UNIVERSITY OF CALIFORNIA

Los Angeles

The Effects of Distance on Lexical Bias:

Sibilant Harmony in Navajo Compounds

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Arts

in Linguistics

by

Andrew Thomas Martin

2005

The thesis of Andrew Thomas Martin is approved.

Bruce Hayes

Colin Wilson

Kie Ross Zuraw, Committee Chair

University of California, Los Angeles

2005

TABLE OF CONTENTS

LIST OF EXHIBITS

ACKNOWLEDGMENTS

The questions addressed in this thesis were first suggested to me by my committee chair, Kie Zuraw, to whom I owe special thanks for her unflagging enthusiasm and unfailing ability to shed light on seemingly baffling problems. I also wish to thank to my committee members Bruce Hayes and Colin Wilson for their support and consistently insightful comments. In addition, this work greatly benefited from discussions with the participants of the UCLA Phonology Seminar, who make doing phonology here not only productive, but fun.

ABSTRACT OF THE THESIS

The Effects of Distance on Lexical Bias:

Sibilant Harmony in Navajo Compounds

by

Andrew Thomas Martin

Master of Arts in Linguistics

University of California, Los Angeles, 2005

Professor Kie Ross Zuraw, Chair

Sibilant harmony in Navajo requires a sibilant in a prefix to agree in anteriority with a following sibilant either in another prefix or in the root. In this thesis I present evidence that the process is much more complex when both sibilants are in different roots which are part of a single compound: compounds with disagreeing sibilants are only sometimes repaired by harmony, and compounds whose sibilants disagree underlyingly are statistically underrepresented in the lexicon. Navajo thus solves the markedness problem of sibilant disagreement with three different strategies in different words: tolerance, when the sibilants are allowed to surface faithfully; repair, when one of the sibilants is forced to agree with the other; and avoidance, in which the offending input is excluded from the lexicon.

Furthermore, these effects are present only when the sibilants are in adjacent syllables or closer; compounds whose sibilants are more distant show little or no bias towards agreement. In this thesis I argue that the distance between two sibilants cannot simply be measured in segments for the purposes of these effects, and that prosodic structure—specifically syllabification—may play a role. I also claim that underrepresented compound types are the result of faithful outputs which are in free variation with the null parse (Prince and Smolensky 1993). Using the Stochastic Optimality Theory framework, I show that a grammar can be constructed which correctly predicts both the alternation rates and the statistical bias of Navajo compounds.

1. Introduction

1.1. Distance in phonology

Many phonological processes in which one segment somehow affects another appear to apply only to adjacent elements—voicing assimilation, for example, typically spreads the voice feature of a consonant to an adjacent consonant only, and has no effect on more distant consonants. The prevalence of this locality condition on such processes has led some researchers to attempt to characterize *all* such segmental processes as requiring adjacency; what appear to be long-distance effects, it is argued, are actually the result of elements being adjacent on a more abstract level (Archangeli and Pulleyblank 1987; Steriade 1987; Halle et al. 2000). Long-distance sibilant harmony, for example, can be described as a strictly local process if sibilants are represented on a separate tier from other consonants—thus, sibilants separated by an arbitrary number of non-sibilants are, at the level of the sibilant tier, adjacent (Shaw 1991; McDonough 1991).

Recent work has shown that adjacency is also relevant at the level of statistical biases in the lexicon. Frisch et al. (2004) examined the effects of the OCP constraint in Arabic roots, which bans triliteral roots like **ssm* in which the first two consonants are the same (Greenberg 1950; McCarthy 1986, 1988). They argue that this categorical prohibition is merely one factor in a larger pattern of similarity avoidance: roots with consonants that are similar, even if not identical, are legal, but statistically underrepresented. Furthermore, this gradient OCP constraint applies even to consonants that are non-adjacent: thus, roots like *smt*, in which the first and third consonants are similar, are attested, but rarer than chance would predict. The OCP effect is weaker, however, for these non-adjacent pairs: roots like *smt* are not as rare as roots like *stm*. This is problematic for the autosegmental view, in which intervening consonants shouldn't matter—if long-distance interaction is possible because distant consonants are actually adjacent (on the appropriate tier), then intervening material should either have no effect on the interaction (if it does not appear on the relevant tier), or block it entirely (if it does appear on the relevant tier). The two possibilities are illustrated in (1).

(1) Autosegmental view of long-distance interaction

(a) Transparency

$$\begin{array}{ccccc} \mathbf{C_1} & \mathrm{V} & \mathrm{C}_2 & \mathrm{V} & \mathbf{C_3} \\ | & & & & | \\ [\mathrm{F}] & & & & [\mathrm{F}] \end{array}$$

(b) Blocking

$$\begin{array}{ccccc} \mathbf{C_1} & \mathrm{V} & \mathrm{C}_2 & \mathrm{V} & \mathbf{C_3} \\ | & & | & & | \\ [\mathrm{F}] & & [\mathrm{F}] & & [\mathrm{F}] \end{array}$$

The third possibility, a gradient long-distance effect, shows that this picture must be enriched somehow—the grammar must have access to more information than simply whether or not two features are adjacent on a given tier.

There is a growing body of literature on statistical lexical biases in several languages that exhibit the gradient effects of distance on long-distance phonological dependencies (Frisch et al. 2004 for Arabic; Frisch 2000 for Thai; Berkley 1994, 2000, Hay et al. 2003 for English). In this paper I will examine gradient distance effects in Navajo sibilant harmony, and argue that gradient distance can also have an effect on active processes: specifically, on the probability that a Navajo compound will undergo sibilant harmony.

In addition to showing that active processes as well as static generalizations may be affected by gradient distance, Navajo also provides evidence for the mechanism by which lexical biases come to exist in a language's lexicon. Distance between sibilants in a Navajo compound affects not only the chance that the compound will undergo harmony, but also the chance that the compound will be formed at all. This suggests that phonological constraints operate at the level of word formation, and can influence which words speakers choose to admit into their lexicons.

The paper is structured as follows: the remainder of this section gives a brief overview of the proposal and introduces some of the statistical techniques and notation used throughout. Section 2 presents evidence that distance in Navajo sibilant harmony is indeed gradient, and examines evidence that syllabification is crucial in the grammar's evaluation of distance. Section 3 shows that an underlying bias towards sibilant agreement is present in compounds, in addition to the bias due to alternations. Finally, section 4 argues for an Optimality theoretic grammar of Navajo in which the MPARSE constraint (Prince and Smolensky 1993) is stochastically ranked with respect to faithfulness and markedness constraints in a way that accounts for the complex behavior of compounds.

1.2. Summary of proposal

I will present evidence that a Navajo compound with two sibilants differing in anteriority, each of which belongs to a separate root (e.g., [**ts**^{h}é-**tʃ**ééʔ] 'amber', lit.

‘stone-resin’), only sometimes undergoes harmony to repair the disagreeing sibilants. When the sibilants are in adjacent syllables, however, the chance that a word will change one of the sibilant’s [anterior] specification (e.g., **/tsʰ**é-**tʃ**éé?/ → [**tʃʰ**é-**tʃ**éé?]) is higher than for words in which the sibilants are in non-adjacent syllables.

It is not just the rate of harmony that is increased by distance—compounds whose sibilants agree underlyingly are more frequent than chance would predict if the sibilants are in adjacent syllables, while agreement for non-adjacent syllables is at chance. This underlying bias is evidence that a constraint enforcing sibilant agreement influences Navajo speakers’ formation or retention of words.

Compounds with disagreeing sibilants can therefore be either excluded from the lexicon or repaired through harmony; both processes are sensitive to distance. I will account for these facts by means of an OT grammar in which three constraint types—faithfulness, markedness, and MPARSE—are stochastically ranked with respect to each other. The stochastic nature of the ranking ensures that different rankings among the constraints are possible with different probabilities; I will show that with the correct ranking values for each constraint, the grammar produces variation with the same frequencies that are observed in the lexicon.

1.3. Note on statistical methods

This paper deals largely with linguistic patterns that are probabilistic rather than categorical. This section serves as an introduction to some of the methods, assumptions, and notation I will use throughout the paper.

1.3.1. *How should chance be calculated?*

The data cited in this paper consist primarily of statistical tendencies. For the most part, I will be examining the probability with which two sibilants agree in a Navajo compound. Imagine that we find that sibilants agree in anteriority in 61.6% of such compounds. What does this number mean? Is it higher or lower than we expect? The answer to this depends crucially on how we calculate "chance"—in this case, the number of sibilant pairs that would be identical if there were no cooccurrence constraint. Any significant deviation from chance suggests that there may be some non-random factor influencing sibilant agreement.

The calculation of chance is an inherently hypothetical enterprise. If sibilants *were* to combine at random, how often *would* they be identical? In this section I distinguish two ways of calculating this value which differ in how they interpret this hypothetical scenario, and discuss the usefulness of each in evaluating cases like sibilant cooccurrence.

One way to compute chance would be to calculate the relative frequencies of each sibilant type, and then use that to calculate the predicted cooccurrence frequencies of any sibilant pair. The table in (2) gives the frequencies of both

[+anterior] and [-anterior] sibilants in Navajo (computed from all words that contain only one sibilant).

(2) Navajo sibilant relative frequencies

[+anterior]	**[-anterior]**
47.7% (953/1998)	**52.3%** (1045/1998)

From these numbers, we can calculate that any two sibilants taken at random have a $(.477^2) + (.523^2) = 50.1\%$ chance of being identical. I will refer to chance calculated in this way as *raw chance*, since it relies only on the distribution of consonants in the language as a whole, without taking into account the exact makeup of the sample under consideration (in this case, compounds with two sibilants each).

Another way to compute chance would be as follows. Given a set of sibilant pairs (i.e., words containing two sibilants each), we might ask: if we were to randomly combine *exactly these* sibilants, how often would they agree? In a specific sample of words, sibilants may not be distributed in the same proportions as in the entire language. I will refer to chance calculated over only the sample in question as the *conditional chance*.

In the sibilants example, conditional chance can be computed using the following table:

(3) Sibilant distribution in compounds

	1st sibilant	2nd sibilant
[+anterior]	57.3%	56.4%
[-anterior]	42.7%	43.6%

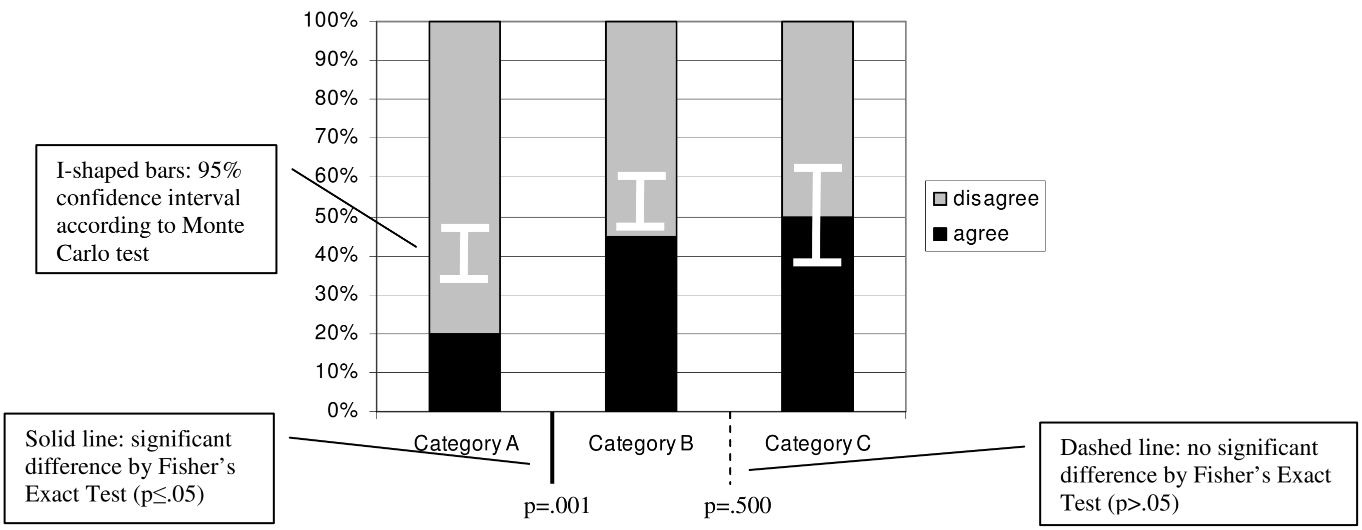
The chance that two sibilants *from this sample* agree (holding their position in the pair constant) is $(.573 \times .564) + (.427 \times .436) = 50.9\%$, still lower than the 61.6% observed rate, but not much different than the raw chance of 50.1%. Because the conditional chance value takes into account the distribution of consonants within a set of words, this is the value that will be reported as "chance" throughout this paper.

1.3.2. *Testing for significance: the Monte Carlo procedure*

Once a chance agreement rate has been calculated, we need a test that will tell us whether it differs significantly from the observed agreement rate. Rather than use an established formula that directly returns a p-value, I will follow Kessler (2001) in using a Monte Carlo procedure, so named because of its use of randomness. In order to determine what would happen if consonants were combined at random, the Monte Carlo test does just that—it actually combines the consonants at random a large number of times (I will use 10,000 repetitions). Given a set of consonant pairs, an algorithm creates a set of new pairs by randomly assigning each first member to a second member, until all consonants are paired up. Then the agreement rate under the new pairing is calculated. This process is then repeated a total of 10,000 times. The result is a distribution of agreement rates that can be compared to the observed rate. If, for example, only 400 of the randomly generated agreement rates are as low as or lower than the observed rate, the observed rate can be said to significantly differ from chance, with an estimated p of .040.

Consonant agreement rates will generally be represented using stacked bar graphs in this paper. Chance for each case will be shown by a white I-shaped confidence interval within the relevant bar. The top and bottom of each confidence interval represent the upper and lower limits within which 95% of the rates generated in the Monte Carlo test fall. In other words, the test only produced rates lower than the interval 2.5% of the time, and higher than the interval 2.5% of the time. If the observed rate falls outside of the confidence interval, it differs significantly from chance (see example chart in (4)).

(4) Example of chart notation



Another piece of notation is used to show whether the agreement rates in chart-adjacent categories are significantly different from one another according to Fisher's Exact Test. If two categories have significantly different agreement rates, their x-axis labels are separated by a solid line; if there is no significant difference, a dashed line is used. Thus, in the example chart (4) above, rates for categories A and B are significantly different, while there is no significant difference between the rates for categories B and C. The exact *p*-values that result from the Fisher's Exact Test are reported beneath each line.

2. Gradient distance effects in Navajo

2.1. Navajo sibilant harmony

2.1.1. *Overview*

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. The two sets of consonants are summarized in the table below.

(5) Navajo sibilant classes

[+anterior]	[-anterior]
s	ʃ
z	ʒ
tsʰ	tʃʰ
ts	tʃ
ts'	tʃ'

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. The examples in (6) demonstrate the alternations in prefixed forms (sibilants are in bold).

(6) Examples of sibilant harmony (Fountain 1998)[1]

(a) /ji-**s**-léé**ʒ**/ → [ji-**ʃ**-tɬéé**ʒ**] 'it was painted'

(b) /ji-**s**-ti**z**/ → [ji-**s**-ti**z**] 'it was spun'

Typically, assimilation proceeds from the root to the prefixes. In compounds, which contain multiple roots, sibilant harmony does not necessarily spread from one root to another, meaning that such words can contain disagreeing sibilants:

(7) Exceptions to sibilant harmony in compounds (Young, Morgan, and Midgette 1992)

(a) **tʃ**éí-**ts'**iin 'thorax, chest bone' (/tʃéí/ 'heart' + /ts'iin/ 'bone')

(b) **ts^h**é-**tʃ**ééʔ 'amber' (/ts^{h}é/ 'stone' + /tʃéé/ 'resin')

(c) naa-ni**ʃ**-**ts^h**oh 'big job' (/niʃ/ 'work' + /ts^{h}oh/ 'big')

It is not, however, true that root consonants are *never* affected by the process, as Sapir and Hojier (1967) point out. They cite compounds like those in (8), in which alternations occur.

(8) Alternations in compounds

(a) **ts**aa-néé**z** 'mule' (/tʃaa/ 'ear' + /nééz/ 'long')

(b) xo**s**-**ts'**óóz 'type of cactus' (/xoʃ/ 'cactus' + /ts'óóz/ 'slender')

Sapir and Hojier are aware that exceptions to sibilant harmony in compounds are plentiful, but do not form any rigorous generalization that would predict when exceptions are possible. They do, however, make the anecdotal observation that assimilation "occurs less often when the consonants are at a greater distance" (14-15).

[1] A note on transcriptions: IPA is used throughout for Navajo examples, with acute accents marking high tones (low tones are unmarked). In order to accommodate accent marks, nasal vowels are indicated with a hook below the relevant symbol (e.g., [ą] for IPA [ã]).

The remainder of this section presents evidence that the Sapir and Hojier claim is true for Navajo compounds—exceptions to harmony are more likely when the sibilants in question are further apart.

2.1.2. *Data collection*

The data described here are taken from an online lexicon of Navajo (Burke and Lachler 2000), an electronic version of Young, Morgan and Midgette (1992). This lexicon consists of a list of 1,130 roots, as well as roughly 20,000 words and phrases derived from these roots. From the lexicon, a word list was created by searching for all words that contained exactly two sibilants. This resulted in a total of 589 words. A preliminary investigation of this list showed that where one of the sibilants is in an affix and the other in a root, anteriority agreement is close to 100%. Because this "ceiling effect" would make it hard to discern the effects of distance on agreement, words in which one of the sibilants is in an affix were removed from the data set, leaving 313 words.

Another factor that interferes with the effects of distance is the exceptionless generalization that two sibilants in the same root must agree in anteriority (McDonough 1991). The examples in (9) show that this is true whether the sibilants are in the same or different syllables in the surface form:

(9) Sibilants within the same root

Same syllable		Adjacent syllables	
tá-**tʃ**óó**ʒ**	‘glottis’	wó-**tʃ’**o**ʃ**-í	‘rotten tooth’
ɬee-**tʃ’**o**ʃ**	‘earthworm’	tɬ’oh-**ts’**ó**z**í	‘rush grass’
naʔní-**ʒ**oo**ʒ**	‘bridge’	beeʔél-**ʒ**á**ʃ**-í	‘sandpaper’

To control for the effects of this constraint, words in which both sibilants belong to the same root (e.g., those in (9)) were removed from the data set, leaving a total of 211 words. All remaining words are thus compounds, with each sibilant occurring in a different root.[2] Some representative examples follow in (10) (roots are underlined).

(10) Examples of compounds with two sibilants (one per root)

(a) **tsʰ**ą́ą́-hal-**ts’**aaʔ — ‘pepper grass’ (lit. ‘belly-hollow’)
(b) **tsʰ**ee-**ts’**iin — ‘tailbone’ (lit. ‘tail-bone’)
(c) k’ii**ʃ**-**ʒ**in-ii — ‘black alder’ (lit. ‘alder-black’)
(d) ii-**tʃ’**ąąh-ii-**tsʰ**oh — ‘large moth’ (lit. ‘one that has fits-large’)
(e) naha-k’í**z**-ii-ɬ-**tʃʰ**í’í-gíí — ‘red cricket’ (lit. ‘one that squeaks-red’)
(f) **tsʰ**é-**z**éí — ‘gravel’ (lit. ‘rock-crumbs’)
(g) **tsʰ**in-aa-bąą**s** — ‘wagon’ (lit. ‘wood-hoop’)

The remainder of the paper will be concerned with these 211 words—compounds with two sibilants, each of which is in a different root.

One further note regarding how words were selected: where multiple versions of the same word were listed (i.e. slightly differing pronunciations listed under the same meaning), only one version was included, as long as the alternants did not crucially differ with respect to sibilant type, position, or distance (e.g., [**tsʰ**é-**tʃ**ééʔ] vs. [**tsʰ**é-**tʃ**eex] ‘amber’). Where multiple alternants *were* crucially different (e.g.,

[2] Note that this does not mean that all of the compounds consist of only two roots; compounds may be a combination of three or more roots, as long as some roots contain no sibilants.

[**tʃ**^h^oʔk'i**s**] vs. [**tʃ**^h^oʔk'i**s**í] 'one-testicled,' which differ in distance), all were listed as separate words. This includes words for which both disagreeing and harmonizing alternants were listed as possibilities (e.g., [**ts**^h^éʔá**ʒ**ííh] vs. [**tʃ**^h^éʔá**ʒ**ííh] 'rock sage').

For each of these words, the two sibilants were classified as [+anterior] or [-anterior], and the distance between them was classified according to how many (if any) syllables intervene between them (a possible alternative segmental characterization of distance will be considered in section 2.3). I considered placing words in which the sibilants are string-adjacent (e.g., [a-tɬ'i**ʒ**-**ts**é] 'gallstone') in a separate category, but decided to include these with adjacent-syllable cases because (a) string-adjacent consonants in Navajo are necessarily in adjacent syllables, and (b) the agreement rates for string-adjacent words and other adjacent-syllable words were not significantly different.

Finally, the rate of agreement between the sibilants in each word was calculated and checked for correlation with the distance between them. The results of this procedure are discussed in the following subsection.

2.1.3. *Results*

None of the compounds in the data set contain sibilants that are in the same syllable; this is a consequence of Navajo syllable structure (only (C)V(V)(C) are legal syllables) and the fact that very few roots are vowel-initial (Young et al. 1993 lists only one vowel-initial root that contains a sibilant: /ááʒ/ 'several go on foot').

Sibilants could end up in the same syllable only if a sibilant-final root combined with a vowel-initial (and sibilant-final) root: i.e., CV.S-VS. None of the compounds, however, have this shape. The chart in (11) shows how the data breaks down according to distance between sibilants (syllable boundaries are marked in examples):

(11) Sibilant distance categories

Distance	Example word	% of total
adjacent σ	**ts**ʰé.**s**ǫʔ 'glass'	**66.4%** (140/211)
1 σ intervening	ná**ʃ**.dóí.**ts**ʰoh 'mountain lion'	**24.2%** (51/211)
2 σ intervening	**ts**i.diɬ.há.**ʃ**ii 'scorpion'	**8.5%** (18/211)
3 σ intervening	**ts**ʰą́.ho.di.niih.**ts**ʰoh 'typhoid fever'	**0.9%** (2/211)

The number of words with very distant sibilants (two or three intervening syllables) is small, and the agreement rates for these categories are not significantly different than the rate for words in which the sibilants are separated by one syllable. I will therefore collapse these three categories into one, which includes all words with sibilants in non-adjacent syllables.

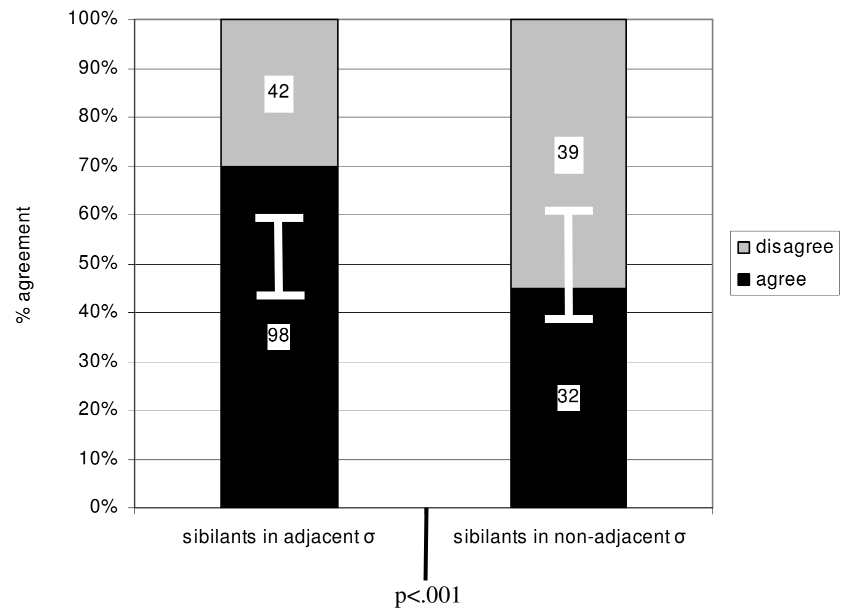
I will thus treat all words as belonging to one of two categories: those in which the sibilants belong to adjacent syllables, and those in which the sibilants are in non-adjacent syllables. The table in (12) gives examples of words for each of these environments.

(12) Examples of relevant environments

Sibilants are…	Possible structures	Examples	
		agree	disagree
in adj σ	...V**S**.**S**V… …**S**V(C).**S**V… …**S**V(C).CV**S**… …V**S**.CV**S**…	[ɬee**ʃ-tʃ’**ih] ‘ashes’ (lit. ‘dirt-breeze’) [**ts’**in-**tsʰ**ah] ‘bone awl’ [-**tʃ**ééh-t’ii**ʒ**] ‘ear wax’	[péé**ʃ-tsʰ**ah] ‘steel awl’ [**ts’**ah-**tʃ’**il] ‘coppermallow’ (lit. ‘sage-plant’) [tl’í**z**í-ɬ-**tʃ’**í-ʔí] ‘reddish goat’ (lit. ‘goat-red-one’)
in non-adj σ	…**S**V(C).CV(C).**S**V… …**S**V(C).CV(C).CV**S**… …V**S**.CV(C).CV**S**…	[**tsʰ**ít-ii-ɬ-**tsʰ**ooí] ‘Rocky Mountain goldfinch’ (lit. ‘bird that is yellow’) [léé-**tʃʰ**ąą-yá**ʒ**í] ‘puppy’ (lit. ‘animal-dog-small’)	[**tʃʰ**ití-**tsʰ**oh] ‘truck’ (lit. ‘car-big’) [**tʃʰ**ití-ɬ-**tsʰ**oii] ‘school bus’ (lit. ‘car that is yellow’)

The chart in (13) shows the agreement rates for these two environments. The numbers given on the chart indicate the number of words from the sample whose sibilants agree or disagree.

(13) Navajo sibilant pair agreement



Sibilants in adjacent syllables agree more often than is predicted by chance, while more distant sibilant pairs do not tend to agree more than they would if randomly combined. Thus, distance affects sibilant agreement, but gradiently, not categorically. In the next section I turn to the question of exactly how distance is measured by Navajo speakers: does prosodic structure play a role, as I have assumed, or is the segmental string sufficient for computing distance?

2.2. How is distance measured?

In the previous section I assigned Navajo compounds to categories based on how many syllables intervened between their sibilants. But does the grammar calculate distance in the same way? It could be that the grammar simply counts syllables, treating all consonant pairs in adjacent syllables as equidistant, but the number of intervening segments could also be crucial. Or perhaps the elapsed time in milliseconds between two consonants (averaged over many tokens) is the relevant measure. It may be that all of these elements contribute to the effects of distance on agreement. In this section I examine evidence that syllabification plays a role in the gradient distance effect described in the previous section.

2.3. Evidence from Navajo for the role of syllables

Several factors make it difficult to determine the role syllable boundaries play in Navajo sibilant agreement. The ideal method would be to isolate words in which the sibilants are separated by only a vowel, and compare same-syllable and adjacent-syllable cases (SV.SV versus SVS.CV)—however, this is impossible in Navajo compounds. Due to the dearth of vowel-initial (or SC-initial) roots, there simply aren't any compounds in which each root contains a sibilant and both sibilants are in the same syllable. The words that would enable us to distinguish between adjacent-syllable and same-syllable environments in Navajo compounds just don't exist (at least not in the Young et al. dictionary).

Given this problem, we might try the same technique, but choose another intervening string to hold constant; instead of sibilant pairs separated by a vowel, we could look at sibilant pairs separated, say, by VCV, and distinguish between adjacent-syllable and non-adjacent-syllable cases. Unfortunately, the small number of such cases make statistical analysis impossible—there are only thirteen such words overall.

Even if vowel length is ignored, there are not enough cases to see a significant effect of syllabification.

Although we cannot use minimally different sets of words to determine the effects of syllable boundaries in Navajo, the next best thing might be to hold constant the number of segments intervening between the two sibilants—if syllable boundaries are irrelevant, then syllabification should not correlate with agreement rates among all words separated by the same number of segments. For example, if we take all words with sibilants pairs separated by exactly three[3] segments, and divide them into adjacent and non-adjacent syllable cases, there should be no difference in the agreement rates. There is, however, a difference:

(14) Sibilant pairs with three intervening segments

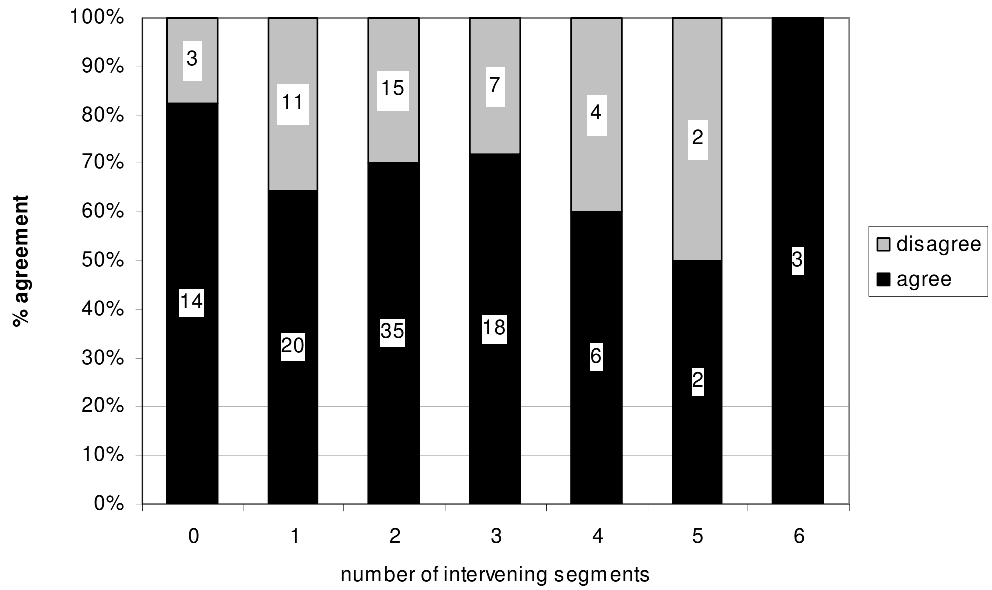
Distance	Agree rate
adjacent syllables	**72.0%** (18/25)
non-adjacent syllables	**27.3%** (3/11)

Distribution is significant (p=.016)

Syllable affiliation thus appears to play a role in modulating long-distance sibilant agreement in Navajo. But does distance in segments also play a role? If so, then when syllabic distance is held constant, there should be a trend of decreasing agreement as segmental distance increases. The chart in (15) shows that this is not the case for those words in which the sibilants are in adjacent syllables.

[3] This number was chosen because it is the only distance for which more than ten words exist in each category (adjacent and non-adjacent syllables). For other distances, there are simply too few words to make statistical comparison meaningful.

(15) Segmental distance: adjacent syllables only



There is no such obvious trend, and Fisher's Exact Test confirms that none of the pairwise differences between columns in the chart are statistically significant.[4]

Furthermore, if we compare three types of adjacent-syllable cases: string-adjacent cases (VS.SV), "close" cases (in which both sibilants are onsets: SV(C).SV) and "far" cases (one onset, the other a following coda: SV(C).CVS) cases within the adjacent syllable environment, no difference is found:[5]

[4] Similar results hold for the non-adjacent syllable cases.

[5] The fourth possibility, where both sibilants are codas, is represented by only one word, and so it has been omitted from this comparison.

(16) Sibilant pairs in adjacent syllables: close vs. far

Distance	Agree rate
String-adjacent	**82.4%** (14/17)
Close (onset-onset)	**67.3%** (68/101)
Far (onset-coda)	**71.4%** (15/21)

Distribution is not significant ($\chi^2(2)=0.268$, $p<1$)

There is thus no evidence that the number of intervening segments correlates with agreement rate; if segmental distance does play a role in determining agreement, its effect is apparently weaker than that of intervening syllable boundaries.

2.4. Evidence from English for the role of syllables

The evidence that syllable boundaries play a role in Navajo sibilant agreement is largely circumstantial, both because the crucial cases (same-syllable sibilant pairs) are missing and because the number of words overall is small. Further evidence that bears on this question can be found in similarity avoidance biases in other languages, however. This section briefly discusses one such example, using data from liquid cooccurrence in English.

2.4.1. *Liquid cooccurrence in English*

Like Arabic, the English lexicon exhibits a bias against words with similar consonants in close proximity (Berkley 1994, 2000; Frisch 1997). Monosyllabic words like *sit*, for example, in which the initial and final consonants share a place of articulation, are statistically underrepresented (Berkley 2000). This constraint grows weaker with increasing distance; words like *swift*, in which the two coronals are separated by several segments, are not as rare as words like *sit,* in which only a single vowel separates the coronals. In this section I consider a special case of this OCP effect: the avoidance of identical liquids in the same word. Words that contain two identical liquids, like *rare* or *lull*, are underrepresented when compared to words with two different liquids, like *lair* or *real*. Like Navajo sibilant agreement, this cooccurrence restriction is stronger the closer the two liquids are in the word.

The data discussed below comes from the CELEX English lemma database (EPL.CD), a list of 52,447 English words (inflected forms are not included) which is based on a (mostly written) corpus of about 18 million words (Baayen et al. 1995). In order to determine the effects of syllabification on the liquid cooccurrence restriction, I examined the set of monomorphemes[6] that contain exactly two liquids[7] each (e.g., *syllable*, *rare*, *mongrel*). From this set of words I took only those in which the liquids were separated by a single short vowel (i.e., [ə], [ɪ], [æ], [ʌ], [ɔ], [ɛ], [ʊ]). In some of these words, like *calorie* and *relax*, the two liquids are in onsets of successive

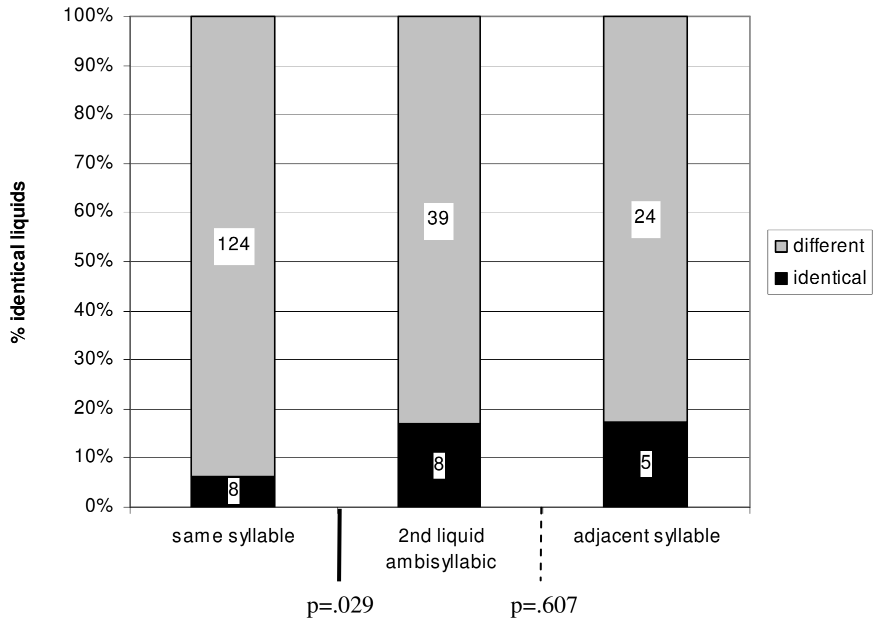
[6] The list of monomorphemes was constructed by Hay, Pierrehumbert and Beckman using the CELEX database by classifying words whose morphological status is left unmarked in CELEX and adding the monomorphemes thus derived to the words explicitly listed as monomorphemic (Hay et al. 2004). I am grateful to Janet Pierrehumbert for supplying me with the CELEX entry numbers for this list, which allowed me to reconstruct it using my own licensed copy of CELEX. I further modified the list, which represents British English pronunciation, by adding [ɹ] where it would be pronounced in my own West Coast American English dialect.

syllables, while in others, like *mirror* and *cathedral*, both liquids are in the same syllable.

Of course, the syllabic affiliation of intervocalic consonants in English is not without controversy—it is often argued that the [l] in a word like *rally* is in fact ambisyllabic, both the coda of the first syllable and the onset of the second (Kahn 1976; Hooper 1978; Vennemann 1981; Gussenhoven 1986; Rubach 1996). In cases where the second liquid is ambisyllabic, determining which category it should belong to is thus problematic. I therefore created a third category, including words in which the liquids are separated by a short vowel and the second is possibly ambisyllabic. I assumed Kahn's (1976) definition of ambisyllabicity: an intervocalic liquid is ambisyllabic if the following syllable is stressed (the preceding syllable may be stressed, or, somewhat more controversially, unstressed). The liquid identity rates that result from dividing this set of words into three categories—unambiguously in the same syllable, unambiguously in adjacent syllables, and ambisyllabic—are summarized in chart (17).

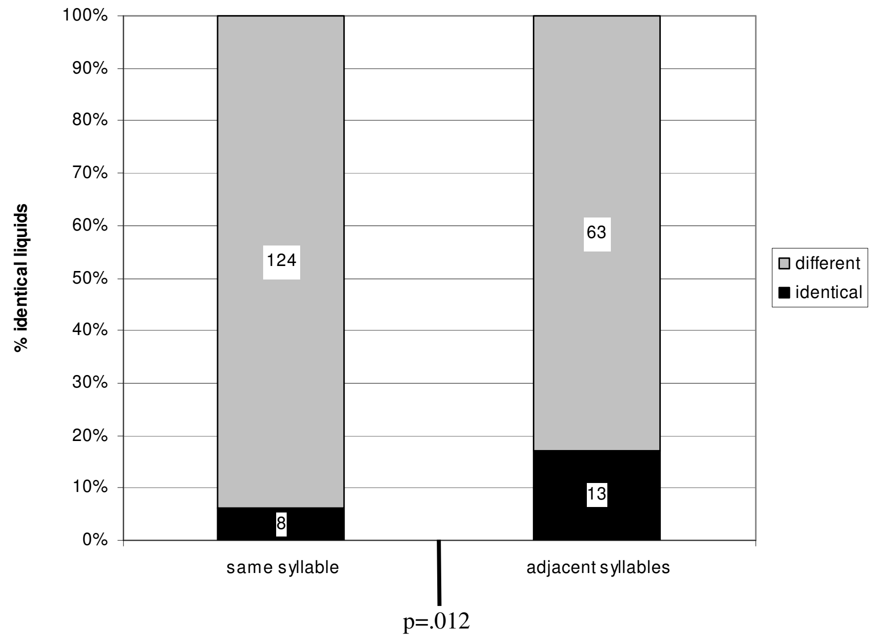
[7] I count [ɚ] as a liquid for these purposes.

(17) English monomorphemes: liquids separated by only a vowel



Although we might expect ambisyllabic cases to pattern with same-syllable words, the figures in (17) do not support such a view. If ambisyllabicity is a real phenomenon, it apparently has little to no effect on liquid cooccurrence. Because there is no significant difference between the adjacent-syllable and ambisyllabic cases, I will collapse them into a single category. Comparing the resulting two categories—same-syllable and adjacent-syllable—reveals the difference shown in (18).

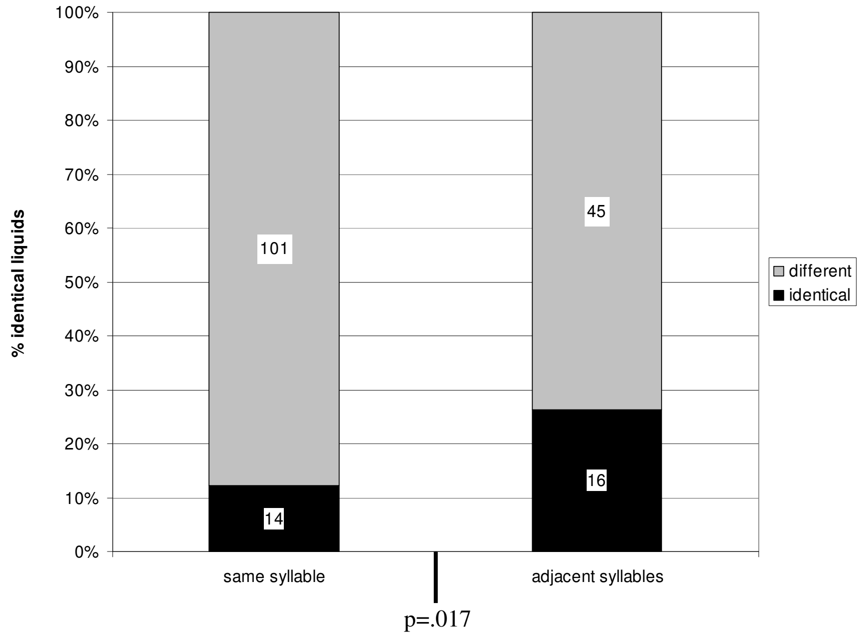
(18) English liquid pairs: effect of syllabification



Liquid pairs in English are permitted to be identical more often if they are in different syllables, even if only a single short vowel separates them. This is not necessarily due to their syllabification, however—there is an alternative explanation for the facts in (18). In the same-syllable words, the second liquid will necessarily be a coda; in the adjacent-syllable words, the second liquid must be an onset. English /l/, however, has different allophones in these environments: "light" [l] in onsets and a velarized "dark" [ɫ] in codas. If the OCP effect is for whatever reason stronger with dark [ɫ] than light [l], it could explain the lower agreement rate in the same-syllable cases.

The theory that /l/ allophony is behind the agreement rate differences in (18) makes several predictions: first, the difference should disappear or be lessened if the data is restricted to those words in which the second liquid is /r/. This is not the case, as is shown in (19).

(19) English liquid pairs with /r/ as second member: effect of syllabification



The difference between the two categories is actually larger, and still significant, when the words with dark [ɫ] are eliminated, although agreement rates for both categories are higher due to a stronger constraint against /l..l/ than against /r..r/.

This theory would also predict that in languages with no light-dark /l/ distinction, the same syllabification effect should not be found. In Spanish, however, the facts are much the same as in English, despite the lack of a dark [ɫ] allophone. A gradient cooccurrence restriction in the Spanish lexicon causes agreeing liquid[8] pairs to be underrepresented. Syllable boundaries make a difference: out of 1,130 Spanish words containing two liquids separated by a single vowel, the liquids agree in [lateral]

[8] I.e., /l/, /ɾ/, /r/, and orthographic *ll*, which was historically a lateral (and still is in some dialects).

15.5% of the time if they are in the same syllable, and 22.4% of the time in adjacent syllables (p=.004).[9]

Of course, ruling out /l/ allophony in English as the sole factor responsible for the agreement differences in (18) does not establish that syllable affiliation is therefore solely responsible. It is plausible that cross-linguistically, prevocalic liquids have different phonetic properties than liquids that do not precede a vowel. The two allophones of /l/ in English may simply be an extreme example of this; without carefully studying the characteristics of liquids in onsets and codas, we cannot conclude that the only difference in the same-syllable and adjacent-syllable words is the intervening syllable boundary. If, for example, coda liquids induce more coarticulation on the preceding vowel than onset liquids, their phonetic cues may be more likely to interfere with those of a liquid in the onset preceding the vowel. This could result in a stronger cooccurrence restriction in an RVR.CV word than an RV.RV word.

So although the data from Navajo and English rule out a naïve, segment-counting definition of distance, and suggest that prosodic structure may play a role, we cannot conclusively establish that distance is counted in syllables without more research into long-distance consonant interactions involving a wide variety of segment types.

---

[9] The Spanish data come from a corpus-derived wordlist created by LLuís Padró (available at http://www.lsi.upc.es/~padro/freqs/spanish-freqs.tgz). Data was limited to words occurring at least twice in the corpus (which consists of over 5 million words); words ending in a verbal infinitive suffix (*-ir*/*-er*/*-ar*) were eliminated.

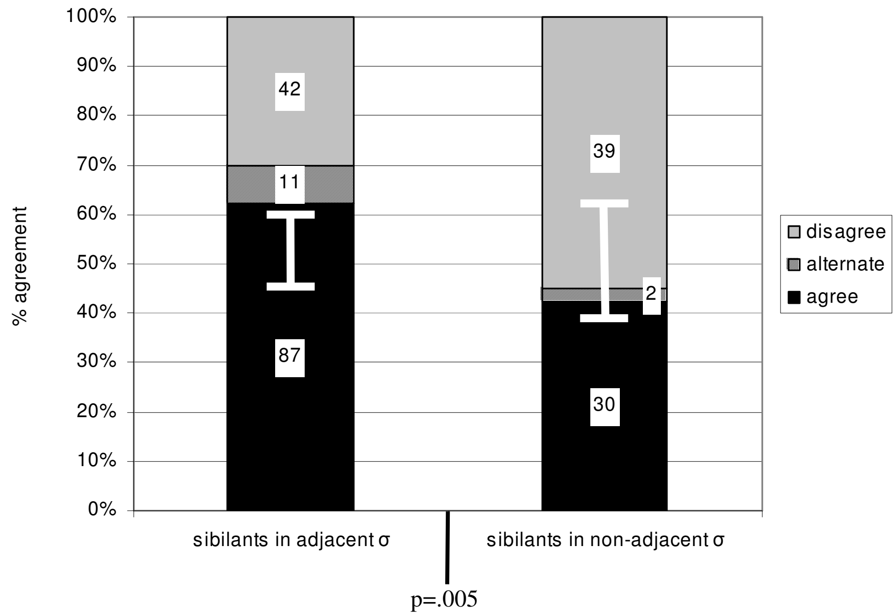
In the next section I return to the case of Navajo sibilant agreement, and address the question of the causes of the agreement bias in compounds.

3. The Origins of Lexical Bias

In §2.1.1 I showed that in compounds like [**ts**aa-néé**z**] 'mule', alternations are possible—root sibilants may surface unfaithfully in order to satisfy the harmony requirement. We can thus ask what is causing the elevated agreement rates in adjacent syllables: are alternations simply more frequent when the sibilants are closer, or is the bias already present underlyingly? The answer, as it turns out, is both.

The underlying specifications of the sibilants in these compounds can be determined by looking at how the individual roots surface in forms where there are no other sibilants present. Doing this reveals that while some of the elevated agreement rate in adjacent syllables is due to alternations, the *underlying* agreement rate is still above chance. The underlying agreement rates are summarized in (20). The category marked "alternate" represents words whose sibilants disagree underlyingly, but which undergo repair so that they agree in the surface form (e.g. /tʃaa+nééz/ → [**ts**aa-néé**z**] 'mule').

(20) Navajo compounds: underlying vs. surface agreement



This chart shows that alternations are responsible for only part of the agreement bias (and a small part at that). In addition, the confidence intervals show that in the adjacent-syllable words, where agreement is elevated, the alternations do not raise the agreement rate above chance—it would be above chance without the repairs. That is, there is a bias towards compounds whose roots *already* contain agreeing sibilants; this bias would be present even if there were no sibilant harmony at all in compounds.

This fact is striking—it suggests that Navajo speakers are biased against forming compounds that would violate a constraint against disagreeing sibilants, even though the language has a well-attested, productive means of repairing such violations. In other words, they apply one of three strategies when confronted with a possible compound with disagreeing sibilants: (1) tolerate the disagreement and pronounce it faithfully, (2) apply sibilant harmony to enforce agreement, or (3) simply avoid using the word at all. Perhaps surprisingly, all three strategies are employed for different words. The theoretical ramifications of this fact will be discussed in §4.

The numbers in (20) reveal another interesting fact about the effects of distance: it is not just that the adjacent-syllable compounds are more likely to agree underlyingly—adjacent-syllable words are also more likely to undergo repair. Out of the 53 adjacent-syllable words whose sibilants disagree (and thus are candidates for repair), 11 actually undergo alternations (20.8%). In contrast, out of the 41 disagreeing non-adjacent-syllable words, only 2 are repaired (4.9%). This difference in alternation rates is significant according to Fisher's Exact Test ($p=.025$). Thus, the effects of distance are twofold: having disagreeing sibilants in adjacent syllables makes a compound less likely to be formed in the first place, and once formed, makes it more likely to undergo repair.

Ohala (1981) has suggested that lexical biases arise as a result of listeners misperceiving marked structures as unmarked ones—these changes in individual words accumulate over time so that eventually words with marked structures are underrepresented in the lexicon. The data here, however, show that this is an unlikely explanation for the agreement bias in Navajo, as the historical change would have to affect the roots themselves, which most of the time appear in contexts other than the 211 compounds that are the subject of this paper. The Navajo constraint on compound formation more closely resembles children's avoidance of words that contain difficult sounds (Ferguson and Farwell 1975; Leonard et al. 1981; Schwartz and Leonard 1982)—a clearly synchronic result of phonotactic constraints.

4. Formal Account

There are typically three ways that languages can respond to inputs containing marked structure: repair the structure, tolerate it, or ban the input altogether. In Optimality Theory, the first option, repair, is usually represented as a ranking of markedness over faithfulness; toleration is modeled using the reverse ranking, faithfulness over markedness. Cases in which a marked input is not mapped to any output are more problematic, but several proposals have been advanced to deal with such cases: notably the MPARSE constraint (Prince and Smolensky 1993) and the CONTROL component of Orgun and Sprouse (1999).

Navajo compounds are remarkable in that all three of the above strategies are used to deal with the same markedness problem, namely sibilants that disagree in [anterior]. Furthermore, which strategy will be used for a given word is unpredictable; their application is probabilistic, with each being employed a certain percent of the time. Because of this, I will model the Navajo facts using a stochastic OT grammar, in which constraint rankings may change probabilistically. The next section introduces the constraints that comprise the grammar; following that, §4.2 describes the stochastic OT framework and §4.3 explains how the Navajo grammar is constructed.

4.1. The constraints

4.1.1. *Markedness and faithfulness*

The Navajo grammar will contain three types of constraints: faithfulness constraints that prevent changes to a sibilant's [anterior] specification, markedness constraints that compel multiple sibilants to agree in [anterior], and MPARSE, which specifically prohibits the null parse. Definitions for the first two constraint types are given in (21).

(21) Navajo constraints

***Faithfulness*:**
IDENT-IO[anterior](root): An input segment in a root must have the same specification for [anterior] as its corresponding ouput segment.

***Markedness*:**
CORRESPOND-CC(sibilant): Two [+strident] segments in an output must be in correspondence.
AGREE[anterior](adjacent σ): A consonant must have the same specification for [anterior] as all corresponding consonants in adjacent syllables.
AGREE[anterior]: A consonant must have the same specification for [anterior] as all corresponding consonants.

The faithfulness constraint is straightforward: it is what prevents repair in those cases where compounds with disagreeing sibilants are tolerated. Note that it is specific to roots—a parallel constraint referring to affixes is presumably very low ranked, so that harmony always proceeds from root to affix.[10]

[10] This means that there is no need to specify directionality in cases where prefixes harmonize with roots (Navajo has no suffixes). There is a need for directionality, however, in compounds: all of the compounds that harmonize do so regressively except for one ([**tʃ**ánéʒí] 'sash garter' from /**tʃ**áád+nééz/). There must therefore be other constraints responsible for direction of assimilation that do not rely on a root-affix asymmetry.

The markedness constraints are based on work by Suzuki (1998) and Rose and Walker (2004). The CORRESPOND-CC constraint establishes a relation between all sibilants in a word. It is then up to the AGREE constraints to enforce similarity between corresponding sibilants. I will assume that CORRESPOND-CC is undominated, and so will not consider candidates in which multiple sibilants are not in correspondence. The rankings of the Agree constraints, on the other hand, are crucial, as they are relativized for distance; it is their ranking that will establish the difference between adjacent- and non-adjacent-syllable cases. This choice is arbitrary—one could assume that CORRESPOND-CC is relativized for distance, and a single AGREE constraint is undominated, but because this would not change the predictions of the analysis, I will ignore this possibility.

The AGREE constraints form a *stringency hierarchy* (Prince 1997). AGREE[anterior] forces any two consonants to agree, no matter what their distance, while AGREE[anterior](adjacent σ) refers to the subset of these consonant pairs that are in adjacent syllables. Thus, any candidate that violates AGREE[anterior](adjacent σ) must also violate the more general AGREE[anterior]. The constraints could be defined differently: AGREE[anterior](adjacent σ), which refers only to sibilants pairs in adjacent syllables, and AGREE[anterior](non-adjacent σ), which refers only to sibilant pairs in non-adjacent syllables. This is the approach taken by Suzuki (1998), who points out that in that case constraints referring to shorter distances must be universally ranked over those referring to longer distances, in order to avoid incorrect typological predictions (namely, long-distance interactions that are stronger when consonants are further apart).

4.1.2. *The null parse*

The constraints described in the previous section can account for toleration or repair of disagreeing sibilants, but cannot model Navajo speakers' predilection for forming compounds whose sibilants already agree. This avoidance of potential inputs that would contain marked material resembles *absolute ungrammaticality*: in English, for example, when presented with an input like /*obnoxious*+*er*/, which violates a restriction on the number of syllables in a stem that *-er* can attach to, the grammar neither repairs the stem (by, say, shortening it), nor tolerates the markedness, but instead simply refuses to offer a legal output. Prince and Smolensky (1993) propose that in these cases the optimal candidate is the *null parse*, an unpronounceable object that essentially informs the speaker that the given input has no legal output—some other input (perhaps periphrastic) must be used to express the desired meaning. The null parse, according to Prince and Smolensky, violates no constraints except MPARSE (which is only violated by the null parse); ranking MPARSE correctly can predict which inputs will map onto the null parse and thus be absolutely ungrammatical. In the case of /*obnoxious*+*er*/, the tableau would look like (22).

(22) The null parse and absolute ungrammaticality

/obnoxious+er/	STEM≤σσ	MAX	MPARSE
obnoxiouser	*!		
☞ NULL			*
noxiouser		*!	

In the case of Navajo compounds, we do not want the grammar to declare every input with disagreeing sibilants as ungrammatical—speakers merely disprefer such inputs. I propose to model this by claiming that a speaker's reluctance to accept an input corresponds in the grammar to free variation between an overt output for that input and the null parse. In other words, when the grammar cannot choose between an overt candidate and the null parse, the overt candidate may be pronounced, but will be dispreferred, meaning that the word will be more likely to fall out of use (or not be formed in the first place) than another input that is mapped unambiguously to a single output. The situation is illustrated in the tableau in (23) (remember that CORRESPOND-CC is undominated and so the sibilants in each candidate are assumed to be in correspondence).

(23) The null parse in free variation with another candidate

/tsʰé + tʃéé/	IDENT-IO[ant]	MPARSE	AGREE[ant]
☞ tsʰé-tʃééʔ			*
☞ NULL		*	
tʃʰé-tʃééʔ	*!		

The wavy line between MPARSE and AGREE[ant] indicates that the constraints are variably ranked. Because of this variability, the overt output [tsʰé-tʃééʔ] is optimal only some of the time—the null parse is also sometimes optimal. The next section explores the implications of this analysis, and describes a stochastic version of Optimality Theory that makes explicit the notion of variable ranking among constraints.

4.2. Stochastic OT

Boersma and Hayes (2001) use OT constraints to predict the well-formedness of two allophones of /l/ in English words. They observe that in words like *Greeley*, there is free variation between dark and light varieties of /l/, while in words like *leaf* or *feel* there is a strong preference for either light or dark. Furthermore, well-formedness judgments by native speakers correspond to these frequencies; there is not a particularly strong preference for *Gree*[l]*ey* over *Gree*[ɫ]*ey*, but [l]*eaf* is considered much better than [ɫ]*eaf*.

They account for this fact by means of a stochastic OT grammar (Boersma 1997), in which constraints are each assigned a ranking value which is perturbed by random noise each time the grammar is run. Cases of free variation are those in which the ranking values of two constraints are close enough that their ranking may switch, so that different outputs may result each time the grammar is used. In their account, the frequency with which a given output may emerge as optimal correlates with its well-formedness. The intermediate well-formedness of *Gree*[l]*ey* is thus a consequence of its being in free variation with another output.

I propose to extend this idea to the Navajo case—compounds with disagreeing sibilants are somewhat ill-formed (and thus dispreferred) because they are in free variation, not with another pronunciation, but with the null parse. In the next section, I will implement this by constructing a grammar in which MPARSE is stochastically ranked with respect to faithfulness and markedness constraints, so that for inputs with disagreeing sibilants in adjacent syllables, the null parse will win some of the time.

4.3. The Navajo grammar

In order to model the complex behavior of Navajo compounds, I will make several simplifying assumptions. Ideally, the grammar should be able to replicate the lexical bias: some words would always be repaired, while others would always be allowed to surface with disagreeing sibilants, in the same proportions seen in the lexicon. The Boersma and Hayes stochastic OT model as it stands cannot do this—it predicts that multiple winners for a given input will be realized as free variation for that input, not as lexical variation across different inputs. Rather than attempt to develop a new type of stochastic grammar, however, I will simply model the lexical variation in Navajo as if it were free variation. I feel I am justified in this for two reasons. First, Sapir and Hojier (1967) claim that there is in fact variation for individual words: a word with disagreeing sibilants will only sometimes be repaired—the probability of repair increases in faster speech and when the sibilants are closer together (this is a general property of sibilant harmony, not restricted to compounds). Second, it seems plausible that the lexical variation could be the result of "fossilized" free variation; words that are rarely repaired could be memorized by learners as never undergoing repair, while more frequently repaired words could be memorized in their repaired form (or, alternatively, both allomorphs could be stored for such an input).[11] Thus, a grammar that predicts free variation may be a model of an earlier stage of the language, or of recently coined, not-yet-established words.

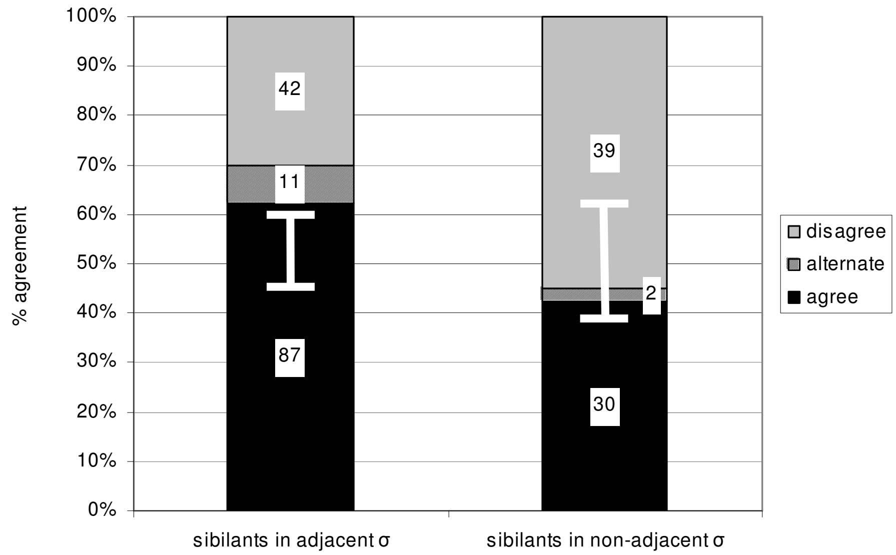
I will thus assume that the learner of Navajo is presented with two patterns of variation: one for compounds with sibilants in adjacent syllables and another for compounds with sibilants in non-adjacent syllables. In the absence of data from actual native speaker behavior, I will assume that this variation is identical to the lexical variation: if, say, 20% of words with disagreeing sibilants in the dictionary are repaired, I will assume that any word with disagreeing sibilants undergoes repair 20% of the time. Of course it is possible that future research will reveal this assumption to be incorrect; for the time being, however, it gives us a starting point to explore what kinds of grammars might result in the observed patterns. The next section reviews the facts of Navajo sibilant agreement in compounds and discusses how to evaluate the success of a hypothetical grammar.

4.3.1. *The facts*

The chart in (24) shows the underlying and surface agreement rates in compounds (reproduced from (20)).

[11] Another possibility: the bias could be an artifact of the process by which the Young et al. (1993) dictionary was compiled—words that undergo repair frequently were more likely to be recorded in their repaired form.

(24) Navajo compounds: underlying vs. surface agreement

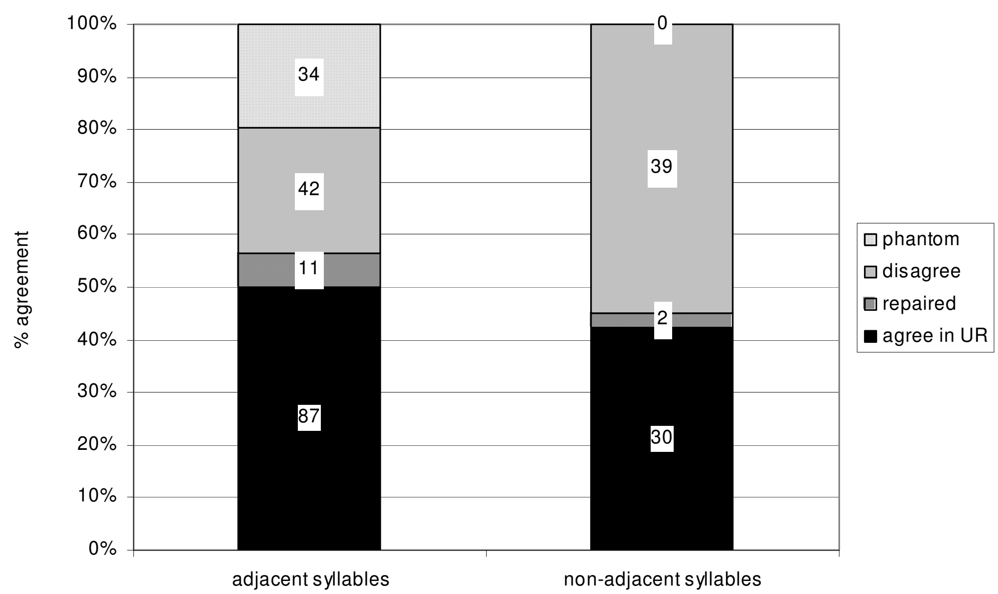


The success of the grammar should be measured by what it does with inputs containing disagreeing sibilants: most should surface faithfully, some should be repaired, and some should be mapped to the null parse. Determining the ratio of faithful to unfaithful mappings is straightforward: it should mirror the same ratio in the lexicon. But the appropriate number of null parse outputs is harder to determine. We obviously cannot count the number of nonexistent compounds in the language and attempt to model that number.

To solve this problem, I will make use of another simplifying assumption: potential inputs are evenly divided between those with agreeing sibilants and those with disagreeing sibilants. The justification for assuming this comes from the raw chance figure calculated earlier: based on the relative frequencies of both sibilant types in the language as a whole, two sibilants taken at random have a 50.1% chance of agreeing. In my model, then, the space of potential compounds has no bias—the underlying bias towards agreement results from some disagreeing compounds being eliminated in favor of the null parse. I will refer to these missing inputs as *phantom* inputs.

The number of phantom compounds that the grammar should produce can be easily calculated, assuming that the number of potential disagreeing compounds is the same as the number of observed agreeing compounds—no constraints penalize compounds whose sibilants agree underlyingly, so all should emerge faithfully. The revised picture of the Navajo lexicon, including phantoms, is shown in (25).

(25) Potential compound space



Because agreement in non-adjacent-syllable words is at chance, I assume that there are no phantom compounds of that type.

The chart in (25) gives us a way to evaluate a potential grammar: given a set of inputs evenly divided between agreers and disagreers, the grammar should produce outputs in roughly the same proportions as those in the chart. In the remainder of this section I will show that, given the set of constraints defined in §4.1, the Gradual Learning Algorithm is capable of learning a ranking that correctly predicts these frequencies.

The Gradual Learning Algorithm (GLA) is designed to learn stochastic OT grammars. It is error-driven: on receiving a training datum in the form of an input-output pair, the algorithm uses its current grammar to generate an output for the input. If the resulting output does not match the training data, all constraints that prefer the incorrect output are incrementally demoted, and all constraints that prefer the correct output are promoted.

In order to find a constraint ranking that correctly predicted the frequencies in the Navajo compound data, I gave the GLA a set of training data structured as in (26).[12] The data consisted of a set of inputs, the possible outputs for each as well as the frequency of each output, and the constraint violations for each output candidate.

(26) GLA training data

Input	Output	Frequency	IDENT-IO	MPARSE	AGREE	AGREE(adj σ)
sV.ʃV	sV.ʃV	48.3%			*	*
	ʃV.ʃV	12.6%	*			
	NULL	39.1%		*		
sV.CV.ʃV	sV.CV.ʃV	95%				*
	ʃV.CV.ʃV	5%	*			
	NULL	0%		*		

[12] The algorithm was implemented using the OTSoft software package (Hayes et al. 2003).

The data also included information about underlyingly agreeing inputs, but because in those cases the winners violate none of the constraints, they harmonically bound the losers and so provide no ranking information to the learner.

This training data was designed with the assumption, mentioned earlier, that what learners encounter in Navajo is free variation, not lexical variation. However, the ranking that the algorithm arrives at is roughly the same if the actual set of Navajo compounds (with no variation for individual words) is used as training data. Because there is no crucial difference, the results I will discuss here are those obtained using the training data in (26).

The data has one more unusual property: it makes use of negative evidence. Some of the candidate outputs are the null parse, which means that the learner is relying on information about nonexistent words in order to decide on a grammar. I discuss the implications of this for real-world learnability in §4.3.2.

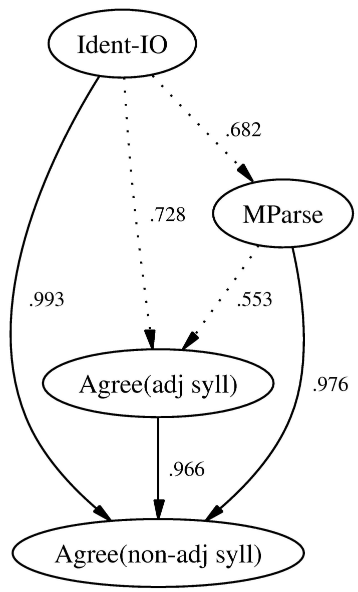
When the GLA was given the training data and allowed to learn for 100,000 cycles (with an initial plasticity of 2 and final plasticity of .00002), the ranking values in (27) were the result.

(27) Results of Gradual Learning Algorithm: ranking values

Constraint	Ranking Value
IDENT-IO	102.732
MPARSE	101.395
AGREE(adj σ)	101.015
AGREE(non-adj σ)	95.872

The Hasse diagram in (28) shows how the constraints are ranked relative to one another; arrows point from constraints with higher ranking values to those with lower values. The number next to each arrow gives the probability that the upper constraint will dominate the lower constraint on any given run of the grammar.

(28) Results of Gradual Learning Algorithm: Hasse diagram



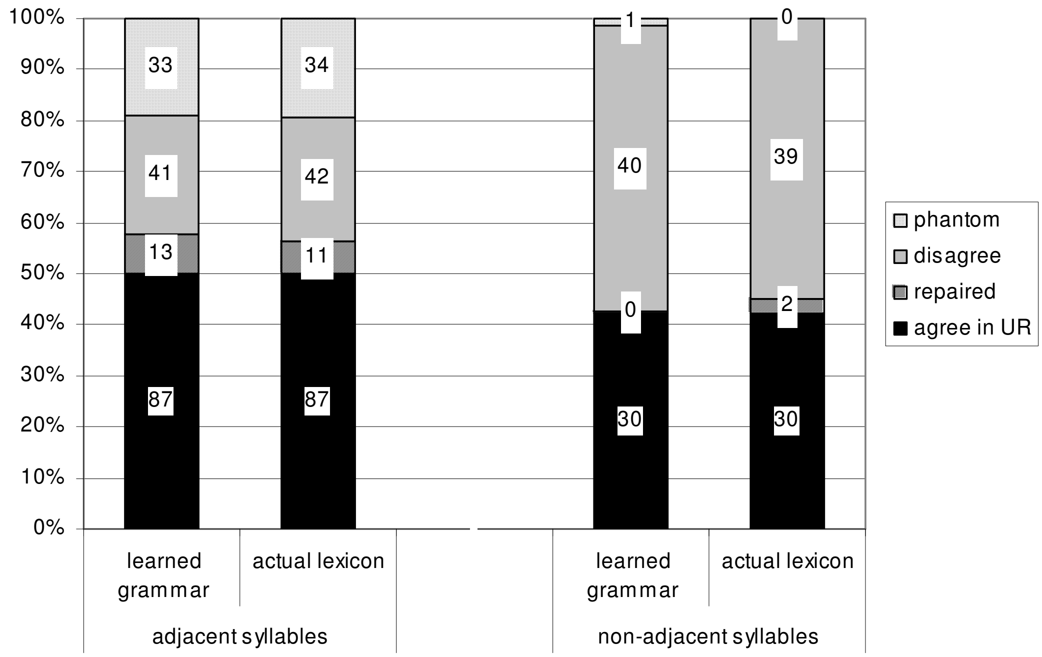
The chart in (29) compares the frequencies produced by this grammar for various outputs as compared to the frequencies in the training data.

(29) Grammar versus training data

Input	Output	Input Frequency	Frequency Generated by Grammar
sV.ʃV	sV.ʃV	48.3%	47.0%
	ʃV.ʃV	12.6%	15.2%
	NULL	39.1%	37.8%
sV.CV.ʃV	sV.CV.ʃV	95%	96.9%
	ʃV.CV.ʃV	5%	0.1%
	NULL	0%	2.5%

The next chart shows this comparison graphically by demonstrating the shape of the lexicon predicted by the grammar (assuming it was given the same number of potential compounds as the actual lexicon) and juxtaposing it with the frequencies observed in the real lexicon.

(30) Grammar versus lexicon



The grammar matches the input frequencies quite well overall, with the exception of underpredicting the rate of alternation and overpredicting the rate of optimal null parses in the non-adjacent-syllable words. But because the number of compounds is quite small (211), and because I've made a number of simplifying assumptions, I do not feel that matching the lexical frequencies exactly is particularly important. What is important is the overall picture— a grammar can be constructed using a small number of stochastically-ranked constraints that emulates the tripartite behavior of

Navajo speakers when confronted with AGREE violations in compounds, and furthermore tolerates violations much more readily when the sibilants are more distant.

4.3.2. *The null parse and learnability*

I have proposed that the underrepresentation of words with marked structures be modeled as free variation between the marked output and the null parse. However, as noted above, this poses a problem for learnability. In order to deduce the correct ranking of MPARSE, the learner must have access to negative evidence—the frequency with which the null parse is chosen over its competitors.

This is a general problem for statistical learning of this sort—a learner cannot realize that something is occurring less often than expected without knowing what the expected frequency is. Ideally, the learning algorithm would compute these expected values using values that are observable—the frequencies of sibilant types in general, for example. The expected values could then be used to rank MPARSE appropriately. The development of such an algorithm is beyond the scope of this paper, however, and is left as a topic for future research.

5. Conclusion

Navajo compounds tell us three new things about the effects of distance on lexical biases: first, distance cannot simply be measured in segments—a more complex metric, possibly involving prosodic structure, is required. Second, distance can gradiently affect an active process like sibilant harmony—the process applies even to distant sibilants, but does so less often than with closer sibilants. Third, the effects of distance can be seen in the underlying distribution of compounds.

Taken together, the second and third points suggest that there is a single mechanism behind active processes and underlying biases. I have argued that this mechanism consists of a set of stochastically ranked OT constraints. In this model, each of the three possible responses to disagreement—toleration, repair, and avoidance—corresponds to one possible constraint ranking:

(31) Possible rankings

(a)	Faith, MPARSE » Mark	toleration
(b)	Mark, MPARSE » Faith	repair
(c)	Faith, Mark » MPARSE	avoidance

In standard, non-stochastic OT, the constraints I have described would establish a typology of responses to a given markedness problem: some languages would tolerate the problem, others would repair it, and in others absolute ungrammaticality would result. The Navajo compound data demonstrates the need for stochastic ranking, as all three possibilities are attested for a single type of markedness.

There is still much to be learned concerning the issues raised in this paper: how exactly distance is measured, how lexical biases are learned and represented in the grammar, and how learners can make use of negative evidence. It is hoped that future research into the effects of distance on lexical biases will shed more light on these topics.

REFERENCES

Archangeli, Diana and Doug Pulleyblank. 1987. Maximal and Minimal Rules: Effects of Tier Scansion. *Proceedings of NELS 17*. 16-35.

Berkley, Deborah. 1994. The OCP and gradient data. In James H. Yoon (ed.), *Studies in the Linguistic Sciences* 24: 2. Urbana, Illinois: University of Illinois. 59-72.

Berkley, Deborah. 2000. Gradient OCP Effects. Ph.D. dissertation. Northwestern University.

Boersma, Paul. 1997. How We Learn Variation, Optionality, and Probability. *Institute of Phonetic Sciences Proceedings* 21: 43-58.

Boersma, Paul and Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32: 45-86.

Burke, Sean and Jordan Lachler. 2000. Online analytical lexicon of Navajo. Web-based version of Young, Morgan, and Midgette 1992. http://www.speech.cs.cmu.edu/egads/navajo/

Ferguson, C.A. and C. Farwell. 1975. Words and sounds in early language acquisition: English initial consonants in the first 50 words. *Language* 51: 419-439.

Fountain, Amy. 1998. An Optimality Theoretic approach to Navajo prefixal syllables. Ph.D. dissertation. University of Arizona. Available at Rutgers Optimality Archive (http://roa.rutgers.edu) as ROA-238.

Frisch, Stefan. 1997. Similarity and Frequency in Phonology. Ph.D. dissertation. Northwestern University.

Frisch, Stephan. 2000. Temporally Organized Representations as Phonological Units. In Michael Broe and Janet Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon.* Cambridge: Cambridge University Press. 283-298.

Frisch, Stefan, Janet Pierrehumbert and Michael Broe. 2004. Similarity Avoidance and the OCP. *NLLT* 22: 179-228.

Greenberg, Joseph. 1950. The Patterning of Root Morphemes in Semitic. *Word* 5. 162-181.

Gussenhoven, Carlos. 1986. English Plosive Allophones and Ambisyllabicity. *Gramma* 10.119-141.

Halle, Morris, Bert Vaux, and Andrew Wolfe. 2000. On Feature Spreading and the Representation of Place of Articulation. *Linguistic Inquiry* 31: 3. 387-444.

Hay, Jennifer, Janet Pierrehumbert and Mary Beckman. 2003. Speech Perception, Well-formedness, and the Statistics of the Lexicon. In *Papers in Laboratory Phonology VI.* Cambridge: Cambridge University Press.

Hayes, Bruce, Bruce Tesar and Kie Zuraw. 2003. OTSoft 2.1, software package. http://www.linguistics.ucla.edu/people/hayes/otsoft/

Hooper, Joan. 1978. Constraints on Schwa-deletion in American English. In J. Fisiak (ed.), *Recent Developments in Historical Phonology.* The Hague: Mouton.

Kahn, Daniel. 1976. Syllable-based Generalizations in English Phonology. Ph.D. dissertation. MIT.

Kessler, Brett. 2001. The Significance of Word Lists. Stanford: CSLI Publications.

Leonard, L.B., R.G. Schwartz, B. Morris and K. Chapman. 1981. Factors influencing early lexical acquisition: lexical orientation and phonological composition. *Child Development* 52: 882-887.

McCarthy, John. 1986. OCP Effects: Gemination and Antigemination. *Linguistic Inquiry* 17. 207-263.

McCarthy, John. 1988. Feature Geometry and Dependency: A Review. *Phonetica* 43: 84-108.

McDonough, Joyce. 1991. On the representation of consonant harmony in Navajo. In *Proceedings of the Tenth West Coast Conference on Formal Linguistics*, 319-335.

Ohala, John. 1981. The listener as a source of sound change. In C.S. Masek, R.A. Hendrick, and M.F. Miller (eds.) *Papers from the Parasession on Language and Behavior*. Chicago: Chicago Linguistic Society. 178-203.

Orgun, Cemil Orhan and Ronald Sprouse. 1999. From MPARSE to CONTROL: Deriving Ungrammaticality. *Phonology* 16. 191-224.

Prince, Alan and Paul Smolensky. 1993. Optimality Theory: Constraint Interaction in Generative Grammar. Ms., Rutgers University and University of Colorado.

Rose, Sharon and Rachel Walker. 2004. A Typology of Consonant Agreement as Correspondence. *Language* 80: 3. 475-531.

Rubach, Jerzy. 1996. Shortening and ambisyllabicity in English. *Phonology* 13. 197-237.

Sapir, Edward and Harry Hoijer. 1967. *The phonology and morphology of the Navaho language*: University of California Publications in Linguistics 40. Berkeley: University of California Press.

Schwartz, R.G. and L.B. Leonard. 1982. Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language* 9: 319-336.

Shaw, Patricia. 1991. Consonant Harmony Systems: the Special Status of Coronal Harmony. In Carole Paradis and Jean-Francois Prunet (eds.), *Phonetics and Phonology, Volume 2: The Special Status of Coronals*. San Diego: Academic Press. 125-157.

Steriade, Donca. 1987. Locality Conditions and Feature Geometry. *Proceedings of NELS 17*. 595-617.

Suzuki, Keiichiro. 1998. A Typological Investigation of Dissimilation. Ph.D. dissertation. University of Arizona.

Vennemann, Theo. 1981. Phonology as Non-functional Non-phonetics. *Phonologica 1980*. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.

Young, Robert W., William Morgan Sr. and Sally Midgette. 1992. Analytical Lexicon of Navajo. Albuquerque: University of New Mexico Press.