weights, represented by μ_j . In the implementation of the algorithm I will use, μ is set to zero for all constraints, so that the prior penalizes any nonzero weight, with the size of the penalty increasing with the square of the weight. This pressure towards low constraint weights translates in a bias against highly skewed distributions—because the prior term increases with the square of each constraint weight, it prefers grammars with many low-weighted constraints over grammars with a few high-weighted constraints. This means that if multiple constraints are each capable of explaining a given property of the data, the learner will assign all of the constraints low weights rather than choose one and assign it a high weight. This property of the prior will prove crucial in modeling the English and Navajo data.

The learning function thus embodies a trade-off between a pressure to model the data as accurately as possible and a pressure to have as general (i.e., uniform) a grammar as possible. The value of the free parameter σ determines the relative importance of each of these factors. Modeling the connection between tautomorphemic and heteromorphemic phonotactics will rely crucially on this trade-off.⁶

⁶ The specific learning algorithm used in the learning simulations reported in this paper is Jäger's (2007) Stochastic Gradient Ascent (SGA), which estimates (to an arbitrary degree of precision) the maximum of the learning function in (12). The Gaussian prior is implemented in the SGA by simply decreasing every constraint weight w by an amount $2\sigma w$ after every learning datum (Johnson 2007). Note that the learning bias I discuss in this paper is a property of the learning function itself, and is independent of which algorithm learners actually use to calculate the maximum of that function—the same results can be obtained with the conjugate gradient algorithm used by Hayes and Wilson (2008), for example.

4.3. Testing the MaxEnt learner on simplified English

Let us consider how the learner described in the previous section would handle a schematic representation of the English facts. Imagine a simplified version of English with two salient features: every word is either a monomorpheme or compound, and every word contains exactly one consonant cluster. Each cluster may be non-geminate, which I will represent as τp , or geminate, represented as pp . A morpheme boundary intervening between the consonants of a cluster is indicated by a “+”. This gives us a set of four codes, listed in (13), with which we can label all of the logically possible words in this simplified English.

- (13) Logically possible word types in simplified English
- | | |
|----------|-------|
| τp | pp |
| $\tau+p$ | $p+p$ |

Because geminates are illegal within morphemes in English, only three of these word types are attested: τp , $\tau+p$, and $p+p$.

Now imagine a learner who is exposed to data consisting of 1,000 words of each of the three legal types τp , $\tau+p$, and $p+p$. What kind of generalizations could the learner form? A learner who could reliably identify morpheme boundaries would be able to learn that geminates are illegal within morphemes, legal across morpheme boundaries, and furthermore that where they are legal, geminates and non-geminate clusters are equally frequent. A learner who was blind to morpheme boundaries, however, could only learn that while geminates are attested, they are not as frequent as other types of cluster—in the training data described above, 1,000 words have geminates, and 2,000 do not.

I propose that human learners combine both *structure-sensitive* and *structure-blind* constraints within a single grammar. Speakers of English know that geminates are legal across morpheme boundaries but not within morphemes, but they have also encoded in their grammars the fact that geminates are not as frequent as non-geminate clusters. When all of

these generalizations are combined in a single grammar (precisely how they are combined will be described below), the result is a strong preference for non-geminate clusters within morphemes, and a mild preference for non-geminate clusters across morpheme boundaries.

Two questions are immediately apparent in this scenario. First, why would learners use structure-blind constraints at all, given that structure-sensitive constraints by themselves permit more accurate modeling of the input? Second, even if we allow that learners are furnished with structure-blind constraints, why would they assign them nonzero weights, resulting in a mislearning of the probability distribution apparent in the learning data? In the MaxEnt learning framework, it is possible to assign zero weights to useless constraints, effectively turning them off—in fact, this is what a purely maximum likelihood learner must do if giving those constraints nonzero weights would lower the probability of the data. My proposal thus might appear to involve a learner who is saddled with useless constraints, and who then perversely goes out of its way to use those constraints, to its own detriment.

In answer to the first question, regarding the motivation for having structure-blind constraints in the first place, one possibility is that structure-blind constraints represent a holdover from an early stage of phonotactic learning. There is substantial evidence that infants learn a great deal about the phonotactic patterns in their language before they are able to parse the speech stream into morphemes or even words (Peters 1983, Jusczyk 1997). Generalizations formed at this stage are by necessity structure-blind. Of course, once they master morphology, children are presumably able to make use of structure-sensitive constraints. It is possible, though, that the structure-blind constraints they used at the earlier stage remain in the grammar into adulthood (one could imagine, for example, that there is some cost to removing constraints once they are part of the grammar).

Another possibility is that the mechanism responsible for positing constraints is biased towards more general constraints (see Hayes and Wilson 2008 for a proposal in this vein). A constraint that simply bans geminates is more general than a constraint that bans geminates only across morpheme boundaries, and so might be included in the grammar on that basis. Of course, answering this question more definitively would require a more thorough understanding of how language-learning infants go about constructing phonological constraints, a topic on which little research has yet been done.

The answer to the second question, as to why learners would assign nonzero weights to structure-blind constraints even at the cost of accuracy, is provided by the Gaussian prior that is part of the MaxEnt learning algorithm. Although this prior is well motivated on mathematical grounds (without it, the learner will assign infinite weights to constraints that are never violated in the data), it can cause weights to be assigned to constraints which are not strictly necessary to explain the data. To see why this occurs, I will now describe the results of several learning simulations on the simplified English data.

The training data I will use for all the demonstrations of the learner is given in (14). The numbers of each word type were chosen so that there would be an equal number of monomorphemes and compounds, and an equal number of compounds with geminates and compounds without geminates.

(14) Training data

Word type	Number of words
pt	2,000
t+p	1,000
p+p	1,000

The constraints the learner will start with are given in (15). Note that a plus sign in parentheses indicates an optional morpheme boundary, while the absence of a plus sign (as in *pp) indicates that no morpheme boundary intervenes between the consonants.

(15) Constraints

Structure-blind

- *p(+)p no geminates
- *p(+)t no non-geminate consonant clusters

Structure-sensitive

- *pp no geminates within a morpheme
- *pt no non-geminate consonant clusters within a morpheme

- *p+p no geminates across a morpheme boundary
- *p+t no non-geminate clusters across a morpheme boundary

When the MaxEnt learning algorithm (with the smoothing term) is given these constraints, and exposed to the data in (14), it arrives at the grammar in (16).

(16) Learning results, all constraints⁷

Constraint	Weight
*p(+)p	0.146
*t(+)p	-0.146
*pp	4.211
*tp	-0.323
*p+p	0.038
*t+p	0.292

Word type	Predicted probability	Probability in training data
pp	0.004	0.000
tp	0.483	0.500
p+p	0.251	0.250
t+p	0.261	0.250

The algorithm assigns a high weight to *pp, which is unsurprising due to the lack of pp sequences in the training data. More surprising is the fact that *p(+)p also receives a small but nonzero weight. This is the effect of the Gaussian prior, which is optimized by making

⁷ In all of the learning simulations presented in this paper, the data was presented to the learner three times. The plasticity (the amount by which constraint weights are perturbed with each incoming learning datum) was set to 0.1 for the first presentation, 0.01 for the second, and 0.001 for the third. The Gaussian prior parameter α was set to 0.01 for the first presentation, 0.001 for the second, and 0.0001 for the third. The results reported are averaged over 100 consecutive runs of the algorithm.

the distribution of weights as uniform as possible. Assigning a weight to *p(+)p lowers the probability of pp sequences, which allows the weight on *pp to be lower. The price of this more uniform distribution is accuracy in modeling the data—the weight on *p(+)p also slightly lowers the predicted probability of p+p sequences, making them appear to the learner to be slightly less frequent than t+p sequences.

This bias against compounds with geminates disappears if the structure-blind constraints are removed from the constraint set, and the learner constructs a grammar using only structure-sensitive constraints. The results of this structure-sensitive-only learning, using the same training data described in (14), are shown in (17).

(17) Learning results, structure-sensitive constraints only

Constraint	Weight	Word type	Predicted probability	Probability in training data
*pp	4.440	pp	0.004	0.000
*tp	-0.403	t <p< td=""> <td>0.480</td> <td>0.500</td> </p<>	0.480	0.500
*p+p	0.220	p+p	0.258	0.250
*t+p	0.219	t+p	0.258	0.250

The weight given to *pp in this grammar is 4.440, as compared to the weight of 4.211 that was assigned to the same constraint by the learner using the structure-blind constraints. This shows that a constraint's weight is dependent not just on the properties of the data, but also on the other constraints that are present in the grammar. In this case, *pp gets a higher weight when there is no other constraint that could also explain the absence of pp in the data. When the structure-blind constraints are included in the grammar, this generalization is split between two constraints, *pp and *p(+)p, which allows the weight on *pp to be slightly lower. The result is a grammar that evaluates compounds with and without geminates as

equally probable. The presence of structure-blind constraints thus plays a crucial role in producing the phonotactic leakage observed in actual languages.⁸

5. The evolution of the lexicon

I have argued that learners consistently underpredict the frequency of compounds that violate a stem-internal phonotactic in their language, resulting in a grammar that assigns such compounds a lower well-formedness value than other compounds. In order to explain why such compounds are underrepresented in Navajo and English, we must further assume that speakers are biased by their phonotactic grammar when they create new compounds (or decide to use novel compounds coined by others). Even if earlier versions of these languages had been unbiased, over time generations of learners would have altered the lexicon of each language, making words that violate stem-internal phonotactics less frequent. However, this model raises another problem—if each generation ends up making the language more biased than the previous generation, given enough time we would expect the underrepresented words to die out completely. Why does English allow geminate consonants in compounds at all, given the rapid turnover in vocabulary typically observed in languages over time?⁹

In order to answer this question, I will present the results of a simulation of multi-generational lexical change which incorporates the learning algorithm described in Section 4. I will assume, following Boersma 2007, that lexical change is driven largely by competitions between synonymous lexical items. New words are created or borrowed by speakers, and must compete with existing words that have the same meaning. The tendency of speakers

⁸ Note that even without the structure-blind constraints, the learner still does not perfectly replicate the frequencies in the training data. This is due to the presence of the Gaussian prior, which pushes the learned distribution towards uniformity.

⁹ For example, roughly 85% of the Old English vocabulary is no longer in use (Baugh and Cable 1993), and more than 80% of the Modern English vocabulary consists of words borrowed from other languages (Stockwell and Minkova 2001).

within a speech community to converge on a single way to express a given concept (Lass 1997, Croft 2000, Baronchelli et al. 2006) creates a selection pressure—words that are better at winning these competitions will come to dominate the lexicon (Martin 2007).

The simulation is structured as follows. The speech community is represented by a single agent, who possesses a lexicon and a grammar. Each “generation” of the simulation is divided into two phases: in the first phase, the agent copies the lexicon of the previous generation’s agent, except for the first agent, who begins with an unbiased lexicon containing 2,000 words of type τp , and 1,000 words each of types $p+p$ and $\tau+p$. Each agent then uses a MaxEnt learning algorithm equipped with the structure-blind and structure-sensitive constraints in (15) to learn a grammar using the lexicon as input. In the second phase, the agent is given the option of replacing some of its lexical items with newly generated compounds, with each novel word competing with an existing word. Once these competitions are resolved, the agent’s updated lexicon is used as the input for the next generation’s agent.

The lexicon updating phase takes place in two stages. First, a number of potential compounds are generated by the morphology. Throughout the simulation, compounds with and without geminates are equally likely, representing the combining of existing stems based on the semantic needs of speakers. Next, each potential compound is randomly paired with an existing compound in the lexicon, representing a word synonymous with the novel word, and the two words compete (note that the simulation does not model competitions between monomorphemes and other words). The probability that a word will win this competition ($p_{win}(x)$) is proportional to its phonotactic probability ($p(x)$), as shown in (18).

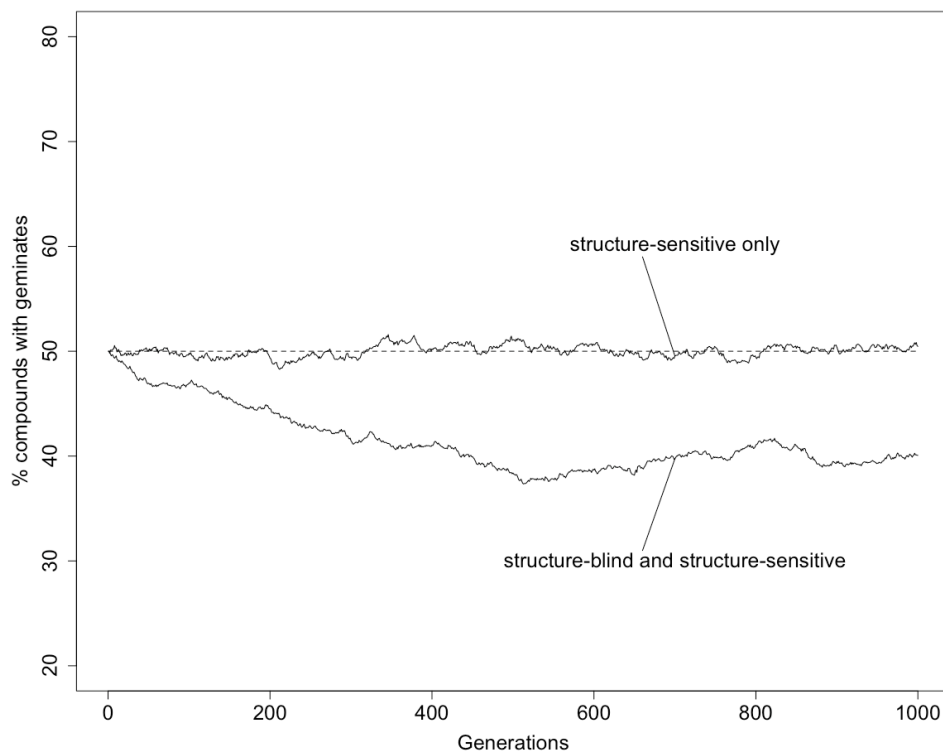
(18) Probability of x winning competition with y

$$p_{win}(x) = \frac{p(x)}{p(x) + p(y)}$$

If a novel word wins a competition, it replaces the existing word; otherwise, the novel word is discarded.

I ran two versions of this simulation for 1,000 generations each. As noted above, the initial agent begins with a lexicon containing 4,000 words. During each generation, 200 new compounds are generated, and allowed to compete with 200 randomly chosen existing compounds. In one version of the simulation, the learner is given structure-sensitive and structure-blind constraints, and in the other version the learner is given only structure-sensitive constraints. The resulting frequencies of compounds with geminates for both versions are shown in (19).¹⁰

(19) Multi-generational learning simulation results



¹⁰ The values depicted in (19) are averaged over ten consecutive runs of each simulation. In order to check that the lexical bias remains stable beyond 1,000 generations, I also performed one run of the simulation for 10,000 generations. With both structure-sensitive and structure-blind constraints, the percentage of compounds with geminates never dropped to zero—between generations 1,000 and 10,000, the mean was 38.5%, the highest occurring percentage was 49.3%, and the lowest 26.8%.

The graph in (19) shows that without the structure-blind constraints, the frequency of geminate compounds randomly varies around 50%, the expected value (shown on the graph with a dashed line) given the initial lexicon and the distribution from which replacement words are drawn. When the structure-blind constraints are added, the relative frequency of geminate compounds drops below 50%, but then levels off. With the structure-blind constraints, compounds with geminates are consistently underrepresented, but are never completely eliminated from the lexicon.

Intuitively, the stabilization in the lexicon can be understood as resulting from the interaction of three factors, which together determine how the lexical statistics change from generation to generation: the phonotactic well-formedness of each word type, the current frequency of each word type, and the probability distribution from which novel words are drawn. The first of these, the well-formedness of each word type, determines the chances of that type winning a competition with another type. As the p+p compounds become rarer, their well-formedness drops, making it harder for them to win subsequent competitions, in a “poor-get-poorer” feedback loop. Although this would seem to doom these compounds to eventual extinction, low frequency also carries an advantage—less frequent types are correspondingly less likely to be faced with a competing word. In short, as the frequency of any word type decreases, its probability of winning competitions drops, but so does its probability of being forced to compete in the first place. This is why, when structure-blind constraints are added in (19), p+p compounds initially drop rapidly in frequency, and then gradually level off, before eventually reaching an equilibrium.

The other crucial parameter which influences the survival of marked compounds types is the probability distribution from which new words are drawn. In the simulations reported above, this is a fixed distribution in which each compound type was equally likely.

No matter how rare geminates become in the actual lexicon, new compounds contain geminates half of the time. If this assumption is altered, and novel words are chosen from a distribution reflecting the current lexical frequencies, then stability collapses, and compounds with geminates will eventually be completely eliminated. In other words, if compounding is driven *exclusively* by phonological factors, then it is indeed a mystery how marked structures survive. But surely this is not the case—morphological operations are at least partly motivated by the semantic needs of speakers, which are presumably blind to phonological considerations.

Thus, the current state of the lexicons of Navajo and English can be seen as a balance between two forces: semantic preferences for certain combinations of morphemes, and phonotactic preferences for certain combinations of sounds. The first drives the English lexicon towards the expected number of geminates, while the second drives the lexicon towards a state with no geminates. The result is a compromise, in which geminates are allowed, but occur at less than the expected rate. Although different languages (or the same language at different times) may enact this compromise to differing degrees, the model predicts that a language in which compounds that violate a categorical stem-internal phonotactic constraint are consistently *overrepresented* would be historically unstable.

6. Conclusion

I have argued that a bias in the human phonotactic learner is responsible for a correlation between tautomorphemic and heteromorphemic phonotactics in Navajo and English. In my model this bias results from two factors: a set of *structure-blind* phonotactic constraints which ignore morphological structure, and Maximum Entropy learning algorithm equipped with a *smoothing term* which penalizes high constraint weights.

I have also shown how this learning bias interacts with the creation and selection of new words, resulting in a persistent lexical bias. The biased lexicons in Navajo and English represent a compromise between the needs of the morphology and the needs of the phonology. The model of lexical change I present here is general in character, and could potentially be applied to a wide range of cases in order to determine how these forces interact in a wider range of languages and phenomena. This research would contribute to our understanding of how phonotactic knowledge participates in the shaping of the lexicon, which in turn forms the basis of the next generation's phonotactic knowledge.

Although the cases discussed in this paper involve an interaction between generalizations at different levels of morphological complexity, the learning model also predicts other types of interaction. In a Maximum Entropy grammar, generalizations stated over constraints which refer to overlapping categories will be interconnected—put simply, when a given structure is underrepresented, similar structures are more likely to also be underrepresented. The model thus makes a very rich set of predictions regarding not just the generalizations that should be attested in natural languages, but relations among the generalizations within a single language. Testing these predictions promises to lead to a deeper understanding of the nature of phonotactic learning.

References

- Abney, Steven. 1997. Stochastic attribute-value grammars. *Computational Linguistics* 23(4). 597-618.
- Baayen, Harald R., Piepenbrock, Richard & van Rijn, Hedderik. 1993. *The CELEX lexical database*. [CD-ROM]. Philadelphia, PA: Linguistics Data Consortium, University of Pennsylvania.

- Baronchelli, Andrea, Felici, Maddalena, Loreto, Vittorio, Caglioti, Emanuele & Steels, Luc. 2006. Sharp transition towards shared vocabularies in multi-agent systems. *Journal of Statistical Mechanics: Theory and Experiment*. P06014.
- Baugh, Albert & Cable, Thomas. 1993. *A History of the English Language*. London: Routledge.
- Berger, Adam L., Della Pietra, Stephen A. & Della Pietra, Vincent J. 1996. A maximum entropy approach to natural language processing. *Computational Linguistics* 22. 39-71.
- Berkley, Deborah. 1994. The OCP and gradient data. *Studies in the linguistic sciences* 24(1/2). 59–72.
- Berkley, Deborah. 2000. *Gradient OCP Effects*. Ph.D Dissertation, Northwestern University.
- Blevins, Juliette. 2004. *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.
- Boersma, Paul. 2007. The evolution of phonotactic distributions in the lexicon. Talk given at Workshop on Variation, Gradience and Frequency in Phonology, Stanford University.
- Clark, Robin & Roberts, Ian. 1993. A computational model of language learnability and language change. *Linguistic Inquiry* 24(2). 299-345.
- Coetzee, Andries W. & Pater, Joe. 2008. Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language & Linguistic Theory* 26(2). 289-337.
- Croft, William. 2000. *Explaining Language Change: An evolutionary approach*. Harlow, Essex: Longman.
- Della Pietra, Stephen A., Della Pietra, Vincent J. & Lafferty, John D. 1997. Inducing features of random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19. 380-393.
- Fischer, Markus. 2005. *A Robbins-Monro type learning algorithm for an entropy maximizing version of stochastic Optimality Theory*. Masters thesis, Humboldt University.
- Fountain, Amy V. 1998. *An optimality theoretic account of Navajo prefixal syllables*. Ph.D. Dissertation, University of Arizona.
- Frisch, Stefan A., Pierrehumbert, Janet B. & Broe, Michael B. 2004. Similarity avoidance and the OCP. *Natural Language & Linguistic Theory* 22(1). 179-228.
- Goldwater, Sharon & Johnson, Mark. 2003. Learning OT constraint rankings using a maximum entropy model. *Proceedings of the Stockholm workshop on variation within Optimality Theory*. 111-120.
- Hale, Mark. 1998. Diachronic syntax. *Syntax* 1(1). 1-18.
- Hammond, Michael. 1999. *The phonology of English: a prosodic optimality-theoretic approach*. Oxford: Oxford University Press.
- Hayes, Bruce P. 1986. Assimilation as spreading in Toba Batak. *Linguistic Inquiry* 17(3). 467-499.
- Hayes, Bruce P. & Wilson, Colin. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39(3). 379-440.
- Hayes, Bruce P., Zuraw, Kie, Siptár, Péter & Londe, Zsuzsa C. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85(4). 822-863.
- Hudson Kam, Carla L. & Newport, Elissa L. 2005. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development* 1(2). 151-195.
- Hudson Kam, Carla L. & Newport, Elissa L. 2009. Getting it right by getting it wrong: When learners change languages. *Cognitive psychology* 59(1). 30-66.

- Jäger, Gerhard. 2007. Maximum Entropy Models and Stochastic Optimality Theory. In Annie Zaenen, Jane Simpson, Tracy Holloway King, Jane Grimshaw, Joan Maling & Chris Manning (eds.), *Architectures, Rules, and Preferences. Variations on Themes by Joan Bresnan*, 467-479. Stanford: CSLI Publications.
- Jaynes, E.T. 1957. Information theory and statistical mechanics. *Physical Review* 106. 620-630.
- Johnson, Mark. 2007. *A gentle introduction to Maximum Entropy Models and their friends*. Talk given at the Northeastern Computational Phonology Meeting, University of Massachusetts.
- Jusczyk, Peter. 1997. *The Discovery of Spoken Language*. Cambridge: MIT Press.
- Kari, James M. 1976. *Navajo verb prefix phonology*. New York: Garland.
- Kawahara, Shigeto, Ono, Hajime & Sudo, Kiyoshi. 2005. Consonant co-occurrence restrictions in Yamato Japanese. *Japanese/Korean Linguistics*. 27-38.
- Kaye, Alan S. 2005. Gemination in English. *English Today* 21(02). 43-55.
- Kessler, Brett. 2001. *The significance of word lists*. Cambridge: MIT Press.
- Kiparsky, Paul. 1968. Linguistic universals and linguistic change. In E. Bach & R. Harms (eds.), *Universals in linguistic theory*, 170-202. New York: Holt, Rinehart and Winston.
- Ladefoged, Peter. 2001. *A Course in Phonetics 3rd Edition*. Fort Worth: Harcourt Brace Jovanovich.
- Lass, Roger. 1997. *Historical Linguistics and Language Change*. Cambridge: Cambridge University Press.
- Legendre, Géraldine, Miyata, Yoshiro & Smolensky, Paul. 1990. 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*. 388-395.
- Lightfoot, David. 1979. *Principles of diachronic syntax*. Cambridge: Cambridge University Press.
- Lightfoot, David. 1991. *How to set parameters: Arguments from language change*. Cambridge: MIT Press.
- Lightfoot, David. 1999. *The development of language: Acquisition, change, and evolution*. Oxford: Blackwell.
- Martin, Andrew T. 2007. *The evolving lexicon*. Ph.D. Dissertation, UCLA.
- Martin, Sven C., Ney, Hermann & Zaphlo, Jörg. 1999. Smoothing methods in maximum entropy language modeling. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing* 1. 545-548.
- McDonough, Joyce. 1991. On the representation of consonant harmony in Navajo. *Proceedings of the West Coast Conference on Formal Linguistics* 10. 319-335.
- McDonough, Joyce. 2003. *The Navajo sound system*. Dordrecht: Kluwer Academic Publishers.
- Oh, Grace E. & Redford, Melissa A. 2009. The effect of boundary recoverability on geminate length in English. *The Journal of the Acoustical Society of America* 125(4). 2568.
- Padgett, Jaye. 1995. *Stricture in Feature Geometry*. Stanford: CSLI Publications.
- Payne, Elinor M. 2005. Phonetic variation in Italian consonant gemination. *Journal of the International Phonetic Association* 35(02). 153-181.
- Peters, Ann. 1983. *The Units of Language Acquisition*. Cambridge: Cambridge University Press.

- Pozdniakov, Konstantin & Segerer, Guillaume. 2007. Similar place avoidance: A statistical universal. *Linguistic Typology* 11. 307-348.
- Prince, Alan & Smolensky, Paul. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, Mass.: Wiley-Blackwell.
- Pycha, Anne. 2010. A test case for the phonetics–phonology interface: gemination restrictions in Hungarian. *Phonology* 27(01). 119-152.
- Ridouane, Rachid. 2007. Gemination in Tashlhiyt Berber: an acoustic and articulatory study. *Journal of the International Phonetic Association* 37(02). 119-142.
- Sapir, Edward & Hoijer, Harry. 1967. *The phonology and morphology of the Navaho language*. Berkeley: University of California Press.
- Sepp, Mary. 2006. *Phonological constraints and free variation in compounding: a corpus study of English and Estonian noun compounds*. Ph.D. Dissertation, City University of New York.
- Smolensky, Paul & Legendre, Géraldine. 2006. *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*. Cambridge, MA: MIT Press.
- Stockwell, Robert & Minkova, Donka. 2001. *English Words: History and Structure*. Cambridge: Cambridge University Press.
- Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science: A Multidisciplinary Journal* 30(5). 945-982.
- Wiltshire, Caroline. 1999. The conspiracy of Luganda compensatory lengthening. In Paul F.A. Kofey (ed.), 131-147. Africa World Press.
- Young, Robert W. & Morgan, William. 1987. *The Navajo language: A grammar and colloquial dictionary*. Albuquerque: University of New Mexico Press.