Modeling the interlanguage: The effect of frequency in the L2 acquisition of English consonant clusters

Emily Moeng, Jennifer Boehm, & Amy Reynolds
University of North Carolina at Chapel Hill
moeng@live.unc.edu; jengriff@live.unc.edu; amyrey@live.unc.edu

This study seeks to determine whether the second language acquisition of non-native linguistic structures is best described within the framework of Optimality Theory, or within the framework of a usage-based model. These two frameworks make different predictions regarding the role of word frequency in second language acquisition. This study examines the productions of high- and low-frequency English words as spoken by native speakers of S'gaw Karen, a language which does not permit coda clusters, who are acquiring English as a second language. Trends suggest that second language acquisition is better described within the framework of a usage-based model, suggesting a word-by-word acquisition of coda clusters, although limitations of the study are also noted in the discussion.

1. Introduction

The goal of this paper is to determine whether the acquisition of new linguistic structures in the intermediate, or "interlanguage", grammar produced during the course of second language acquisition is best described on a word-by-word basis, or as an overall acquisition. We argue that the interlanguage grammar predicted within the framework of a usage-based model (Bybee 2001, Bybee 2006, Larsen-Freeman 2007) would predict the production of new linguistic structures which are positively correlated with word frequency, whereas an interlanguage grammar predicted within the framework of Optimality Theory (Prince & Smolensky 2008, Hancin-Bhatt & Bhatt 1997) would show no correlation between the production of new linguistic structures and word frequency. This study analyzes the development of coda consonant clusters by speakers whose first language (L1) does not allow coda clusters. Specifically, this study looks at the production of the coda clusters [-nd] (as in *fu<u>nd</u>*) and [-ks] (as in *mi<u>x</u>*) by native speakers of S'gaw Karen, a language which does not permit coda clusters.

Studies in the Linguistic Sciences: Illinois Working Papers ***2016****: 35-48*

2. Background

This paper examines the acquisition of English coda clusters by speakers of S'gaw Karen, a language which does not permit coda clusters. This section will provide background on the relevant phonology of S'gaw Karen and English (Section 2.1), and then outline predictions made by two competing theories: Optimality Theory (Section 2.2), and Exemplar Theory (Section 2.3).

2.1. Consonant clusters in the L1 (S'gaw Karen) and in the L2 (English)

While most syllables in S'gaw Karen are of the form CV, they can have complex onset clusters of the form C_1C_2, where C_2 is an approximant. Consonant clusters are not allowed in the coda, but some consonants (nasals and liquids) are allowed to make up a simple coda (Fischer 2013).

English, on the other hand, allows a number of coda clusters in both the onset and in the coda. This study focuses on two English coda clusters: [-ks], as in the word *box*, and [-nd], as in the word *find*. These clusters are specifically chosen because they each exhibit different behaviors in native English speakers. Specifically, the coda cluster [-ks] is produced faithfully nearly 100% of the time, whereas the coda cluster [-nd] is faithfully produced only a portion of the time. Previous studies have found that English speakers will reduce [-nd] codas more often in high-frequency words than in low-frequency words. For example, the high-frequency word *find* will be reduced at a higher rate into *fin(d)* than the low-frequency word *fund* will be reduced to *fun(d)* (Labov 1989, Bybee 2002, Hooper 1976).

2.2. Optimality Theory: Frequency effects and L2 acquisition

Optimality Theory assumes that a grammar consists of two types of constraints (Faithfulness constraints and Markedness constraints), an input, and an output. Faithfulness constraints require an output (phonetic production) to match its input (underlying form), while Markedness constraints require the output to be well-formed in some way (regardless of input). For example, the Faithfulness constraint MAX requires all input segments to also appear in the output, thus preventing the deletion of input segments. In contrast, the Markedness constraint *COMPLEX requires the output to be free of consonant clusters. These two types of constraints are often at odds with one another, with Faithfulness constraints requiring the output to be unchanged from its input, while Markedness constraints require the output to be unmarked in structure. Optimality Theory assumes that a language will have a single ranking of all of its Faithfulness and Markedness constraints.

Unlike a rule-based theory of grammar, Optimality Theory assumes that an input has multiple output candidates, each of which may violate Faithfulness and/or Markedness constraints. Each of these candidates is evaluated so that the candidate which has incurred the least serious violations is chosen as the output surface form. The violation of a high-ranked constraint counts as a more serious violation than the violation of a low-ranked constraint.

Within the framework of Optimality Theory, L2 acquisition occurs when learners copy their original L1 constraint rankings and re-rank these to match the rankings of the target L2. This can occur through the demotion of Markedness constraints (Hancin-Bhatt & Bhatt 1997), and/or through the demotion of Faithfulness constraints (Hayes 1999, Zhang 2013, Hancin-Bhatt 2008). Previous studies have found the use of Optimality Theory useful for accounting for second language phenomena, such as L1 transfer effects on a second language, as well as the emergence of universal preferences during second language acquisition (Hancin-Bhatt & Bhatt 1997, Zhang 2013, Broselow 2004). This paper adds to these studies by determining whether word frequency plays a role in the acquisition of new linguistic structures.

This study assumes a simple stochastic model of Optimality Theory, which allows for variation within speakers due to the addition of noise to the model (Boersma & Hayes 2001). In a stochastic model of Optimality Theory, the L1 (S'gaw Karen), which does not permit coda clusters, would have the ranking shown at the top of Figure 1, in which the Markedness constraint *COMPLEX-CODA prohibiting coda consonant clusters outranks Faithfulness constraints such as DEP (which avoids insertions), MAX (which avoids deletions), and IDENT (which avoids changes). (For simplicity, this paper will be treating these as a bundle of Faithfulness constraints.) Because *COMPLEX-CODA outranks these Faithfulness constraints, and would therefore be a more serious violation than inserting a vowel to avoid a coda cluster (a violation of DEP), or deleting a consonant to avoid a coda cluster (a violation of MAX), or changing a segment to avoid a coda cluster (a violation of IDENT), then no matter the input, a speaker with this ranking will not produce coda clusters.

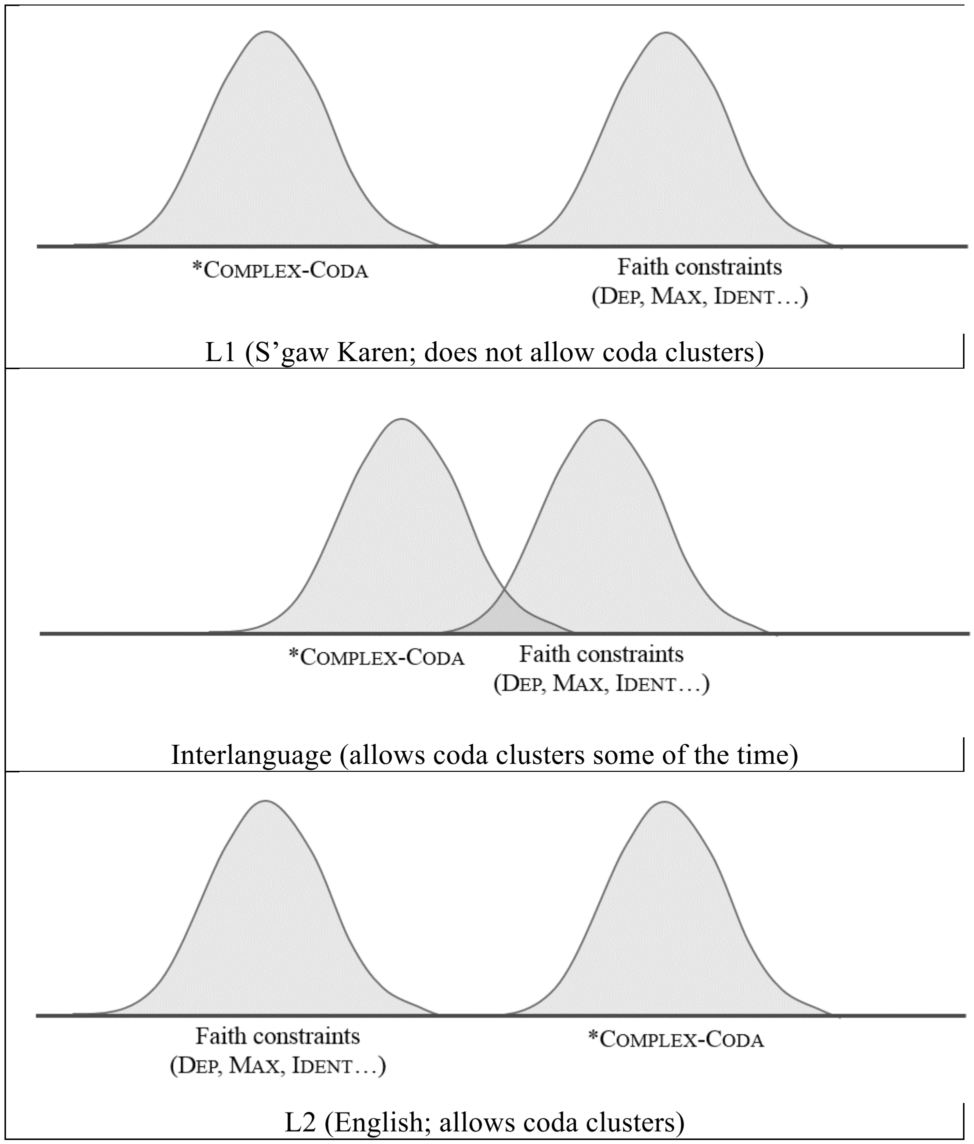


Figure 1. The constraint rankings for the L1 S'gaw Karen, which does not allow coda clusters (top); the constraint rankings for the interlanguage grammar (middle); and the constraint rankings for the L1, English, which does allow coda clusters (bottom).

In contrast to the speaker's initial language (S'gaw Karen), the target language (English) does permit coda clusters. Therefore, in English, at least one Faithfulness constraint now outranks the Markedness constraint *COMPLEX-CODA, as seen in the bottom panel of Figure 1.

As noted earlier, we will be assuming a model in which the acquisition of a second language begins with a copy of the first language's rankings, and then gradually shifts rankings to match those of the target language. In this case, speakers of S'gaw Karen must demote the Markedness constraint *COMPLEX-CODA and/or promote Faithfulness constraints. Therefore, during the process of language acquisition, the interlanguage will have the rankings shown in the middle panel of Figure 1.

In a stochastic model of Optimality Theory, overlapping constraints, as seen in the middle panel of Figure 1, will result in variable productions within a single speaker (Boersma & Hayes 2001). Therefore the image shown depicting the interlanguage in Figure 1 shows the case where most of the time (say, 70% of the time), the speaker does not produce coda clusters, since *COMPLEX-CODA still mostly outranks the Faithfulness constraints. But because these constraints overlap, speakers with this interlanguage grammar will still produce coda clusters 30% of the time.

The exact percentage of words whose coda clusters are reduced will change over time as the speaker shifts his or her constraint rankings to match the rankings of the target language. The important point to note is that in this simple stochastic model, frequent words and infrequent words will be reduced the same amount. That is, if we suppose a speaker fails to produce a coda cluster in a frequent word such as *find* 70% of the time, we would expect this speaker to fail to produce the coda cluster in an infrequent word, such as *fund*, 70% of the time as well.

To summarize, within an Optimality Theoretic model, we would expect word frequency to have no effect on the rate at which a speaker fails to fully produce a coda cluster.[1]

2.3. Usage-based models: Frequency and L2 acquisition

Unlike an Optimality Theoretic model, a usage-based model (Bybee 2002, Pierrehumbert 2001) would predict word frequency to have an effect on coda cluster reduction rates. Within a usage-based model, speakers acquiring a second language would begin with clouds of exemplars, specific instances of words that have been experienced by the listener. When acquiring a second language, these speakers will add exemplars from the second language to their existing exemplar clouds.

[1] A model that allows lexical indexation for frequent and infrequent words would allow for an effect of word frequency, and would predict that frequent words would be reduced more than infrequent words, in a stepwise function. This paper, however, will only be examining the simple case in which lexical indexation for frequency is not allowed.

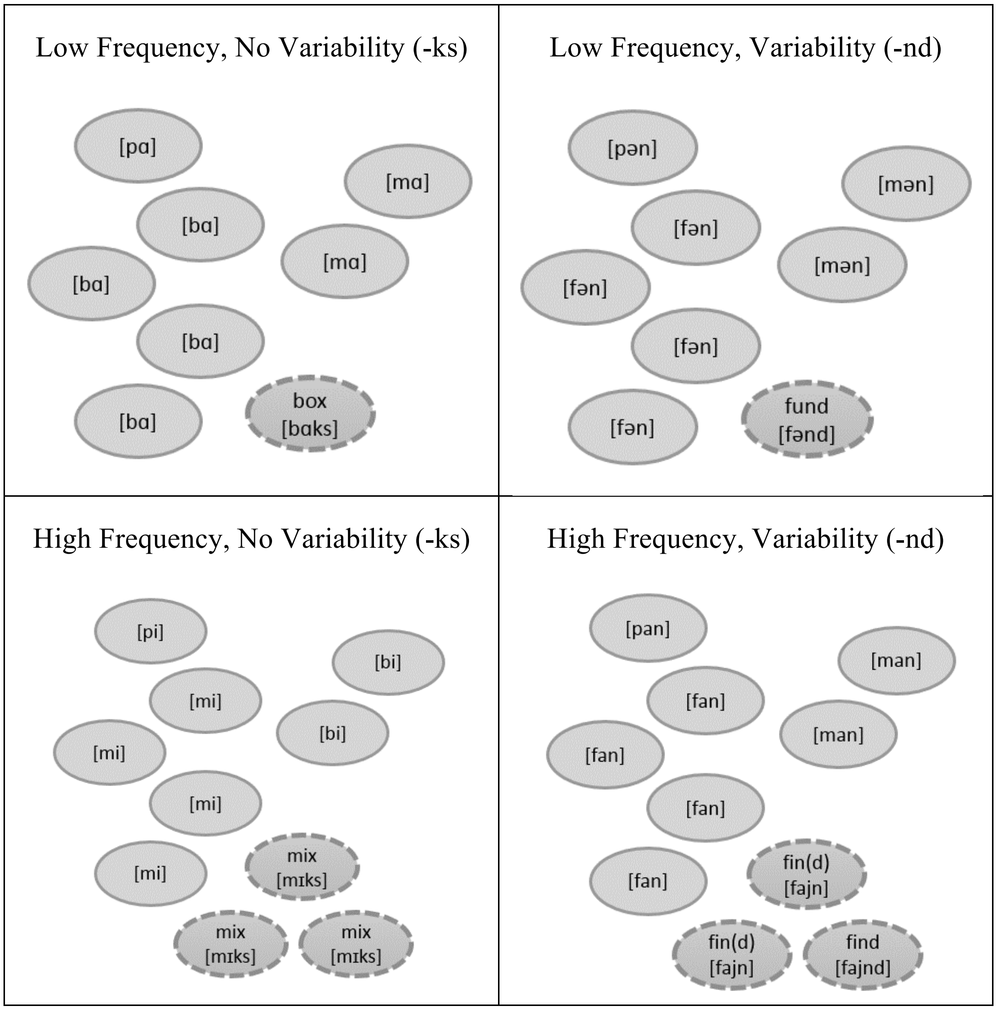


Figure 2. Sample exemplar clouds for four cases. Circles with dashed borders represent exemplars heard from the L2, which allows coda clusters. Circles with solid-line borders represent exemplars from the L1, which does not allow coda clusters.

Figure 2 shows an intermediate stage of second language acquisition for four cases: the second language acquisition of low-frequency words (top row), high-frequency words (bottom row), words which show little to no variability (those with [-ks] coda clusters, left column), and words which do show variability by English speakers (those with [-nd] coda clusters, right column). For each of these cases, speakers begin with exemplar clouds of words in their first language, none of which have coda clusters since S'gaw Karen does not allow them. This is indicated with circles with solid-line borders. Gradually, learners hear exemplars from the L2, English, and add these to their exemplar clouds. These exemplars from English are indicated with circles with dashed-line borders. For the case of words with [-ks] coda clusters, these L2 exemplars are heard by learners as unreduced nearly

100% of the time, since English speakers do not reduce [-ks] coda clusters. However, because English speakers reduce high-frequency words ending with [-nd] coda clusters frequently, a learner would hear these exemplars being reduced quite often, hence the variation seen in the bottom right cell of Figure 1 representing the acquisition of the L2 word *find*. When second language learners then attempt to produce a word like *find* in their L2 (English), they access their exemplar clouds for all instances of *find* that they have heard in order to determine how it should be produced. Since *find*, unlike *mix*, has variation in what they have heard, it is more likely that the second language learner will produce a form where the final consonant cluster is reduced rather than produced. For *mix*, on the other hand, in which the second language learners have not heard any variation, we would expect them to more likely produce the fully produced [-ks] cluster rather than the reduced [-k].

3. Research question and predictions

Because S'gaw Karen does not allow coda consonant clusters, speakers must acquire coda clusters when acquiring English. This study will be comparing the acquisition of two English coda clusters: [-ks] and [-nd]. These clusters are chosen because each exhibits different behaviors with regards to word frequency among native English speakers. [-ks] codas are produced fully nearly 100% of the time, regardless of word frequency, while [-nd] codas exhibit variation dependent on word frequency, being reduced to [-n] codas more often in frequent words than in infrequent words by native English speakers.

<table>
<tr><th>L1 (S'gaw Karen)
No coda consonant clusters</th><th>Interlanguage</th><th>L2 (English)</th></tr>
<tr><td>HighFreqRed = LowFreqRed = 100%</td><td rowspan="2">?

Best described on a word-by-word basis, or with an overall constraint re-ranking?</td><td>HighFreqRed > LowFreqRed

for some clusters ([-nd] codas)</td></tr>
<tr><td>HighFreqRed = LowFreqRed = 100%</td><td>HighFreqRed = LowFreqRed = 0%

for some clusters ([-ks] codas)</td></tr>
</table>

Table 1. This study seeks to determine how to best describe the interlanguage grammar: on a word-by-word basis, or through an overall constraint re-ranking.

S'gaw Karen speakers who are learning English as a second language must transition from reducing coda clusters 100% of the time (Table 1, column 1) to producing coda clusters at rates comparable to native English speakers (Table 1, column 3).

A stochastic Optimality Theoretic account of L2 acquisition (with no frequency-based lexical indexation[2]) predicts no correlation between coda cluster reduction rates and lexical or phrasal frequency, since the constraints relevant to coda clusters are re-ranked, affecting the entire grammar regardless of word frequency, as shown in Table 2. For example, if the interlanguage grammar produces [-nd] coda clusters 30% of the time, this would apply to *all* [nd]-coda lexical items, regardless of the frequency at which it has been heard.

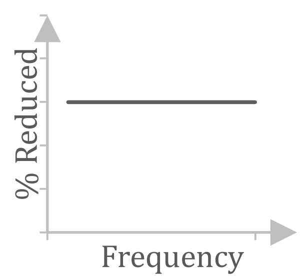
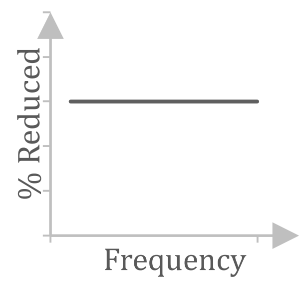
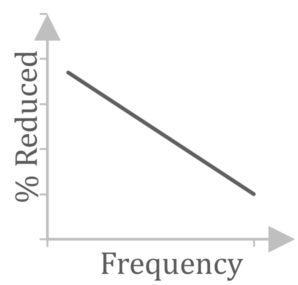
	English speakers show **NO VARIABILITY** ([-ks])	English speakers show **VARIABILITY** ([-nd])
Optimality Theory		
Usage-based model		Unknown

Table 2. Predictions made by Optimality Theory and a usage-based model.

In contrast, a usage-based account of L2 acquisition would predict that the coda clusters in high-frequency words and/or phrases are acquired more

[2] Indexation of high and low-frequency words would predict a stepwise function, or, if lexical class affiliation is not fixed as in Coetzee (2008), would predict a smooth increasing or decreasing curve relation, as is the case for the usage-based account.

quickly than the coda clusters in low-frequency words or phrases (Jurafsky 1996, Bybee 2006). This difference in rate of acquisition would result from the quicker development of high-frequency word exemplar clouds compared to that of low-frequency words. The exemplar clouds for high-frequency words would develop more rapidly and contain more examples wherein coda clusters would be fully produced due to phonological and morphemic contexts. Therefore, as shown in Table 2, a usage-based model would predict that more frequent words ending with a [-ks] coda cluster will be reduced less often, as learners have had more experience with these words and have therefore learned them faster. It is unknown what a usage-based model would predict for words ending with [-nd] coda clusters, since English speakers themselves reduce frequent [-nd] words more often than infrequent [-nd] words. Because learners of English are receiving a higher rate of unreduced exemplars of infrequent [-nd] words compared to frequent [-nd] words, it is unknown how this would balance out with the fact that learners are receiving a larger quantity of the more frequent words. Therefore we make no prediction within a usage-based model concerning the effect of word frequency on reduction rate for the [-nd] words.

4. Methods

4.1. Participants

This study examines non-native English as spoken by speakers of S'gaw Karen, an under-documented Tibeto-Burman language spoken in Myanmar. Due to ongoing conflict in Myanmar, over 18,000 Burmese refugees have settled in the southeast region of the United States since 2005 (USOoRR 2012), yet the language remains widely understudied. Participants were 21 native S'gaw Karen speakers living in the Carrboro area of North Carolina, ranging from the ages of 18-48. 16 participants were female, and 5 were male. Participants had lived in the United States for anywhere from 1 to 9 years and were between 6 and 18 years old when they began to study English, with most of the participants studying English in school while living in Myanmar or Thai refugee camps. Data was also collected on what other languages the participants spoke, with Pwo Karen and Burmese being the most commonly spoken second languages.

4.2. Stimuli

28 English verbs were chosen, half of which ended with a [-ks] coda cluster, and half of which ended with an [-nd] coda cluster. Of those, half were

monomorphemic (e.g. *flex*), and half were bimorphemic (e.g. *cook-s*). Verbs were chosen to have a range of token frequencies. Frequencies were taken from the Corpus of Contemporary American English (Davies 2008), which contains approximately 450 million tokens of English words, and are presented in Table 3.

	Monomorphemic		**Bimorphemic**	
	Word	*Frequency*	*Word*	*Frequency*
[-ks] words	flex	632	cooks	1,364
	box	1,427	attacks	1,533
	fax	2,576	kicks	2,780
	tax	2,888	breaks	7,539
	relax	7,441	picks	9,703
	mix	8,990	speaks	12,455
	fix	11,479	takes	73,032
[-nd] words	fund	4,506	banned	4,570
	defend	12,211	earned	15,124
	mind	18,175	owned	16,557
	send	38,879	signed	24,412
	spend	43,489	joined	28,926
	end	46,741	learned	52,263
	find	193,274	turned	114,067

Table 3. Word frequencies of [-ks] and [-nd] stimulus words. Frequencies are out of a total of approximately 450 million words taken from 189,431 texts (Davies 2008).

Each of the 28 verbs were embedded into two phrases, of the form ***(verb)*** *the* ***(noun)***. For example, *flex* was embedded into the phrase ***flex*** *the ankle* and ***flex*** *the knee*. Each of these 56 phrases was then recorded by one of the experimenters, a native speaker of English, who pronounced each of the [d] portions in the [-nd] clusters with a released [d]. The 56 phrases were normalized for peak amplitude, so that they sounded equally loud when played back, by using Praat, a speech analysis program (Boersma & Weenink 2013).

To ensure that the consonant clusters were identical across all [-nd] words, the [nd ð] portion of one of the recorded phrases, *ear**ned th**e spotlight*, was used to replace all other [nd ð] portions of all other [-nd] phrases. To ensure that the consonant clusters were identical across all [-ks] words, the [ks ð] portion of one of the recorded stimuli, *atta**cks th**e brain*, was used to replace all other [ks ð] portions of all other [-ks] phrases. This ensured that [-nd] words had the same [n] duration (62 ms) and the same [d] closure duration

(15 ms), followed by a visible stop release, and that all [-ks] words had the same [k] closure duration (56 ms) and [s] duration (14 ms).

4.2. Procedure

Participants heard each phrase played one at a time in random order through a Logitech 530 USB headset at a comfortable volume in a quiet room. As the phrase was played, participants also saw the phrase written in English on a screen. This ensured that participants would have orthographical evidence that each of these words contained a coda consonant cluster, even if the participant was not familiar with the particular word. Each list was played twice, resulting in 112 (56 x 2) total phrases. Participants were asked to repeat back the phrase they had heard, matching the pronunciation as closely as possible. They were also told that they could ask to have it played back as many times as they wished.

5. Analysis

The authors scored each produced phrase as "reduced" if the oral stop in each coda cluster was deleted (that is, if the [d] was deleted from an [-nd] cluster, or if the [k] was deleted from a [-ks] cluster). Other "repair" methods were sometimes employed by the speakers to avoid the cluster, such as the insertion of a vowel ("fu-ned" or "fun-de" instead of "fund") or metathesis of the consonants. Since consonant cluster reduction involves the deletion of a segment from the cluster and these alternate repair methods merely move or separate the cluster segments, neither involve true reduction, so they were not considered in the analysis for this study. Phrases in which it was judged to be unclear by the scorer whether the oral stop had been deleted or not were also removed from analysis. All phrases in which the consonant cluster was fully produced were marked as "not reduced". These judgments were made by both listening to the recordings and by viewing the spectrogram and waveform of the audio recording through Praat.

6. Results

Figure 3 shows the percent of all reduced coda clusters across all speakers. Word frequency is shown on a logarithmic scale.

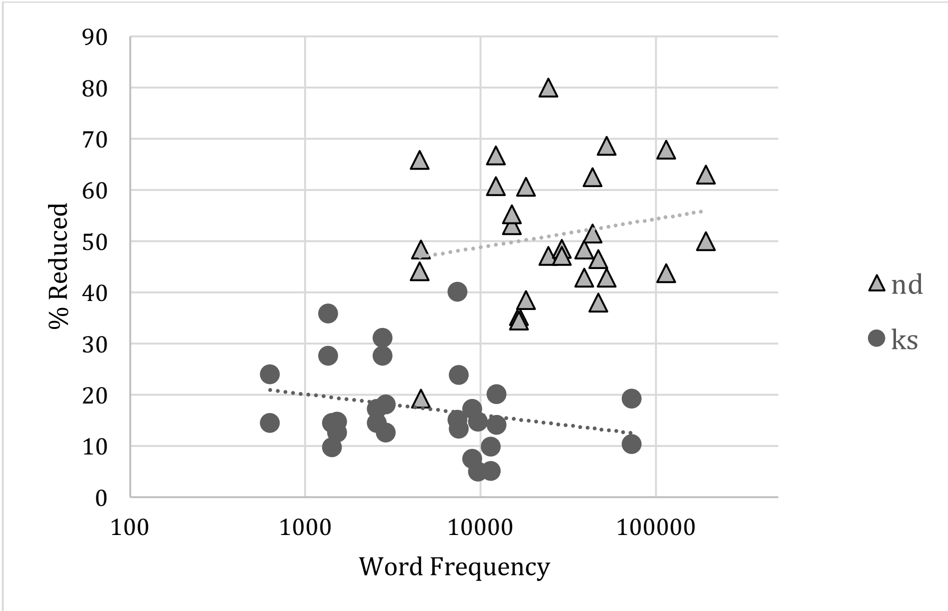


Figure 3. The effect of word frequency on reduction rates. [-ks] words are less reduced the greater the word frequency, whereas [-nd] words are more reduced the greater the word frequency.

As can be seen from Figure 3, the greater the word frequency, the less [-ks] words (indicated with circles) were reduced, as predicted by the usage-based model. However, a logistic regression analysis conducted to predict the reduction rate of [-ks] words using word frequency as a predictor found that this trend was not statistically significant (p = 0.39).

A logistic regression analysis was also conducted to predict the reduction rate of [-nd] words (indicated with triangles) using word frequency as a predictor. It was found that the positive trend between word frequency and reduction rate was also not statistically significant (p = 0.13).

6. Conclusion and discussion

This study found trends supporting a model in which learners acquire new structures on a word-by-word basis, with [-ks] words being reduced less the more frequent the word. However, these trends were not found to be statistically significant, making our results inconclusive.

Although the trends found in this study support a usage-based model, it should be noted that there may be individuals within this study who in fact used an Optimality Theoretic or rule-based acquisition approach whose results were obscured by speakers who adhered to a more usage-based approach. It may be the case that second language learners utilize different approaches as they progress through the phases of second language acquisition. For example, a learner may start by learning a novel phonological sequence such as coda clusters on a word-by-word basis and later forming general constraint rankings or rules which are then applied to the interlanguage as a whole. An Optimality Theoretic or rule-based model's predicted results (with no word frequency effect) may have been obscured in this study by including multiple second language learners at varying stages of acquisition. If a learner were not making a distinction based on word frequency at a later stage of acquisition, the pooling of that learner's results with other learners who were making such a distinction would make it appear that the entire group was distinguishing based on word frequency. We did not test the participants' English proficiency or acquisition levels, making it impossible to distinguish results based on phase of acquisition. Future studies would benefit from including such a metric in their considerations to see whether different methods are used by learners at different stages of acquisition.

It should also be noted that there were a number of limitations to this study, the largest being in the choice of stimuli. Because this study already had a number of requirements (word frequency, coda cluster, number of morphemes, etc.), it was found to be nearly impossible to also control for the surrounding phonetic environment. Because of this, stimuli did not simply vary in word frequency, but also in syllable number and in the identity of the preceding vowel and even in orthography, all of which may have had some effect on cluster reduction rates.

Although there are undoubtedly many weaknesses of the current study, we believe this is a question worth further study, as it compares predictions made by two main theories within linguistics. It may be that this question would be better answered by observing learners acquiring new structures in an artificial language, in which orthography and frequency can be easily controlled for.

REFERENCES

Boersma, Paul & Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32. 45-86.

Boersma, Paul & David Weenink. 2001. Praat, a system for doing phonetics by computer. 341-345.

Broselow, Ellen. 2004. Unmarked structures and emergent rankings in second language phonology. *International Journal of Bilingualism*, 8(1). 51-65

Bybee, Joan L. 2001. Phonology and Language Use. Cambridge: Cambridge University Press.

Bybee, Joan L. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14. 261-290.

Bybee, Joan L. 2006. From usage to grammar: The mind's response to repetition. *Language*. 711-733.

Coetzee, Andries W. 2008. "Phonological variation and lexical frequency."

Davies, Mark. 2008. The Corpus of Contemporary American English: 450 million words, 1990-present. Available online at http://corpus.byu.edu/coca/.

Fischer, Lucia. 2013. Sgaw Karen as spoken by a member of the local North Carolina Karen community: a phonetic analysis and phonemic description. Master's thesis, The University of North Carolina, Chapel Hill, North Carolina.

Hancin-Bhatt, Barbara & Rakesh M. Bhatt. 1997. Optimal L2 syllables. *Studies in Second Language Acquisition* 19(3). 331-378.

Hancin-Bhatt, Barbara. 2008. Second language phonology in Optimality Theory. In Mary L. Zampini & Jette G. Hansen Edwards (eds.), *Phonology and Second Language Acquisition*, 117-146. Amsterdam/Philadelphia: John Benjamins.

Hooper, Joan B. 1976. Word frequency in lexical diffusion and the source of morphological change. In William M. Christie, ed. Current Progress in Historical Linguistics. Amsterdam: North-Holland Publishing Co. p. 95-105.

Jurafsky, Daniel. 1996. A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science* 20(2). 137-194.

Labov, William. 1989. The child as linguistic historian. Language Variation and Change, 1:85-97.

Larsen-Freeman, Diane. 2007. Reflecting on the cognitive social debate in second language acquisition. *The Modern Language Journal* 91. 773-787.

Lieven, Elena, Heike Behrens, Jennifer Speares & Michael Tomasello. 2003. Early syntactic creativity: A usage-based approach. *Journal of Child Language* 30(2). 333-370.

Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Bybee, Joan & Paul Hopper (eds.), Frequency and the Emergence of Linguistic Structure, 137-58.

Prince, Alan & Paul Smolensky. 2008. Optimality Theory: Constraint Interaction in Generative Grammar. John Wiley & Sons.

Zhang, Hang. 2013. The Second Language Acquisition of Mandarin Chinese Tones by English, Japanese and Korean Speakers. PhD diss., The University of North Carolina at Chapel Hill.