“Bienvenido al vecindario”: Inserting Spanglish nonwords into English versus Spanish phonological neighborhoods.

A Dissertation Presented

by

April Pufahl

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Cognitive Science

Stony Brook University

December 2014

Stony Brook University

The Graduate School

April Pufahl

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Arthur G. Samuel – Dissertation Advisor
Professor, Cognitive Science

Susan E. Brennan - Chairperson of Defense
Professor, Cognitive Science

Antonio L. Freitas
Associate Professor, Social and Health Psychology

Marie K. Huffman
Associate Professor, Linguistics

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

"Bienvenido al vecindario": Inserting Spanglish nonwords into

English versus Spanish phonological neighborhoods.

by

April Pufahl

Doctor of Philosophy

in

Cognitive Science

Stony Brook University

2014

If speech perception and production share lexical representations, then properties of these representations should have similar effects in both modalities. Likewise, if the system supporting lexical processing is fundamentally the same regardless of the language implemented, then a given property should produce the same effect across all languages. However, previous research suggests cross-modal and cross-linguistic differences for one property: phonological neighborhood density, a measure of the number of similar-sounding words in a language. These studies have relied on existing words, affording little control over the numerous, and possibly confounding, properties associated with each word. To avoid these potential pitfalls, I created 48 Spanglish nonwords that could be plausible words in both English and Spanish. Critically, nonwords were designed to systematically vary in the way they connect with existing words in each language, i.e., their new neighbors. The phonological neighborhoods these nonwords joined differed not only in overall density but also in the proportion of neighbors that are also neighbors of one another (i.e., the clustering coefficient), and the number of neighbors that share the same onset or offset phoneme. In this way, the present research was designed to assess the role of clustering and position-specific neighbors in driving the phonological neighborhood density effects observed in English and Spanish. Results suggest that, while clustering and overall neighborhood density slows speech processing, offset neighbors facilitate perception and onset neighbors facilitate production. This, coupled with the fact that English neighbors tend to share onsets while Spanish neighbors tend to share offsets, can explain the previously observed cross-modal and cross-linguistic differences.

Keywords: phonological neighborhood density; clustering coefficient; cohort neighbors; rhyme neighbors; spoken word recognition; speech production

Dedication Page

I dedicate this work to my parents, Dennis and Joy Pufahl, for their endless love, support, and encouragement.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ANOVA = analysis of variance

CC = clustering coefficient

CLEARPOND = Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities

CPSAMPA = a modified version of the Extended Speech Assessment Methods Phonetic Alphabet, or X-SAMPA

CVC word = consonant vowel consonant word

df = degrees of freedom

F = F statistic

ges = generalized eta-squared

GLMM = generalized linear mixed model

IPA = international phonetic alphabet

LMM = liner mixed model

MSE = mean squared error

ns = not significant

p = p value

ppts = participants

RT = reaction time

Acknowledgments

There are a number of people without whom this dissertation might not have been written, and to whom I am greatly indebted.

“Mil gracias” to my adviser, Arty Samuel, not only for his guidance on this project, but for his excellent mentorship. He’s been a rock in a sea of ups and downs. His door is always open and he’s always been available, even when he’s an ocean away. I hope I’ve picked up some of his steady determination in breaking tasks down and checking them off one by one.

I’d also like to thank my dissertation committee, Susan Brennan, Tony Freitas, and Marie Huffman, for their enthusiasm and support for this research project. Their diverse expertise has greatly improved this project. Thanks also to Richard Gerrig for seeking me out at a time when I was really struggling. Our meetings gave me the encouragement and structure I needed to set aside my self-doubts and push through work that needed to be done.

Several people have offered technical and experimental support on this project. Thanks to Blair Armstrong for assisting with the calculation of the clustering coefficients. Thanks to Donna Kat for programming support. “Eskerrik asko” to Saioa Larraza for helping with stimuli selection. “Gracias” to Roberto Abreu for being the voice of this experiment. “Eskerrik asko” to the BCBL lab for collecting the Spanish data. Thanks to the research assistants in the Katsam lab, Sydney Attard, Brielle Gonzalez, and Karyn Morra, for helping me collect the pilot and English data.

Finally, I would not be the woman I am today without the love and support of my parents, Dennis and Joy Pufahl. They have always been my number one cheerleaders, and the first to celebrate my accomplishments with pride. Thanks also to Sky Rolnick for asking what

my projects are about and sincerely wanting to know, and for making me celebrate little milestones along the way.

Chapter 1: Introduction

The goal of the present research is to clarify the apparent cross-modal and cross-linguistic differences previously observed for one property of lexical representations: *phonological neighborhood density*. “Phonological” refers to phonemes, the smallest units of sound that affect meaning. Phonemes are defined contrastively. For example, /d/ and /t/ are phonemes in English because “bad” and “bat” have different meanings. Furthermore, “bad” and “bat” are neighbors of one another, since they differ by only one phoneme. Words in dense phonological neighborhoods are those for which many similar sounding words, or neighbors, exist in the language. In this introduction, I first briefly review the reasons why cross-modal and cross-linguistic differences are unexpected for properties of lexical representations. I then review the previous research on phonological neighborhood density and initial work on two factors (clustering and position-specific neighbors) proposed to explain these effects. Finally, I describe how the present research was designed to assess the role of these factors in driving the phonological neighborhood density effects observed in English and Spanish.

In everyday conversation, people act as both listeners and speakers. In order for successful communication to occur, the systems supporting speech perception and speech production must be integrated at some level of representation. Theorists disagree on the degree to which perception and production overlap. Some propose full integration (Allport, 1984; Coleman, 1998; Fowler, 1986; Liberman & Mattingly, 1985; MacKay, 1987; Martin & Saffran, 2002). Others propose separate but connected sublexical representations (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Hickok & Poeppel, 2004; Schwartz, Basirat, Ménard, & Sato, 2012). However, most assume shared lexical (lemma or word-level) representations (but see Levelt et al., 1999; Monsell, 1987; Roelofs, 2003). If perception and production share the same lexical representations, it is logical that properties of those representations should have similar effects in both modalities. Indeed, some properties of words do have the same effect in both perception and production. For example, high frequency words are easier to perceive and produce than low frequency words (perception: e.g., Oldfield & Wingfield, 1965; Solomon & Postman, 1952; production: e.g., Dell, 1990). The same is true for concrete words relative to abstract words (perception: e.g., Kroll & Merves, 1986; production: e.g., Strain, Patterson, & Seidenberg, 1995).

If perception and production share lexical representations, then not only should properties of those representations have similar effects in both modalities, but it is parsimonious to assume that this pattern should hold cross-linguistically. The goal of psycholinguistic research is to delineate the systems supporting language, in this case speech perception and speech production. One of the assumptions embedded in this goal is that, regardless of the specific language implemented, the system is fundamentally the same. Listeners and speakers of different languages are all using the same hardware; therefore, it is likely that they also use the same kinds of representations to solve the problem of perceiving and producing speech. Properties of these representations may differ across languages (for example, some languages include lexical tones while others do not), but it is assumed that a given property will produce the same effect across all languages. Indeed, the high frequency advantage in English speech perception has also been observed in German (Brysbaert et al., 2011) and Spanish (Carreiras, Alvarez, & De Vega, 1993) and the high frequency advantage in English speech production has also been observed in Chinese (Caramazza, Costa, Miozzo, & Bi, 2001) and Dutch (Jescheniak & Levelt, 1994), among others. Similarly, the advantage for concrete words relative to abstract words has also been observed in Dutch (de Groot, 1989). If perception and production share lexical representations, and if those representations are similar across languages, it is logical that properties of those representations should have similar effects not only in both modalities but across all spoken languages as well. However, previous research suggests cross-modal and cross-linguistic differences for one property: phonological neighborhood density.

Phonological Neighborhood Density

Phonological neighborhood density refers to the number of similar-sounding words in a language. These similar-sounding words are called “neighbors” and are traditionally defined as words that differ by changing, adding, or subtracting one phoneme (Luce et al., 1990 as cited in Dell & Gordon, 2003). For example, there are 50 words that sound similar to “cat” in English (e.g., “hat”, “catch”, and “at”), but only 25 that sound similar to “dog” (Marian, Bartolotti, Chabal, & Shook, 2012). Therefore, “cat” belongs to a denser phonological neighborhood than “dog”.

The majority of researchers who have investigated the effects of phonological neighborhood density on speech perception and production have used English speaking subjects, and thus have investigated these effects within an English phonological neighborhood (for a review, see Dell & Gordon, 2003). Those studies demonstrate that, in perception, words with many neighbors are harder to recognize than words with few neighbors (Goldinger, Luce, & Pisoni, 1989; Luce & Pisoni, 1998; Vitevitch & Luce, 1998, 1999). For example, native speakers of English were more accurate typing spoken words from sparse versus dense neighborhoods (Goldinger et al., 1989; Luce & Pisoni, 1998). In fact, prominent theories of spoken word perception all propose competition between neighbors (TRACE: McClelland & Elman, 1986; Shortlist: Norris, 1994; Neighborhood Activation Model: Luce & Pisoni, 1998; PARSYN: Luce, Goldinger, Auer, & Vitevitch, 2000; Distributed Cohort Model: Gaskell & Marslen-Wilson, 1997, 1999, 2002).

However, in production, neighbors appear to behave differently, facilitating rather than inhibiting performance. English words with many neighbors are easier to produce than words with few neighbors. Vitevitch (1997) compared two corpora of spoken words, one of substitution errors and another without errors. After controlling for length and grammatical class, he found that the words in the error corpus were more likely to be words in low density neighborhoods. Subsequent experiments (Harley & Bown, 1998; Vitevitch & Sommers, 2003) used an experimental tip-of-the-tongue paradigm and found that words in low density neighborhoods were more likely to elicit a tip-of-the-tongue state. Using spoonerisms and tongue-twisters, Vitevitch (2002) found that not only were speech errors more likely for words in low density neighborhoods, but pictures of those words were also named more slowly than words in high density neighborhoods. Collectively, the studies done in English suggest that high density makes lexical access more difficult in perception, but easier in production.

Although it would be parsimonious to assume that a lexical property such as phonological neighborhood density would show similar cross-linguistic effects, there is evidence that the effects found in English do not generalize to other languages, such as Spanish. As just noted, for English, there appears to be a high density disadvantage in speech perception and a high density advantage in speech production. The opposite pattern appears to be true for Spanish: a high density advantage in speech perception and a high density disadvantage in speech production. On an auditory lexical decision task, native Spanish-speakers were faster and more accurate when the target word came from a dense, versus sparse, neighborhood (Vitevitch & Rodríguez, 2005). The research on Spanish production has been mixed, partly due to the number of correlated and possibly confounded variables. Vitevitch and Stamer (2006) were the first to report a high density disadvantage for picture naming in Spanish. They later replicated this finding (Vitevitch & Stamer, 2009) by analyzing a subset of the data from the International Picture-Naming Project (E. Bates et al., 2003). Baus, Costa, and Carreiras (2008) were likewise able to replicate this effect. However, they found the opposite pattern, a high density advantage like that found in English, when using a new set of line drawings. Post hoc analyses ruled out the influence of three potential confounds that differentiated the two sets of stimuli (initial syllable structure, distribution of word onsets, and word length in phonemes) leaving no explanation for the contradictory findings.

Recently, Sadat, Martin, Costa, and Alario (2013) argued for a methodological resolution, favoring linear mixed effects modeling over the *t*-tests and ANOVAs used previously. They argued that this avoided the loss of information that accompanies averaging across participants, items, and conditions. Furthermore, they were able to include a number of correlated variables (e.g., name agreement, age of acquisition, lexical frequency, neighborhood frequency) in the model in an attempt to control possible confounds. The results of their picture naming experiment confirmed the presence of a high density disadvantage: Native Spanish-speakers showed longer naming latencies for words in dense, versus sparse, neighborhoods. Using the subset of stimuli that overlapped with those used by Baus et al. (2008), they found that while a t-test showed a high density advantage, the more sophisticated modeling analysis showed a trend towards a high density disadvantage. They argued that these kinds of analyses are required to detect the small effect of phonological neighborhood density and to control for the influence of confounding variables. Given this, in the current research, I used a similar kind of modeling in order to provide the most sensitive measurement. Critically, I also used a new procedure relying on nonwords that avoids many of the inherent confounds, and cross-linguistic differences in stimuli.

The observed cross-linguistic differences between the effects of phonological neighborhood density on speech perception and production in English and Spanish are a puzzle. Languages of course differ in their inventory of phonemes, and their lexical items. However, the architecture of the system, and how lexical and phonemic codes communicate, should not be language-dependent: There is no *a priori* reason to assume that the speech perception and production systems should be fundamentally different as a function of the language spoken. Presumably, the system uses the same kind of representations to solve the problem of perceiving and producing speech regardless of the language. This assumption has led to speculation regarding moderating variables behind the observed differences across languages. One line of reasoning suggests that the *degree of clustering* within the neighborhood is critical, while another suggests that *position-specific neighbors* could be driving effects. In what follows, I will consider the logic behind, and current evidence for, each theory.

Clustering

The first theory suggests that it is not the number of neighbors, but how clustered the neighborhood is, that matters. A quantitative measurement of neighborhood clustering is the clustering coefficient, defined as the proportion of a word's neighbors that are also neighbors with one another (Vitevitch, 2008).

$$clustering\ coefficient = \frac{\#\ connections\ between\ neighbors}{\#\ possible\ connections}$$

Scores range from 0 to 1 with higher scores indicating greater clustering. For example, as shown in *Figure 1* below, the words "badge" and "log" both have 13 phonological neighbors. This means there can be 78 possible connections between these neighbors. This is calculated by taking the number of neighbors times the number of neighbors minus one, all divided by two.

$$\#\, possible\ connections = \frac{\#\, neighbors\ \times (\#\, neighbors - 1)}{2}$$

$$\frac{13\ \times (13 - 1)}{2} = 78$$

Of these 78 possible connections, there are 45 between the neighbors of "badge" and 21 between the neighbors of "log". Therefore, "badge" has a high clustering coefficient of 0.58 while "log" has a low clustering coefficient of 0.27.

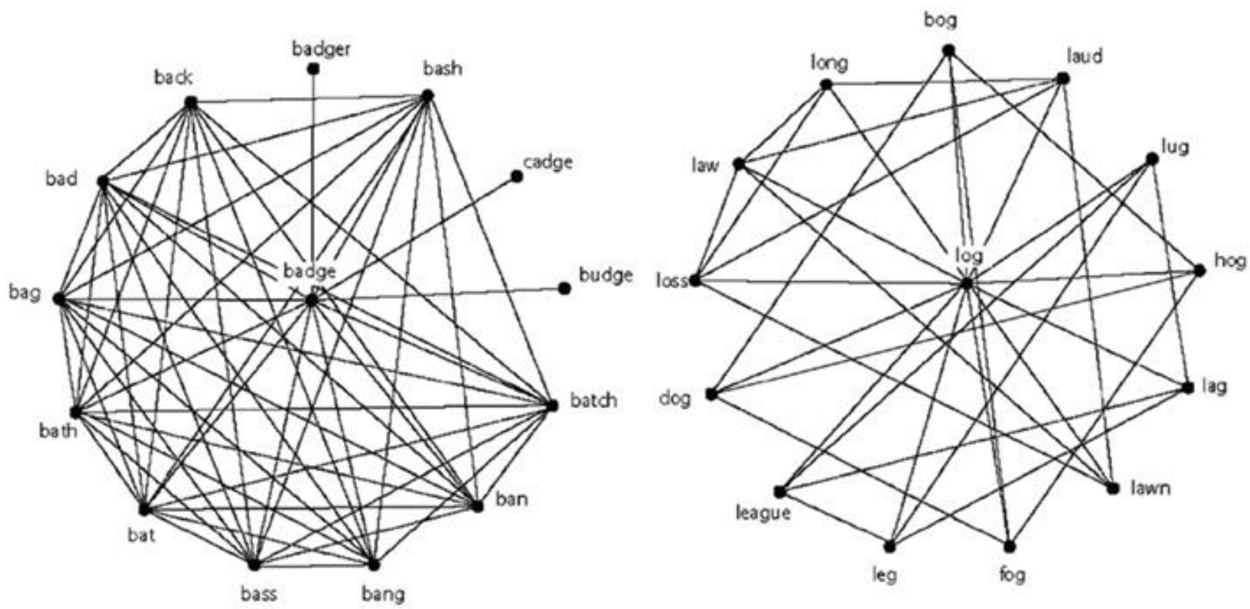


Figure 1: Illustration of the degree of clustering in the neighborhoods to which the words "badge" and "log" belong. Originally published as Figure 2 in Chan & Vitevitch, 2009.

If activation spreads throughout the phonological network along these links between neighbors, then a word with a low clustering coefficient, like "log" will receive more activation than a word with a high clustering coefficient, like "badge". This suggests that words like "log" will be easier to process than words like “badge”. Empirical evidence suggests that this is indeed the case for perception (Chan & Vitevitch, 2009) and production (Chan & Vitevitch, 2010) in English. Chan and Vitevitch (2009) identified a set of CVC words that had either a high (average .35) or low (average .25) clustering coefficient. They found that participants were more accurate typing CVC words heard in white noise when the words had a low (72% accuracy) versus high (58% accuracy) clustering coefficient. Furthermore, participants were faster and more accurate on a lexical decision task for words with a low (888ms, 93.3% accuracy) versus high (900ms, 91.6% accuracy) clustering coefficient. By comparing substitution errors to words matched in length and grammatical class (borrowed from the methodology from Vitevitch, 1997 described above), Chan and Vitevitch (2010) found that the errors were from words with a higher clustering coefficient (average .32) than the matched words (average .29). Using a picture-naming task, they observed that participants were faster to name pictures of words with a low (739ms) versus high (772ms) clustering coefficient.

These studies provide evidence that clustering is an important variable in English. However, the relationship between clustering and the effects of phonological neighborhood density reported in the literature is unclear. While these two variables are correlated, the relationship is weak and largely uninformative. Using the CLEARPOND database (Marian et al., 2012), I selected all words in English (n = 10449) and Spanish (n = 8948) with three or more phonemes and three or more phonological neighbors [note: this is the lexical space of interest in the present research] and computed clustering coefficients. For both languages, there was a small but significant negative correlation between phonological neighborhood density and the clustering coefficient, English, $r(10447) = -0.036$, $p < 0.01$ and Spanish, $r(8946) = -0.127$, $p < .01$. This means that as phonological neighborhood density increases, there tends to be a slight decrease in the degree of clustering between neighbors. Given this weak relationship, it is difficult to know how clustering might have influenced previously reported effects of phonological neighborhood density, since words from dense neighborhoods are not necessarily from clustered neighborhoods and vice versa.

However, it is possible that clustering could play a role in the apparent cross-linguistic differences in the effects of phonological neighborhood density in English and Spanish (Arbesman, Strogatz, & Vitevitch, 2010a). Overall, the languages differ in degree of clustering. The average clustering coefficient for English (0.28) is higher than that for Spanish (0.19), reflecting the higher clustering of words in English (Arbesman, Strogatz, & Vitevitch, 2010b). This difference is less drastic when focusing on words with three or more phonemes and three or more phonological neighbors (i.e., the ones for which clustering should matter the most). Using a subset with only these words (the same sample I selected from the CLEARPOND database described above), I computed an average clustering coefficient of 0.30 for English and 0.28 for Spanish. By this metric, no cross-linguistic differences should be observed. Furthermore, with the lack of a strong relationship between density and clustering, it is premature to suggest how clustering could have affected previous results without knowing more about the degree of clustering for the stimuli used in previous studies of phonological neighborhood density. Since clustering was an uncontrolled variable, further investigation is needed to tease apart the many correlated factors that might be contributing to the divergent results.

Position-Specific Neighbors

The second theory suggests that, since speech unfolds in time, the portion of the neighbor that overlaps is critical. Previous research on position-specific neighbors has focused on comparing cohort vs rhyme neighbors. *Cohort neighbors* include all words in the language that share the same onset phoneme(s). The number of overlapping phonemes required varies by study. *Rhyme neighbors* are those that share the same offset phoneme(s), generally the final vowel and any following consonants. Note that these definitions are broader than that of "neighbor", which requires all but one phoneme to overlap. Since the purpose of the present research was to test whether position-specific effects were driving overall density effects, I did not compare cohort vs rhyme neighbors. Instead, I focused on the subset of neighbors that share the onset phoneme (or onset neighbors) and the subset of neighbors that share the offset phoneme (or offset neighbors). However, I expected effects to be similar for both definitions. Note also that the definition of onset used in the present research differs from typical usage in linguistics, in which "onset" refers to all consonants that appear before a vowel in the initial syllable. For present purposes, "onset" and "offset" simply refer to the initial and final phonemes.

As noted above, prominent theories of spoken word perception all propose competition between neighbors. However, some emphasize competition between cohort neighbors (Shortlist: Norris, 1994; Distributed Cohort Model: Gaskell & Marslen-Wilson, 1997, 1999, 2002) while others include competition between all neighbors (TRACE: McClelland & Elman, 1986; Neighborhood Activation Model: Luce & Pisoni, 1998). These models therefore predict a high density disadvantage in speech perception, particularly for cohort and onset neighbors and perhaps less so for rhyme and offset neighbors.

There is evidence that the cohort/rhyme status of a word's neighbors matters for speech perception in English. Dumay et al. (2012) taught participants new neighbors for English hermit words (i.e., words without neighbors) like "carousel". Critically, these new neighbors were either cohort or rhyme neighbors, overlapping in all but the final/initial phoneme, such as the new words “carousem” and “barousel”. Participants were then asked to detect pauses which had been inserted into the hermit words, a task which previous research (Mattys & Clark, 2002; Mattys, Pleydell-Pearce, Melhorn, & Whitecross, 2005) has shown becomes harder as lexical competition increases. Results indicated that participants were about 30ms slower to detect pauses in hermit words that had gained a new cohort neighbor than those that had gained a rhyme neighbor or did not gain a neighbor. This suggests that cohort competitors engage in lexical competition to a greater degree than rhyme neighbors, which may compete weakly or not at all.

This evidence that cohort neighbors are driving lexical competition is one possible explanation for the inconsistent effects of neighborhood density in speech perception in English versus Spanish. As with clustering, overall, English and Spanish differ in the kinds of neighbors most words have. English neighbors are more likely to be cohort competitors and Spanish neighbors are more likely to be rhyme neighbors (Dumay, Damian, & Bowers, in preparation). If words with mostly cohort neighbors are difficult to perceive, then this is consistent with the high density disadvantage in cohort-biased English and high density advantage in rhyme-biased Spanish.

The effect of the cohort/rhyme status of a word’s neighbors on speech production is less clear. Dell, Burger, and Svec (1997) argued, when it comes to speech production, serial-order matters. The speech production “system must activate the present, deactivate the past, and prepare to activate the future” (Dell et al., 1997, pp 123). If this is true, then the portion of a neighbor that overlaps with the intended word is critical. Overlapping segments are thought to facilitate processing while non-overlapping segments interfere or compete. As a target word unfolds in time, the degree of facilitation and competition from any given neighbor changes. This makes it difficult to predict overall word-level effects. For example, cohort neighbors should help start production quickly, but this facilitation changes to competition as the overlapping segments pass. The remaining non-overlapping segments provide interference that must be deactivated to produce the intended segment. Similarly, rhyme neighbors will make it difficult for the system to select the initial phoneme(s) of the target word. Later, as the target word unfolds, feedback between the overlapping final segments will support processing of those segments. While these predictions are informative at a fine-grained time scale, it is unclear what kind of facilitation/competition they predict for the production of words in their entirety.

Not only are there unclear predictions for the effect of the cohort/rhyme status of a word's neighbors on speech production, but there are also unclear results from the literature. As described above, Dumay et al. (2012) taught participants new neighbors for English hermit words. In addition to the perceptual task (pause detection) previously described, they also included a production task (picture naming). Participants were about 25ms quicker to name pictures of hermit words that gained a rhyme neighbor than those that gained a cohort neighbor or did not gain a neighbor. This suggests that rhyme neighbors have an overall facilitatory effect on speech production. It is unclear if cohort neighbors have no effect or if their combined facilitation and competition cancels out overall.

While Dumay et al. (2012) found a facilitatory effect of rhyme neighbors on speech production, Bien, Baayen, and Levelt (2011) found an inhibitory effect. Native Dutch speakers were slower to produce verbs that had many rhyme neighbors, defined as the number of neighbors in which the first phoneme is the one exchanged. This is in contrast to evidence that it is easier to produce a series of rhyme neighbors like "pick" and "tick" than cohort neighbors like "pick" and "pin" (Sevald & Dell, 1994). However, as Bien, Baayen, and Levelt (2011) note, this is due to the sequential nature of the production task they used, e.g., the competition from the neighbor “tick” when producing the initial word “pick” quickly turns to facilitation when “tick” is the next item produced. In their position-response association task, participants learned to associate two neighboring words with cues that would appear on the left or right side of the screen. This allowed experimenters to cue the production of these words without having to rely on picture naming, as most production studies have done. In one block, participants were cued to say each word 20 times, randomly mixed with distractor trials in which a single digit would appear at center screen to be named.

In addition to the inhibitory effect of rhyme neighbors, Bien et al. (2011) also found a facilitatory effect of cohort neighbors, defined as those that share the first two phonemes: the more cohort neighbors, the shorter the naming latency. It is worth noting that both studies (Bien et al., 2011; Dumay et al., 2012) used naming latency as the primary measure. Presumably, this measure reflects how quickly the production system can get started, and would therefore be most likely to benefit from shared onset, or be hurt by its absence.

Given these unclear predictions and results, it is difficult to predict how cohort/rhyme status position-specific neighbors might moderate the effects of phonological neighborhood density more broadly. But, if as Bien et al. (2011) observed, there is an advantage for producing words with cohort neighbors and a disadvantage for producing words with rhyme neighbors, then this is consistent with the high density production advantage in cohort-biased English and high density disadvantage in rhyme-biased Spanish.

The Present Research

The present research was designed to assess the role of clustering and position-specific neighbors in driving the phonological neighborhood density effects reported in the literature, particularly as they relate to the apparent cross-linguistic differences observed in English and Spanish. Previous studies have focused on existing words in one language or another and therefore have had no control over the numerous, and possibly confounding, properties associated with each pre-existing word. These uncontrolled variables might be responsible for the inconsistent results across studies. To avoid these potential pitfalls, in the present study I focused on a set of nonwords that will be added to each language. I created 48 Spanglish nonwords that could be plausible words in both English and Spanish. Critically, nonwords were designed to systematically vary in the way they connect with existing words in each language, i.e., their new neighbors. The phonological neighborhoods these nonwords joined differed not only in overall density but also in the proportion of neighbors that are also neighbors of one another (i.e., the clustering coefficient), and the number of neighbors that share the same onset or offset phoneme. By adding these nonwords to the lexicons of native speakers of English and Spanish, I tested how these neighborhood characteristics affect perception and production. The appeal of this design is that any observed cross-linguistic differences in performance can be attributed to the fact that these identical items enter a lexical space with systematically different characteristics in each language. By using exactly the same lexical items in each language, the myriad confounds that come with existing words are side-stepped. In addition, each new word was inserted into a neighborhood chosen to have a high degree of clustering or a low one, with predominantly onset or offset neighbors; this was done in both English and Spanish.

The advantages of this approach come with a cost: Stimulus selection required extensive corpus analysis, a suitable task for testing perception needed to be established, and a sizable number of native speakers of both English and Spanish need to be tested on both perception and production. For the production task, manual scoring of each of the many productions was done.

In Chapter 2, I describe the development and testing of a perceptual task that is suitable for the training conditions. The majority of previous research relied on lexical decision, but this task is inappropriate for the newly-created lexical items used in this project. [Fortunately, the standard production task, picture naming, is appropriate for this project since participants will learn the Spanglish nonwords by associating each with an unusual picture.] While some researchers have successfully used a pause detection task, the difference in logic behind their studies and the present study was enough to warrant caution. For example, Dumay et al. (2012) added a new neighbor to a hermit word, and used pause detection to look for the impact of this new neighbor on perception of the original hermit word. In the present study, I instead focused on neighborhoods with two or more neighbors and tested the newly learned Spanglish nonword directly, rather than looking for its impact on its neighbors. As described below, I attempted to replicate effects found with three different sets of stimuli using a pause detection task and a newly created "same/different" task, and found the latter to be a more valid measure.

In Chapter 3, I describe the process of searching large databases of existing words in English and Spanish and their associated neighborhood characteristics, in order to identify possible Spanglish nonword stimuli. The critical neighborhood characteristics in this project (clustering coefficient and onset/offset densities) were not available from previous databases, but have now been calculated.

In Chapters 4 and 5, I describe the methodology and results of the set of experiments that was conducted with native English speakers (in Stony Brook, NY) and native Spanish speakers (in San Sebastian, Spain). Finally, in Chapter 6, I discuss the implications of these findings.

Chapter 2: Task Norming

Lexical decision has been the standard perceptual task used in previous studies of phonological neighborhood density. However, since the present research involved teaching participants a set of Spanglish nonwords, a word/nonword decision task seems inappropriate. A lesser-used option is a pause detection task. This task is presumed to be sensitive to the amount of lexical activation present from competing words (Mattys & Clark, 2002; Mattys et al., 2005). Previous researchers (Dumay & Gaskell, 2007; Gaskell & Dumay, 2003) have successfully used pause detection to detect increased reaction times for hermit words that recently gained a new neighbor as part of the experimental procedure. Furthermore, as described in Chapter 1, this increase in reaction times appears to be specifically driven by adding cohort (but not rhyme) neighbors (Dumay et al., 2012). Critically, the items used in the pause detection task were the original hermit words in English, i.e., pre-existing words that gained a novel neighboring word. In the present research, the items of interest are the newly added Spanglish nonwords, i.e., novel words that connect with two or more pre-existing neighbors. Therefore, it was unclear if pause detection would be sensitive to the characteristics of the neighborhood a new word enters.

As such, I developed a "same/different" task designed to tap into fine-grained phonetic processing. In this task, participants heard two items and judged whether both instances contained exactly the same sounds (i.e., sequence of consonants and vowels). The first item was the target word spoken at a normal speech rate. To make the task challenging, the second item was compressed by 50% using Praat (Boersma & Weenink, 2014), meaning that it was spoken twice as fast as normal. For the "same" trials, participants heard another token of that same target word. For the "different" trials, participants heard a nonword variation of the target word that was created by changing one of the consonant phonemes (usually a change in voicing or place of articulation). For example, the word “experiment” could become “experiNent”. Participants respond by pressing one of two buttons labeled “same” and “different”. The “same” trials are of primary interest, as they should tap into the degree of processing difficulty for the test words. If a given property of lexical representations makes recognition more difficult, then participants should be slower and/or less accurate to respond that the second word in a pair is indeed the same as the first.

Materials and Procedure

In order to test whether a pause detection task and/or the new “same/different” task would be sensitive to the kinds of neighborhood characteristics I wished to manipulate, I adapted stimuli from three previous experiments and attempted to replicate their reported effects, using both tasks.

The first set of stimuli varied in uniqueness point, i.e., the point at which there are no other words in the lexicon that share the initial phonemes, a concept central to cohort models of speech perception (e.g., Marslen-Wilson & Welsh, 1978). Mattys and Clark (2002) used stimuli that varied in uniqueness point in order to manipulate lexical activity at word offset in the hopes that their new pause detection task would be sensitive to this activity. They compared 20 words like “pretzel” that have an early uniqueness point with 20 words like “settle” that have a late uniqueness point. Pauses were inserted after the target word (e.g., “pretzel”) which was embedded at the beginning of a 5-syllable string (e.g., “pret-zel-[150ms pause]-poe-faye-gol”). They found that detecting pauses after late (but not early) unique words was slower compared to nonword matched trials. This supported the hypothesis that the longer period of lexical competition for late versus early unique words was responsible for the slowed response. One goal of the present task norming was to see if this result could be replicated using pauses embedded in the target words themselves following a modification used by Dumay and Gaskell (Dumay & Gaskell, 2007; Gaskell & Dumay, 2003). A second goal was to see if there is a corresponding cost for late unique words on the “same-different” task.

The second set of stimuli were CVC words that varied in clustering coefficient (Chan & Vitevitch, 2009). Thirty-seven words had a low clustering coefficient (average .25) and 38 words had a high clustering coefficient (average .35). They found that participants were more accurate typing words heard in white noise (+24-dB signal-to-noise ratio) when the words had a low (72% accuracy) versus high (58% accuracy) clustering coefficient. Similarly, participants were faster and more accurate on a lexical decision task for words with a low (888ms, 93.3% accuracy) versus high (900ms, 91.6% accuracy) clustering coefficient. As with the uniqueness point manipulation, the goal of the present task norming was to see whether the pause detection and “same-different” tasks showed a facilitatory effect of low versus high clustering.

The third set of stimuli were 72 high and 68 low frequency words used by Balota and Chumbley (1985). Note that four low frequency words from the original set were excluded by accident. In the original study, participants were faster producing high versus low frequency words. Frequency effects are the gold standard of lexical processing: They tend to be large and robust, so these items were included in the hope that they would provide a clear test of the sensitivity of the pause detection and “same-different” tasks.

For the pause detection task, 200ms of silence was inserted as close to the end of each target word as possible, being careful not to create a new stop consonant. Pauses were intended to sound artificial, and not like a natural slowed speech rate. The stimuli for all three manipulations (uniqueness point, clustering, and frequency) were combined to produce one large set of stimuli. Participants heard all 254 stimuli twice (both with and without a pause), randomized in two lists. For the first list, half of the stimuli were presented with a pause and half without a pause. Stimuli were then switched from pause to no pause and vice versa for the second list. Participants responded by pressing one of two buttons, labeled “pause” and “no pause”. Accuracy and reaction time were measured, with reaction times measured from pause onset for both pause present and pause absent trials (for the absent trials, the measurement was from the point where the pause had been inserted in the corresponding pause-present stimulus).

The “same/different” task was as described above. Again, participants heard all 254 target words twice (one “same” trial and one “different”), randomized in two lists, as was done for the pause detection task. On each trial, participants first heard the word spoken at a normal speech rate. For “same” trials, this was followed by a different production of the word (no phonemes changed) at 50% compression. For “different” trials, this was followed by the word at 50% compression with one consonant phoneme changed. Participants responded by pressing one of two buttons, labeled “same” and “different”. Accuracy and reaction time were measured, with reaction times measured from word offset on all trials.

Participants

Sixty-six native English speakers from the Stony Brook University subject pool completed both tasks in a one-hour session, with the order of the tasks counterbalanced. Data for two participants on the “same/different” task was lost due to program error. Additionally, one participant performed near chance (50%) on the pause detection task with responses indicating that they were simply alternating their responses rather than performing the task as instructed. As such, their data was dropped from analysis. The final dataset included 65 participants for the pause detection task and 63 for “same different” task.

Analysis

Reaction times for incorrect responses and responses <100ms or >2000ms were excluded from analysis. I used by participants and by items ANOVAs (analysis of variance) to analyze both the reaction time and percent accuracy data, which has long been the standard practice in the field. I repeated these analyses using LMMs (linear mixed models) to analyze the reaction time data and GLMMs (generalized linear mixed models) to analyze the percent accuracy data. Recently, researchers (Baayen, Davidson, & Bates, 2008; Baayen, 2008; Jaeger, 2008; Janssen, 2012) have argued that this type of modeling is most appropriate when dealing with the language-as-fixed-effect fallacy (Clark, 1973). Overall, results were largely consistent regardless of the type of analysis used. In general, when there was a significant effect, the typical pattern was a significant by participants ANOVA, a non-significant by items ANOVA, and LMM/GLMMs that approached or reached significance. This pattern reflects the greater variance due to items than due to participants. Note that mixed models are considered more conservative (reducing the chance of a Type I error), since they model random effects by participants and by items simultaneously and do not average responses across either random effect.

For the mixed models, random intercepts were included for participants and items. Adding random slopes for any of the fixed effects did not increase the model fits or change their interpretation. Similarly, transforming the reaction time data, either by base e log, base 10 log, or square root, had no effect on results. Therefore, the dependent variable was untransformed reaction times in milliseconds. I used log likelihood ratio tests to assess the fit of the various models compared to a null model including only random intercepts (Baayen, 2008).

All analyses were conducted in R (R Core Team, 2014). The package ez (Lawrence, 2013) was used to run the ANOVAs. The formatted ANOVA tables in this document were created with package afex (Singmann & Bolker, 2014) and exported with package xtable (Dahl, 2014). The package gplots (Warnes et al., 2014) was used to generate the barplots. The packages lme4 (D. Bates, Maechler M, Bolker, & S, 2014) and lmerTest (Kuznetsova, Bruun Brockhoff, & Haubo Bojesen Christensen, 2014) were used to fit and compare the LMMs and GLMMs. The package texreg (Leifeld, 2013) was used to create and export the formatted tables of the statistical models.

For the pause detection task, I predicted that any significant differences would be seen in the reaction times, but not necessarily percent accuracy (overall, participants performed near ceiling on this task). In particular, I predicted that the pause absent trials may be the most valid, since the presence of the pause might interrupt natural processing.

Similarly, for the "same/different" task, the focus was on the reaction times for the "same" trials (since the presence of a different phoneme might interrupt natural processing). I predicted that participants should be faster to respond "same" when a word has characteristics that make it easier to process (early uniqueness point, low clustering coefficient, or high frequency). Performance on "different" trials might show a similar pattern, but that was unclear. In general, results for percent accuracy should either parallel those for reaction times or show no effect.

In what follows, I will present the predictions and results for each set of stimuli and for each task.

Uniqueness Point - Predictions and Results

For the uniqueness point stimuli, on both tasks participants should be faster when responding to early versus late unique words since early unique words have less lexical competition as the word unfolds. This prediction was confirmed, with the pause absent trials and the same trials driving the overall effects. On the pause absent trials, there was a 28ms advantage for early (vs late) unique words. Similarly, on the “same” trials, there was a 22ms advantage for early (vs late) unique words.

For the accuracy data, participants were less likely to notice the different phoneme in early unique words (a 15% drop in percent accuracy). This could be a type of perceptual restoration effect in which the unambiguous activation of the early unique word is robust enough to overcome the small phonetic mismatch on a “different” trial. Finally, participants were more accurate responding to early unique words during pause present (a 2% advantage) and “same” trials (a 2% advantage). This result is consistent with the reaction time advantage on these trials, suggesting a processing advantage for both speed and accuracy.

See *Figure 2* and *Figure 3* and *Table 1*, *Table 2*, *Table 3*, and *Table 4* below for details.

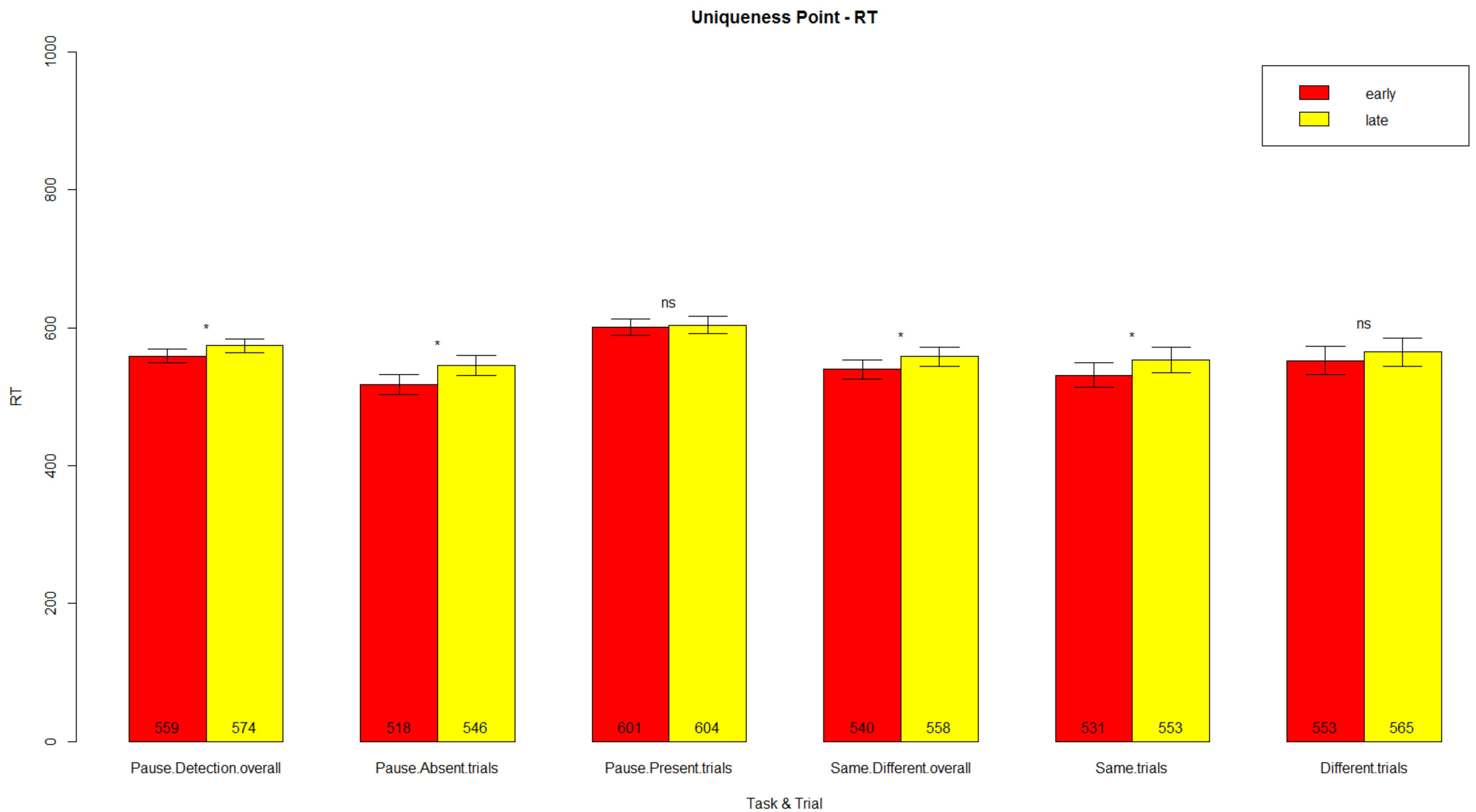


Figure 2: Mean reaction times on the pause detection and "same different" tasks for the set of stimuli that varied in uniqueness point. Standard errors are represented in the figure by the error bars attached to each column.

Table 1. ANOVA Results for the Effect of Uniqueness Point on Reaction Times					
Task & Trial	*df*	MSE	*F*	ges	*p*
Pause Detection overall (by ppts)	1, 64	1113.34	6.76 *	.005	.01
Pause Detection overall (by items)	1, 38	743.91	3.25 +	.08	.08
Pause Absent trials (by ppts)	1, 64	3578.29	7.01 *	.01	.01
Pause Absent trials (by items)	1, 38	2316.60	3.09 +	.08	.09
Pause Present trials (by ppts)	1, 64	2705.08	0.18	.0003	.68
Pause Present trials (by items)	1, 38	1628.67	0.08	.002	.78
Same/Different overall (by ppts)	1, 62	2361.77	3.58 +	.003	.06
Same/Different overall (by items)	1, 38	2105.11	2.88 +	.07	.10
Same trials (by ppts)	1, 62	3680.46	2.76	.003	.10
Same trials (by items)	1, 38	2777.08	1.97	.05	.17
Different trials (by ppts)	1, 62	6764.21	0.50	.0009	.48
Different trials (by items)	1, 37	7770.30	0.35	.009	.56

***p < 0.001, **p < 0.01, *p < 0.05, +p < 0.1 based on unrounded *p* values

Note: The table reports the degrees of freedom (*df*), mean squared error (MSE), *F* value, generalized eta squared (*ges*), and *p* value for each ANOVA.

Table 2. LMM Results for the Effect of Uniqueness Point on Reaction Times

	Pause Detection overall	Pause Absent trials	Pause Present trials	Same Different overall	Same trials	Different trials
(Intercept)	559.39 (14.00)***	520.37 (17.72)***	601.64 (14.78)***	544.18 (21.05)***	536.34 (22.24)***	578.47 (27.14)***
unique (late)	16.20 (8.52)$^{+}$	27.79 (15.14)$^{+}$	4.85 (12.04)	20.01 (13.79)	20.03 (16.12)	9.46 (24.64)
AIC	65953.10	33853.96	31884.57	55616.70	32427.51	23211.22
BIC	65985.51	33882.97	31913.45	55648.10	32456.20	23238.26
Log Likelihood	-32971.55	-16921.98	-15937.29	-27803.35	-16208.75	-11600.61
Num. obs.	4827	2448	2379	3941	2293	1648
Num. groups: Participants	65	65	65	63	63	63
Num. groups: Items	40	40	40	40	40	39
Variance: Participants (Intercept)	10391.99	12917.15	9512.72	21797.35	23007.50	25160.51
Variance: Items (Intercept)	327.19	1396.14	847.81	1122.75	1291.84	3766.30
Variance: Residual	48069.41	54805.37	35635.11	74286.06	74553.68	67799.57

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{+}p < 0.1$

Note: For the fixed effects, the table displays estimates (β) followed by standard errors in parentheses.

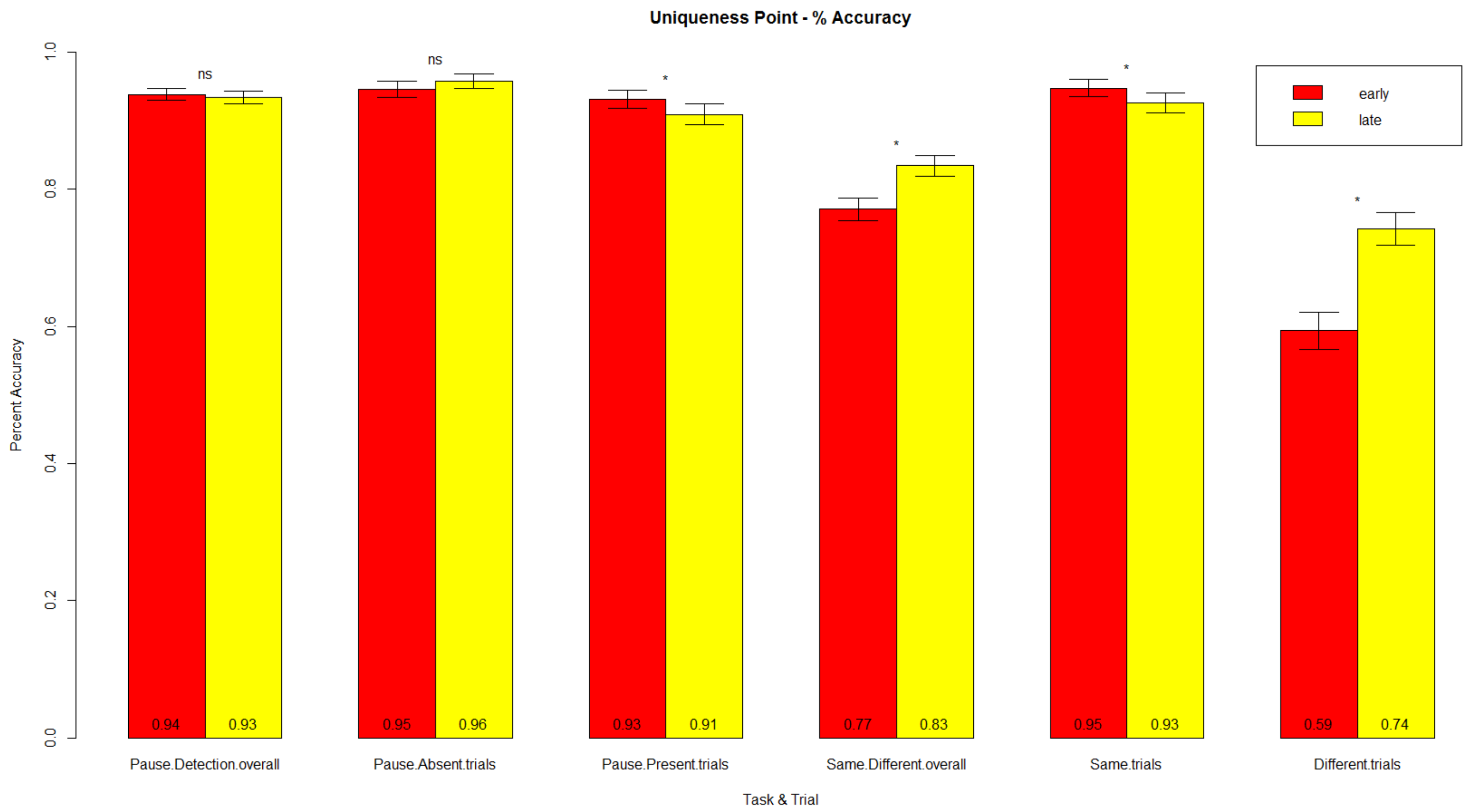


Figure 3. Mean percent accuracy on the pause detection and “same different” tasks for the set of stimuli that varied in uniqueness point. Standard errors are represented in the figure by the error bars attached to each column.

Table 3. ANOVA Results for the Effect of Uniqueness Point on Percent Accuracy

Task & Trial	*df*	MSE	*F*	ges	*p*
Pause Detection overall (by ppts)	1, 64	0.00	0.72	.002	.40
Pause Detection overall (by items)	1, 38	0.00	0.18	.005	.67
Pause Absent trials (by ppts)	1, 64	0.00	2.55	.006	.12
Pause Absent trials (by items)	1, 38	0.00	1.35	.03	.25
Pause Present trials (by ppts)	1, 64	0.00	5.92 *	.02	.02
Pause Present trials (by items)	1, 38	0.00	1.14	.03	.29
Same/Different overall (by ppts)	1, 62	0.00	36.12 ***	.16	<.0001
Same/Different overall (by items)	1, 38	0.03	1.55	.04	.22
Same trials (by ppts)	1, 62	0.00	4.21 *	.02	.04
Same trials (by items)	1, 38	0.00	2.08	.05	.16
Different trials (by ppts)	1, 62	0.01	75.56 ***	.25	<.0001
Different trials (by items)	1, 38	0.11	2.06	.05	.16

***p < 0.001, **p < 0.01, *p < 0.05, +p < 0.1 based on unrounded *p* values

Note: The table reports the degrees of freedom (*df*), mean squared error (MSE), *F* value, generalized eta squared (*ges*), and *p* value for each ANOVA.

Table 4. GLMM Results for the Effect of Uniqueness Point on Percent Accuracy

	Pause Detection overall	Pause Absent trials	Pause Present trials	Same Different overall	Same trials	Different trials
(Intercept)	2.98 (0.15)***	3.38 (0.23)***	3.01 (0.22)***	1.61 (0.27)***	3.27 (0.23)***	0.75 (0.56)
unique (late)	-0.08 (0.18)	0.28 (0.25)	-0.37 (0.27)	0.38 (0.38)	-0.35 (0.26)	1.12 (0.78)
AIC	2378.00	922.47	1369.42	4287.19	1130.99	1899.59
BIC	2404.22	945.92	1392.86	4313.25	1154.29	1922.88
Log Likelihood	-1185.00	-457.24	-680.71	-2139.60	-561.50	-945.80
Num. obs.	5187	2594	2593	4993	2501	2492
Num. groups: Participants	65	65	65	63	63	63
Num. groups: Items	40	40	40	40	40	40
Variance: Participants (Intercept)	0.39	0.91	0.42	0.10	0.56	0.77
Variance: Items (Intercept)	0.19	0.22	0.48	1.31	0.37	5.81
Variance: Residual	1.00	1.00	1.00	1.00	1.00	1.00

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{+}p < 0.1$

Note: For the fixed effects, the table displays estimates (β) followed by standard errors in parentheses.

Clustering – Predictions and Results

For the clustering stimuli, on both tasks participants should be faster when responding to words with low versus high clustering coefficients, since perceiving words from highly clustered neighborhoods should require more lexical activity than perceiving words from less clustered neighborhoods.

However, reaction time results were in the unexpected direction on both tasks. Participants responded faster to words with high clustering coefficients than those with low clustering coefficients. On the pause detection task, this effect was driven by the pause absent trials (a 33ms high clustering advantage). On the “same/different” task, this effect was driven by the “different” trials (a 33ms high clustering advantage). However, recall that the primary predictions were for “same” trials because predictions for “different” trials were difficult to make.

Results for percent accuracy were in the expected direction, indicating more accurate performance for words with low clustering coefficients. Participants were less likely to notice the different phoneme in words with high clustering coefficients (a 13% drop in percent accuracy), and this effect drove an overall disadvantage for words with high clustering coefficients on the “same/different” task.

See *Figure 4* and *Figure 5* as well as *Table 5*, *Table 6*, *Table 7*, and *Table 8* below for details.

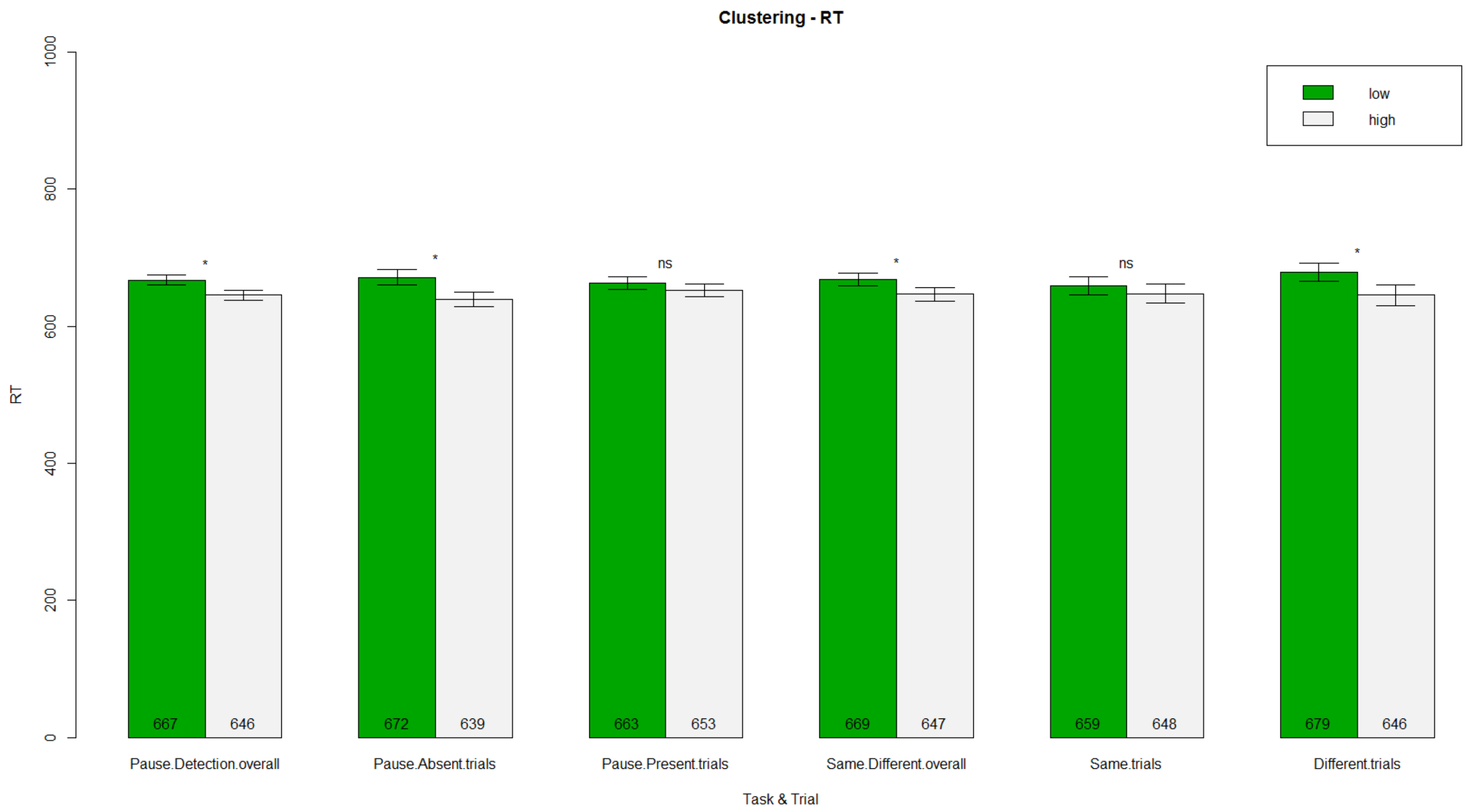


Figure 4. Mean reaction times on the pause detection and “same different” tasks for the set of stimuli that varied in clustering. Standard errors are represented in the figure by the error bars attached to each column.

Table 5. ANOVA Results for the Effect of Clustering on Reaction Times

Task & Trial	*df*	MSE	*F*	ges	*p*
Pause Detection overall (by ppts)	1, 64	747.00	19.42 ***	.01	<.0001
Pause Detection overall (by items)	1, 73	4089.08	2.11	.03	.15
Pause Absent trials (by ppts)	1, 64	1521.53	22.92 ***	.02	<.0001
Pause Absent trials (by items)	1, 73	8374.11	2.27	.03	.14
Pause Present trials (by ppts)	1, 64	1282.06	2.04	.002	.16
Pause Present trials (by items)	1, 73	3198.21	0.66	.009	.42
Same/Different overall (by ppts)	1, 62	1838.96	6.26 *	.004	.02
Same/Different overall (by items)	1, 73	3462.63	2.01	.03	.16
Same trials (by ppts)	1, 62	3637.44	0.25	.0003	.62
Same trials (by items)	1, 73	5487.33	0.75	.01	.39
Different trials (by ppts)	1, 62	2469.25	14.22 ***	.01	.0004
Different trials (by items)	1, 73	8199.75	1.12	.02	.29

***p < 0.001, **p < 0.01, *p < 0.05, +p < 0.1 based on unrounded *p* values

Note: The table reports the degrees of freedom (*df*), mean squared error (MSE), *F* value, generalized eta squared (*ges*), and *p* value for each ANOVA.

Table 6. LMM Results for the Effect of Clustering on Reaction Times

	Pause Detection overall	Pause Absent trials	Pause Present trials	Same Different overall	Same trials	Different trials
(Intercept)	668.88 (15.99)***	673.53 (19.27)***	664.27 (15.72)***	673.37 (21.01)***	667.03 (23.11)***	690.64 (22.74)***
clustering (high)	-21.43 (14.51)	-32.38 (20.83)	-9.78 (12.81)	-17.65 (13.28)	-8.31 (16.55)	-27.02 (17.86)
AIC	125162.29	63760.58	61196.59	107662.77	57475.49	50156.76
BIC	125197.90	63792.78	61228.69	107697.48	57507.04	50187.67
Log Likelihood	-62576.14	-31875.29	-30593.30	-53826.38	-28732.74	-25073.38
Num. obs.	9155	4621	4534	7649	4070	3579
Num. groups: Participants	75	75	75	75	75	75
Num. groups: Items	65	65	65	63	63	63
Variance: Participants (Intercept)	3547.64	7263.39	2416.76	2573.77	3755.04	4339.44
Variance: Items (Intercept)	9686.36	9847.20	10664.96	22239.70	24869.85	22827.39
Variance: Residual	48533.04	53359.23	39705.68	72515.14	73800.50	65639.88

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{+}p < 0.1$

Note: For the fixed effects, the table displays estimates (β) followed by standard errors in parentheses.

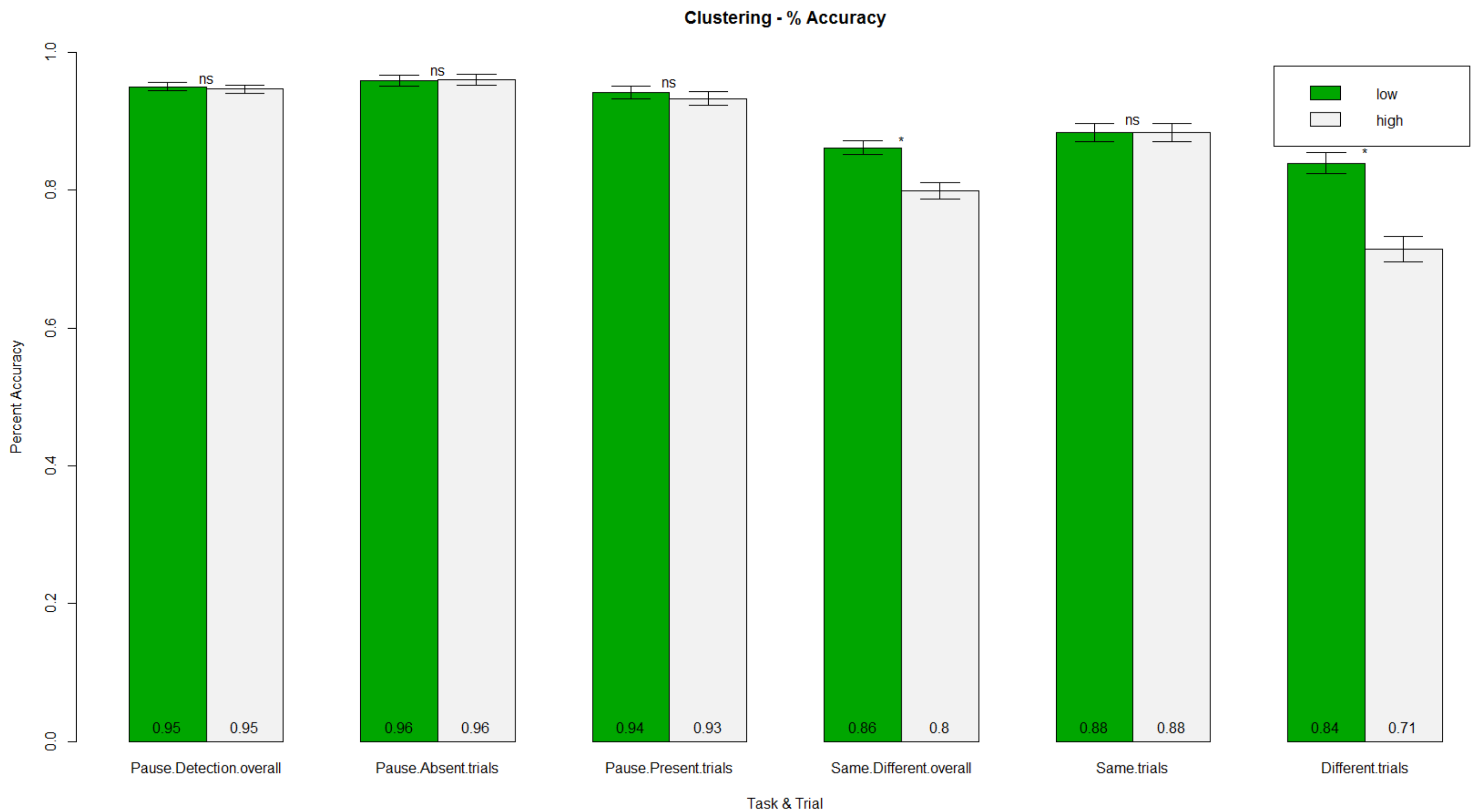


Figure 5. Mean percent accuracy on the pause detection and "same different" tasks for the set of stimuli that varied in clustering. Standard errors are represented in the figure by the error bars attached to each column.

Table 7. ANOVA Results for the Effect of Clustering on Percent Accuracy

Task & Trial	*df*	MSE	*F*	ges	*p*
Pause Detection overall (by ppts)	1, 64	0.00	0.54	.002	.47
Pause Detection overall (by items)	1, 73	0.00	0.44	.006	.51
Pause Absent trials (by ppts)	1, 64	0.00	0.07	.0003	.79
Pause Absent trials (by items)	1, 73	0.00	0.04	.0006	.84
Pause Present trials (by ppts)	1, 64	0.00	2.45	.006	.12
Pause Present trials (by items)	1, 73	0.00	0.90	.01	.35
Same/Different overall (by ppts)	1, 62	0.00	165.35 ***	.19	<.0001
Same/Different overall (by items)	1, 73	0.02	3.84 +	.05	.05
Same trials (by ppts)	1, 62	0.00	0.00	<.0001	.96
Same trials (by items)	1, 73	0.01	0.00	<.0001	.99
Different trials (by ppts)	1, 62	0.00	204.98 ***	.24	<.0001
Different trials (by items)	1, 73	0.06	4.91 *	.06	.03

***p < 0.001, **p < 0.01, *p < 0.05, +p < 0.1 based on unrounded *p* values

Table 8. GLMM Results for the Effect of Clustering on Percent Accuracy

	Pause Detection overall	Pause Absent trials	Pause Present trials	Same Different overall	Same trials	Different trials
(Intercept)	3.22 (0.13)***	3.66 (0.20)***	3.17 (0.16)***	2.16 (0.17)***	2.55 (0.19)***	2.32 (0.30)***
clustering (high)	-0.08 (0.11)	0.07 (0.21)	-0.19 (0.16)	-0.51 (0.22)*	-0.12 (0.22)	-0.92 (0.39)*
AIC	3781.70	1538.95	2173.47	7480.77	3015.59	3439.30
BIC	3810.41	1564.89	2199.42	7509.34	3041.38	3465.08
Log Likelihood	-1886.85	-765.47	-1082.74	-3736.39	-1503.80	-1715.65
Num. obs.	9696	4844	4852	9323	4666	4657
Num. groups: Participants	75	75	75	75	75	75
Num. groups: Items	65	65	65	63	63	63
Variance: Participants (Intercept)	0.08	0.33	0.22	0.85	0.73	2.71
Variance: Items (Intercept)	0.51	0.81	0.65	0.17	0.62	0.76
Variance: Residual	1.00	1.00	1.00	1.00	1.00	1.00

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{+}p < 0.1$

Note: For the fixed effects, the table displays estimates (β) followed by standard errors in parentheses.

Frequency – Predictions and Results

For the frequency stimuli, on both tasks participants should be faster when responding to frequent versus infrequent words. Generally, frequent words are perceived faster and more accurately than infrequent words. However, frequent words tend to come from dense neighborhoods – so it is possible there could be more competition. The original experiment from which these stimuli were borrowed did not include measures beyond word frequency and orthographic length, so there are many potential confounds.

On the pause detection task, there was a small 1% advantage in percent accuracy for high frequency words on the pause absent trials. There were no other significant effects.

On the "same/different" task, results were in the expected direction. On the "same" trials, participants were faster (a 12ms advantage) and more accurate (a 3% advantage) when responding to the high versus low frequency words. Furthermore, participants were less likely to notice the different phoneme changed in the high frequency words (a 5% drop in accuracy). This difference could be a type of perceptual restoration effect in which the strong lexical activation of the high frequency word overcomes the small phonetic mismatch on a "different" trial.

See *Figure 6* and *Figure 7* and *Table 9, Table 10, Table 11,* and *Table 12* below for details.

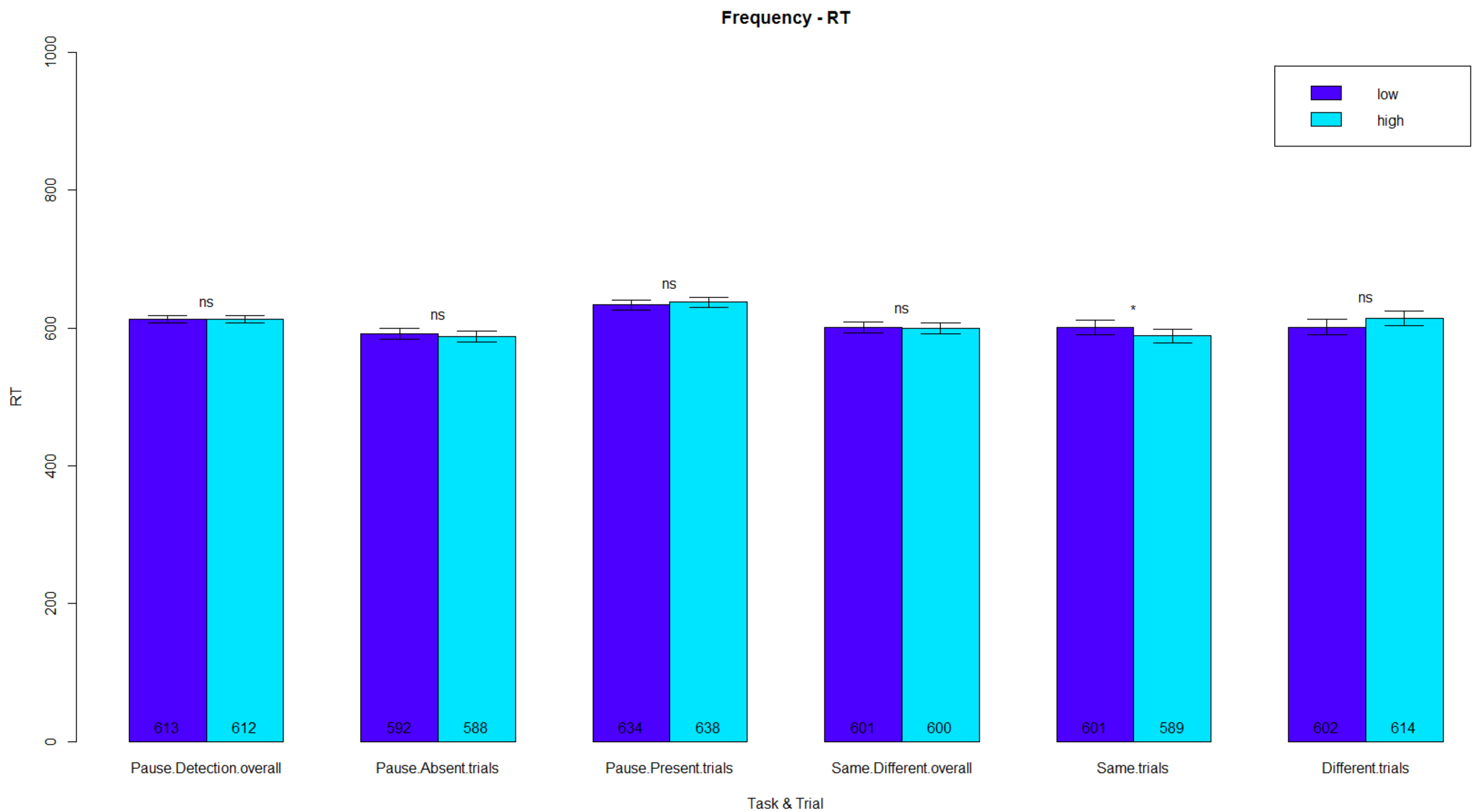


Figure 6. Mean reaction times on the pause detection and "same different" tasks for the set of stimuli that varied in word frequency. Standard errors are represented in the figure by the error bars attached to each column.

Table 9. ANOVA Results for the Effect of Frequency on Reaction Times

Task & Trial	*df*	MSE	*F*	ges	*p*
Pause Detection overall (by ppts)	1, 64	488.44	0.15	<.0001	.70
Pause Detection overall (by items)	1, 138	3804.99	0.01	<.0001	.94
Pause Absent trials (by ppts)	1, 64	975.46	0.74	.0005	.39
Pause Absent trials (by items)	1, 138	9526.00	0.07	.0005	.79
Pause Present trials (by ppts)	1, 64	706.23	0.21	.0001	.64
Pause Present trials (by items)	1, 138	2344.69	0.19	.001	.66
Same/Different overall (by ppts)	1, 62	650.88	0.72	.0002	.40
Same/Different overall (by items)	1, 138	5732.04	0.06	.0005	.80
Same trials (by ppts)	1, 62	1225.05	5.57 *	.002	.02
Same trials (by items)	1, 138	7133.88	0.39	.003	.54
Different trials (by ppts)	1, 62	1511.93	1.71	.0009	.20
Different trials (by items)	1, 138	12395.47	1.52	.01	.22

***p < 0.001, **p < 0.01, *p < 0.05, +p < 0.1 based on unrounded *p* values

Table 10. LMM Results for the Effect of Frequency on Reaction Times

	Pause Detection overall	Pause Absent trials	Pause Present trials	Same Different overall	Same trials	Different trials
(Intercept)	613.75 (14.71)***	593.54 (17.55)***	635.61 (14.11)***	607.85 (21.09)***	612.40 (23.35)***	623.93 (22.19)***
frequency (high)	-1.80 (10.44)	-4.46 (16.40)	1.85 (8.25)	-0.40 (12.72)	-13.37 (14.00)	20.86 (17.75)
AIC	232228.91	118102.15	113537.05	191905.46	109167.67	82488.95
BIC	232267.60	118137.45	113572.20	191943.06	109202.43	82522.36
Log Likelihood	-116109.45	-59046.07	-56763.52	-95947.73	-54578.83	-41239.47
Num. obs.	16957	8598	8359	13625	7721	5904
Num. groups: Participants	140	140	140	140	140	140
Num. groups: Items	65	65	65	63	63	63
Variance: Participants (Intercept)	3398.96	8582.78	1637.97	4870.53	5448.50	8900.49
Variance: Items (Intercept)	10423.40	11017.17	10667.32	22775.51	27972.08	20874.40
Variance: Residual	50167.86	50611.52	44232.58	73600.13	76114.57	63090.32

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{+}p < 0.1$

Note: For the fixed effects, the table displays estimates (β) followed by standard errors in parentheses.

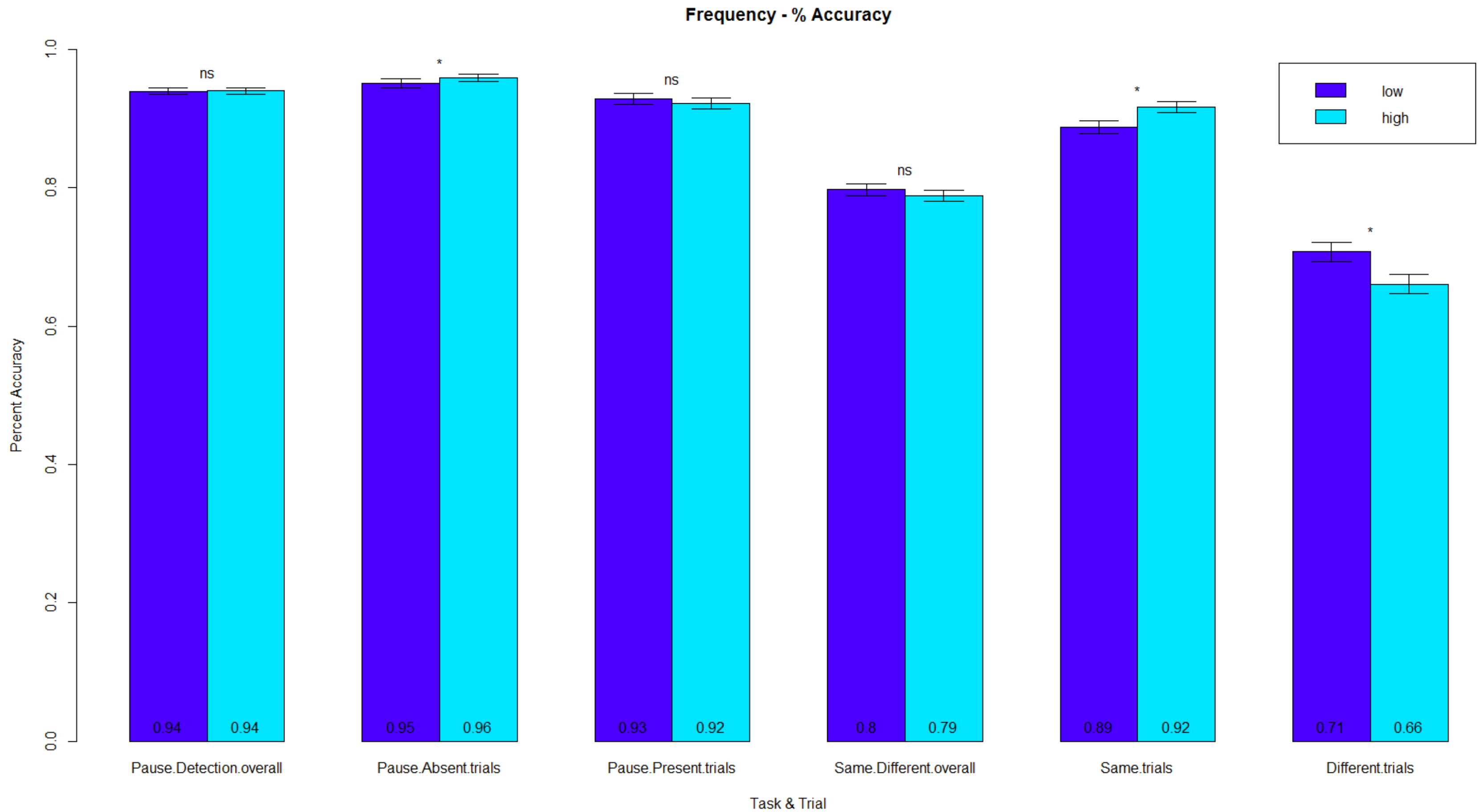


Figure 7. Mean percent accuracy on the pause detection and "same different" tasks for the set of stimuli that varied in word frequency. Standard errors are represented in the figure by the error bars attached to each column.

Table 11. ANOVA Results for the Effect of Frequency on Percent Accuracy

Task & Trial	*df*	MSE	*F*	ges	*p*
Pause Detection overall (by ppts)	1, 64	0.00	0.03	<.0001	.87
Pause Detection overall (by items)	1, 138	0.00	0.01	<.0001	.92
Pause Absent trials (by ppts)	1, 64	0.00	2.45	.004	.12
Pause Absent trials (by items)	1, 138	0.00	2.76 +	.02	.10
Pause Present trials (by ppts)	1, 64	0.00	1.15	.004	.29
Pause Present trials (by items)	1, 138	0.00	0.52	.004	.47
Same/Different overall (by ppts)	1, 62	0.00	3.54 +	.006	.06
Same/Different overall (by items)	1, 138	0.02	0.12	.0009	.73
Same trials (by ppts)	1, 62	0.00	21.87 ***	.04	<.0001
Same trials (by items)	1, 138	0.01	2.42	.02	.12
Different trials (by ppts)	1, 62	0.00	32.65 ***	.04	<.0001
Different trials (by items)	1, 138	0.09	0.85	.006	.36

***p < 0.001, **p < 0.01, *p < 0.05, +p < 0.1 based on unrounded *p* values

Table 12. GLMM Results for the Effect of Frequency on Percent Accuracy

	Pause Detection overall	Pause Absent trials	Pause Present trials	Same Different overall	Same trials	Different trials
(Intercept)	3.03 (0.11)***	3.45 (0.16)***	2.90 (0.13)***	1.64 (0.14)***	2.59 (0.17)***	1.40 (0.27)***
frequency (high)	0.04 (0.10)	0.21 (0.12)+	-0.06 (0.14)	0.04 (0.19)	0.49 (0.20)*	-0.25 (0.35)
AIC	7806.08	3024.46	4555.14	15588.15	4772.07	7102.94
BIC	7837.30	3052.91	4583.59	15619.21	4800.35	7131.23
Log Likelihood	-3899.04	-1508.23	-2273.57	-7790.08	-2382.03	-3547.47
Num. obs.	18123	9056	9067	17396	8689	8707
Num. groups: Participants	140	140	140	140	140	140
Num. groups: Items	65	65	65	63	63	63
Variance: Participants (Intercept)	0.21	0.12	0.41	1.12	1.08	4.02
Variance: Items (Intercept)	0.49	0.99	0.48	0.10	0.55	0.77
Variance: Residual	1.00	1.00	1.00	1.00	1.00	1.00

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{+}p < 0.1$

Note: For the fixed effects, the table displays estimates (β) followed by standard errors in parentheses.

Conclusion

Table 13 presents a summary of the results of the task norming experiments. Overall, the "same/different" task proved to be a more valid measure of these lexical properties (i.e., uniqueness point, clustering, and frequency) than the pause detection task. This was especially true of the "same" trials, as expected. See the summary of task norming results in *Table 13*. Therefore, the "same/different" task was used to measure perceptual performance in the experiments to follow, with the focus on the "same" trials.

Table 13. Summary of Task Norming Results

	Pause Detection overall	Pause Absent trials	Pause Present trials	Same Different overall	Same trials	Different trials
Uniqueness Point - RT	early*	early*	ns	early*	early*	ns
Uniqueness Point - %	ns	ns	early*	late*	early*	late*
Clustering - RT	high*	high*	ns	high*	ns	high*
Clustering -%	ns	ns	ns	low*	ns	low*
Frequency -RT	ns	ns	ns	ns	high*	ns
Frequency -%	ns	high*	ns	ns	high*	low*

* $p < 0.1$ on one or more statistical tests

Note: The level listed is the one for which there was an advantage (i.e., a faster reaction time or higher percent accuracy.

Chapter 3: Creation of the Spanglish Nonword Stimuli

I created 48 Spanglish nonwords that could be plausible words in both English and Spanish. Critically, nonwords were designed to systematically vary in the way they connect with existing words in each language, i.e., their new neighbors. The phonological neighborhoods these nonwords joined systematically differed not only in overall density but also in the proportion of neighbors that are also neighbors of one another (i.e., the clustering coefficient), and the number of neighbors that share the same onset or offset phoneme. In this way, the present research was designed to assess the role of clustering and position-specific neighbors in driving the phonological neighborhood density effects observed in English and Spanish.

Generating Nonword Candidates

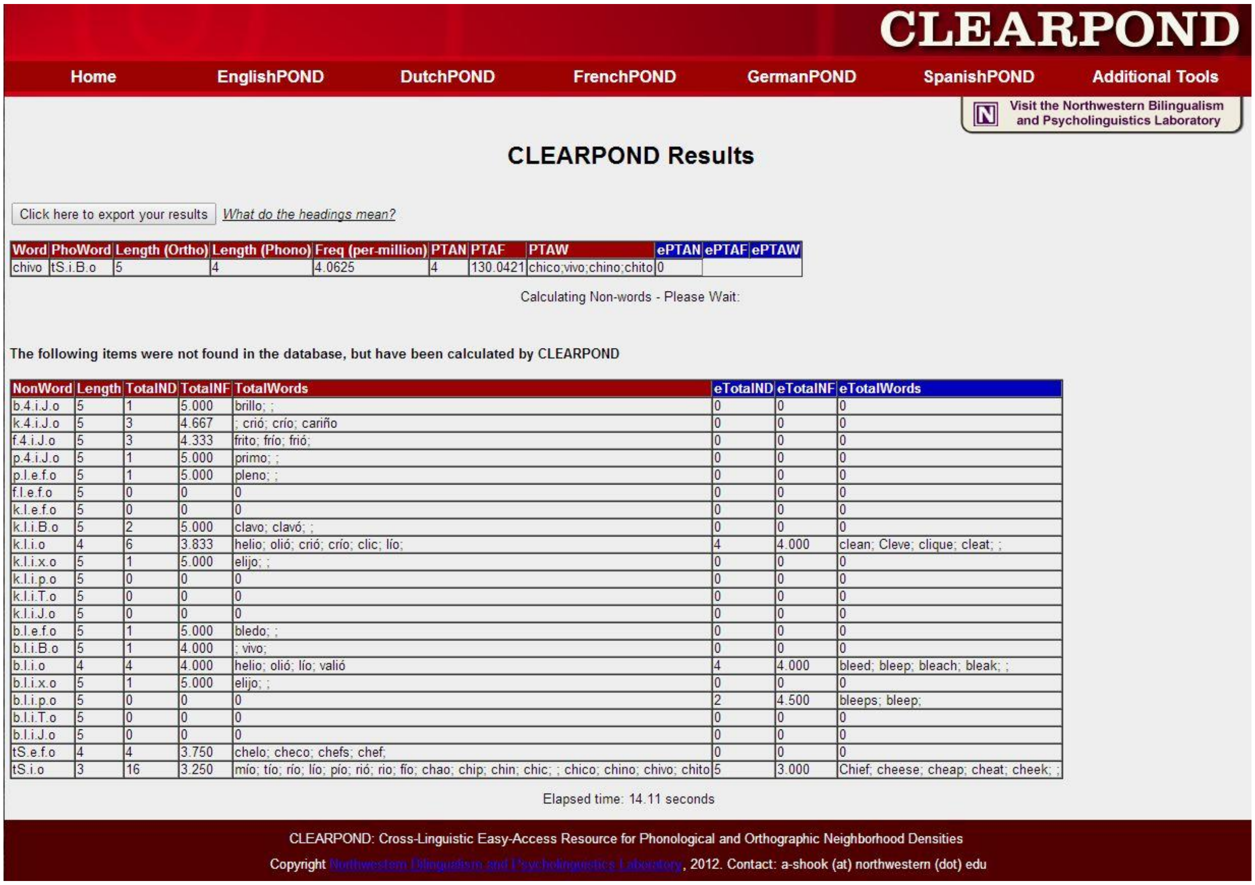
To create a list of possible nonword candidates, real words in both languages (from the CLEARPOND database, Marian, Bartolotti, Chabal, & Shook, 2012) were used as seed words in Wuggy, a program that generates orthographic nonwords (Keuleers & Brysbaert, 2010). Any duplicates were removed. These orthographic nonwords were then transcribed in the International Phonetic Alphabet (IPA) based on how they would be produced in Spanish, the more orthographically transparent and phonetically restrictive of the two languages. This was done using the Spanish Phonetic Transcription Converter available online at http://learn-foreign-language-phonetics.com/spanish-phonetic-transcription-converter.php?site_language=english . Again, any duplicates were removed. The IPA transcriptions were then converted to the format used by the CLEARPOND database to represent phonetic transcription (CPSAMPA, a modified version of the Extended Speech Assessment Methods Phonetic Alphabet, or X-SAMPA; Marian et al., 2012). See *Table 14* below for an example of these steps for the Spanish seed word "grifo" ("faucet" in English). The Spanglish nonword /blio/ was selected from the set of possible nonwords created from this procedure, with the selection process guided by the way that a possible nonword would fit into both the Spanish and the English lexicons.

Table 14. Example Nonwords Generated from the Spanish Seed Word “grifo”.

Orthographic Nonword Generated in Wuggy	**Spanish IPA Transcription**	**CPSAMPA Transcription**
bri-ño	bɾiɲo	b.4.i.J.o
cri-ño	kɾiɲo	k.4.i.J.o
fri-ño	fɾiɲo	f.4.i.J.o
pri-ño	pɾiɲo	p.4.i.J.o
ple-fo	plefo	p.l.e.f.o
fle-fo	flefo	f.l.e.f.o
cle-fo	klefo	k.l.e.f.o
cli-bo	kliβo	k.l.i.B.o
cli-ho	klio	k.l.i.o
cli-jo	klixo	k.l.i.x.o
cli-vo	kliβo	-
cli-po	klipo	k.l.i.p.o
cli-zo	kliθo	k.l.i.T.o
cli-ño	kliɲo	k.l.i.J.o
ble-fo	blefo	b.l.e.f.o
bli-bo	bliβo	b.l.i.B.o
bli-ho	blio	b.l.i.o
bli-jo	blixo	b.l.i.x.o
bli-vo	bliβo	-
bli-po	blipo	b.l.i.p.o
bli-zo	bliθo	b.l.i.T.o
bli-ño	bliɲo	b.l.i.J.o
che-fo	ʧefo	tS.e.f.o
chi-bo	ʧiβo	tS.i.B.o
chi-ho	ʧio	tS.i.o

Example of 25 orthographic nonwords generated in Wuggy for the Spanish seed word “grifo”, along with the Spanish IPA transcription and CPSAMPA transcription used to query the CLEARPOND database. The Spanglish nonword /blio/ was created from this set (highlighted in yellow). The items highlighted in red are duplicates (same phonetic transcription). The item highlighted in green is a real word in Spanish, “chivo” as noted in top line of the CLEARPOND output in *Figure 8* below.

The CPSAMPA transcriptions of each potential nonword were then queried in the CLEARPOND database available online at http://clearpond.northwestern.edu/spanishpond.html . Items that were real words in either language were removed. Additionally, the CLEARPOND database was used to gather measures of phonological neighborhood size and mean phonological neighborhood frequency as well as the list of phonological neighbors for each potential nonword, in each language. See *Figure 8* below for an example of the CLEARPOND output for the potential nonwords in Table 14, using the Spanish lexicon. A comparable output was produced for these nonwords using CLEARPOND's English lexicon.



CLEARPOND

Home | EnglishPOND | DutchPOND | FrenchPOND | GermanPOND | SpanishPOND | Additional Tools

Visit the Northwestern Bilingualism and Psycholinguistics Laboratory

CLEARPOND Results

Click here to export your results | *What do the headings mean?*

Word	PhoWord	Length (Ortho)	Length (Phono)	Freq (per-million)	PTAN	PTAF	PTAW	ePTAN	ePTAF	ePTAW
chivo	tS.i.B.o	5	4	4.0625	4	130.0421	chico;vivo;chino;chito	0		

Calculating Non-words - Please Wait:

The following items were not found in the database, but have been calculated by CLEARPOND

NonWord	Length	TotalND	TotalNF	TotalWords	eTotalND	eTotalNF	eTotalWords
b.4.i.J.o	5	1	5.000	brillo; ;	0	0	0
k.4.i.J.o	5	3	4.667	; crió; crío; cariño	0	0	0
f.4.i.J.o	5	3	4.333	frito; frío; frió;	0	0	0
p.4.i.J.o	5	1	5.000	primo; ;	0	0	0
p.l.e.f.o	5	1	5.000	pleno; ;	0	0	0
f.l.e.f.o	5	0	0	0	0	0	0
k.l.e.f.o	5	0	0	0	0	0	0
k.l.i.B.o	5	2	5.000	clavo; clavó; ;	0	0	0
k.l.i.o	4	6	3.833	helio; olió; crió; crío; clic; lío;	4	4.000	clean; Cleve; clique; cleat; ;
k.l.i.x.o	5	1	5.000	elijo; ;	0	0	0
k.l.i.p.o	5	0	0	0	0	0	0
k.l.i.T.o	5	0	0	0	0	0	0
k.l.i.J.o	5	0	0	0	0	0	0
b.l.e.f.o	5	1	5.000	bledo; ;	0	0	0
b.l.i.B.o	5	1	4.000	; vivo;	0	0	0
b.l.i.o	4	4	4.000	helio; olió; lío; valió	4	4.000	bleed; bleep; bleach; bleak; ;
b.l.i.x.o	5	1	5.000	elijo; ;	0	0	0
b.l.i.p.o	5	0	0	0	2	4.500	bleeps; bleep;
b.l.i.T.o	5	0	0	0	0	0	0
b.l.i.J.o	5	0	0	0	0	0	0
tS.e.f.o	4	4	3.750	chelo; checo; chefs; chef;	0	0	0
tS.i.o	3	16	3.250	mío; tío; río; lío; pío; rió; rio; fío; chao; chip; chin; chic; ; chico; chino; chivo; chito	5	3.000	Chief; cheese; cheap; cheat; cheek; ;

Elapsed time: 14.11 seconds

CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities

Copyright Northwestern Bilingualism and Psycholinguistics Laboratory, 2012. Contact: a-shook (at) northwestern (dot) edu

Figure 8. Example of the CLEARPOND output for the potential nonwords created from the Spanish seed word "grifo".

Clustering Measurement

Nonwords with fewer than two neighbors in either language were eliminated, since it is impossible to have clustering when there is only one neighbor. To compute clustering coefficients for each nonword in each language, I first calculated the total number of connections possible between neighbors of a given nonword, i.e., the number of neighbors times the number of neighbors minus one, all divided by two. I then counted the actual number of connections between neighbors. The clustering coefficient is the actual number of connections divided by the total number of connections possible (Vitevitch, 2008). Values range from 0 to 1, with higher values indicating greater clustering. For example, as shown in *Figure 9* below, the Spanglish nonword /blio/ has four neighbors in Spanish: helio, olió, lío, and valió. [Note that stress is not considered in these computations and is therefore a potential confounding factor. Also, note that “helio” is a neighbor because the “h” is silent, and “valio” is a neighbor because ‘b’ and ‘v’ are produced as the same phoneme in this position in Spanish.] Of the six possible connections between those words, three exist, giving a clustering coefficient of 3/6 or 0.5. That same nonword has four neighbors in English: bleed, bleep, bleach, and bleak. Of the six possible connections between those words, all six exist (i.e., each is a neighbor of the others) giving a clustering coefficient of 6/6 or 1. This pattern is typical of these languages, as English tends to be more clustered than Spanish.

Position-Specific Neighbors Measurement

Since the purpose of the present research was to test whether position-specific effects were driving overall density effects, I focused on the subset of neighbors that share the onset/offset phoneme (rather than cohort/rhyme neighbors typically used in the literature). Onset and offset densities for each nonword in each language were defined as the number of neighbors that share the same onset or offset phoneme. It is important to note that because all the possible set of candidates were neighboring words, by definition, they only varied by one added, subtracted, or substituted phoneme. Therefore, these neighboring words shared more than just the onset or offset phoneme. This definition of onset density fits with the traditional definition of cohort neighbors, for which the overlapping first phoneme is most critical. However, this is not the case for offset density. Its definition allows for deviations from the traditional definition of rhyme neighbors, for which the final vowel and any subsequent consonants are critical. For the majority of the Spanglish nonwords (39/48), the final phoneme is a vowel, and therefore this more traditional definition of rhymes is maintained overall. However, for the Spanglish nonwords that end with a consonant, the definition of offset density does include some neighbors that would not traditionally be considered rhymes.

An example of onset and offset neighborhood densities is illustrated in *Figure 9* below. The Spanglish nonword /blio/ shares its onset with one Spanish neighbor, "valió", since 'v' and 'b' are both realized as the same phoneme in this word initial position in Spanish. Furthermore, /blio/ shares its offset with all four Spanish neighbors. The pattern in English is the opposite: /blio/ shares its onset with all four English neighbors but does not share its offset with any neighbors. Again, the pattern illustrated is the one most typical of these languages: English neighbors tend to share onsets while Spanish neighbors tend to share offsets.

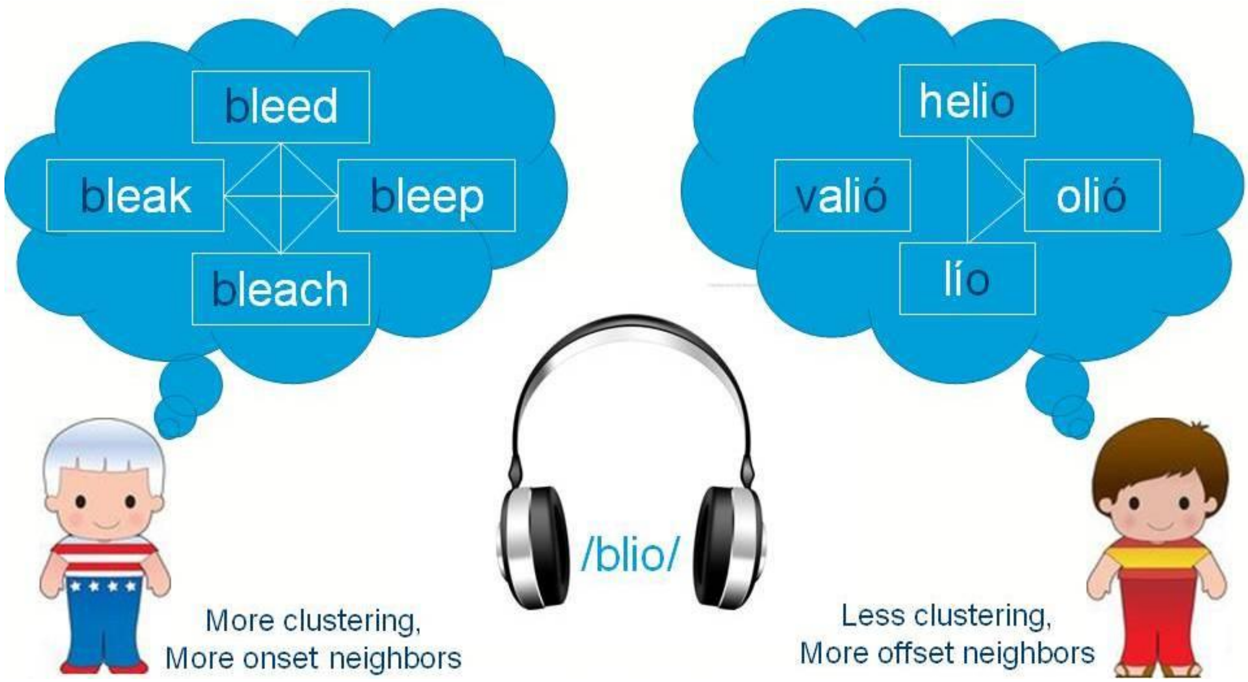


Figure 9. Illustration of the English and Spanish phonological neighborhoods for the Spanglish nonword /blio/.

Spanglish Nonword Stimuli

Forty-eight critical Spanglish nonwords were chosen from the list of possible nonword candidates. The goal was to have the items systematically differ in the ways they connect with existing words in each language. Half the items enter a space that is more clustered in English than Spanish, while for the other half the opposite is true. Within this split, half of the items show the same onset/offset pattern in both languages, with half of those items having more onset than offset neighbors and the other half having more offset than onset neighbors. Again, for the other half, the opposite is true, with items showing contrasting onset/offset patterns in each language, as with the example nonword /blio/ shown in *Figure 9*. See *Table 15* below for an illustration of the design.

Table 15. Illustration of the Experimental Design of the Spanglish Nonword Stimuli.

Example Nonword	**English Neighborhood**		**Spanish Neighborhood**	
/blio/	More clustering	More onset	Less clustering	More offset
/anu/	More clustering	More offset	Less clustering	More onset
/ʧike/	More clustering	More onset	Less clustering	More onset
/eni/	More clustering	More offset	Less clustering	More offset
/kjo/	Less clustering	More onset	More clustering	More offset
/iti/	Less clustering	More offset	More clustering	More onset
/suθ/	Less clustering	More onset	More clustering	More onset
/isi/	Less clustering	More offset	More clustering	More offset

Note that the clustering comparison is across languages (i.e., the nonword enters a neighborhood that is either more clustered in English or in Spanish). The onset/offset comparison is within a language (e.g., the nonword /blio/ has more cohort than rhyme neighbors in English and more rhyme than cohort neighbors in Spanish).

For a full list of the stimuli and associated measures, see the Appendix.

Chapter 4: Method

Participants

Forty-eight native speakers of English were recruited from the subject pool at Stony Brook University. Additionally, 48 native speakers of Spanish were recruited from the subject pool at the Basque Center for Cognition, Brain and Language. Since this population is typically bilingual (Basque and Spanish) or multilingual, in order to be eligible for this study, participants must have indicated that they acquired Spanish before the age of 3. For 43 of the 48 participants, Spanish was also the language they acquired first. Furthermore, participants must have indicated high levels of proficiency for Spanish and low levels for English. Participants rated themselves on a scale of 1 (low proficiency) to 10 (high proficiency). Average ratings were 9.4 for speaking, 9.3 for understanding, 9.2 for writing, and 9.2 for reading Spanish. Thirty-four of the 48 participants also indicated some proficiency in English. Their average age of acquisition was 7.9 with a range of 3-15. Their average ratings were 4.1 for speaking, 4.9 for understanding, 4.4 for writing, and 5.1 for reading English.

Additional participants were recruited, but their data had to be replaced for one of several reasons. Four participants were unable to attend the second session due to weather or other unanticipated issues. One participant did not follow task instructions. Five participants scored less than 70% accuracy on the picture association task (the new word training task – see below). Finally, data for six participants were lost due to program or experimenter error on one or more tasks.

Materials

Real words: In order to compare nonword performance in each language with that for real words, I selected 52 line drawings of common objects from the International Picture Naming Project (Szekely et al., 2004). I computed clustering coefficients and onset and offset densities for the picture labels in each language, using the same method described in Chapter 3 for the nonword stimuli. While this set of stimuli could not be as well controlled as the nonwords (that was, of course, the whole point of teaching new lexical items that could be controlled this way), I attempted to select a range of values on each measure that was comparable. I then created "different" items used in the "same/different" perceptual task by changing one consonant phoneme to create a nonword in that language. This secondary set of stimuli provides a within-subject comparison of naming for existing lexical items to naming for the newly-learned words.

Spanglish nonwords: For details on the creation of the 48 Spanglish nonword stimuli, see Chapter 3. For the "different" items used in the "same/different" perceptual task, a Spanglish nonword neighbor was chosen from the items remaining after the 48 critical stimuli were selected. These were neighbors by substitution of one phoneme. For example, for the critical item /blio/, the Spanglish nonword /blia/ was chosen to be the "different" item. The "different" items were required to be similar enough to the target word to make the judgment challenging. However, note that the analyses focus on the "same" trials, as the task norming has shown that these are the critical trials for tapping lexical effects.

Unusual objects: The critical Spanglish items were introduced into the listeners' lexicons by associating each item with the picture of an unusual object. This picture could later be used to cue production of the newly learned Spanglish nonword. Forty-eight color pictures of unusual objects were borrowed from previous research (Dumay et al., 2012; Leach & Samuel, 2007;

Samuel & Larazza, in preparation). The items were chosen to be ones for which there is no common label, such as those shown in *Figure 10* below.

Figure 10. Sample pictures of the unusual objects used in the picture association learning task and picture naming task.

Audio Recordings: In recording the critical Spanglish nonwords, the goal was to have a blend of an English and a Spanish accent because these items were taught to native speakers of each language. This was in addition to selecting only nonword stimuli with vowels and consonants that are common to the two languages.

A native Spanish speaker, who was highly proficient in English, recorded all English, Spanish, and Spanglish stimuli used in the experiment. The speaker was a male undergraduate student from the same college-age population as the participants. He was a linguistics major familiar with reading and producing words transcribed in the International Phonetic Alphabet. Although originally from the Dominican Republic (he moved to the US at age 14), he was coached to produce real Spanish words in the manner typically produced in Spain, e.g., producing the final phoneme in “cruz” as /θ/ rather than /s/. When necessary, he listened to and imitated the productions of a Basque-Spanish bilingual living in the region from which the Spanish population was recruited.

Although lexical stress was not used in computing the neighbors, during recording the speaker matched the stress pattern of the majority of the neighboring words in both languages, or produced a neutral stress pattern. Multiple takes of the Spanglish nonword stimuli were recorded, once in an “English mode” immediately after recording the English word stimuli and another in a “Spanish mode” immediately after recording the Spanish word stimuli. For each, the speaker tried to pretend the Spanglish nonwords were real words in that language. After listening carefully to both sets of recordings, it was clear the “Spanish mode” set best represented the target IPA transcriptions. Additionally, since Spanish is the more restrictive of the two languages, particularly in the vowel space, it seemed native Spanish speakers would be less likely to accept English-accented nonwords than native English speakers would be to accept Spanish-accented nonwords. Thus, the tokens used in the experiment were those recorded in “Spanish mode”.

All words were filtered to remove background noise and converted to a sampling rate of 44 kHz (Goldwave, version 5.70). A second token of each word and nonword, as well as all the “different” tokens, were compressed by 50% using Praat (Boersma & Weenink, 2014) for use in the “same/different” task.

For a full list of the stimuli and associated measures, see the Appendix.

Procedure

Participants were recruited for two one-hour sessions, scheduled on separate days, up to a week apart. Each participant was tested individually.

On the first day, they began by completing the “same/different” task on real words in their native language. Wearing headphones, participants listened to pairs of spoken words, played one after the other (ISI = 500ms). Their task was to judge whether both instances contained exactly the same sounds (i.e., sequence of consonants and vowels). The first word was spoken at a normal speech rate and the second word (or nonword) was spoken twice as fast to make the task challenging. Participants were told that when they heard the word a second time, sometimes one sound that is part of that word would be different, for example "experiment" could become something like "experiNent". Participants indicated whether the word contained all the same sounds, or contained a different sound, by pressing one of two response buttons. They were asked to respond quickly and accurately. The program advanced to the next trial (ITI = 1 second) after a response was made or after 3 seconds if no response was made. Responses and reaction times were recorded. Each of the 4 practice items and 48 experimental items were heard twice (one "same" and one "different" trial) randomized in two blocks.

Next, participants completed a picture-naming task on the same target words in their native language. Their exposure to these words during the "same-different" task should have increased the likelihood that participants would produce the desired label for each picture. Participants wore a headset with headphones and a microphone positioned close to their lips. They saw a line drawing of an object appear on the computer screen and were instructed to name the picture as quickly as possible. All productions were recorded using DMDX (Forster & Forster, 2003) and naming accuracy and latencies were measured individually using CheckVocal (Protopapas, 2007).

The last task on Day 1 was the picture-association learning task. Each of the 48 nonwords was randomly paired with one of the pictures of unusual objects. On Day 1, participants learned half (24) of the picture-nonword pairings. On each trial, the computer displayed two pictures side by side, while the nonword associated with one of the pictures was presented over headphones. Participants indicated which picture matched the nonword by pressing one of two response buttons (left button = left picture, right button = right picture). Once a response was made, the incorrect picture disappeared to provide feedback regarding the correct picture-nonword pairing. The incorrect picture displayed was randomly chosen trial-by-trial from the other unusual pictures. Participants heard each nonword 24 times, randomized within blocks. Responses and reaction times (measured from word offset) were recorded. Samuel and Larraza (in revision) have shown that, with this learning procedure, participants produce essentially perfect choice behavior after approximately 10 exposures to each picture. This was replicated in the present research.

On the second day, participants began with the picture-association learning task in order to learn the other half (24) of the picture-nonword pairings. This was followed by the "same/different" task and picture-naming task using all 48 nonword stimuli following the same procedure used with real words.

Chapter 5: Results

Analysis

The central theoretical question of the present research was whether clustering and/or position-specific neighbors could be driving the previously observed effects of phonological neighborhood density in English and in Spanish. As such, the fixed effects of interest were phonological neighborhood density, the clustering coefficient, the number of neighbors sharing the onset phoneme, and the number of neighbors sharing the offset phoneme. *Table 16* below summarizes the various models used for analysis. When modeling the data, I first assessed how well overall phonological neighborhood density predicted performance (model 1). This allowed for a direct comparison with previous results. I then modeled each of the remaining fixed effects alone (models 2, 4, and 6 for the clustering coefficient, onset neighbors, and offset neighbors respectively) to see if they were good predictors of overall behavior. The critical models were those in which phonological neighborhood density was combined with these fixed effects (models 3, 5, and 7 for the clustering coefficient, onset neighbors, and offset neighbors respectively). In these combined models, the two fixed effects (e.g., phonological neighborhood density and the clustering coefficient) act as controls for one another, with the better predictor explaining more of the variance. A comparison of the estimates from these statistical models was used to determine the direction (facilitation or inhibition) and predictive power of these effects.

Table 16. Summary of the fixed and random effects included in each of the models used for the analysis.

	Null model	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**
# Neighbors		✓		✓		✓		✓
Clustering Coefficient			✓	✓				
# Onset Neighbors					✓	✓		
# Offset Neighbors							✓	✓
Items (Intercept)	✓	✓	✓	✓	✓	✓	✓	✓
Participants (Intercept)	✓	✓	✓	✓	✓	✓	✓	✓

The dependent measures of interest were reaction times on "same" trials in the "same/different" (perceptual) task and naming latencies on the picture naming (production) task. These were collected for pre-existing lexical items in each language, and for the newly learned Spanglish nonwords (referred to as English or Spanish nonwords depending on the lexicon the items joined). Note that the real word stimuli could not be controlled as well as the nonwords, which is why the latter were designed to vary systematically along the fixed effects of interest. Reaction times for incorrect responses and responses <100ms or >2000ms were excluded from analysis. Additionally, naming latencies were only included when the correct item was produced. For the Spanglish nonwords, this meant producing each phoneme correctly (because if a different phoneme was learned, then a different neighborhood, with different characteristics, would be activated).

LMMs were used to model the data. Random intercepts were included for participants and items. When I attempted to include random slopes for the fixed effects, some of these models failed to converge. For the majority that did converge, adding random slopes did not increase the fit of the models or change their interpretation. Even for those that did increase the fit, the interpretation did not change (i.e., estimates remained largely unchanged). Similarly, transforming the reaction time data, either by base e log, base 10 log, or square root, had no effect on results. Therefore, the dependent variable was untransformed reaction times in milliseconds, since this simplifies the interpretation of the model estimates. I report the model estimates (β), standard errors (SEs), *t* values, and *p* values.

All analyses were conducted in R (R Core Team, 2014). The packages lme4 (D. Bates et al., 2014) and lmerTest (Kuznetsova et al., 2014) were used to fit LMMs. The package texreg (Leifeld, 2013) was used to create and export the formatted tables of the statistical models.

Hypotheses

For phonological neighborhood density, I predicted a replication of the most common pattern of results reported for each language, particularly for the real word stimuli. That is, for English words, I predicted a high density disadvantage in speech perception and a high density advantage in speech production. And, for Spanish words, I predicted a high density advantage in speech perception and a high density disadvantage in speech production. However, as reviewed in Chapter 1, previous results have been rather inconsistent, with both facilitation and inhibition being observed. The existence of both patterns suggests that there are multiple controlling variables, and the present research was designed to test two of these possibilities (clustering and position-specific neighbors).

For the clustering coefficient, previous research predicts a high clustering disadvantage in both modalities. However, since the rationale behind clustering depends on spreading activation, it is difficult to predict *a priori* how far and how strongly activation will spread among links between neighbors. Additionally, the previous research on the clustering coefficient has focused on simple CVC words, which may or may not generalize to the more complex stimuli used in the present research.

For position-specific neighbors, previous research suggests an onset neighbor disadvantage in speech perception and an offset neighbor advantage in speech production. The rationale behind the effect of position-specific neighbors taps into one of the fundamental characteristics of speech – that it unfolds over time. As such, there is a strong anticipatory component to perceiving speech. Theories have typically described the system as one that activates and eliminates possible word candidates based on the speech that has already unfolded. Likewise, in production, the system prepares for the soon-to-be produced items while deactivating those already produced. Due to the nature of the tasks used in the present research, I predicted the largest effects from onset neighbors. For tasks that use single word perception, like the "same/different" task used in the present research, the lack of prior noise or need for word segmentation means the word's onset is a clear and reliable cue to perception. It should therefore have a strong effect on overall performance. In this case, the shared onsets presumably increase the number of word candidates under consideration during perception, leading to an inhibitory effect. This is in contrast to a facilitatory effect in production, in which onset neighbors support processing of the shared onset. This should help production start quickly and reduce naming latencies.

Perception of Existing English Words

Results did not replicate the high density disadvantage previously observed for the perception of English words. The pattern of results suggests a tiny 0.55ms increase in reaction time per neighbor (model 1 $\beta = 0.55$, SE = 0.96, $t(46) = 0.57$, $p = 0.57$). There was a hint that clustering could be driving some of this small disadvantage, but this was not significant (model 3, $\beta = 124.91$, SE = 107.24, $t(49) = 1.17$, $p = 0.25$).

Like overall neighborhood density, onset neighbors were not good predictors of overall behavior. However, the data suggests that offset neighbors could be beneficial for perception. The combined model 7 estimated a 3.91ms increase per neighbor ($\beta = 3.91$, SE = 2.56, $t(46) = 1.52$, $p = 0.13$) combined with a 4.52ms decrease per offset neighbor ($\beta = -4.52$, SE = 3.21, $t(46) = -1.41$, $p = 0.17$).

See *Figure 11* and *Table 17* below for details. *Figure 11* displays the reaction times for perceiving English words on the "same/different" task as a function of the four fixed effects of interest (# phonological neighbors, clustering coefficient, # onset neighbors, and # offset neighbors). Each colored circle represents the reaction time for one of the 48 native speakers of English on one of the 48 English word stimuli. The x-axis reflects the range of possible values for each fixed effect. For example, phonological neighborhood density of the English words ranged from 3 to 49 neighbors, with the onset and offset densities comprising a subset of these neighbors. By definition, the clustering coefficient ranges from 0 to 1. As stated above, reaction times greater than 2000ms (or less than 100ms) were dropped prior to analysis, so the y-axis reflects this range. Both regression and lowess lines were added to each graph to highlight the overall pattern of responses. The regression line is the trend line from a linear regression model of the data. The lowess line is a localized regression line created by fitting simple regression models to subsets of the data. In this way, a lowess line can capture any curvature to the data that a simple regression line would not reflect. This same format was used for the subsequent figures in this chapter.

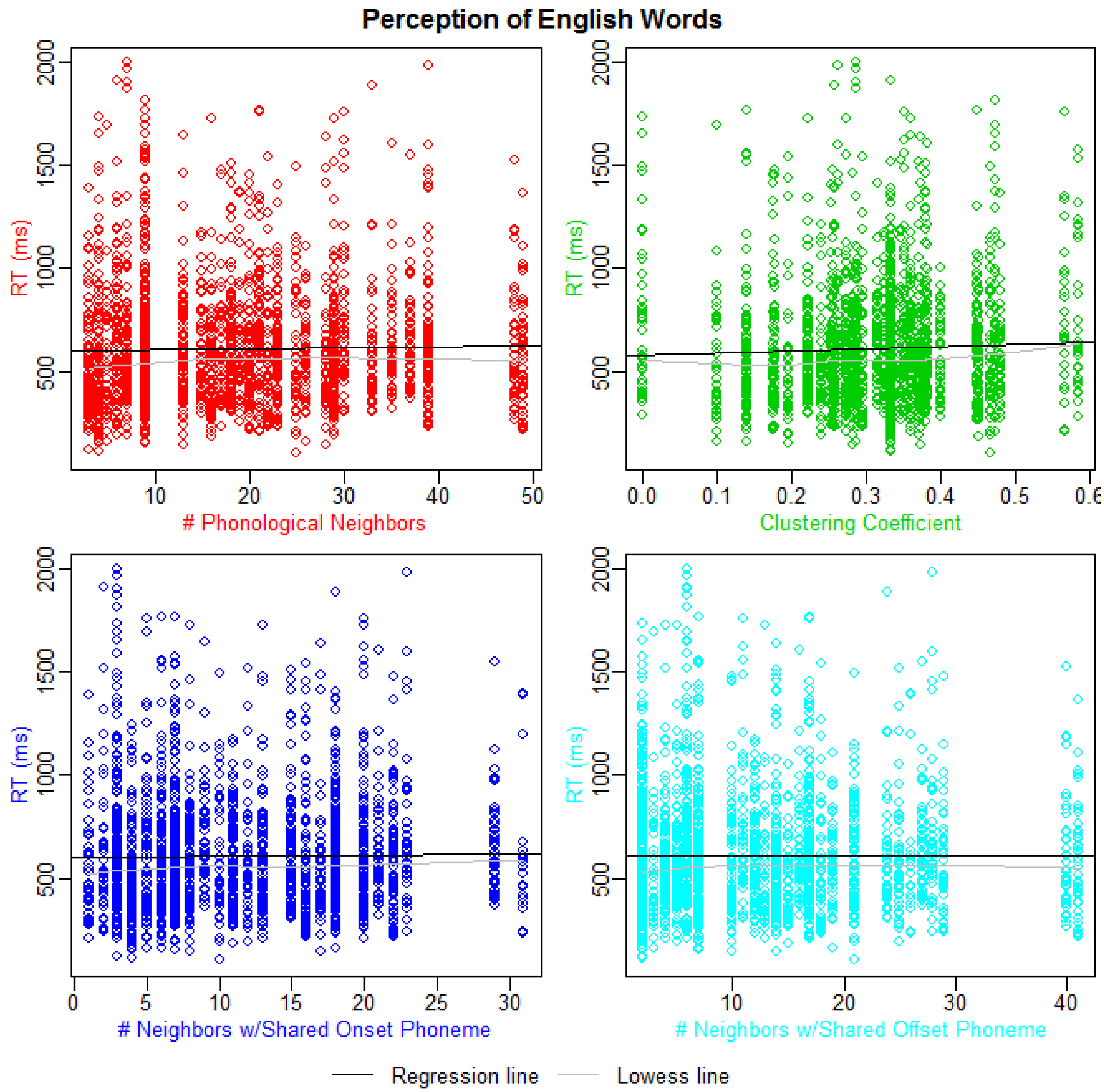


Figure 11. Reaction times for perceiving English words on the "same/different" task.

Table 17. LMMs for the Perception of English Words on the "Same/Different" Task.

	Null model	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**
Intercept	615.92***	605.45***	574.91***	569.97***	609.92***	607.75***	615.56***	599.16***
	(24.78)	(30.85)	(41.25)	(43.21)	(31.29)	(31.44)	(29.24)	(30.84)
# Neighbors		0.55		0.37		1.18		3.91
		(0.96)		(0.96)		(1.96)		(2.56)
Clustering Coefficient			131.52	124.91				
			(106.02)	(107.24)				
# Onset Neighbors					0.50	-1.22		
					(1.61)	(3.27)		
# Offset Neighbors							0.03	-4.52
							(1.21)	(3.20)
AIC	29394.83	29396.51	29395.31	29397.17	29396.73	29398.37	29396.83	29396.57
BIC	29417.46	29424.79	29423.59	29431.11	29425.02	29432.31	29425.12	29430.51
Log Likelihood	-14693.42	-14693.26	-14692.65	-14692.58	-14693.37	-14693.19	-14693.42	-14692.28
Num. obs.	2115	2115	2115	2115	2115	2115	2115	2115
Num. groups: Items	48	48	48	48	48	48	48	48
Num. groups: Participants	48	48	48	48	48	48	48	48
Variance: Items (Intercept)	5314.38	5265.82	5140.06	5117.59	5296.86	5251.57	5314.24	4989.48
Variance: Participants (Intercept)	22840.25	22843.10	22832.12	22834.50	22843.22	22839.21	22840.39	22836.85
Variance: Residual	57193.18	57193.72	57186.16	57186.60	57193.77	57192.94	57193.18	57196.33

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{+}p < 0.1$

Note: For the fixed effects, the table displays estimates (β) followed by standard errors in parentheses.

Production of Existing English Words

Results did not replicate the high density advantage previously observed for the production of English words. In fact, the pattern of results suggests a disadvantage, a pattern previously reported for Spanish (Vitevitch & Stamer, 2006). There was a non-significant 1.29ms increase in naming latency per neighbor (model 1 $\beta = 1.29$, SE = 0.97, $t(47) = 1.33$, $p = 0.19$).

As was the case with overall neighborhood density, the pattern of results suggests a disadvantage for both onset and offset neighbors. Estimates from the combined models (5 and 7) indicate that the overall (non-significant) effect was driven by the onset neighbors. Model 5 estimated a 2.30ms increase in naming latencies per onset neighbor (compared to the 2.42ms estimate from model 4) and only a 0.08ms increase per neighbor (compared to the 1.29ms estimate from model 1), indicating the majority of the variance was explained by the subset of neighbors sharing the onset. The reverse pattern was true of the offset neighbors. Model 7 estimated a 0.18ms increase per offset neighbor (compared to the 1.53ms estimate from model 6) and a 1.16ms increase per neighbor (compared to the 1.29ms estimate from model 1), indicating the majority of the variance was explained by the overall number of neighbors.

Contrary to expectations, there was a hint that clustering could be advantageous for production, but this was not significant (model 3, $\beta = -161.90$, SE = 106.44, $t(47) = -1.52$, $p = 0.14$).

See *Figure 12* and *Table 18* below for details.

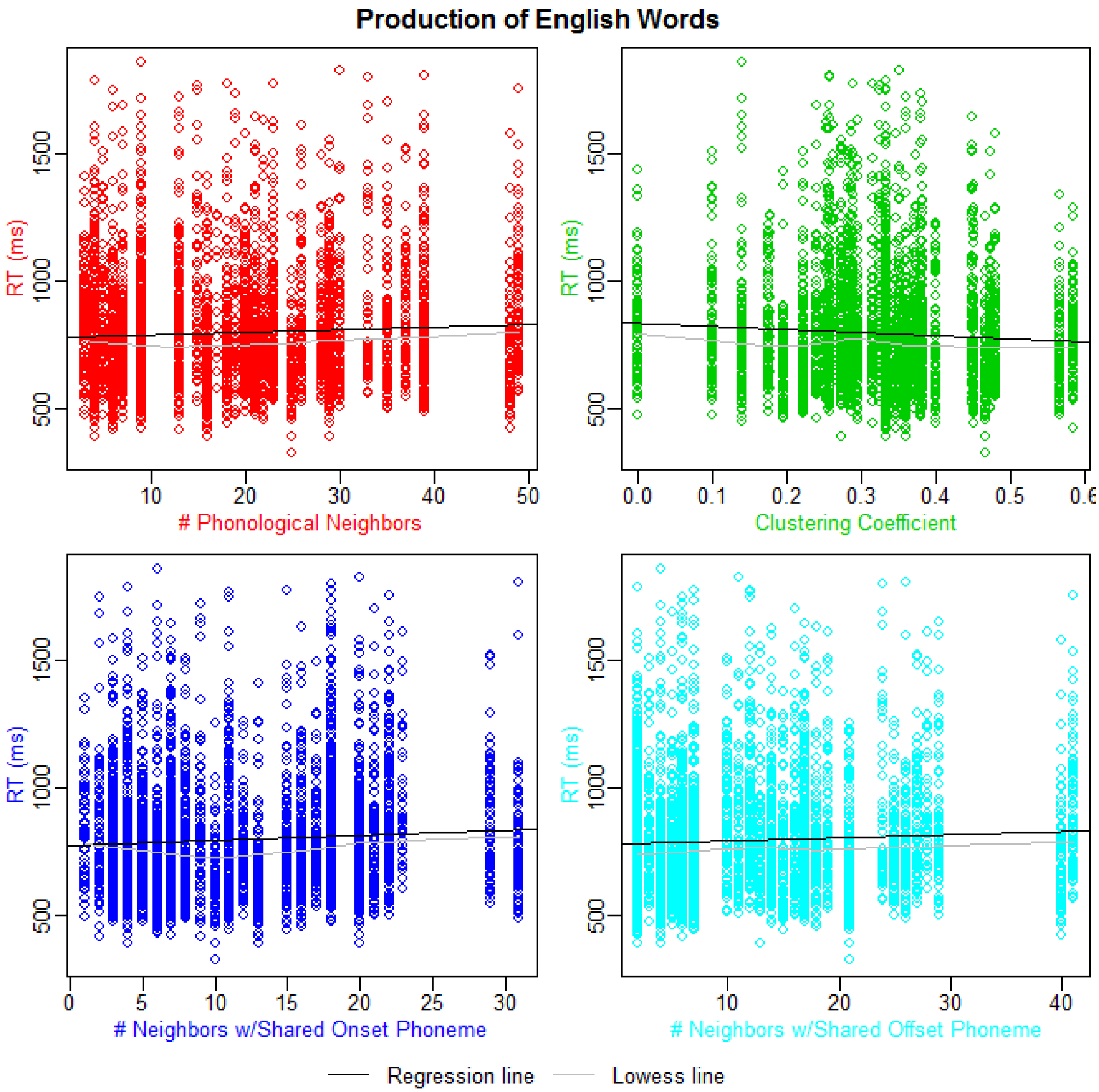


Figure 12. Naming latencies for producing English words on the picture naming task.

Table 18. LMMs for the Production of English Words on the Picture Naming Task.

	Null model	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Intercept	807.34***	782.50***	850.14***	828.94***	778.52***	778.35***	787.60***	782.75***
	(16.85)	(25.07)	(37.73)	(39.25)	(25.34)	(25.66)	(22.96)	(25.47)
# Neighbors		1.29		1.52		0.08		1.16
		(0.97)		(0.96)		(1.98)		(2.65)
Clustering Coefficient			-136.58	-161.90				
			(107.85)	(106.44)				
# Onset Neighbors					2.42	2.30		
					(1.60)	(3.27)		
# Offset Neighbors							1.53	0.18
							(1.22)	(3.32)
AIC	53093.99	53094.27	53094.41	53094.00	53093.78	53095.78	53094.45	53096.26
BIC	53119.10	53125.65	53125.80	53131.67	53125.17	53133.44	53125.84	53133.93
Log Likelihood	-26543.00	-26542.13	-26542.21	-26541.00	-26541.89	-26541.89	-26542.23	-26542.13
Num. obs.	3935	3935	3935	3935	3935	3935	3935	3935
Num. groups: Items	48	48	48	48	48	48	48	48
Num. groups: Participants	48	48	48	48	48	48	48	48
Variance: Items (Intercept)	6543.27	6294.52	6325.95	5993.54	6228.28	6227.91	6320.42	6294.00
Variance: Participants (Intercept)	6546.69	6548.29	6545.14	6546.81	6547.63	6547.69	6549.12	6548.41
Variance: Residual	39676.48	39676.08	39675.56	39674.91	39675.86	39675.86	39676.10	39676.08

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{+}p < 0.1$

Note: For the fixed effects, the table displays estimates (β) followed by standard errors in parentheses.

Perception of Existing Spanish Words

Results did not replicate the high density advantage previously observed for the perception of Spanish words. Instead, I observed a high density disadvantage. There was a significant 2.64ms increase in reaction time per neighbor (model 1 $\beta = 2.64$, SE = 1.34, $t(46) = 1.97$, $p = 0.05$).

There was a trend suggesting that onset neighbors could be driving this disadvantage (model 4 $\beta = 3.81$, SE = 2.04, $t(46) = 1.87$, $p = 0.07$). However, the combined model (5) indicates that overall neighborhood density is the better predictor (even though n.s.), since the estimate for onset neighbors dropped from 3.81 to 0.46.

There was also a hint that offset neighbors were driving the observed effects (model 6 $\beta = 2.17$, SE = 1.43, $t(46) = 1.52$, $p = 0.14$). Critically, the combined model (7) shows a 7.79ms decrease in reaction time per offset neighbor ($\beta = -7.79$, SE = 5.37, $t(46) = -1.45$, $p = 0.15$) combined with a 9.82ms increase per neighbor ($\beta = 9.82$, SE = 5.13, $t(46) = -1.92$, $p = 0.06$). This indicates an overall high density disadvantage but an advantage from offset neighbors. It is possible that this offset boost was responsible for the high density advantage reported in previous studies, particularly since Spanish neighbors tend to share offsets.

The clustering coefficient was not a good predictor of overall behavior (model 2 $\beta = 17.17$, SE = 40.12, $t(46) = 0.43$, $p = 0.67$). As shown in the green data displayed in the upper right corner of *Figure 13*, both the regression and lowess lines are essentially flat, indicating no differences in reaction time as a function of clustering.

See *Figure 13* and *Table 19* below for details.

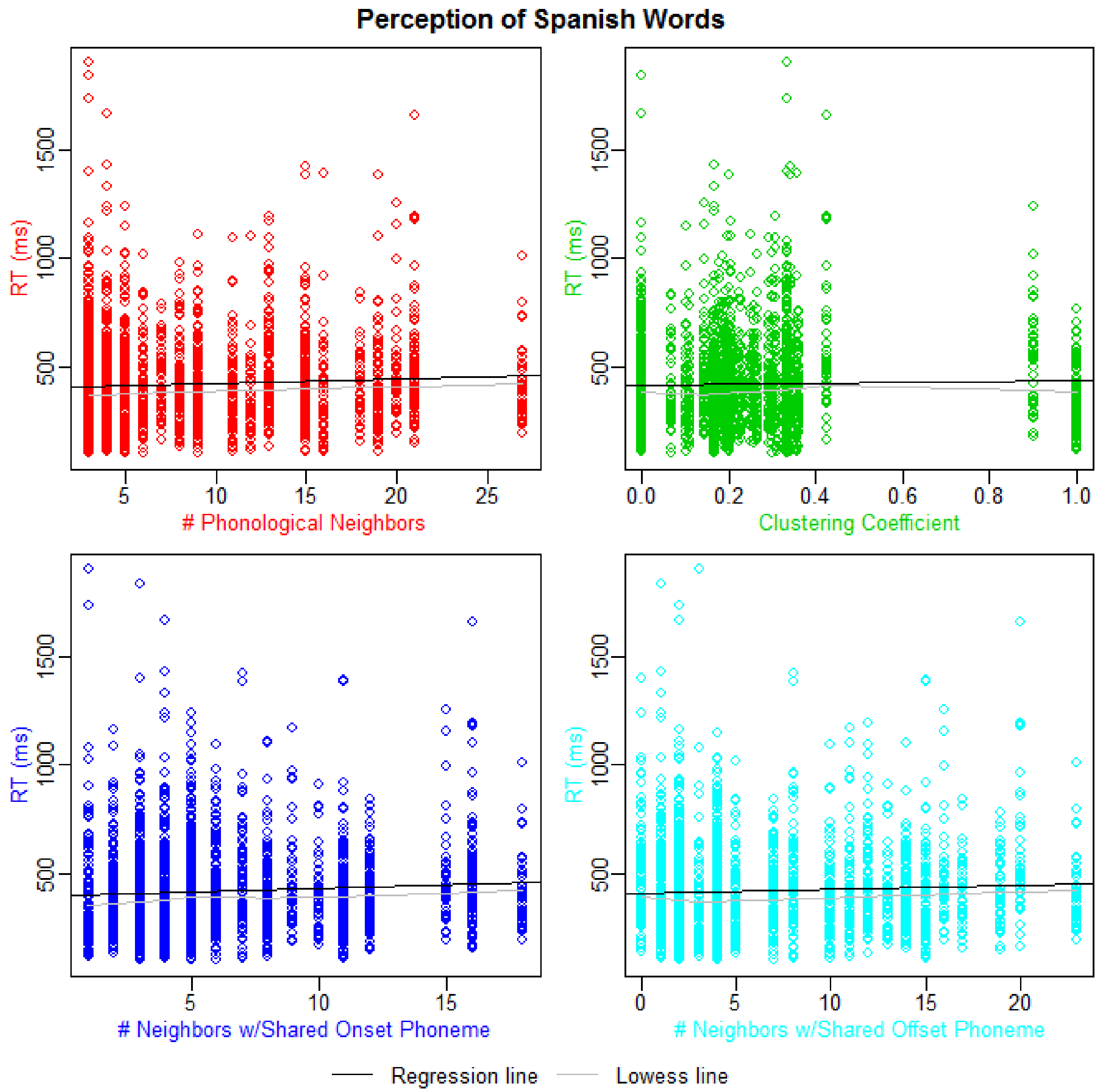


Figure 13. Reaction times for perceiving Spanish words on the "same/different" task.

Table 19. LMMs for the Perception of Spanish Words on the "Same/Different" Task.

	Null model	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**
Intercept	417.33^{***}	393.96^{***}	413.11^{***}	390.13^{***}	393.65^{***}	393.62^{***}	402.44^{***}	383.92^{***}
	(20.31)	(23.39)	(22.56)	(25.14)	(23.82)	(23.79)	(22.47)	(24.20)
# Neighbors		2.64^{*}		2.63^{*}		2.35		9.82^{+}
		(1.34)		(1.34)		(3.87)		(5.12)
Clustering Coefficient			17.17	15.96				
			(40.12)	(38.54)				
# Onset Neighbors					3.81^{+}	0.46		
					(2.04)	(5.86)		
# Offset Neighbors							2.17	-7.79
							(1.43)	(5.37)
AIC	28185.00	28183.26	28186.81	28185.09	28183.63	28185.26	28184.74	28183.21
BIC	28207.66	28211.60	28215.15	28219.09	28211.96	28219.26	28213.08	28217.21
Log Likelihood	-14088.50	-14086.63	-14088.41	-14086.55	-14086.81	-14086.63	-14087.37	-14085.60
Num. obs.	2137	2137	2137	2137	2137	2137	2137	2137
Num. groups: Items	48	48	48	48	48	48	48	48
Num. groups: Participants	48	48	48	48	48	48	48	48
Variance: Items (Intercept)	3124.15	2834.02	3108.30	2820.20	2858.81	2833.29	2946.57	2682.42
Variance: Participants (Intercept)	16046.42	16049.67	16045.16	16048.42	16055.36	16050.41	16047.99	16052.87
Variance: Residual	27843.31	27844.01	27843.59	27844.30	27844.15	27844.03	27843.71	27844.63

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{+}p < 0.1$

Note: For the fixed effects, the table displays estimates (β) followed by standard errors in parentheses.

Production of Existing Spanish Words

Results did replicate the high density disadvantage previously observed for the production of Spanish words. There was a significant 6.29ms increase in naming latency per neighbor (model 1 $\beta = 6.29$, SE = 2.26, $t(47) = 2.78$, $p = 0.008$).

Both onset and offset neighbors were good predictors of overall behavior (see models 4 and 6). Critically, the combined model (5) indicated a significant 10.21ms decrease in naming latency per onset neighbor ($\beta = -10.21$, SE = 10.02, $t(49) = -1.02$, $p = 0.31$) combined with a 12.42ms increase per neighbor ($\beta = 12.42$, SE = 6.43, $t(47) = 1.93$, $p = 0.06$). This indicates an overall high density disadvantage but an advantage from onset neighbors. This provides support for the hypothesis that onset neighbors support processing of the shared onset, in turn helping production start quickly and reducing naming latencies.

It seems that offset neighbors were driving the overall high density disadvantage. There was a significant 6.68ms increase in naming latency per offset neighbor, (model 6 $\beta = -6.68$, SE = 2.33, $t(46) = 2.87$, $p = 0.006$). The combined model (7) suggests that offset neighbors are a better predictor than overall neighborhood density, since the estimate for offset neighbors was reduced only slightly (from 6.68 to 5.90) while the estimate for neighborhood density was reduced substantially (from 6.29 to 0.77).

The clustering coefficient was not a good predictor of overall behavior (model 2 $\beta = 31.99$, SE = 74.38, $t(50) = 0.43$, $p = 0.67$). As shown in the green data displayed in the upper right corner of *Figure 14*, both the regression and lowess lines are essentially flat, indicating no differences in naming latency as a function of clustering.

See *Figure 14* and *Table 20* below for details.

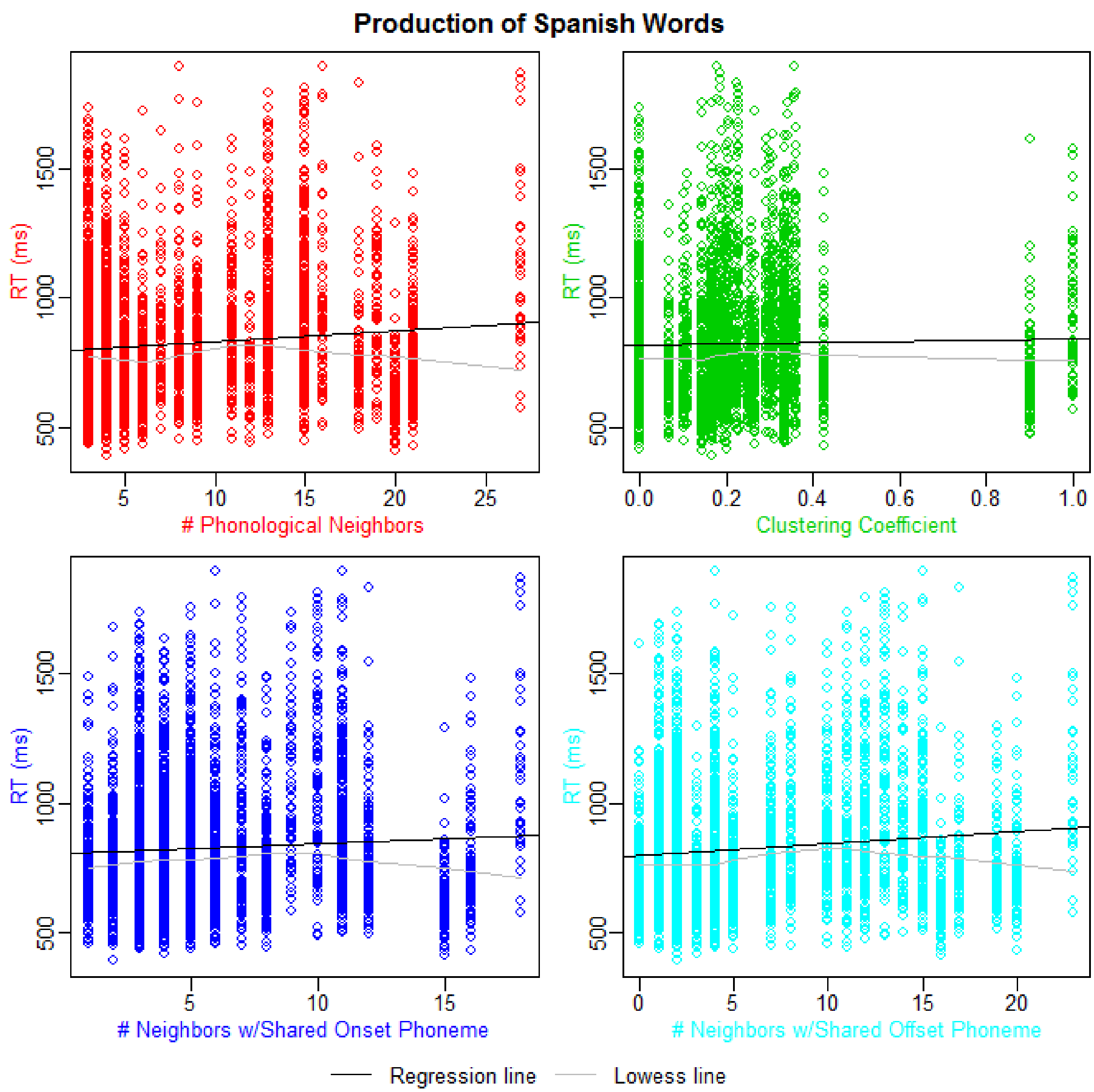


Figure 14. Naming latencies for producing Spanish words on the picture naming task.

Table 20. LMMs for the Production of Spanish Words on the Picture Naming Task.

	Null model	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**
Intercept	840.55^{***}	785.04^{***}	833.01^{***}	781.01^{***}	790.84^{***}	793.97^{***}	794.76^{***}	793.25^{***}
	(19.53)	(27.28)	(26.20)	(31.17)	(29.31)	(28.26)	(24.47)	(29.96)
# Neighbors		6.29^{**}		6.25^{**}		12.42^{+}		0.77
		(2.26)		(2.26)		(6.43)		(8.84)
Clustering Coefficient			31.99	18.72				
			(74.38)	(70.09)				
# Onset Neighbors					8.05^{*}	-10.21		
					(3.62)	(10.02)		
# Offset Neighbors							6.68^{**}	5.90
							(2.33)	(9.16)
AIC	50409.81	50404.48	50411.62	50406.41	50406.99	50405.47	50404.07	50406.07
BIC	50434.68	50435.57	50442.71	50443.72	50438.08	50442.78	50435.16	50443.38
Log Likelihood	-25200.90	-25197.24	-25200.81	-25197.20	-25198.50	-25196.73	-25197.04	-25197.03
Num. obs.	3708	3708	3708	3708	3708	3708	3708	3708
Num. groups: Items	48	48	48	48	48	48	48	48
Num. groups: Participants	48	48	48	48	48	48	48	48
Variance: Items (Intercept)	10141.42	8785.10	10100.90	8777.30	9350.85	8451.38	8652.92	8659.15
Variance: Participants (Intercept)	7295.85	7294.11	7293.35	7292.64	7300.69	7285.89	7290.97	7291.32
Variance: Residual	43674.43	43661.69	43674.50	43661.40	43658.79	43670.37	43664.98	43664.49

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{+}p < 0.1$

Note: For the fixed effects, the table displays estimates (β) followed by standard errors in parentheses.

Perception of English Nonwords

As we have seen, the results for existing words in both languages were rather messy, with some outcomes matching prior findings but many others mismatching. As noted in Chapter 1, a fundamental and recurring problem in this literature is the natural confounding of any variables with the factors of interest. This is why, in the current study, the approach was to eliminate as many confounding variables as possible by introducing carefully chosen new "words" into the English and Spanish lexicons. With items selected this way, the results for the nonwords did replicate the high density disadvantage previously observed for the perception of English words. There was a significant 3.28ms increase in reaction time per neighbor (model 1 $\beta = 3.28$, SE = 1.20, $t(48) = 2.74$, $p = 0.009$).

Both onset and offset neighbors were good predictors of overall behavior (see models 4 and 6). There was a nearly significant 3.49ms increase in naming latency per onset neighbor, (model 4 $\beta = 3.49$, SE = 1.79, $t(45) = 1.95$, $p = 0.057$). However, the combined model (5) suggested that overall neighborhood density is a better predictor than onset neighbors, since the estimate for neighborhood density was reduced only slightly (from 3.49 to 3.47) while the estimate for onset neighbors was reduced substantially (from 3.28 to -0.36). Clearly these two factors are explaining much of the same variance (as is to be expected, since onset neighbors are simply a subset of the overall number of neighbors).

There was a significant 2.76ms increase in naming latency per offset neighbor, (model 6 $\beta = 2.76$, SE = 1.38, $t(48) = 2.00$, $p = 0.052$). Critically however, the combined model (7) suggested a 3.40ms decrease in reaction time per offset neighbor ($\beta = -3.40$, SE = 3.17, $t(47) = -1.07$, $p = 0.29$) combined with a 6.05ms increase per neighbor ($\beta = 6.05$, SE = 2.84, $t(47) = 2.13$, *p* = 0.04). This indicates an overall high density disadvantage but an advantage from offset neighbors. It also mirrors the pattern observed for perception of existing words in Spanish.

The clustering coefficient was not a good predictor of overall behavior (model 2 β = -13.91, SE = 19.74, $t(44) = -0.71$, $p = 0.49$). As shown in the green data displayed in the upper right corner of *Figure 15*, both the regression and lowess lines are essentially flat, indicating no differences in reaction time as a function of clustering.

See *Figure 15* and *Table 21* below for details.

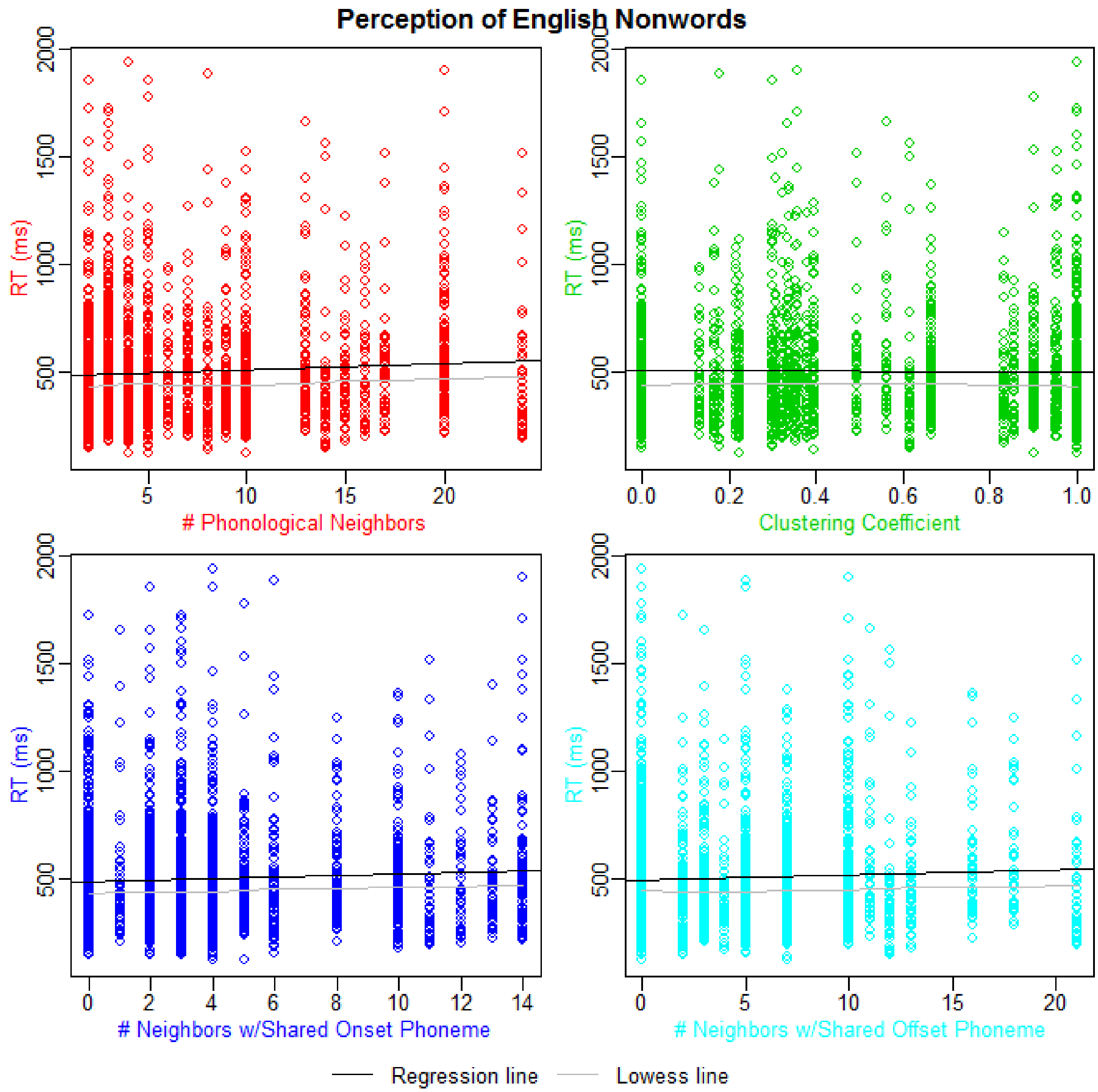


Figure 15. Reaction times for perceiving English nonwords on the "same/different" task.

Table 21. LMMs for the Perception of English Nonwords on the “Same/Different” Task.

	Null model	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Intercept	506.71***	482.96***	513.94***	487.02***	492.31***	483.07***	493.42***	479.28***
	(21.66)	(23.15)	(23.96)	(25.60)	(22.77)	(23.17)	(22.56)	(23.33)
# Neighbors		3.28**		3.21**		3.47+		6.05*
		(1.19)		(1.20)		(1.86)		(2.84)
Clustering Coefficient			-13.91	-6.89				
			(19.74)	(18.63)				
# Onset Neighbors					3.49+	-0.36		
					(1.79)	(2.70)		
# Offset Neighbors							2.76*	-3.40
							(1.38)	(3.17)
AIC	29045.19	29040.16	29046.70	29042.03	29043.54	29042.15	29043.33	29041.03
BIC	29067.83	29068.46	29074.99	29075.98	29071.83	29076.10	29071.63	29074.99
Log Likelihood	-14518.59	-14515.08	-14518.35	-14515.01	-14516.77	-14515.07	-14516.67	-14514.52
Num. obs.	2120	2120	2120	2120	2120	2120	2120	2120
Num. groups: Items	48	48	48	48	48	48	48	48
Num. groups: Participants	48	48	48	48	48	48	48	48
Variance: Items (Intercept)	1538.35	1183.22	1504.09	1173.16	1327.13	1184.40	1348.39	1115.85
Variance: Participants (Intercept)	19894.06	19900.40	19899.55	19903.51	19896.28	19900.49	19898.59	19899.75
Variance: Residual	47675.32	47670.69	47677.95	47672.21	47682.10	47669.71	47667.59	47677.41

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{+}p < 0.1$

Note: For the fixed effects, the table displays estimates (β) followed by standard errors in parentheses.

Production of English Nonwords

Overall, participants were not very good at producing the newly learned Spanglish nonwords. This was despite the fact that they all performed near ceiling on the picture association task after the first eight repetitions. This left only 731 observations of a possible 2304 (48 items time 48 participants) for analysis. Because of this, it is difficult to put too much faith in the results on this task, but see *Figure 16* and *Table 22*.

Phonological neighborhood density was not a good predictor of overall behavior (model 1 $\beta = -0.90$, SE = 1.97, $t(34) = -0.46$, $p = 0.65$). However, the pattern of data suggests a tiny high density advantage of subtracting 0.90ms per neighbor.

Similar to overall phonological neighborhood density, neither onset nor offset densities were good predictors of naming latencies. Though non-significant, results from the combined models (5 and 7) suggest an advantage from onset neighbors and a disadvantage from offset neighbors. Naming latencies decreased by 3.13ms for each onset neighbor (model 5 $\beta = -3.13$, SE = 4.71, $t(42) = -0.66$, $p = 0.51$). Naming latencies increased by 1.63ms for each offset neighbor (model 7 $\beta = 1.63$, SE = 5.42, $t(36) = 0.30$, $p = 0.77$).

The clustering coefficient was not a good predictor of overall behavior, however the pattern of data suggest a high clustering disadvantage (model 2 $\beta = 39.63$, SE = 31.82, $t(40) = 1.25$, $p = 0.22$). That is, naming latencies tended to increase as clustering increased.

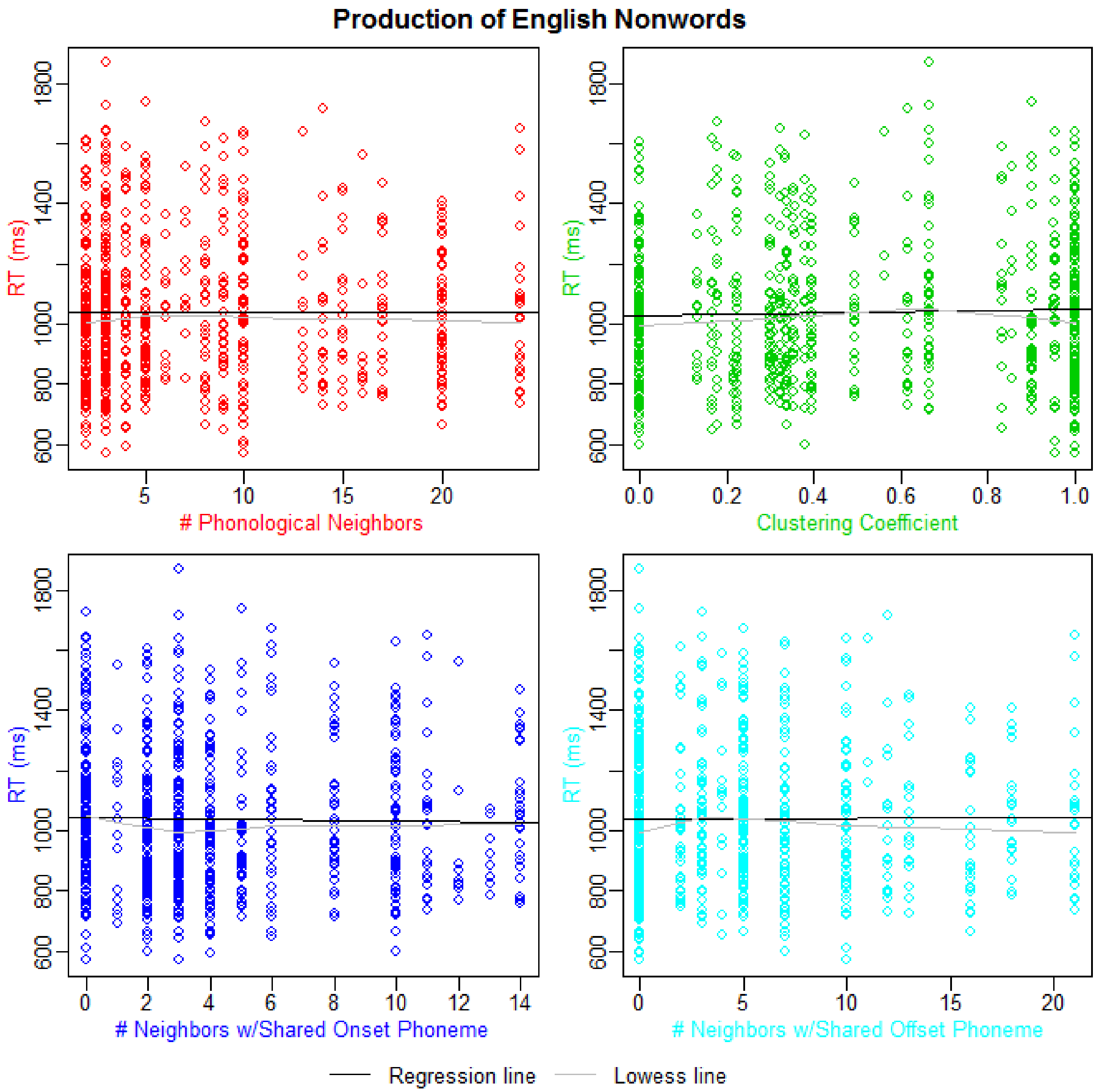


Figure 16. Naming latencies for producing English nonwords on the picture naming task.

Table 22. LMMs for the Production of English Nonwords on the Picture Naming Task.

	Null model	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**
Intercept	1061.38***	1068.09***	1041.62***	1046.69***	1071.24***	1069.54***	1064.66***	1069.99***
	(16.48)	(22.00)	(22.94)	(28.37)	(20.79)	(22.09)	(19.69)	(22.92)
# Neighbors		-0.90		-0.60		0.68		-2.26
		(1.97)		(1.98)		(3.08)		(4.93)
Clustering Coefficient			39.63	38.43				
			(31.82)	(32.09)				
# Onset Neighbors					-2.32	-3.13		
					(3.00)	(4.71)		
# Offset Neighbors							-0.65	1.63
							(2.16)	(5.41)
AIC	10066.63	10068.42	10067.08	10068.99	10068.03	10069.98	10068.54	10070.33
BIC	10085.01	10091.40	10090.06	10096.56	10091.00	10097.55	10091.51	10097.90
Log Likelihood	-5029.32	-5029.21	-5028.54	-5028.50	-5029.02	-5028.99	-5029.27	-5029.17
Num. obs.	731	731	731	731	731	731	731	731
Num. groups: Items	48	48	48	48	48	48	48	48
Num. groups: Participants	48	48	48	48	48	48	48	48
Variance: Items (Intercept)	2884.56	2896.30	2866.73	2872.74	2879.87	2868.52	2902.58	2867.38
Variance: Participants (Intercept)	5573.26	5598.91	5651.49	5667.02	5618.94	5615.21	5590.76	5593.54
Variance: Residual	50217.97	50187.42	50086.10	50070.75	50160.14	50163.79	50195.76	50197.79

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{+}p < 0.1$

Note: For the fixed effects, the table displays estimates (β) followed by standard errors in parentheses.

Perception of Spanish Nonwords

The pattern of results did replicate the high density advantage previously observed for the perception of Spanish words, however they did not reach significance. Results suggested a 3.06ms decrease in reaction time per neighbor (model 1 β = -3.06, SE = 2.84, $t(43) = -1.08$, p = 0.29).

Onset neighbors were not a significant predictor of overall behavior (model 5 β = 0.83, SE = 5.54, $t(45) = 0.15$, p = 0.88). Results indicated a tiny 0.83ms increase per onset neighbor.

There was a hint that offset neighbors were driving the observed effects (model 6 β = -4.55, SE = 2.95, $t(44) = -1.54$, p = 0.13). The combined model (7) suggested a 4.63ms decrease in reaction time per offset neighbor (β = -4.63, SE = 4.25, $t(44) = -1.09$, p = 0.28) combined with a 0.11ms increase per neighbor (β = 0.11, SE = 4.03, $t(43) = 0.03$, p = 0.98). This result adds support to the view that it is this boost from offset neighbors that was responsible for the high density advantage reported in previous studies, particularly since Spanish neighbors tend to share offsets.

The clustering coefficient was a significant predictor of overall behavior. Results indicated a high clustering disadvantage (model 3 β = 56.82, SE = 24.33, $t(45) = 2.34$, p = 0.02).

See *Figure 17* and *Table 23* below for details.

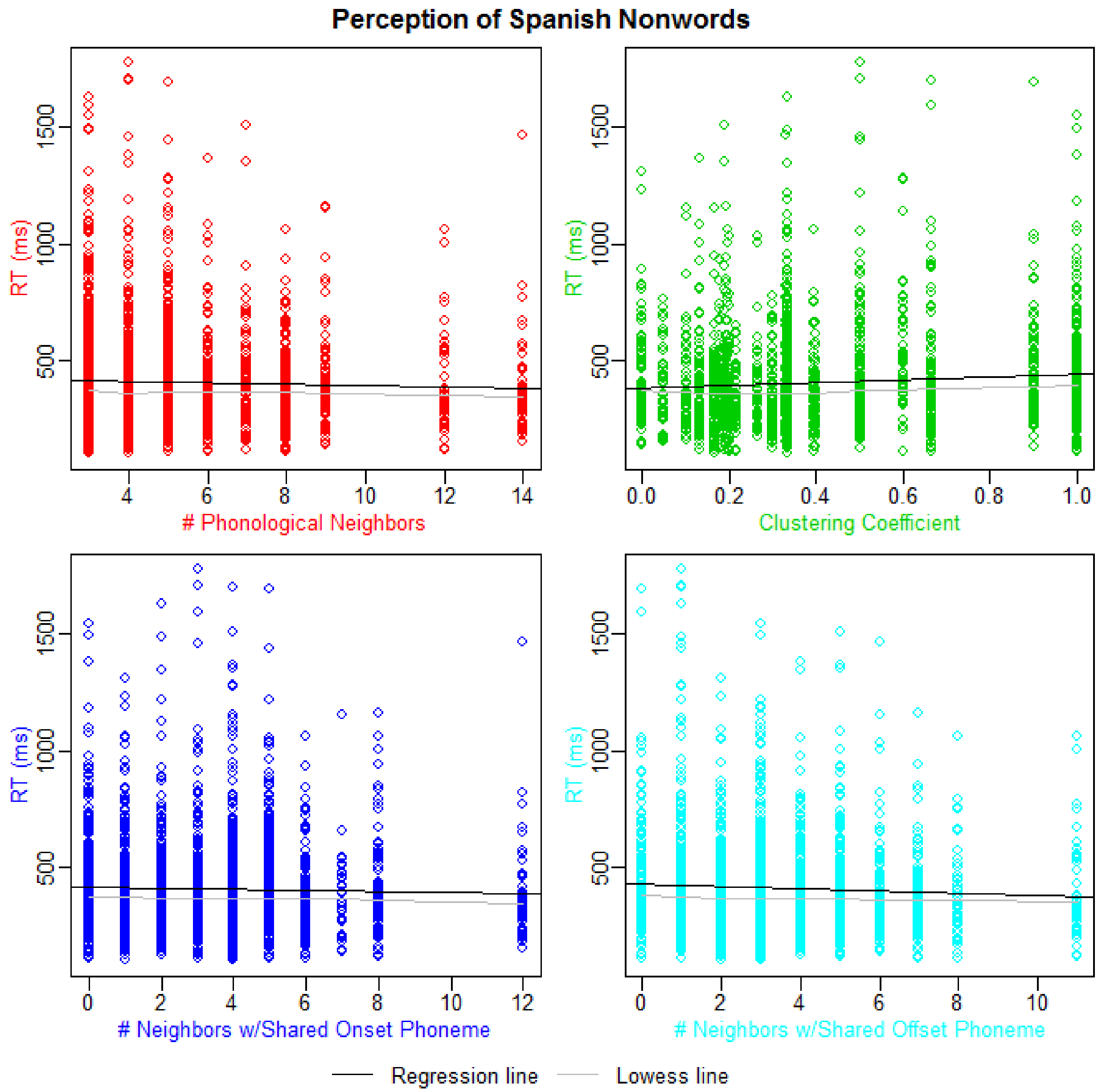


Figure 17. Reaction times for perceiving Spanish nonwords on the "same/different" task.

Table 23. LMMs for the Perception of Spanish Nonwords on the "Same/Different" Task.

	Null model	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Intercept	411.92***	427.21***	387.24***	393.86***	419.84***	428.10***	426.95***	426.68***
	(17.89)	(22.82)	(20.14)	(26.39)	(20.16)	(23.57)	(20.33)	(22.68)
# Neighbors		-3.06		-1.09		-3.78		0.11
		(2.84)		(2.81)		(5.61)		(4.03)
Clustering Coefficient			59.63*	56.82*				
			(23.27)	(24.33)				
# Onset Neighbors					-2.39	0.83		
					(2.82)	(5.54)		
# Offset Neighbors							-4.55	-4.63
							(2.95)	(4.25)
AIC	27655.44	27656.29	27651.26	27653.11	27656.72	27658.27	27655.12	27657.12
BIC	27677.99	27684.48	27679.45	27686.94	27684.91	27692.10	27683.31	27690.95
Log Likelihood	-13823.72	-13823.15	-13820.63	-13820.56	-13823.36	-13823.13	-13822.56	-13822.56
Num. obs.	2076	2076	2076	2076	2076	2076	2076	2076
Num. groups: Items	48	48	48	48	48	48	48	48
Num. groups: Participants	48	48	48	48	48	48	48	48
Variance: Items (Intercept)	1544.69	1489.55	1264.15	1257.50	1512.87	1487.37	1429.49	1429.35
Variance: Participants (Intercept)	13052.42	13051.56	13040.88	13041.11	13052.09	13051.47	13049.21	13049.19
Variance: Residual	32441.78	32441.77	32441.64	32441.71	32440.80	32442.14	32443.62	32443.65

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, +$p < 0.1$

Note: For the fixed effects, the table displays estimates (β) followed by standard errors in parentheses.

Production of Spanish Nonwords

As was the case in English, overall, participants were not very good at producing the newly learned Spanglish nonwords. This was despite the fact that they all performed near ceiling on the picture association task after the first eight repetitions. While the Spanish population performed slightly better than the English population, there were only 906 observations of a possible 2304 (48 items x 48 participants) for analysis.

The pattern of data suggests a non-significant high density disadvantage (model 1 β = 5.12, SE = 5.09, $t(42) = 1.01$, $p = 0.32$). Naming latencies increased by 5.12ms per neighbor.

There was a significant effect of onset neighbors. Critically, the combined model (5) indicated a significant 21.36ms decrease in naming latency per onset neighbor (β = -21.36, SE = 9.97, $t(47) = -2.14$, $p = 0.04$) combined with a significant 24.54ms increase per neighbor (β = 24.54, SE = 10.25, $t(49) = 2.39$, $p = 0.02$). This indicates an overall high density disadvantage but an advantage from onset neighbors. As was the case for the production of Spanish words, this result provides support for the hypothesis that onset neighbors support processing of the shared onset, in turn helping production start quickly and reducing naming latencies.

Offset densities were not a good predictor of naming latencies (model 7 β = 0.13, SE = 8.28, $t(50) = 0.02$, $p = 0.99$).

Finally, the pattern of data suggests a high clustering disadvantage (model 3 β = 60.99, SE = 45.96, $t(45) = 1.33$, $p = 0.19$). That is, naming latencies tended to increase as clustering increased.

See *Figure 18* and *Table 24* for details.

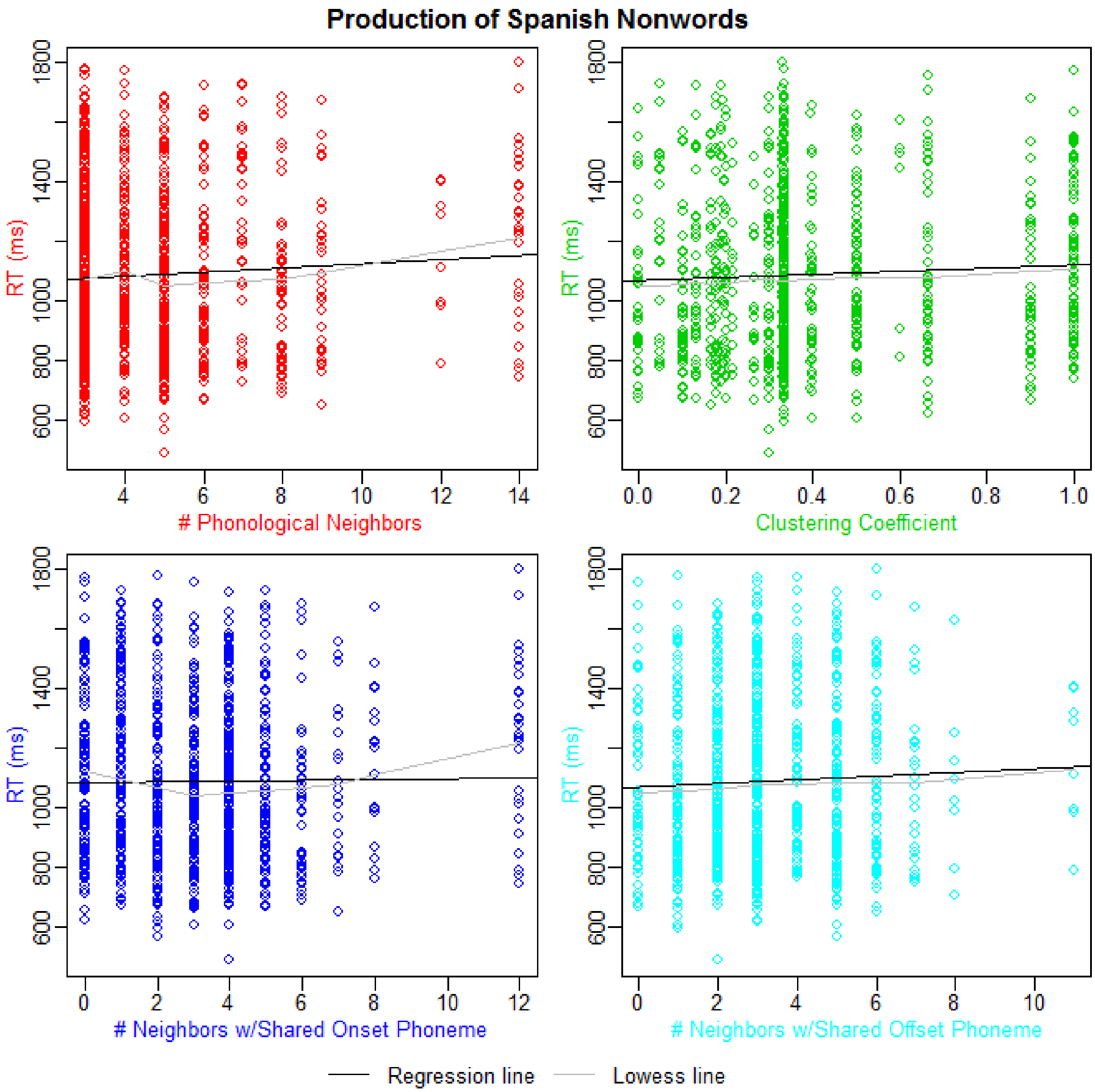


Figure 18. Naming latencies for producing Spanish nonwords on the picture naming task.

Table 24. LMMs for the Production of Spanish Nonwords on the Picture Naming Task.

	Null model	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Intercept	1101.72***	1076.23***	1084.97***	1041.03***	1102.74***	1048.55***	1087.77***	1076.19***
	(18.15)	(30.96)	(25.42)	(40.18)	(24.33)	(32.08)	(26.68)	(31.08)
# Neighbors		5.12		7.25		24.54*		5.04
		(5.09)		(5.24)		(10.25)		(7.11)
Clustering Coefficient			41.48	60.99				
			(44.74)	(45.96)				
# Onset Neighbors					-0.31	-21.36*		
					(5.02)	(9.97)		
# Offset Neighbors							4.22	0.13
							(5.96)	(8.28)
AIC	12665.43	12666.43	12666.58	12666.71	12667.43	12664.13	12666.93	12668.43
BIC	12684.67	12690.48	12690.63	12695.57	12691.47	12692.99	12690.98	12697.29
Log Likelihood	-6328.72	-6328.22	-6328.29	-6327.36	-6328.71	-6326.07	-6328.47	-6328.22
Num. obs.	906	906	906	906	906	906	906	906
Num. groups: Items	48	48	48	48	48	48	48	48
Num. groups: Participants	47	47	47	47	47	47	47	47
Variance: Items (Intercept)	3398.14	3230.01	3263.31	2964.59	3398.56	2511.97	3316.79	3229.90
Variance: Participants (Intercept)	7535.32	7460.16	7455.05	7304.33	7538.68	7387.73	7504.99	7460.36
Variance: Residual	62317.44	62346.17	62342.58	62396.50	62315.97	62417.63	62328.19	62346.15

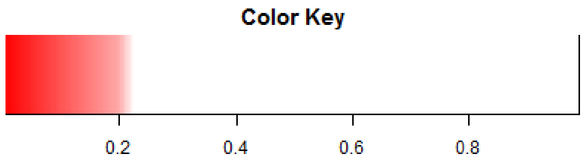
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{+}p < 0.1$

Note: For the fixed effects, the table displays estimates (β) followed by standard errors in parentheses.

Chapter 6: Discussion

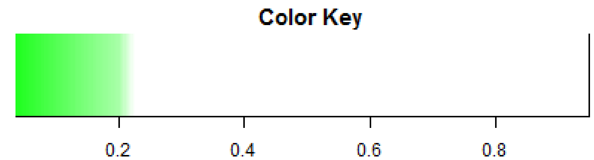
The present research was motivated by the apparent cross-modal and cross-linguistic differences previously reported for one property of linguistic representations: phonological neighborhood density. These differences have been puzzling because they contradict two basic assumptions about language. One assumption is that perception and production share lexical representations. Therefore, a given lexical property should have similar effects in both modalities. This is how word frequency behaves: frequent words have a processing advantage in both perception and production (see Chapter 1 for review). A second assumption is that the system supporting language is fundamentally the same regardless of the language implemented. Continuing the example, word frequency behaves the same way in Spanish as it does in English. This reasoning motivated the search for possible alternative explanations (clustering and position-specific neighbors) for the observed effects of phonological neighborhood density in English and Spanish.

The present results allow us to keep the above assumptions intact by providing evidence that position-specific neighbors are driving the contradictory effects of phonological neighborhood density observed cross-modally and cross-linguistically. In *Figure 19* below, I've summarized the pattern of results for words and nonwords, across modality and population by including a heatmap of the significance of the various model estimates. Contrary to previous results, the present results suggest a high density disadvantage, regardless of modality or language. All the significant findings indicate that processing slowed down as the number of neighbors increased. This is clearly seen in the number of red squares in the first four columns of *Figure 19*. However, it is worth noting that for the results with the Spanglish nonwords, the observed, though largely non-significant, pattern matches what was to be expected based on previous results (see the last four rows in the first column of *Figure 19*). That is, as predicted, the present results for nonwords replicated the high density disadvantage in English speech perception (a 3.28ms increase per neighbor) and high density advantage in English speech production (a 0.90ms decrease per neighbor), as well as the high density advantage in Spanish speech perception (a 3.06ms decrease per neighbor) and high density disadvantage in Spanish speech production (a 5.12 increase per neighbor).



P Value for Positive Estimates (increased RTs)

MODEL ESTIMATES



P Value for Negative Estimates (decreased RTs)

	Neighbors	Neighbors.Clustering.	Neighbors.OnsetNeighbors.	Neighbors.OffsetNeighbors.	Clustering	Clustering.Neighbors.	OnsetNeighbors	OnsetNeighbors.Neighbors.	OffsetNeighbors	OffsetNeighbors.Neighbors.
Perception of English Words	0.5456	0.3654	1.181	3.905	131.52	124.9099	0.5034	-1.215	0.02758	-4.516
Production of English Words	1.292	1.5153	0.08189	1.1584	-136.58	-161.9036	2.415	2.29695	1.529	0.1799
Perception of Spanish Words	2.642	2.633	2.3539	9.824	17.17	15.963	3.81	0.4649	2.168	-7.788
Production of Spanish Words	6.29	6.249	12.42	0.7737	31.99	18.724	8.051	-10.21	6.677	5.9022
Perception of English Nonwords	3.277	3.212	3.4655	6.046	-13.91	-6.891	3.49	-0.3571	2.755	-3.399
Production of English Nonwords	-0.8991	-0.6005	0.6843	-2.26	39.63	38.4291	-2.324	-3.1271	-0.6483	1.631
Perception of Spanish Nonwords	-3.055	-1.091	-3.7847	0.1076	59.63	56.817	-2.393	0.8347	-4.553	-4.6347
Production of Spanish Nonwords	5.122	7.253	24.543	5.0445	41.48	60.986	-0.3148	-21.359	4.224	0.1293

Figure 19. Summary of model estimates.
For the column headings, note that the variable after the period was other variable controlled in the model. For example, Neighbors.Clustering and Clustering.Neighbors were the estimates from model 3. Neighbors.Clustering is the estimate for the number of neighbors when controlling for the clustering coefficient.

Effects for the clustering coefficient replicated the high clustering disadvantage reported previously (Chan & Vitevitch, 2009, 2010). However, the effect does not appear to be very robust as only one significant effect was found, for the perception of Spanish nonwords (though others approached significance). Nor does it appear to explain the overall effect of phonological neighborhood density. Finally, there was also an unexpected, though non-significant, high clustering advantage observed for the production of English words.

However, position-specific neighbors do appear to explain the apparent cross-modal and cross-linguistic differences observed for phonological neighborhood density. In particular, onset neighbors aid production and offset neighbors aid perception. Since neighbors in English tend to share onsets, this is a likely explanation for the high density advantage previously observed for English speech production. Similarly, since neighbors in Spanish tend to share offsets, this is a likely explanation for the high density advantage in Spanish speech perception. In particular, the significant effects for onset neighbors are consistent in English and Spanish. Onset neighbors slow down perception of Spanish words and English nonwords and speed up production of Spanish words and nonwords. The significant effects for offset neighbors were less clear: speeding up perception of English nonwords and slowing down production of Spanish words.

The position-specific effects reported here are consistent with the cohort effects observed by Dumay et al. (2012). As reviewed in Chapter 1, they observed that participants were about 30ms slower to perceive words that gained a cohort neighbor. This is consistent with the present findings that onset neighbors are driving the high density disadvantage typically reported for English perception (whereas offset neighbors appear to aid perception). However, the present results are inconsistent with the rhyme effects observed by Dumay et al. (2012). They observed that participants were about 25ms faster to produce words that gained a rhyme neighbor. The present results suggest a disadvantage for producing words with many offset neighbors, however, this was only significant for the production of real words in Spanish.

While the present results for production are inconsistent with those observed by Dumay et al. (2012) in English they are consistent with those observed by Bien et al. (2011) in Dutch. As reviewed in Chapter 1, their results suggest that rhyme neighbors inhibit production and cohort neighbors facilitate production. This parallels the pattern for offset and onset neighbors observed in the present research. I observed facilitation from offset neighbors for the production of Spanish words and nonwords. I also observed inhibition from onset neighbors for the production of English and Spanish words, though this was not significant in English.

The present results add clarity to a contradictory literature on the topic of phonological neighborhood density. One of the challenges of psycholinguistic research is that natural languages are not well suited for experimentation. Any property one might want to study is confounded with other properties. The back and forth regarding the effects of phonological neighborhood density on Spanish production is an excellent example of this problem. As reviewed in Chapter 1, Vitevitch and Stamer (2006) first reported a high density disadvantage for picture naming in Spanish. Baus, Costa, and Carreiras (2008) were able to replicate this effect, not only with the original Spanish stimuli but also with German translations of the stimuli, for which there was no difference in phonological neighborhood density. That is to say, they found a high density disadvantage for German picture naming when the “high” and “low” density words did not actually vary in phonological neighborhood density - they were instead arbitrarily grouped by the density of their Spanish translations. Baus et al. then found the opposite pattern (a high density advantage) for Spanish picture naming using a new set of line drawings, which they argued were better controlled. In response, Vitevitch and Stamer (2009) replicated their effect with English translations of this new set of Spanish stimuli (which, like the German translations, did not actually vary in phonological neighborhood density). They then went on to find a high density advantage using a larger dataset from the international picture naming project. The takeaway from this back and forth is that stimuli selection is critical. However, to date, it has been unclear what possible confounds might be driving the contradictory results. The present results suggest that position-specific neighbors might have been the confounding variable at play.

Further evidence for the difficulty of experimental control in psycholinguistic research comes from Sadat et al. (2013). As described in Chapter 1, their methodology allowed them to simultaneously model a number of correlated variables (e.g., name agreement, age of acquisition, lexical frequency, and neighborhood frequency) in the model in an attempt to control possible confounds. Their results confirmed the presence of a high density disadvantage for picture naming in Spanish. The present results add support for this conclusion using a different approach. Instead of using a mathematical approach to deal with the confounds inherent in natural language, the present experiment inserted new words into each lexicon in a way that was systematically controlled with regard to the variables of interest.

Theoretical Explanations

Phonological neighbors, in general, should be detrimental to performance because of the confusability they introduce during processing. Across multiple areas of cognition, theories posit, and the evidence supports, parallel activation of similar representations during processing (e.g., McClelland, Rumelhart, & PDP Research Group, 1986). For speech processing, the parallel activation of phonological neighbors should reduce the likelihood that the desired word's representation is the one perceived or produced. Even during error-free processing, parallel activation should increase processing time. This is supported by the present results.

One defining characteristic of spoken language is that it unfolds in time. As a result, the parallel activation of phonologically similar representations is dynamic and time-sensitive (Dell, Burger, et al., 1997). When phonologically similar representations are activated in parallel, the non-overlapping portions are most problematic for processing. Overlapping portions should facilitate, or at least not interfere. Therefore, it is logical that the degree of facilitation and competition from these co-activated word candidates should vary as a function of the position-specific location of phonological overlap between neighboring words. Critically, the non-overlapping segments are the ones that increase confusability. Therefore, the location of those non-overlapping segments is critical.

Word onsets are particularly important for perception. For example, both Shortlist (Norris, 1994) and TRACE (McClelland & Elman, 1986) create a list of lexical candidates organized according to the onset segment. Similarly, according to the Distributed Cohort Model (Gaskell & Marslen-Wilson, 1997, 1999, 2002) incoming speech activates all lexical candidates that share the onset segment. Candidates are eliminated if subsequent segments do not match. Experimentally, the visual world paradigm (e.g., Allopenna, Magnuson, & Tanenhaus, 1998) has been used to demonstrate robust cohort competitor effects. These experiments demonstrate the anticipatory nature of speech perception. As the initial phonemes in the word “beaker” unfold, participants are equally likely to look at a picture of a beaker or a picture of a beetle (a cohort neighbor). By comparison, looks to a picture of a speaker (a rhyme neighbor) arose later and were not as likely (a peak fixation probability of 0.1 for rhyme neighbors compared to 0.2 for cohort neighbors in a display of 4 pictures). Given this, it is not surprising that the present results suggest that neighbors with shared onsets are driving the observed high density disadvantage in speech perception. Offset neighbors, by contrast, appear to aid perception, or at least not have the same detrimental effect as onset neighbors.

In perception, the confusability resulting from the activation of phonological neighbors makes selection difficult. Multiple lexical representations at least partially match the phonological input. In production, this kind of phonological selection is not at play. Perception begins broadly, ready for any utterance, whereas production begins narrowly, with a specific message intended to be uttered. While phonological activation is the driving force in perception, semantic activation is the driving force in production. In production, semantic, not phonological, neighbors make processing difficult. This is why semantic errors are common in production. Phonology, by comparison, plays a lesser role due to the nature of the task itself.

Phonological confusability in production is thought to arise from links between lexical and sublexical (phonological) representations. Dell and Gordon (2003) proposed an interactive model to account for the high density advantage observed in English production. Their two-step model involves 1) the activation of the lexical (lemma) representation and 2) the phonological representations associated with it. The links between lexical and phonological representations are interactive such that activation spreads bidirectionally between levels. A target word is first activated semantically, so the bulk of the competition arises from words with semantic rather than phonological similarity. For example, when producing the word "cat", the semantic neighbor "dog" will also be weakly activated. In step two, the phonological representations for /k/, /æ/, and /t/ will be activated (as well as /d/, /ɑ/, and /g/ to a lesser extent). Activation from these phonemes then feeds back into the lexicon, activating the many phonological neighbors of "cat". These in turn, send activation back down to the phonological level, strengthening the activation of the overlapping phonemes. In this way, phonological neighbors boost activation of the shared phonological representations, and this activation helps distinguish the target word from words that are semantically, but not phonologically, related.

The present results support interactive activation models like Dell and Gordon's (2003) which posit facilitation from overlapping segments. Lexical representations of neighboring words will increase activation for the shared phonological segments, thereby facilitating processing of those shared segments. The present research extends these models by providing evidence that the position of this overlap is critical. The naming latency task used in the present research should be most sensitive to onsets, since the measurement is how quickly the production system can get started. Therefore I hypothesized that onset neighbors should facilitate production. The present results supported this.

Additional support for the facilitative effect of shared onsets in production comes from investigations of malapropisms. Malapropisms are word substitution errors in which the substituted word is a real word that is unrelated in meaning but similar in pronunciation to the intended word. Of 397 word substitution errors Fay and Cutler (1977) analyzed, 183 were malapropisms, making this a common production error. Overwhelmingly, the word produced in error was the nearest cohort competitor, i.e. a neighbor that shares onset phonemes, of the intended word, for example “equivalent” and “equivocal”. This suggests that words with shared onsets are highly activated during production. While this activation does sometimes lead to errors, it should also facilitate processing of the shared onset phonemes.

Alternative Accounts

The interactive activation and competition account developed by Chen and Mirman (2012) provides an alternative explanation for when and how neighbors should facilitate or inhibit processing. Though similar to the interactive model proposed by Dell and Gordon (2003) described above, this model also includes inhibitory connections between lexical representations. These connections are weighted by a sigmoid function so that weakly active neighbors exert less influence than strongly activated neighbors. In this account, weak and strong neighbors are defined by the number of semantic features they share in common. In this way, their model accounts for both phonological neighbors (through bidirectional connections between lexical and sublexical representations) and semantic neighbors (through weighted connections between lexical representations). Their simulations demonstrated that the net effect of co-activated representations depends on these semantic connections: weakly activated semantic neighbors lead to a net facilitatory effect whereas strongly activated semantic neighbors lead to a net inhibitory effect. Though the present research did not use semantic similarity as a measurement, others (e.g., Mirman & Magnuson, 2008) have demonstrated that this is a reliable predictor of word recognition performance.

Another possible explanation for the cross-modal differences observed for phonological neighborhood density stems from clarifying the level of processing at which these effects occur. While neighborhood density appears to be an important property of lexical representations, phonotactic probability appears to be an important property of sublexical representations. Phonotactic probability refers to the frequency with which segments and series of segments occur within words of a given language. As such, it is positively correlated with neighborhood density, since words with many phonological neighbors are also those that contain segments that frequently occur in that language. In a series of experiments, Vitevitch and Luce (2005, 1998, 1999) demonstrated that the lexical status of the stimuli was the critical factor determining competition or cooperation. When the items were real words, and could therefore activate lexical representations, words in low density neighborhoods with low phonotactic probability were repeated faster than those in high density neighborhoods with high phonotactic probability. When the items were nonwords, and therefore largely only activated sublexical representations, nonwords in high density neighborhoods with high phonotactic probability were repeated faster than those in low density neighborhoods with low phonotactic probability. The authors concluded that these two correlated properties act at different levels of processing: neighborhood density has a lexical locus with increased density leading to increased competition between word representations whereas phonotactic probability has a sublexical locus with increased probability leading to increased facilitation between sublexical representations. In summary, they argued that what appeared to be a dissociation between perception and production at the lexical level is instead a dissociation between levels of processing.

Given the above reasoning, one would have predicted opposite effects observed for words and nonwords in the present research. Competition at the lexical level should have led to a net high density disadvantage for the production of real words. By contrast, facilitation at the sublexical level should have led to a net high density advantage for the production of nonwords. However, the present results do not support this. Instead, results were largely consistent across words and nonwords. As neighborhood density increased, so did naming latencies. However, naming latencies decreased as the neighbors sharing onsets increased. It is possible that this contradiction is due to participants accepting these nonwords as new lexical items. Indeed this was a goal of the study and the picture association task used was meant to encourage this. The repetition task used by Vitevitch and Luce (2005, 1998, 1999) likely did not encourage participants to think of these nonwords as new lexical items they learned during the experiment. Additionally, the present research did not include measures of phonotactic probability so it is unclear if this would have exerted an effect above and beyond that of overall neighborhood density.

Future Directions

In addition to exploring the above alternative accounts, future research could explore the role of position-specific neighbors using stimuli from previous experiments on phonological neighborhood density. Such experiments would be a critical replication of the present effects. Furthermore, future research should consider the role of phonological stress in moderating the strength of competition and facilitation between neighbors. Unfortunately, the databases used for the present experiment did not include phonological stress in the phonetic transcriptions, so this could not be included as a potential confound. Phonological stress may be of particular importance in languages like Spanish, which include many accented words. For example, the Spanglish nonword /blio/ might interact more strongly with its neighbors “valió” and “olió” than its neighbors “helio” and “lío” if spoken with final accented stress. This is similar to Chen and Mirman's (2012) idea of weakly and strongly activated neighbors, using stress pattern instead of semantic similarity.

It would also be interesting to see if more robust effects from position-specific neighbors emerge with broader definitions of neighbors than those used in the present research. As noted previously, since the purpose of the present research was to test whether position-specific effects were driving overall density effects, I focused on the subset of neighbors that share the onset phoneme and the subset of neighbors that share the offset phoneme. However, a broader definition might be necessary, particularly in perception, since the full word form is unclear when the utterance begins to unfold. Therefore, restricting the analysis to competition/facilitation from only words with all but one overlapping phoneme, means eliminating partially overlapping words that might also be influencing behavior. To this end, Bien et al.'s (2011) approach of including multiple measures of neighbors at various positions (e.g., all the words that share the first phoneme, first two phonemes, etc.) is desirable. Finally, a more time sensitive perceptual task, like those using the visual word paradigm, might be able to tease apart the effects of overlapping and non-overlapping segments in real time, rather than observing their net effects on whole word perception.

Conclusion

Overall, the present research supports the idea that phonological neighbors are bad for processing. Each additional neighbor adds milliseconds to participants' response time. However, not all neighbors are alike. In particular, neighbors sharing the onset phoneme can aid production, and words sharing the offset phoneme can aid perception. Results for clustering suggest a high clustering disadvantage; however this effect did not appear to be driving the overall effects of phonological neighborhood density. These results are most consistent with interactive activation models of speech processing.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language, 38*(4), 419–439. doi:10.1006/jmla.1997.2558

Allport, D. A. (1984). Speech production and comprehension: One lexicon or two. In W. Prinz & A. F. Sanders (Eds.), *Cognition and Motor Processes* (pp. 209–228). Berlin: Springer-Verlag.

Arbesman, S., Strogatz, S. H., & Vitevitch, M. S. (2010a). Comparative analysis of networks of phonologically similar words in English and Spanish. *Entropy*, *12*(3), 327–337. doi:10.3390/e12030327

Arbesman, S., Strogatz, S. H., & Vitevitch, M. S. (2010b). The structure of phonological networks across multiple languages. *International Journal of Bifurcation and Chaos*, *20*(03), 679–685. doi:10.1142/S021812741002596X

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R* (Vol. 2). Cambridge University Press. doi:10.1558/sols.v2i3.471

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. doi:10.1016/j.jml.2007.12.005

Balota, D., & Chumbley, J. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language*, *24*(1), 89–106. doi:10.1016/0749-596X(85)90017-8

Bates, D., Maechler M, Bolker, B., & S, W. (2014). _lme4: Linear mixed-effects models using Eigen and S4_. Retrieved from http://cran.r-project.org/package=lme4

Bates, E., D’Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., … Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*(2), 344–80. doi:10.3758/BF03196494

Baus, C., Costa, A., & Carreiras, M. (2008). Neighbourhood density and frequency effects in speech production: A case for interactivity. *Language and Cognitive Processes*, *23*(6), 866–888. doi:10.1080/01690960801962372

Bien, H., Baayen, R. H., & Levelt, W. J. M. (2011). Frequency effects in the production of Dutch deverbal adjectives and inflected verbs. *Language and Cognitive Processes*, *26*(4-6), 683–715. doi:10.1080/01690965.2010.511475

Boersma, P., & Weenink, D. (2014). Praat: Doing phonetics by computer. Retrieved from http://www.praat.org/

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: a review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, *58*(5), 412–24. doi:10.1027/1618-3169/a000123

Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1430–1450. doi:10.1037//0278-7393.27.6.1430

Carreiras, M., Alvarez, C. J., & De Vega, M. (1993). Syllable frequency and visual word recognition in Spanish. *Journal of Memory and Language*, *32*(6), 766–780. doi:10.1006/jmla.1993.1038

Chan, K. Y., & Vitevitch, M. S. (2009). The influence of the phonological neighborhood clustering coefficient on spoken word recognition. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(6), 1934–49. doi:10.1037/a0016902

Chan, K. Y., & Vitevitch, M. S. (2010). Network structure influences speech production. *Cognitive Science*, *34*(4), 685–97. doi:10.1111/j.1551-6709.2010.01100.x

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review*, *119*(2), 417–30. doi:10.1037/a0027175

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335–359. doi:10.1016/S0022-5371(73)80014-3

Coleman, J. (1998). Cognitive reality and the phonological lexicon: A review. *Journal of Neurolinguistics*, *11*(3), 295–320. doi:10.1016/S0911-6044(97)00014-6

Dahl, D. B. (2014). xtable: Export tables to LaTeX or HTML. Retrieved from http://cran.r-project.org/package=xtable

De Groot, A. M. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(5), 824–845. doi:10.1037//0278-7393.15.5.824

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, *5*(4), 313–349. doi:10.1080/01690969008407066

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: a functional analysis and a model. *Psychological Review*, *104*(1), 123–47. doi:10.1037/0033-295X.104.1.123

Dell, G. S., & Gordon, J. K. (2003). Neighbors in the lexicon: Friends or foes? In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp. 9–37). Berlin: Mouton de Gruyter.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. a. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*(4), 801–38. doi:10.1037/0033-295X.104.4.801

Dumay, N., Damian, M. F., & Bowers, J. S. (2012). The Impact of Neighbour Acquisition on Phonological Retrieval. In *53rd Annual Meeting of the Psychonomic Society*. Minneapolis, MN. Retrieved from http://www.psychonomic.org/past-future-meetings

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, *18*(1), 35–39. doi:10.1111/j.1467-9280.2007.01845.x

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 116–124. doi:10.3758/BF03195503

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*, 3–28.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*(2), 105–132. doi:10.1016/s0010-0277(03)00070-2

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, *12*(5), 613–656. doi:10.1080/016909697386646

Gaskell, M. G., & Marslen-Wilson, W. D. (1999). Ambiguity, competition, and blending in spoken word recognition. *Cognitive Science*, *23*(4), 439–462. doi:10.1016/s0364-0213(99)00011-7

Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology*, *45*(2), 220–266. doi:10.1016/s0010-0285(02)00003-8

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, *28*(5), 501–518. doi:10.1016/0749-596X(89)90009-0

Harley, T., & Bown, H. (1998). What causes a tip-of-the- • tongue state? Evidence for lexical neighbourhood effects in speech production. *British Journal of Psychology*, *89*(1), 151–174. doi:10.1111/j.2044-8295.1998.tb02677.x

Hickok, G. S., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(1-2), 67–99. doi:10.1016/j.cognition.2003.10.011

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. doi:10.1016/j.jml.2007.11.007

Janssen, D. P. (2012). Twice random, once mixed: Applying mixed models to simultaneously analyze random effects of language and participants. *Behavior Research Methods*, *44*(1), 232–47. doi:10.3758/s13428-011-0145-1

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 824–843. doi:10.1037//0278-7393.20.4.824

Keuleers, E., & Brysbaert, M. (2010). Wuggy: a multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–33. doi:10.3758/BRM.42.3.627

Kroll, J. F., & Merves, J. S. (1986). Lexical access for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(1), 92–107. doi:10.1037//0278-7393.12.1.92

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2014). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). Retrieved from http://cran.r-project.org/package=lmerTest

Lawrence, M. A. (2013). ez: Easy analysis and visualization of factorial experiments. Retrieved from http://cran.r-project.org/package=ez

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: when adults learn new words. *Cognitive Psychology*, *55*(4), 306–53. doi:10.1016/j.cogpsych.2007.01.001

Leifeld, P. (2013). {texreg}: Conversion of Statistical Model Output in {R} to {\LaTeX} and {HTML} Tables. *Journal of Statistical Software*, *55*(8), 1–24. Retrieved from http://www.jstatsoft.org/v55/i08/

Levelt, W. J. M., Roelofs, A., & Meyer, a S. (1999). A theory of lexical access in speech production. *The Behavioral and Brain Sciences*, *22*(1), 1–38; discussion 38–75. doi:10.1017/S0140525X99001776

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36. doi:10.1016/0010-0277(85)90021-6

Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics*, *62*(3), 615–25. doi:10.3758/BF03212113

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*(1), 1–36. Retrieved from http://journals.lww.com/ear-hearing/Abstract/1998/02000/Recognizing_Spoken_Words__The_Neighborhood.1.aspx

MacKay, D. G. (1987). *The Organization of Perception and Action: A Theory for Language and Other Cognitive Skills* (p. 233). New York: Springer-Verlag. Retrieved from http://mackay.bol.ucla.edu/publications.html

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PloS One*, *7*(8), e43230. doi:10.1371/journal.pone.0043230

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*(1), 29–63. doi:10.1016/0010-0285(78)90018-X

Martin, N., & Saffran, E. (2002). The relationship of input and output phonological processing: An evaluation of models and evidence to support them. *Aphasiology*, *16*(1/2), 107–150. doi:10.1080/02687040143000447

Mattys, S. L., & Clark, J. (2002). Lexical activity in speech processing: Evidence from pause detection. *Journal of Memory and Language*, *47*, 343–359. doi:10.1016/S0749-596X(02)00037-2

Mattys, S. L., Pleydell-Pearce, C. W., Melhorn, J. F., & Whitecross, S. E. (2005). Detecting silent pauses in speech: A new tool for measuring on-line lexical and semantic processing. *Psychological Science*, *16*(12), 958–64. doi:10.1111/j.1467-9280.2005.01644.x

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. doi:10.1016/0010-0285(86)90015-0

McClelland, J. L., Rumelhart, D. E., & PDP Research Group. (1986). *Parallel distributed processing, volume 1: Explorations in the microstructure of cognition: Foundations*.

Explorations in the microstructure of cognition (Vol. 2). MIT Press. Retrieved from http://mitpress.mit.edu/books/parallel-distributed-processing

Mirman, D., & Magnuson, J. S. (2008). Attractor dynamics and semantic neighborhood density: processing is slowed by near neighbors and speeded by distant neighbors. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*(1), 65–79. doi:10.1037/0278-7393.34.1.65

Monsell, S. (1987). On the relation between lexical input and output pathways for speech. In A. Allport, D. G. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language Perception and Production: Relationships Between Listening, Speaking, Reading, and Writing* (pp. 273–311). London: Academic Press.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189–234. doi:10.1016/0010-0277(94)90043-4

Oldfield, R., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*(4), 273–281. doi:10.1080/17470216508416445

Protopapas, A. (2007). CheckVocal: a program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, *39*(4), 859–62. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18183901

R Core Team. (2014). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from http://www.r-project.org/

Roelofs, A. (2003). Modeling the relation between the production and recognition of spoken word forms. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities.* (pp. 115–158). Berlin: Mouton de Gruyter.

Sadat, J., Martin, C. D., Costa, A., & Alario, F.-X. (2013). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology*, *68*, 33–58. doi:10.1016/j.cogpsych.2013.10.001

Samuel, A. G., & Larraza, S. (n.d.). Does listening to non-native speech impair native speech perception?

Schwartz, J.-L., Basirat, A., Ménard, L., & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, *25*(5), 336–354. doi:10.1016/j.jneuroling.2009.12.004

Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition*, *53*(2), 91–127. doi:10.1016/0010-0277(94)90067-1

Singmann, H., & Bolker, B. (2014). afex: Analysis of Factorial Experiments. Retrieved from http://cran.r-project.org/package=afex

Solomon, R., & Postman, L. (1952). Frequency of usage as a determinant of recognition thresholds for words. *Journal of Experimental Psychology*, *43*, 195–201. doi:10.1037/h0054636

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1140–54. doi:10.1037/0278-7393.21.5.1140

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., … Bates, E. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language*, *51*(2), 247–250. doi:10.1016/j.jml.2004.03.002

Vitevitch, M. S. (1997). The neighborhood characteristics of malapropisms. *Language and Speech*, *40 (Pt 3)*, 211–28. doi:10.1177/002383099704000301

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 735–747. doi:10.1037//0278-7393.28.4.735

Vitevitch, M. S. (2008). What can graph theory tell us about word learning and lexical retrieval? *Journal of Speech, Language and Hearing Research*, *51*, 408–423. doi:10.1044/1092-4388(2008/030)

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, *9*(4). doi:10.1111/1467-9280.00064

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic Phonotactics and Neighborhood Activation in Spoken Word Recognition. *Journal of Memory and Language*, *40*(3), 374–408. doi:10.1006/jmla.1998.2618

Vitevitch, M. S., & Rodríguez, E. (2005). Neighborhood density effects in spoken word recognition in Spanish. *Journal of Multilingual Communication Disorders*, *3*(1), 64–73. doi:10.1080/14769670400027332.Neighborhood

Vitevitch, M. S., & Sommers, M. S. (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory & Cognition*, *31*(4), 491–504. doi:10.3758/BF03196091

Vitevitch, M. S., & Stamer, M. K. (2006). The curious case of competition in Spanish speech production. *Language and Cognitive Processes*, *21*(6), 760–770. doi:10.1080/01690960500287196

Vitevitch, M. S., & Stamer, M. K. (2009). The influence of neighborhood density (and neighborhood frequency) in Spanish speech production: A follow-up report. In *Spoken Language Laboratory Technical Report No. 1*. Retrieved from http://kuscholarworks.ku.edu/dspace/handle/1808/5500

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W. H. A., Lumley, T., … Venables, B. (2014). gplots: Various R programming tools for plotting data. Retrieved from http://cran.r-project.org/package=gplots

Appendix

List of word and nonword stimuli

KEY
More clustering in English than Spanish
More clustering in Spanish than English
More onset than offset neighbors (within language)
More offset than onset neighbors (within language)

English Word	English IPA	English Dif IPA	# Neighbors	Clustering Coefficient	# Onset Neighbors	# Offset Neighbors	Spanish Word	Spanish IPA	Spanish Dif IPA	# Neighbors	Clustering Coefficient	# Onset Neighbors	# Offset Neighbors
cheese	tʃiz	tʃið	25	0.4667	10	21	queso	keso	kefo	18	0.2222	12	17
chest	tʃest	tʃesp	20	0.4789	7	17	pecho	petʃo	ketʃo	11	0.2909	6	10
letter	letɚ	lenɚ	20	0.1737	16	17	carta	karta	karka	15	0.1429	11	14
nest	nest	nesk	21	0.4476	7	17	nido	niðo	niβo	13	0.3077	5	12
pot	pɑt	fɑt	39	0.2551	22	29	olla	oʎa	oɲa	19	0.1988	11	15
tail	teɪl	θeɪl	49	0.3333	22	41	cola	kola	kora	15	0.2286	10	13
thumb	θʌm	θʌn	21	0.5667	5	17	dedo	deðo	geðo	12	0.197	8	11
beard	biəɹɚd	miəɹɚd	9	0.4722	3	6	barba	barβa	barfa	6	0.2667	5	5
girl	gɝl	dɝl	22	0.2944	16	18	niña	niɲa	nima	4	0.1667	2	2
nail	neɪl	zeɪl	48	0.3466	20	40	clavo	klaβo	klafo	8	0.1786	6	4
nurse	nɝs	nɝʃ	13	0.2821	8	10	enfermera	emfermera	emfemera	3	0	3	1
rocket	ɹɑkɪ̈t	ɹɑpɪ̈t	7	0.2857	3	6	cohete	koete	koepe	5	0.1	5	4
shell	ʃel	θel	29	0.399	11	21	concha	kontʃa	tontʃa	3	0	3	2
shirt	ʃɝt	ʃɝp	16	0.45	11	15	camisa	kamisa	kanisa	3	0.3333	3	2
wood	wʊd	wʊg	15	0.3238	11	12	tabla	tabla	kabla	3	0	2	2
boat	boʊt	boʊp	37	0.2718	29	27	barco	barko	bargo	9	0.1667	5	7
bridge	bɹɪdʒ	bɹɪtʃ	9	0.5833	7	2	puente	pwente	pwende	3	0.3333	1	2
fish	fɪʃ	fɪs	16	0.3583	13	4	pez	peθ	keθ	3	0.3333	1	3
hoof	huf	huθ	13	0.359	9	6	pata	pata	pada	27	0.1852	18	23
map	mæp	mæb	35	0.3815	21	16	mapa	mapa	maka	13	0.359	7	12
roof	ɹuf	ɹuʃ	23	0.2885	18	12	techo	tetʃo	θetʃo	9	0.1944	4	8
shovel	ʃʌvl	θʌvl	6	0.3333	6	2	pala	pala	paɲa	21	0.2571	12	19
bench	bentʃ	gentʃ	9	0.2222	7	4	banca	banka	banga	6	0.0667	6	4
bird	bɝd	dɝd	39	0.2578	31	26	pájaro	paxaro	pagaro	3	0	3	1
bomb	bɑm	bɑn	29	0.2734	20	13	bomba	bomba	domba	3	0	2	2
bride	bɹaɪd	bɹaɪg	22	0.2121	16	16	novia	noβja	nofja	4	0	3	2
floor	flɔɹ	fɹɔɹ	19	0.3801	18	7	piso	piso	tiso	19	0.2047	16	14
globe	gloʊb	kloʊb	7	0.2381	6	2	mundo	mundo	munto	4	0	3	3
lion	laɪən	waɪən	7	0.1429	5	5	león	leon	reon	6	0.1333	4	4
mouse	maʊs	maʊʃ	18	0.3791	15	14	ratón	raton	radon	5	0.2	4	3
picture	pɪktʃɚ	tɪktʃɚ	4	0.3333	4	2	cuadro	kwadro	kwabro	3	0	3	1
plate	pleɪt	pleɪk	18	0.3137	15	7	plato	plato	plako	8	0.1071	8	5
present	pɹɛzənt	bɹɛzənt	3	0.3333	3	2	regalo	reɣalo	reɣamo	4	0.1667	4	1
purse	pɝs	pɝʃ	23	0.253	18	14	bolsa	bolsa	bolfa	4	0	4	2
rope	ɹoʊp	ɹoʊk	33	0.2746	25	15	cuerda	kwerda	kwerga	4	0	2	2
wig	wɪg	wɪd	30	0.3494	20	11	peluca	peluka	peluta	3	0.3333	3	2
witch	wɪtʃ	wɪdʒ	28	0.373	17	14	bruja	bruxa	bruga	3	0	3	1
box	bɑks	gɑks	29	0.3695	12	25	caja	kaxa	kaβa	21	0.4238	16	20
glass	glæs	glæz	9	0.1389	6	7	vaso	baso	faso	20	0.1421	15	16
knot	nɑt	nɑp	39	0.2605	23	28	nudo	nuðo	nuβo	11	0.3091	5	10
llama	lɑmə	lɑnə	6	0.3333	2	6	llama	ʎama	ʎaβa	15	0.3429	7	8
turkey	tɝki	tɝgi	7	0.2381	4	5	pavo	paβo	pafo	16	0.3583	11	15
canoe	kənu	kəmu	3	0.3333	1	2	canoa	kanoa	kamoa	4	1	4	4
moose	mus	muʃ	26	0.2862	18	19	alce	alθe	alfe	5	0.3	5	4
wheat	wit	wiθ	33	0.2576	18	24	trigo	triɣo	kriɣo	4	0.3333	4	4
block	blɑk	blɑp	17	0.1765	13	10	cubo	kuβo	kufo	13	0.2051	9	11
crib	kɹɪb	kɹɪd	4	0	3	2	cuna	kuna	kuma	7	0.1905	5	7
monkey	mʌnki	mʌnti	4	0.3333	4	2	chango	tʃaŋgo	tʃaŋko	4	1	1	4
baby	beɪbi	beɪdi	9	0.1944	7	7	bebé	beβe	befe	9	0.25	8	2
cross	kɹɔs	kɹɔʃ	9	0.2222	8	5	cruz	kruθ	gruθ	5	0.9	5	0
dress	dɹɛs	dɹɛʃ	9	0.1389	6	4	vestido	bestiðo	festiðo	3	0.3333	3	0
queen	kwin	kwim	5	0.1	5	3	reina	rejna	rejma	4	0.1667	4	1

IPA	Different IPA	# Neighbors	Clustering Coefficient	# Onset Neighbors	# Offset Neighbors	# Neighbors	Clustering Coefficient	# Onset Neighbors	# Offset Neighbors
		English				Spanish			
ais	ail	7	0.9524	0	7	8	0.2143	4	7
eni	emi	4	0.8333	0	4	3	0.3333	1	2
ili	iθi	13	0.5641	3	11	3	0	1	2
omi	oni	7	0.8571	0	7	7	0.1905	4	5
utu	ulu	5	0.3	0	5	3	0	0	3
θus	θis	10	0.9556	0	10	5	0.6	1	5
ali	afi	10	0.9556	0	10	14	0.3297	12	6
anu	alu	3	0.6667	0	3	3	0.3333	2	1
kisi	kili	5	0.3	4	5	4	0.1667	4	2
olis	opis	2	1	0	2	6	0.2	5	3
puli	puti	15	0.3524	10	13	3	0.3333	3	2
ʧini	ʧiθi	8	0.3929	4	7	5	0.3	4	2
blio	blia	4	1	4	0	4	0.5	1	4
dupe	dume	3	1	3	0	4	0.3333	2	4
gika	dika	3	1	3	0	4	0.5	1	4
sise	bise	3	1	3	0	3	0.3333	1	3
tuta	tuto	4	1	4	0	6	0.1333	4	5
wila	wina	5	0.9	5	0	6	0.2667	3	6
dili	dila	10	0.3778	10	7	3	0.3333	2	1
file	pile	3	1	3	0	8	0.1786	6	5
ite	ile	3	1	3	0	7	0.0476	5	4
lune	lume	5	0.9	5	0	5	0.1	4	3
pite	piʧe	3	1	3	0	9	0.1944	8	7
ʧike	ʧipe	4	1	4	0	9	0.1667	7	6
alus	alis	2	0	0	2	3	0.3333	2	3
fus	fun	20	0.3368	10	16	4	1	0	4
gin	pin	20	0.3947	8	18	3	1	0	3
isi	ini	6	0.1333	2	5	4	0.3333	1	3
jun	jin	24	0.3225	11	21	3	1	0	3
θip	θin	14	0.6154	3	12	3	0.6667	0	3
fumi	fusi	3	0	1	3	5	0.9	5	0
iti	ipi	10	0.2222	4	7	3	0.3333	2	1
muði	musi	3	0.3333	2	3	5	0.9	5	0
ondi	undi	2	0	0	2	3	0.6667	3	0
supi	sumi	9	0.1667	6	7	4	0.3333	4	2
unu	umu	3	0.6667	0	3	5	1	5	0
kjo	sjo	3	0.6667	3	0	3	1	0	3
liʧa	liʧo	2	0	2	0	8	0.3929	6	8
luta	luða	2	0	2	0	6	0.1333	4	5
muso	luso	2	0	2	0	12	0.197	8	11
nila	nilo	2	0	2	0	5	0.3	2	5
pikta	pikto	2	0	2	0	5	0.4	4	5
biʧi	biti	9	0.2222	8	5	3	0.3333	3	2
fili	bili	13	0.3077	13	10	4	0.5	3	1
piki	pini	16	0.2167	12	10	5	0.5	5	1
suti	suki	8	0.1786	6	5	3	0.3333	3	2
suθ	puθ	17	0.4926	14	5	5	0.6	4	1
tui	tua	20	0.3579	14	10	4	0.6667	4	1