

## RESEARCH

# A new test for exemplar theory: Varying versus non-varying words in Spanish

Anne Pycha

University of Wisconsin, Milwaukee, US

[pycha@uwm.edu](mailto:pycha@uwm.edu)

We used a Deese-Roediger-McDermott false memory paradigm to compare Spanish words in which the phonetic realization of /s/ can vary (word-medial positions: *bu[s]to* ~ *bu[h]to* ‘chest’, word-final positions: *remo[s]* ~ *remo[h]* ‘oars’) to words in which it cannot (word-initial positions: *[s]opa* ~ *\*[h]opa* ‘soup’). At study, participants listened to lists of nine words that were phonological neighbors of an unheard critical item (e.g., *popa*, *sepa*, *soja*, etc. for the critical item *sopa*). At test, participants performed free recall and yes/no recognition tasks. Replicating previous work in this paradigm, results showed robust false memory effects: that is, participants were more likely to (falsely) remember a critical item than a random intrusion. When the realization of /s/ was consistent across conditions (Experiment 1), false memory rates for varying versus non-varying words did not significantly differ. However, when the realization of /s/ varied between [s] and [h] in those positions which allow it (Experiment 2), false recognition rates for varying words like *busto* were significantly higher than those for non-varying words like *sopa*. Assuming that higher false memory rates are indicative of greater lexical activation, we interpret these results to support the predictions of exemplar theory, which claims that words with heterogeneous versus homogeneous acoustic realizations should exhibit distinct patterns of activation.

**Keywords:** Spanish; false memory; variation; exemplar theory

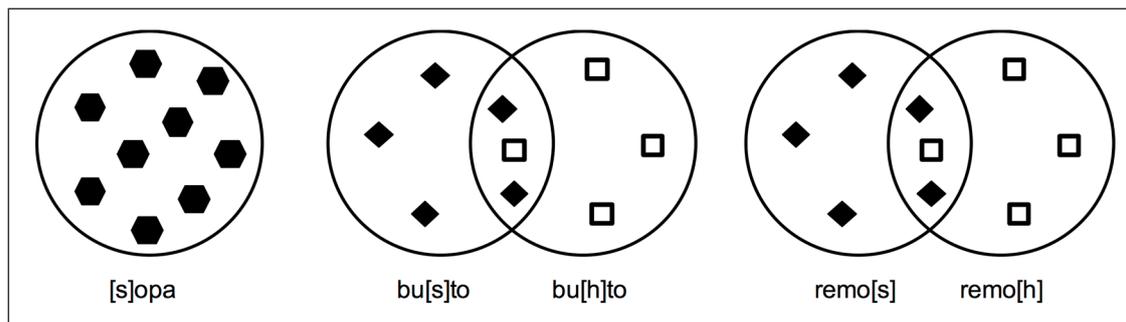
## 1 Introduction

Exemplar theory’s fundamental claim is that listeners store each instance of a word in memory, retaining details such as social context and talker identity (Pierrehumbert 2016). Previous research has provided substantial evidence for this claim, primarily by focusing on voices. For example, several studies have shown that listeners respond to words more quickly and accurately when they are spoken in a familiar voice, compared to an unfamiliar one (Church & Schacter 1994; Nygaard, Sommers & Pisoni 1994; Goldinger 1996; Bradlow, Nygaard & Pisoni 1999). These findings make sense only if listeners retain voice information, and are at odds with traditional, abstract notions of lexical representations in which phonetic details are discarded (e.g., Chomsky & Halle 1968). Given its emphasis on such details, one of the challenges for exemplar theory has been to provide a mechanism by which listeners can nevertheless make generalizations and recognize highly variable speech input as words. In several proposals, this is accomplished with clustering (K. Johnson 1997, 2006; Sumner et al. 2014; Pierrehumbert 2016). The basic idea is that exemplars which are similar to each other, such as those containing the same sequence of segments or those produced by the same talker, cluster together in acoustic space. During word recognition, stored exemplars become activated according to how acoustically similar they are to the speech input. Thus, for example, any stored exemplar will become activated if it is acoustically similar to an input such as [kæt], and will activate even

more strongly if it contains similar voice details. Johnson (1997) conducted simulations to show that this approach can correctly classify speech input according to linguistic categories (such as vowels) as well as social categories (such as talker gender). The clustering mechanism thus predicts that listeners can successfully recognize the input [kæt] as the word *cat*, and that they will do so even better when it is produced by a familiar voice.

The clustering mechanisms of exemplar theory do not merely mimic the capacity for generalization that is inherent to abstractionist theories, but make their own unique predictions. As spelled out in the work of Goldinger (see also Hintzman 1986, 1998), the first prediction is that the activation level for a heard word correlates with the number of matching exemplars that are stored in memory. Thus, the activation level (or “echo intensity”) for a frequent word like *cat* should be high because [kæt] is acoustically similar to many stored exemplars, while the activation level for an infrequent word like *fob* should be low because [fɒb] is similar to only a few exemplars. The second prediction is that the activation content – that is, the identity of exemplars that are activated at any given moment – changes according to whether the matching exemplars are relatively heterogeneous versus homogeneous. Thus, the activation content (or “echo content”) for a frequent word like *cat* should be relatively generic because its exemplars are heterogeneous, having been produced in many different contexts by many different voices; when they are all simultaneously activated, the result is a blend of these contexts and voices. Meanwhile, the activation content for an infrequent word like *fob* should be relatively specific because its exemplars are homogeneous, having been produced in only a few contexts by a few voices; when they are activated, the result is overwhelmingly faithful to these contexts and voices. In a series of shadowing experiments, Goldinger (1998) provided support for these predictions by showing that, for high-frequency words compared to low-frequency words, reaction times were faster (due to greater activation levels), but imitation of target voices was poorer (due to more generic activation content).

In the current study, we broaden the investigation beyond voices, and test the unique predictions of exemplar theory by focusing instead on positional variation (Sumner et al. 2014). Positional variation occurs when speakers can optionally realize a segment in more than one way, but only when that segment occupies a particular position. One example comes from American English, where speakers can reduce voiceless coronal stops to glottalized realizations in word-final position,  $ba[t] \sim ba[ʔt̚] \sim ba[ʔ]$  (Deelman & Connine 2001; Sumner & Samuel 2005). Crucially, speakers do not reduce [t] in word-initial position,  $[t^h]ip \sim *[ʔ]ip$ . Another example, and the focus of the current study, comes from “lowland” varieties of Spanish, where speakers can reduce voiceless alveolar fricatives to glottal fricatives or  $\emptyset$  in word-final position,  $remo[s] \sim remo[h] \sim remo[\emptyset]$  ‘oars’; and also in word-medial position before a consonant  $bu[s]to \sim bu[h]to \sim bu[\emptyset]to$  ‘chest’ (Lipski 1984, 1994). Crucially, in most varieties of Spanish, speakers do not reduce [s] in word-initial position,  $[s]opa \sim *[h]opa \sim *opa$  ‘soap’ (for exceptions, see E. L. Brown & Cacoullos 2003; E. K. Brown & Brown 2012). The key point of interest is that exemplar theories, and in particular their clustering mechanisms, predict a fundamental difference between non-varying words like *sopa* and varying words like *busto* and *remos*. Specifically, the exemplars of *sopa* are relatively homogeneous by virtue of being realized with [s] alone; therefore, they should form a single cluster. Meanwhile, the exemplars of *busto* and *remos* are relatively heterogeneous because they are sometimes realized with [s] and sometimes with [h]; they should therefore form two overlapping but separate clusters, as depicted in Figure 1. (For simplicity, we omit consideration of  $[\emptyset]$  variants, but our overall predictions would be similar if we had included it; for related discussion see File-Muriel & Brown 2011).



**Figure 1:** In the left panel, the word *sopa* forms a single cluster of exemplars realized with [s] (filled hexagons). In the middle panel, the word *busto* forms two clusters, one of exemplars realized with [s] (filled diamonds) and one of exemplars realized with [h] (empty squares). In the right panel, the word *remos* forms two clusters, one of exemplars realized with [s] (filled diamonds) and one of exemplars realized with [h] (empty squares).

Based upon the clustering depicted in Figure 1, exemplar theory makes two specific predictions that we test in the current study. The first prediction concerns situations in which speech input is restricted to [s] variants alone (here, this will be a laboratory situation). Given that the activation level of a word increases directly with the number of previously-stored exemplars (Goldinger 1998), the theory predicts that speech input with [s] should activate a non-varying word like *sopa* more strongly than a varying word like *busto* or *remos*. This is because the input [s]*opa* activates roughly all of the exemplars for *sopa*, since all previously-heard tokens of this word were realized with [s] and, therefore, all exemplars are acoustically similar to [s]*opa*. On the other hand, the input *bu*[s]*to* activates only a fraction of the exemplars for *busto*, since only a fraction of the previously-heard tokens of the word were realized with [s] and, therefore, only a fraction of exemplars are acoustically similar to *bu*[s]*to*. The same argument holds for the input *remo*[s] and the word *remos*. This can be visualized in Figure 1, where [s]*opa* has many exemplars in its cluster, but *bu*[s]*to* and *remo*[s] have relatively few exemplars in their respective clusters.

The second prediction concerns situations in which speech input is unrestricted and contains both [s] and [h] variants (here, this will be a laboratory situation). Note that in such a situation, we no longer expect an advantage for *sopa*, because the number of activated exemplars becomes equivalent across the word types. That is, speech input consisting of both *bu*[s]*to* and *bu*[h]*to* would activate roughly all of the exemplars for *busto* (since these are the two principal variants of the word), just as input consisting of [s]*opa* activates roughly all of the exemplars for *sopa* (since this is the sole principal variant of the word). This can be visualized in Figure 1, where the [s]*opa* cluster contains as many exemplars as the *bu*[s]*to* plus *bu*[h]*to* clusters combined. The same logic applies to *remo*[s] plus *remo*[h]. Nevertheless, we still expect *sopa* to behave differently. Given that the activation content of a word changes according to the relative heterogeneity of its exemplars (Goldinger 1998), exemplar theory predicts that a non-varying word like *sopa* should produce a relatively specific activation pattern because its exemplars are homogenous, while varying words like *busto* and *remos* should produce a relatively generic activation pattern because its exemplars are heterogeneous. Again, this can be visualized in Figure 1, where the single circle representing the [s]*opa* cluster takes up a smaller amount of acoustic space and represents a more specific pattern, compared to the two circles representing the *bu*[s]*to* plus *bu*[h]*to* clusters, which take up a larger amount of acoustic space and collectively represent a blended pattern. The same logic applies to the *remo*[s] plus *remo*[h] clusters. Previous research allows to refine this prediction even further. In an experiment focusing

on voices, input from multiple talkers (generic) produced stronger lexical activation than input from a single talker (specific) (Roediger et al. 2004). This suggests that, when word frequency is held constant, generic activation content contributes to overall higher activation levels. Adapting this finding to phonological variation, we predict that input from multiple variants, as in *bu[s]to* plus *bu[h]to*, should produce stronger lexical activation than input from a single variant, as in *[s]opa*.

The exemplar diagrams in Figure 1 suggest that [h] realization occurs roughly 40% of the time. This is, of course, an over-simplification, because a large number of different factors have been shown to influence speakers' productions of [s] versus [h], including the variety of Spanish in question (Lipski 1984, 1994; Minnette Fox 2006) as well as the segment's position within the word (Terrell 1979; Lafford 1989; E. L. Brown & Cacoullos 2003; E. K. Brown & Brown 2012). For example, rates of [h] realization can be as high as 92% for speakers from Puerto Rico, and as low as 36% for speakers from speakers from Colombia; furthermore, for these and other documented varieties, rates differ for segments in medial compared to final position (E. K. Brown 2009). (Presumably, rates also differ from one word to the next but, to our knowledge, no study of Spanish has published the relevant statistics for individual lexical items; for statistics on variation in English, see Patterson & Connine 2001; Patterson, LoCasto & Connine 2003). Because these factors affect the relative proportion of [s] and [h] exemplars, we would also expect them to modulate the degree of activation difference between *sopa* words, versus *busto* and *remos* words. Crucially, however, we do not expect them to alter our two basic predictions. For any listener of any Spanish variety who has heard almost exclusively [s] in word-initial position, but some combination of [s] and [h] in medial position as well as final position, we predict stronger activation of *sopa* words in [s]-restricted situations, but stronger activation of *busto* and *remos* words in unrestricted situations.

The two predictions that we have laid out are important because they are unique to exemplar theory (Ranbom & Connine 2007; note that ideas related to "echo intensity" are developed in the work of Connine and colleagues, but their framework does not incorporate any notion of "echo content"; Connine, Ranbom & Patterson 2008; Pinnow & Connine 2014), and distinguish it from other theories of variant recognition that retain the notion of a single, abstract representation for each word, such as inference models (Gaskell & Marslen-Wilson 1996, 1998; Pitt 2009; a related idea is investigated in Gow 2001) and underspecification models (Lahiri & Marslen-Wilson 1991; Lahiri & Reetz 2010). Testing these predictions, however, presents a methodological problem. While words like *sopa*, *busto*, and *remos* differ along multiple dimensions (which we discuss more fully in Section 2), the crucial difference for our purposes lies in the position of /s/. Unfortunately, segment position is known to exert a large influence in most word recognition tasks, for reasons unrelated to positional variability. Initial segments in particular play a disproportionately important role in activating representations, compared to medial or final segments (Marslen-Wilson & Welsh 1978; Nooteboom 1981 and many others; e.g., Gaskell & Marslen-Wilson 1997). And, word-initial privilege makes one of the same predictions that exemplar theories make, namely that an input containing [s] should activate words like *sopa*, where [s] is initial, more strongly than words like *busto* or *remos*, where [s] is medial or final. Thus, positional effects obtained in most tasks would confound any claims about differences between varying versus non-varying words.

To address this problem, the current study uses a Deese-Roediger-McDermott (DRM) false memory paradigm (seminal papers include Deese 1959; Roediger & McDermott 1995; for reviews, see Gallo 2006; 2010). In this paradigm, participants see or hear lists of words that are associates or neighbors of a critical item, and subsequently try to remember those

words. The key result is that participants often (falsely) remember the critical item, even though they never heard it. For example, the words *rack*, *pack*, *bake*, *book*, *bag*, *bat*, etc., are all phonological neighbors of the critical item *back*, differing from it by the substitution of one phoneme. After hearing such a list, listeners falsely remembered the unheard word *back* on average 65 to 70% of the time (Sommers & Lewis 1999). Several studies have reported similar results, for both serial recall and yes/no recognition tasks, and established the robustness of false phonological memories (Wallace, Stewart & Malone 1995; Wallace et al. 1995; Schacter, Verfaellie & Anes 1997; Wallace et al. 1998; Wallace et al. 2001; McDermott & Watson 2001; Westbury, Buchanan & Brown 2002; Watson, Balota & Roediger 2003; Amberg, Yamashita & Wallace 2004; Garoff-Eaton, Kensinger & Schacter 2007; Ballardini, Yamashita & Wallace 2008; Ballou & Sommers 2008). In our study, we asked participants to listen to lists of neighbors such as *popa*, *sepa*, *soja*, etc. ('stern', 'I know (SUBJUNCTIVE)', 'soy', respectively; for the critical item *sopa*, *gusto*, *vasto*, *bulto*, etc. ('taste', 'vast', 'bundle'; for the critical item *busto*), and *demos*, *ramos*, *retos*, etc. ('demonstrations', 'branches', 'challenges'; for the critical item *remos*). Our key question was whether rates of false recall and/or recognition differed for critical items such as *sopa* versus *busto* and *remos*.

Two aspects of the DRM false memory paradigm are crucial for our current goal: first, we can interpret false memory rates as reflecting lexical activation of the critical item, and second, both initial and non-initial segments contribute to this activation in an equivalent manner. Both aspects are explained by the concept of converging neighborhood activation (for a different view of false memories, see Kroll et al. 1996; Reinitz 2001). Previous work has demonstrated that when a listener hears speech input such as [popa] 'stern', she activates the representation for the target *popa*, but also the representations for phonological neighbors that differ in the substitution of one phoneme, such as those for *sopa*, *papa*, *poca* ('soap', 'potato', 'little (FEM)') and so on (Goldinger, Luce & Pisoni 1989; Goldinger et al. 1989; Luce & Pisoni 1998; Luce et al. 2000; Stockall, Stringfellow & Marantz 2004; Magnuson et al. 2007). In the DRM false memory paradigm, such activation occurs repeatedly, and eventually converges on the single word which is a neighbor to all of the items on a given list, namely the critical item (*sopa*). It is this converging activation which creates the experience of having heard a word (Collins & Loftus 1975; Sommers & Lewis 1999; Roediger, Balota & Watson 2001).

Additional support for this notion comes from studies demonstrating that reduced amounts of converging activation result in lower rates of false memories. Sommers and Lewis (1999), for example, manipulated the similarity of list words to the critical item. Neighbors like *rack*, *pack*, *bake*, *book*, *bag*, *bat*, etc., are strongly associated with the critical item *back* (because their consonants and vowels are highly confusable with [b], [æ], [k]) and, as noted above, produced high rates of false remembering. By contrast, neighbors like *shack*, *yak*, *ban*, *batch*, *beak*, *bike*, etc. are only weakly associated to *back* (because their consonants and vowels are less confusable with [b], [æ], [k]), and produced significantly lower rates of false remembering. Lowering the amount of converging activation by other methods, e.g., by decreasing the number of heard associates on a list, also results in lower rates of false memories (Wallace et al. 1995; Robinson & Roediger 1997; see also Roediger et al. 2001), showing that we can interpret rates of false recall and/or recognition as reflecting levels of spreading activation for the representation of the critical item.

Furthermore, the critical item in the DRM false memory paradigm receives equivalent activation from initial and non-initial segments. This is because the list of heard words includes neighbors that differ from the critical item by a single segment, and the position of this segment varies. For example, the critical item *sopa* [sopa] receives activation from

spoken *popa* [popa] (initial C substitution), *sepa* [sepa] (medial V substitution), and *soja* [soxa] (medial C substitution). Analogously, the critical item *busto* [busto] receives activation from spoken *gusto* [gusto] (initial C substitution), *vasto* [basto] (medial V substitution), and *bulto* [bulto] (medial C substitution). Because activation occurs indirectly from diverse neighbors, rather than directly from the speech signal, initial segments do not disproportionately affect activation compared to non-initial segments. Indeed, Westbury, Buchanan and Brown (2002) showed that lists of English phonological neighbors with initial CV overlap (*bade, bane, beige, etc.*) and lists with final VC overlap (*make, wake, sake, etc.*) produced equivalent false memory rates for the critical item (*bake*). Thus, a key advantage of the false memory approach is that we can attribute any differences between false memories for *sopa* versus *busto* and *remos* to their non-varying versus varying status, rather than to task-dependent asymmetries in the role played by [s] in initial position.

Although the basic predictions outlined above could be investigated in any number of languages with positional variation, Spanish offers a particularly interesting test case for exemplar theory because of the diversity of factors that have been shown to influence speakers' productions of [s] versus [h]. In addition to those factors we have already mentioned, gender plays a role (Ma & Herasimchuk 1971; Mack 2011), as does the contextual environment for disambiguation (Poplack 1980; Hundley 1987). Since theories with abstract representations, such as inference and underspecification models, do not have the mechanisms in place to incorporate these non-linguistic factors, some authors have already argued for an exemplar-based theory of the lexicon specifically by invoking Spanish production data (Bybee 2000; E. K. Brown 2009). Despite the richness of the production literature, however, only a small handful of studies have investigated Spanish [s] ~ [h] variation in the perception literature (Hammond 1978; Widdison 1995; Figueroa 2000; Boomershine 2006; Mack 2011; Carlson 2012; Schmidt 2013, 2015). One of these studies employs an exemplar framework: Boomershine (2006) showed that variants such as *a[s]tro* ~ *a[h]tro* 'astro' have the longest reaction times in shadowing and lexical decision tasks (compared to other variants such as *cancío[n]* ~ *cancío[ŋ]* 'song', and *po[r]que* ~ *po[l]que* 'why'), and also yielded the highest accuracy rates in a dialect identification task. Boomershine makes a brief statement to the effect that her study supports exemplar theory, although she does not describe in any detail how its mechanisms would predict this particular pattern of results. As far as we are aware, no other perceptual study of Spanish [s] ~ [h] variation investigates exemplar theory. Thus, we chose to focus on Spanish in order to fill a somewhat obvious gap in the literature: given that Spanish [s] ~ [h] variation provides a strong body of evidence in support of exemplar theory within the production domain, we felt it was important to employ Spanish [s] ~ [h] variation as a test of the predictions of exemplar theory within the perception domain.

Experiment 1 tested our first prediction, which is that speech input restricted to the variant [s] should activate words like *sopa* more strongly than words like *busto* or *remos*. To this end, listeners heard lists of words in which the target variant was always realized as [s]. In the list for the critical item *sopa*, for example, participants heard *popa*, [s]epa, [s]oja, etc. In the list for *busto*, participants heard *gu[s]to*, *va[s]to*, *bulto*, etc., and in the list for *remos*, participants heard *demo[s]*, *ramo[s]*, *reto[s]*, etc. To preview, results showed a trend in support of the prediction, although this trend was not significant.

Experiment 2 tested our second prediction, which is that unrestricted speech input which includes both the [s] and [h] variants should produce activation for words like *busto* and *remos* which is greater than the activation that the [s] variant alone produces for words like *sopa*. To this end, Experiment 2 was identical to Experiment 1 except that, for varying critical items, listeners heard lists of words in which the target variant was realized approximately half the time as [s], and half of the time as [h]. For example, in the list for

*busto*, participants heard *gu[s]to*, *va[h]to*, *bulto*, etc., and in the list for *remos*, participants heard *demo[s]*, *ramo[h]*, *reto[s]*, etc. Meanwhile, the list for *sopa* contained only [s] variants, as in Experiment 1. To preview, although results do not support this prediction for words like *remos*, they do support it for words like *busto*.

## 2 Experiment 1: Restricted speech input with [s] variants alone

Experiment 1 compared false recall and recognition rates for critical items with initial, medial, or final /s/. The neighbors for these critical items were all pronounced with [s]. The prediction is that false recall and recognition rates should be higher for critical items with /s/ in initial position, compared to other positions.

### 2.1 Method

As critical items, we selected eighteen Spanish words. As shown in Table 1, six of these words contained /s/ in initial position, six contained /s/ in medial position before a consonant, and six contained /s/ in final position. All critical items were bisyllabic, with at least nine phonological neighbors, and all were either nouns or adjectives.

As will be apparent, while the critical items in each condition crucially differed in the position of /s/, they also differed in syllable structure as well as morphological composition. For syllable structure, the critical items with initial /s/ use CV.CV, those with medial /s/ use (C)VC.CV, and those with final /s/ use (C)V.(C)VC. To a large extent, these differences arise directly from the phonological environments that prohibit variation between [s] and [h] (syllable-initial) versus those that license it (syllable-final). The solution to this problem would have been to impose a maximal CVC.(C)VC structure on all critical items, but this was not feasible. Spanish has a highly restricted inventory of consonants in the syllable coda, and only /s, l, r, d/ and nasals are commonly attested in this position (Hualde 2005: 76). Thus, the number of Spanish words with CVC.(C)VC structure is relatively small to begin with; of those that do exist, many are formed with the addition of plural *-s* (e.g., *sopas* ‘soups’, *bustos* ‘chests’), an unsuitable addition to critical items that already contained /s/ in a different position. In addition, every critical item required a minimum of nine phonological neighbors, and only a handful of CVC.(C)VC words in Spanish meet this criteria. Thus, constraints on the Spanish lexicon precluded the imposition of a uniform syllable structure. For morphological composition, the critical items in the initial and medial /s/ conditions are singular forms, while those in the final /s/ condition are plural forms. Again, this is due to constraints on the lexicon: very few Spanish words end with a non-plural /s/.

Previous work in the DRM false memory paradigm suggests that differences in syllable and morphological structure should not affect our results. Two studies have used critical items with a variety of syllable structures, such as CVC (*dog*, *face*, *fat*, etc.), CCVC (*flag*, *glass*, *sleep*, etc.), and CVCC (*hand*, *cold*, *hard*, etc.) (McDermott & Watson 2001; Watson et al. 2003), and reported false recognition results comparable to those for studies in which the critical items

**Table 1:** Critical items used in the design of Experiments 1 and 2.

Initial /s/		Medial /s/		Final /s/	
Critical item	Gloss	Critical item	Gloss	Critical item	Gloss
<i>sano</i>	‘healthy’	<i>asta</i>	‘flagpole’	<i>amos</i>	‘owners’
<i>seca</i>	‘drought’	<i>busto</i>	‘chest’	<i>aves</i>	‘birds’
<i>silo</i>	‘silo’	<i>casta</i>	‘caste’	<i>días</i>	‘days’
<i>sito</i>	‘located’	<i>pasto</i>	‘pasture’	<i>hoyos</i>	‘holes’
<i>sola</i>	‘alone (FEM)’	<i>pesto</i>	‘pesto’	<i>picas</i>	‘pikes’
<i>sopa</i>	‘soup’	<i>poste</i>	‘post’	<i>remos</i>	‘oars’

were exclusively CVC, such as Sommers and Lewis (1999). A third study used critical items containing both simple onsets as well as onset clusters, and also reported comparable results (Schacter et al. 1997: 337). In the morphological domain, Pycha (2017) compared critical items that were either morphologically simple (*rise, fade*, etc.) versus morphologically complex (*lies, paid*, etc.) and reported no differences between these two conditions. While more research is needed to investigate these issues further, it would appear that neither syllable structure nor morphological structure exert demonstrable effects on false memory rates.

The lexical statistics for the eighteen critical items are displayed in Table 2.

Although our intent was to balance frequency, number of neighbors, and neighborhood frequency across the three conditions, constraints on the Spanish lexicon prevented us from doing so. The language simply does not contain a large number of words that meet the basic criteria for being a critical item (contains /s/, bisyllabic, noun or adjective, at least nine neighbors), and thus did not provide the option of selecting words as a function of their lexical characteristics. In order to control for their influence, we used a composite – the frequency-weighted neighborhood probability rule, or FW-NPR – that reflects all three of these characteristics and included it as a factor in our statistical analysis. Adapted from Luce and Pisoni (1998: 12), we calculated the FW-NPR as (Frequency of target) divided by (Frequency of target) + (Summed frequencies of neighbors). For example, *sopa* has a raw frequency of 25.82 per million words, and therefore a log frequency of 1.43. And, *sopa* has eleven neighbors, whose log frequencies sum to 18.67. Therefore, the FW-NPR for *sopa* is  $(1.43)/(1.43 + 18.67) = 0.07$ . Luce and Pisoni (1998) show that FW-NPR values for American English words correlate closely with the outcomes of word recognition studies, suggesting that this metric serves a valid predictor of the effects of lexical characteristics on activation of word representations.

For each critical item, we constructed a list of nine phonological neighbors that differed by the addition, deletion, or substitution of a single phoneme. Sample lists are displayed in Table 3, and the complete list of stimuli is in the Appendix.

Note that in Spanish orthography, the letters *b* and *v* at the beginning of a word indicate the same sound, namely [b]. Thus, *busto, vasto, and visto* all begin with [b]. The letter *h* is silent, as in *hemos* ['e.mos] (Hualde 2005: 7). Accent marks indicate stress, e.g., *gustó* [gus.'to] has stress on the final syllable. In words with no accent mark and an open final syllable, stress generally falls on the penultimate syllable, e.g., *gusto* ['gus.to] 'he/she/you liked'. In words with no accent mark and a closed final syllable, stress generally falls on the final syllable except in plural words, where stress falls on the same syllable as in the singular form, e.g., *demo* ['de.mo], *demos* ['de.mos] (Hualde 2005: 222–223).

As is apparent in Table 3, all neighbors occurred as isolated words. This allowed us to adhere to the definition of “neighbor” as a word that differs by a single phoneme, which is a crucial feature of the DRM false memory tasks. However, this strategy also presented some

**Table 2:** Lexical statistics (means and standard deviations) for critical items used in the design of Experiments 1 and 2, from the Clearpond database (Marian et al. 2012). Frequencies are reported as  $\log_{10}(\text{frequency}+1)$ . Neighbors is a count of words that differ from the critical item by the addition, deletion, or substitution of one phoneme. Neighbor frequency is the mean frequency of all neighbors of a critical item. FW-NPR is the frequency-weighted neighborhood probability rule, adapted from Luce and Pisoni (Luce & Pisoni 1998: 12); see text.

	Frequency	Neighbors	Neighbor frequency	FW-NPR
Initial /s/	1.14 (0.71)	15.67 (3.01)	1.97 (0.40)	0.037 (0.022)
Medial /s/	0.52 (0.28)	12.67 (2.34)	2.04 (0.62)	0.027 (0.028)
Final /s/	0.99 (0.89)	11.50 (1.64)	2.01 (0.42)	0.041 (0.037)

disadvantages. First, it created an admittedly artificial presentation for the nouns, which would usually be accompanied by a determiner, such as *lo*, *la*, *los*, or *las* ‘the’. Second, it excluded any exemplars with reduced or resyllabified /s/. In connected Spanish speech, /ss/ sequences across word boundaries reduce to [s], and word-final /s/ resyllabifies when the following word begins with a vowel. For example, *los sacos* ‘the bags’ is pronounced [lo.sa.cos], and *pocos amigos* ‘few friends’ is pronounced [po.ko.sa.mi.yos] (Hualde 2005: 87, 97–98). Although researchers have offered evidence that an individual exemplar may indeed contain more than one word (Bybee 2001, 2002), we are not aware of any study that has examined the perceptual consequences for the individual words involved. Thus, while it is possible that a phrase such as *pocos amigos* could form a single exemplar, particularly if it occurs frequently, we simply do not yet know how this impacts the storage of the individual words *pocos* and *amigos* (similarly for phrases such as *los sacos*). Future research may shed light on this issue. In the meantime, we note that our use of isolated speech did help to avoid potential ambiguities in perception. As Hualde (2005: 87) points out, resyllabification of /s/ creates homophony between phrases such as *las alas* [la.sa.las] ‘wings’ and *la salas* [la.sa.las] ‘you salt it’, such that the listener may not always know which form the speaker intended. In the context of the current study, no such ambiguity was possible.

The lexical characteristics for the 162 neighbors are displayed in Table 4.

As with the critical items, our intent was to balance these characteristics across the three conditions, but constraints on the Spanish lexicon prevented us from doing so. Following the reasoning described above, then, we also included the mean FW-NPR for neighbors in the statistical analysis of our results.

To create the fillers, we selected twelve critical items that did not contain /s/, and created a list of nine phonological neighbors for each, yielding a total of 108 words.

## 2.2 Recording

The speaker for the recording was an educated, middle-aged, female native speaker of the Puerto Rican variety of Spanish, living and working in the United States, and unaware of the purpose of the experiment. She recorded the list words as well as the critical items for

**Table 3:** Sample lists of phonological neighbors used in Experiments 1 and 2.

<i>sopa</i>		<i>busto</i>		<i>remos</i>	
Neighbor	Gloss	Neighbor	Gloss	Neighbor	Gloss
<i>copa</i>	‘cup’	<i>bastó</i>	‘it was enough’	<i>demos</i>	‘demonstrations’
<i>popa</i>	‘stern’	<i>bulto</i>	‘bundle’	<i>hemos</i>	‘we have’
<i>ropa</i>	‘clothing’	<i>busco</i>	‘I look’	<i>ramos</i>	‘branches’
<i>sepa</i>	‘I know’	<i>gusto</i>	‘taste’	<i>reímos</i>	‘we laugh’
<i>soda</i>	‘soda’	<i>gustó</i>	‘he/she/you liked’	<i>renos</i>	‘reindeers’
<i>sofá</i>	‘sofa’	<i>justo</i>	‘just’	<i>reos</i>	‘defendants’
<i>soja</i>	‘soybean’	<i>susto</i>	‘scare’	<i>retos</i>	‘challenges’
<i>sopla</i>	‘he/she/you blows’	<i>vasto</i>	‘vast’	<i>rezos</i>	‘prayers’
<i>sosa</i>	‘soda’	<i>visto</i>	‘seen’	<i>vemos</i>	‘we see’

**Table 4:** Lexical statistics (means and standard deviations) for the list words used in Experiments 1 and 2, from the Clearpond database (Marian et al. 2012). See legend for Table 2.

	Frequency	Neighbors	Neighbor frequency	FW-NPR
Initial /s/	1.20 (0.66)	12.31 (5.30)	1.71 (0.58)	0.075 (0.034)
Medial /s/	1.22 (0.83)	11.63 (7.32)	1.72 (0.61)	0.089 (0.037)
Final /s/	1.23 (0.89)	9.56 (3.30)	1.71 (0.61)	0.086 (0.029)

Experiments 1 and 2 during the same session. Importantly, she was able to produce both [s] and [h] variants of the list words with ease, because both occur in her native variety of Spanish (Ma & Herasimchuk 1971; Poplack 1980; E. K. Brown 2009). She recorded the words in random order, in a sound-proof booth using a head-mounted microphone, producing the [s] variants during the first part of the session, then the [h] variants. The audio recording was digitized at a sampling rate of 44.1 kHz, and segmented into individual files using the Praat program (Boersma & Weenink 2014).

### 2.3 Procedure

The thirty lists of phonological neighbors (eighteen target lists plus twelve filler lists) were divided into two sets of fifteen (A and B), each containing three lists from the initial /s/ condition, three from the medial /s/ condition, three from the final /s/ condition, and six from the fillers. Each participant was randomly assigned to set A or set B.

During the experiment, participants were seated in a quiet setting in front of a computer equipped with a mouse, keyboard, and high-quality headphones. Spanish-speaking research assistants (one who spoke a Mexican variety without [s] ~ [h] variation, and another who spoke a Colombian variety without [s] ~ [h] variation) greeted the participants and explained the procedure in Spanish. Printed instructions on the computer screen, also in Spanish, guided participants through each step.

In the study phase, participants listened to fifteen lists of nine spoken words. Each word on a list was played individually, followed by 1 second of silence before the onset of the next word. After each list, participants did a free recall task, in which they were given 45 seconds to type as many words as they could remember from the list, in any order. After 45 seconds, they proceeded to the next list. The overall order of the fifteen lists, as well as the order of the nine words within each list, was randomized for each participant.

In the test phase, after listening to all fifteen lists, participants did a recognition task in which they listened to an individual spoken word, and made a yes/no judgment as to whether they had heard the word previously in the experiment. There were 92 items in the recognition task, which included 45 “old” words that the participant had actually heard (three from each of the fifteen studied lists), plus 47 “new” words that the participant had not heard. The unheard words included the fifteen critical items from the participant’s own set. In addition, the unheard words included 32 foils, which were created by using words from the opposite set (that is, for participants assigned to set A, the foils came from set B, and vice-versa). The foils consisted of eight critical items from the other set (one from each of eight unheard lists, which included two list from initial /s/ condition, two from medial /s/ condition, two from final /s/ condition, and two fillers), and twenty-four neighbor words from other sets (three from each of eight unheard lists, again with equal numbers from each condition). As with the heard lists, critical items in the recognition task were pronounced with [s], e.g., [s]opa, bu[s]to, remo[s]. The order of items in the recognition task was randomized for each participant.

### 2.4 Participants

#### 2.4.1 Setting: Multiple varieties of Spanish in contact

Experiments 1 and 2 both took place on a university campus in a small city in the mid-western region of the United States. This setting placed constraints on our participant pool, because only a limited number of Spanish speakers were available to complete the experimental tasks in our on-campus laboratory. As such, it was not feasible to limit recruitment to speakers from a single region of the world. On the other hand, this setting also offered an advantage, because the Spanish-speaking population in the U.S. is notably diverse, representing over twenty different countries of origin. While the majority

of Spanish speakers in the U.S. have origins in Mexico, where [s] ~ [h] variation is not typical, significant minorities of speakers have origins in Puerto Rico (about 5 million in U.S.) and Cuba (about 2 million in U.S.), where [s] ~ [h] variation is the norm. Given this diversity, it is reasonable to suppose that many, if not most, Spanish speakers living in the U.S. have substantial exposure to [s] ~ [h] variation, even if they do not speak such a variety themselves.

Of course, Spanish speakers living in the U.S. also have exposure to American English. It seems unlikely, however, that this fact affects our basic predictions, which are based upon the premise that similar-sounding exemplars cluster together. Spanish and English have different lexicons, so exemplars from these two languages are not typically similar-sounding (e.g., Spanish [remos] sounds totally different from English [o<sup>w</sup>.ɪz]). A possible exception concerns cognate words like Spanish *sopa* [sopa] ‘soup’ and English *soup* [sup] but even in these cases, segmental differences are present, and these would seem to mitigate against storing such exemplars in the same cluster.

To address the diversity of Spanish varieties within our participant pool, we formulated a questionnaire that asked each participant to indicate their age, gender, country of origin (or if they were born in the U.S., the country of their parents’ origin), and number of years in the U.S. The questionnaire also provided simple examples of [s] ~ [h] variation, and asked participants to answer a multiple-choice question indicating whether they a) spoke a variety of Spanish with that type of variation, b) could easily understand such a variety, even if they did not speak it themselves, or c) had difficulty understanding such varieties. The questionnaire was written in Spanish, and participants completed it after the experimental tasks. During data analysis, questionnaire data allowed us to consider the extent to which country of origin affected our results.

#### 2.4.2 Participants in Experiment 1

Forty-nine native speakers of Spanish served as participants in Experiment 1. A large majority (38) had Mexican origins (sixteen of these had grown up in Mexico, and twenty-two had grown up in the United States with parents of Mexican origin). In addition, two participants were from Colombia, two were from the United States with parents of Puerto Rican origin, one was from Chile, one from Panama, and one from Spain. Four participants did not state their country of origin.

On the questionnaire, thirty-one of the participants indicated that they could easily understand varieties of Spanish with [s] ~ [h] variation. Four of them spoke such a variety themselves, while fourteen indicated they had difficulty understanding such varieties. Thirty-four participants were female, and fifteen were male. Their ages ranged from 20 to 55, with a mean of 32 (11.90). They were all living in the United States at the time of the study, and spoke English in addition to Spanish. The experiment took approximately forty-five minutes of each participant’s time, and they received cash compensation.

### 2.5 Results of Experiment 1

#### 2.5.1 Recall results for Experiment 1

The recall task yielded a total of 4,127 responses. Thus, on average, listeners responded with 5.61 items per list ( $= 4,127 / (15 \text{ lists per participant} * 49 \text{ participants})$ ). Once the fillers were removed, there were 2,427 target responses. Participants sometimes typed the same word twice for one list, resulting in 206 duplicates (68 in the initial /s/ condition, 72 in the medial /s/ condition, and 66 in the final /s/ condition), which we removed.

During data analysis, we attended to the fact that Spanish orthography has several instances in which different letter or letter combinations map to the same spoken sound. Both *b* and *v* indicate [b] (when initial) or [β] (when medial), so we treated

these interchangeably and accepted e.g., *cavo* for intended *cabo* ‘cape’, as well as *labo* for intended *lavo* ‘I wash’. Both *ll* and *y* indicate [ʎ] (or alternatively [j], depending upon dialect), so we treated these interchangeably and accepted e.g., *caya* for intended *calla* ‘he/she/you shuts up’, as well as *cullo* for intended *cuyo* ‘whose’. Both *ce* and *s* indicate [s], so we treated these interchangeably and accepted e.g., *caceta* for intended *caseta* ‘stand’, as well as *sesto* for intended *cesto* ‘basket’. Finally, orthographic *h* has no corresponding phonetic realization, so we accepted e.g., *arta* for intended *harta* ‘full’.

Apart from these alternative spellings, there were 12 typos that did not produce a Spanish word (2 in initial /s/, 6 in medial /s/, and 4 in final /s/), which we removed. In addition, there were four responses that did not contain appropriate coding, due to experimenter error. After removal of fillers, duplicates, typos, and errors, 2,205 responses were included in the final analysis.

Following the methodology used in previous studies on false recall (Roediger & McDermott 1995; Sommers & Lewis 1999) we classified a response as “veridical” if it corresponded to a word that actually occurred on the list (e.g., *sopa*, *papa*, *poca*, etc.), “intrusion” if it did not occur on the list (e.g., random intrusions such as *buena* ‘good (FEM)’ *cosa* ‘thing’, etc.), and “critical item” if it corresponded to the critical item (e.g., *sopa*). To calculate veridical proportions, again following previously established methodology, we divided the number of veridical responses per list by nine, which was the number of words that actually occurred on each list. For example, if a participant provided four veridical responses for a list, the proportion of veridical responses would be  $0.44 = 4/9$ . For intrusion proportions, we divided the number of intrusion responses by the total number of responses per list. For example, if a participant provided one intrusion response for a list, plus four veridical responses and one critical item, the proportion of intrusion responses would be  $0.17 = 1/(1 \text{ intrusion} + 4 \text{ veridical} + 1 \text{ critical item})$ . Finally, the critical item proportion was calculated as 0 if the participant did not respond with the critical item, and 1 if they did. The mean proportion of veridical, intrusion, and critical item responses given across the three conditions is shown in Table 5.

To analyze the results, we used a mixed logit model in which we counted each response as a “success” and each possible lack of response as a “failure”, following Jaeger (2008). For example, if a participant provided five veridical responses for a list, we counted five successes plus four failures, where failures indicate the four words that the participant heard but did not recall. If a participant provided two intrusion responses, we counted two successes, and the remaining responses (veridical plus critical item) as failures. If a participant provided the critical item as a response, we counted one success, and zero failures. Although many previous studies on the DRM false memory paradigm analyze results with ANOVA, these traditional test pose documented problems for categorical outcome variables (Jaeger 2008), and also do not allow the inclusion of random effects.

We used the `glmer()` function from the `lme4` package in R, with predictor variables of response type (Veridical vs. Intrusion vs. Critical item) and condition (Initial /s/ vs. Medial /s/ vs. Final /s/), and with two random intercepts: one for the critical item’s FW-NPR and one for the mean FW-NPR of corresponding neighbors on the list. For example, following the calculations described in Section 2.1, the FW-NPR of the critical item

**Table 5:** Proportions of response types (means, standard deviations) in recall task for Experiment 1.

	Veridical	Intrusion	Critical item
Initial /s/	0.44 (0.17)	0.18 (0.21)	0.30 (0.46)
Medial /s/	0.38 (0.15)	0.21 (0.22)	0.23 (0.42)
Final /s/	0.37 (0.16)	0.26 (0.21)	0.24 (0.43)

*busto* is 0.03, and this was used as one random intercept. The nine heard neighbors on the *busto* list (*gusto*, *vasto*, *bulto*, etc.) have a mean FW-NPR of 0.11, and this was used as another random intercept. Models that also included random intercepts for participant failed to converge.

We used treatment coding. “Initial /s/” served as the baseline for position, because previous studies on phonological false memories have focused almost exclusively on critical items of this type (i.e., non-varying words), and our key motivation was to test how “Medial /s/” and “Final /s/” would deviate from this baseline. “Critical item” served as the baseline for response type, because this allowed us to test for a basic false memory effect, in which we expect a higher rate of responses to critical items compared to random intrusions. This baseline also allowed us to make our crucial comparison between Initial /s/ versus Medial and Final /s/ specifically for recall of critical items (rather than for recall of veridical items or intrusions). Results of the model are displayed in Table 6.

The model shows that participants exhibited a significant false memory effect, but that the predicted differences between words with initial /s/ versus those with medial or final /s/, while trending in the expected direction, did not reach significance. The false memory effect is apparent because the odds of a response decreased in the Intrusion condition compared to the Critical Item baseline, by a factor of approximately 0.58 ( $= e^{-0.55}$ ). In other words, participants were much less likely to recall a random intrusion such as *buena* ‘good (FEM)’ compared to a critical item such as *sopa*. Also, as expected, the odds of a response increased in the Veridical condition compared to the Critical Item baseline, by a factor of approximately 1.86 ( $= e^{0.62}$ ). That is, participants were more likely to recall a word they had actually heard, such as *popa*, *sepa*, or *soja*, than a critical item they had not heard. Rates of recall for medial /s/ and final /s/ did not differ significantly from the baseline, which indicates that participants were equally likely to falsely recall critical items such as *sopa*, *busto*, and *remos*. The model also indicated one interaction. Compared to the baseline, the odds of a response increased in the Final /s/ Intrusion condition by a factor of approximately 2.08 ( $= e^{0.73}$ ).

A separate analysis that included only the thirty-eight participants of Mexican origin yielded an identical pattern of results. And, a separate analysis that included only participants who could speak or easily understand [s] ~ [h] varieties also yielded an identical pattern of results. As in the main model, both of these separate analyses revealed decreased odds in the Intrusion condition, increased odds in the Veridical condition, and an interaction in the Final /s/ Intrusion condition. Both analyses showed a trend toward greater false recall in the Initial /s/ condition compared to Medial and Final /s/ conditions, but the trend did not reach significance in either analysis.

**Table 6:** Results of mixed logit regression model for recall task in Experiment 1.

	Estimate	Std. Error	z	Pr(> z )
(Intercept)	-0.85	0.18	-4.72	$2.33 \times 10^{-6}$ *
Condition (Medial /s/)	-0.35	0.27	-1.32	0.19
Condition (Final /s/)	-0.28	0.26	-1.05	0.30
Response (Veridical)	0.62	0.19	3.31	$9.21 \times 10^{-4}$ *
Response (Intrusion)	-0.55	0.20	-2.72	0.01*
Condition (Medial /s/): Response (Veridical)	0.09	0.28	0.32	0.75
Condition (Final /s/): Response (Veridical)	-0.05	0.27	-0.18	0.86
Condition (Medial /s/): Response (Intrusion)	0.54	0.29	1.82	0.07
Condition (Final /s/): Response (Intrusion)	0.73	0.29	2.51	0.01*

**2.5.2 Recognition results for Experiment 1**

The recognition task yielded 4,805 responses for analysis (= 92 items \* 49 participants). As with the recall task, “veridical” items actually occurred on a list while “critical items” were the unheard words used to construct the lists. “Intrusion” items did not occur on any list that the participant heard; differently from the recall task, these occurred as foils drawn from the experimental set that the participant was not assigned to, as described in Section 2.3. To calculate veridical proportions, we divided the number of “yes” responses by the total number of items of this type. For example, participants gave judgments to three words from each of the fifteen lists that they heard previously at study; if they responded “yes” to two of these, their veridical proportion for this list was 0.67 (= 2/3). Participants also gave judgments to four intrusions from each of eight lists that they did not hear; if they responded “yes” to one of these, their intrusion proportion was 0.25 (= 1/4). Finally, participants gave judgments to one critical item from each of fifteen lists that they heard previously at study; if they responded “yes” to it, their critical item proportion was 1 (= 1/1). The mean proportion of “yes” responses given to veridical words, intrusions, and critical items is shown in Table 7.

We analyzed the results using a mixed logit model similar to that used for the recall task. We counted each “yes” response as a success and each “no” response as a failure. Unfortunately, models with a full interaction between item type (Veridical vs. Intrusion vs. Critical item) and condition (Initial /s/ vs. Medial /s/ vs. Final /s/) failed to converge. Therefore, we ran an analysis on a subset of the data that included critical item responses only (excluding veridical responses and intrusions), with a random intercept for critical item FW-NPR. (The list FW-NPR, which calculates an aggregate reflecting a list of words as a whole, is not relevant for the recognition task, in which participants are cued with individual spoken words rather than being asked to perform free recall from an entire list). We used treatment coding such that “Initial /s/” served as the baseline for position. Table 8 displays the results of the model.

Results show that rates of recognition for medial /s/ and final /s/ did not differ significantly from the baseline, which indicates that participants were equally likely to falsely recall critical items such as *sopa*, *busto*, and *remos*.

A separate analysis that included only the thirty-eight participants of Mexican origin yielded an identical pattern of results. And, a separate analysis that included only participants who could speak or easily understand [s] ~ [h] varieties also yielded an identical pattern of results. As in the main model, rates for recognition for medial /s/ and final /s/ did not differ significantly from the baseline.

**Table 7:** Proportion of “yes” responses (means, standard deviations) in recognition task for Experiment 1.

	Veridical	Intrusion	Critical item
Initial /s/	0.68 (0.47)	0.26 (0.44)	0.72 (0.45)
Medial /s/	0.67 (0.47)	0.35 (0.48)	0.73 (0.44)
Final /s/	0.64 (0.48)	0.25 (0.44)	0.72 (0.45)

**Table 8:** Results of mixed logit regression model, on critical items only, for recognition task in Experiment 1.

	Estimate	Std. Error	z	Pr(> z )
(Intercept)	0.98	0.29	3.42	0.00*
Condition (Medial /s/)	0.09	0.40	0.22	0.82
Condition (Final /s/)	0.00	0.40	0.01	0.99

### 2.5.3 Summary and discussion for Experiment 1

Experiment 1 tested the exemplar-based hypothesis that spoken words restricted to [s] variants should activate representations for words like *sopa* more strongly than representations for words like *busto* or *remos*. Results support the hypothesis only insofar as the data exhibits a trend in the appropriate direction for the recall task. These data show that participants falsely recalled words with initial /s/, such as *sopa*, approximately 30% of the time. But they falsely recalled words with medial and final /s/, such as *busto* and *remos*, approximately 23% or 24% of the time. These differences are in the predicted direction, but our analysis indicated they were not significant. Also, the data exhibited no trends in the appropriate direction, significant or otherwise, for the recognition task.

While the lack of significance could mean that exemplar theory makes a fundamentally incorrect prediction, it could also be due to limitations of the current study. In particular, while the majority of participants (31 out of 49) indicated that they could understand varieties of Spanish in which non-initial [s] varies with [h], they did not speak such a variety themselves. Previous research has shown that participants can rapidly adapt to new dialects during listening tasks (e.g., Dahan, Drucker & Scarborough 2008; Sumner & Samuel 2009), but the premise of the current study was different in the sense that we assumed a set of stored exemplars affected by a lifetime of listening experience. It is possible that, in their lifetimes, our participants simply had not heard enough exemplars containing [h] to produce significant representational differences between words like *sopa* versus *busto* and *remos*. Schmidt (2015) has shown that Spanish-speaking participants who use only [s], but who have many personal contacts using both [s] and [h], perform differently on consonant identification tasks than those participants who have fewer such contacts, suggesting that it would be worthwhile to collect more detailed data on social contacts, and on [s] versus [h] usage, from participants in future studies.

## 3 Experiment 2: Speech input with [s] and [h] variants

Like Experiment 1, Experiment 2 compared false recall and recognition rates for critical items with initial, medial, or final /s/. Unlike Experiment 2, neighbors for the critical items with medial or final /s/ were pronounced sometimes with [s], and sometimes with [h]. The prediction is that speech input with [s] and [h] variants should activate varying words (those with medial or final /s/, such as *busto* and *remos*) to a greater degree than speech input with the [s] variant alone can activate non-varying words (those with initial /s/, such as *sopa*).

### 3.1 Method

The stimuli and procedure were identical to Experiment 1, as was the speaker who produced the recording. The only difference was that, in each list of words in the medial and final /s/ conditions, approximately half of the words containing /s/ were realized with [s], and half with [h]. For example, the nine neighbors on the list for critical item *busto* were realized as *ba[h]tó*, *bulto*, *bu[s]co*, *gu[h]to*, *gu[s]tó*, *ju[h]to*, *su[s]to*, *va[s]to*, *vi[h]to*. The nine neighbors for *remos* were realized as *demo[s]*, *hemo[h]*, *ramo[s]*, *reímo[h]*, *reno[s]*, *reo[h]*, *reto[s]*, *rezo[h]*, *vemo[s]*. Variants were assigned randomly by alphabetizing the neighbors, and then assigning [s] or [h] in alternating order; words without target /s/, such as *bulto*, were skipped. The lists of words in the initial condition were all realized with [s], as in Experiment 1. For the recognition task, targets were presented with the same variant that the participant had heard during study, while critical items were presented with [s].

Note that it is possible that our recording procedure created a somewhat artificial listening context for the participants of Experiment 2, who heard the same speaker alternate

between words with [s] and words with [h] in a balanced fashion, a situation that is unlikely to occur in everyday circumstances. While using multiple speakers (e.g., one who pronounced [s] variants only, and another who pronounced [h] variants only) might have created a more natural-seeming context, it would have confounded the interpretation of the results, because participants would conceivably respond in different ways to the different voices – indeed, the research on voice information cited in Section 1 strongly suggests that such a scenario would occur. Because we were interested in isolating the effects of variation on word activation, it was important not to confound these with the effects of voices, and so we used a single speaker to present the stimuli.

### 3.2 Participants

Experiment 2 took place at the same university campus as Experiment 1. Thus, while it was subject to the same constraints on participant recruitment described in Section 2.4.1, we continued to operate under the assumption that the United States represents a contact situation in which many, and possibly most, Spanish speakers have substantial exposure to [s] ~ [h] variation.

Forty-seven native speakers of Spanish, none of whom participated in Experiment 1, participated in the study. They came from diverse backgrounds, although as we discuss in Section 3.4, their results patterned uniformly. Eleven participants were from Spain, nine from Colombia, and eight from Mexico (seven who grew up there, and one who grew up in the United States with parents of Mexican origin). In addition, five participants were from Puerto Rico (four who grew up there, and one who grew up in the United States with parents of Puerto Rican origin), three from Dominican Republic, three from Venezuela, two from Guatemala, one from Peru, and one from Honduras. Four participants did not state their country of origin.

The participants completed the same questionnaire used in Experiment 1. On it, twenty-nine of the participants indicated that they could easily understand varieties of Spanish with [s] ~ [h] variation. Eleven of them spoke such a variety themselves, while seven indicated they had difficulty understanding such varieties. Twenty-nine participants were female, and eighteen were male. Their ages ranged from 18 to 63, with a mean of 28 (8.93). They were all living in the United States at the time of the study, and spoke English in addition to Spanish. The experiment took approximately forty-five minutes of each participant's time, and they received cash compensation.

### 3.3 Results for Experiment 2

#### 3.3.1 Recall results for Experiment 2

The recall task yielded a total of 4,349 responses. Thus, on average, listeners responded with 6.17 items per list ( $= 4,349 / (15 \text{ lists per participant} * 47 \text{ participants})$ ). Once the fillers were removed, there were 2,506 target responses. Duplicate responses were removed (127 in the initial /s/ condition, 129 in the medial /s/ condition, and 101 in the final /s/ condition), and so were typos that did not produce a Spanish word (27 in initial /s/ condition, 26 in medial /s/ condition, 15 in final /s/ condition), leaving a total of 2,081 responses for the final analysis. We coded responses and calculated proportions in the same manner as described for Experiment 1. The mean proportion of veridical, intrusion, and critical item responses given across the three conditions is shown in Table 9.

To analyze the results, we followed the same procedure described for the recall task in Experiment 1; namely, a mixed logit model with predictor variables of response type (Veridical vs. Intrusion vs. Critical item) and condition (Initial /s/ vs. Medial /s/ vs. Final /s/), and random intercepts for the critical item's FW-NPR and for the mean FW-NPR of corresponding neighbors on the list. Models that also included random intercepts for

participant failed to converge. As before, we used treatment coding with “Initial /s/” and “Critical Item” as baselines. The results of this model are in Table 10.

The model shows that participants exhibited a significant false memory effect, because the odds of a response decreased in the Intrusion condition compared to the Critical Item baseline, by a factor of approximately 0.56 ( $= e^{-0.58}$ ). Also, as expected, the odds of a response increased in the Veridical condition compared to the Critical Item baseline, by a factor of approximately 2.51 ( $= e^{0.92}$ ). The model indicates no significant difference in false memory rates for critical items with medial /s/ compared to the baseline, which suggests that participants were equally likely to falsely recall critical items such as *sopa* and *busto*. The model does indicate a significant difference in false memory rates for critical items with final /s/ compared to the baseline, but in a direction contrary to our predictions. The odds of a response decreased in the Final /s/ condition compared to the Initial /s/ baseline, by a factor of approximately 0.54 ( $= e^{-0.62}$ ), which shows that participants were less likely to falsely recall a critical item such as *remos* compared to *sopa*. The model also indicated one interaction. Compared to the baseline, the odds of a response increased in the Final /s/ Intrusion condition by a factor of approximately 7.61 ( $= e^{2.03}$ ).

### 3.3.2 Recognition results for Experiment 2

The recognition task yielded 4,324 responses for analysis (= 92 items \* 47 participants). The mean proportion of “yes” responses given to veridical words, intrusions, and critical items is shown in Table 11.

As in Experiment 1, we analyzed the results using a mixed logit model with predictor variables of item type (Veridical vs. Intrusion vs. Critical item) and condition (Initial /s/

**Table 9:** Proportions of response types (means, standard deviations) in recall task for Experiment 2.

	Veridical	Intrusion	Critical item
Initial /s/	0.49 (0.18)	0.16 (0.18)	0.28 (0.45)
Medial /s/	0.41 (0.15)	0.18 (0.18)	0.30 (0.46)
Final /s/	0.30 (0.17)	0.46 (0.27)	0.18 (0.38)

**Table 10:** Results of mixed logit regression model for recall task in Experiment 2.

	Estimate	Std. Error	z	Pr(> z )
(Intercept)	-0.96	0.19	-5.11	0.00*
Condition (Medial /s/)	0.07	0.26	0.26	0.79
Condition (Final /s/)	-0.62	0.29	-2.13	0.03*
Response (Veridical)	0.92	0.20	4.66	$3.14 \times 10^{-6}$ *
Response (Intrusion)	-0.58	0.21	-2.77	0.01*
Condition (Medial /s/): Response (Veridical)	-0.40	0.28	-1.43	0.15
Condition (Final /s/): Response (Veridical)	-0.17	0.30	-0.56	0.58
Condition (Medial /s/): Response (Intrusion)	0.04	0.30	0.15	0.88
Condition (Final /s/): Response (Intrusion)	2.03	0.32	6.42	$1.37 \times 10^{-10}$ *

**Table 11:** Proportion of “yes” responses (means, standard deviations) in recognition task for Experiment 2.

	Veridical	Intrusion	Critical item
Initial /s/	0.73 (0.44)	0.27 (0.45)	0.59 (0.49)
Medial /s/	0.70 (0.46)	0.37 (0.48)	0.75 (0.44)
Final /s/	0.68 (0.47)	0.17 (0.38)	0.56 (0.50)

vs. Medial /s/ vs. Final /s/), with a random intercept critical item FW-NPR. Models with participant as a random factor failed to converge. We used treatment coding with “Initial /s/” and “Critical Item” as baselines. The results of this model are in Table 12.

The model shows that participants exhibited a significant false memory effect, because the odds of a response decreased in the Intrusion condition compared to the Critical Item baseline, by a factor of approximately 0.33 ( $= e^{-1.12}$ ). Also, as expected, the odds of a response increased in the Veridical condition compared to the Critical Item baseline, by a factor of approximately 2.89 ( $= e^{1.06}$ ).

Consistent with our predictions, the model indicates a significant increase in false memory rates for critical items with medial /s/ compared to the baseline. The odds of a response increased in the Medial /s/ condition compared to the baseline, by a factor of approximately 2.32 ( $= e^{0.84}$ ), which shows that participants were more likely to falsely recall critical items such as *busto* compared to *sopa*. The model does not, however, reveal such a finding for critical items with final /s/. The model also indicates one interaction. Compared to the baseline, the odds of a response decreased in the Medial /s/ Veridical condition by a factor of approximately 0.36 ( $= e^{-1.03}$ ).

### 3.4 Summary and discussion for Experiment 2

Experiment 2 tested the hypothesis that speech input containing [s] and [h] will activate representations for words like *busto* and *remos* more than input containing only [s] will activate representations for words like *sopa*. Results support the hypothesis, but only for words with medial /s/ in the recognition task. Thus, our key finding is that rates of false recognition were significantly greater for critical items like *busto* compared to the baseline *sopa*.

Although a potential caveat comes from the diversity of our participant group, which we noted in Section 3.2, our result was consistent across different sub-groups of participants. This is shown in Table 13, which displays rates of false recognition for critical items (veridical words and intrusions are omitted), broken down by participants’ level of

**Table 12:** Results of mixed logit regression model for recognition task in Experiment 2.

	Estimate	Std. Error	z	Pr(> z )
(Intercept)	0.11	0.26	0.44	0.66
Condition (Medial /s/)	0.84	0.37	2.29	0.02*
Condition (Final /s/)	0.17	0.36	0.47	0.64
Response (Veridical)	1.06	0.29	3.64	$2.77 \times 10^{-4}$ *
Response (Intrusion)	-1.12	0.26	-4.27	$1.95 \times 10^{-4}$ *
Condition (Medial /s/): Response (Veridical)	-1.03	0.41	-2.50	0.01*
Condition (Final /s/): Response (Veridical)	-0.58	0.41	-1.40	0.16
Condition (Medial /s/): Response (Intrusion)	-0.33	0.38	-0.87	0.39
Condition (Final /s/): Response (Intrusion)	-0.69	0.39	-1.78	0.07

**Table 13:** Proportion of “yes” responses to critical items in the recognition task for Experiment 2. “Speak” refers to participants who speak a variety of Spanish with variation between [s] and [h], “Understand” refers to those who don’t speak such a variety but understand it with ease, and “Difficulty” refers to those who don’t speak such a variety and have difficulty understanding it.

	Speak	Understand	Difficulty	Mean
Initial /s/	0.52	0.63	0.50	0.59
Medial /s/	0.76	0.71	0.83	0.75
Final /s/	0.61	0.55	0.44	0.56

familiarity with [s] ~ [h] variation. Regardless of whether participants spoke a variety of Spanish with this variation, could understand it with ease, or could understand it only with difficulty, the pattern is the same: namely, higher rates of false recognition for critical items with medial /s/, such as *busto*, compared to those with initial /s/, such as *sopa*.

Additional support comes from Table 14, which displays rates of false recognition broken down by participants’ countries of origin. With just a couple of exceptions, regardless of where participants originated, the pattern is the same: again, higher rates of false recognition for critical items with medial /s/, such as *busto*, compared to those with initial /s/, such as *sopa*.

The upshot is that a diverse participant group yielded a relatively consistent pattern of results, strengthening our conclusion that speech input containing [s] and [h] will activate representations for words like *busto* more than input containing [s] will activate representations for words like *sopa*.

This conclusion raises two further questions. First, why do we see the effect only in the medial /s/ condition, and not final /s/? Second, why do we see the effect only in recognition, not recall? Both questions have potentially straightforward answers, which we turn to in the following sections.

### 3.4.1 Medial versus final conditions

The different patterns for the medial versus final /s/ conditions probably have their origins in perceptual factors; specifically, in the inaccurate perception of [h] variants. We can see this by examining the recall data in Table 9, which shows that, for the final /s/ condition, rates of veridical recall (0.30) were lower than the intrusion rate (0.46). In other words, participants were more likely to write down a random intrusion than to accurately remember a word that they had actually heard on the list. This is a highly aberrant pattern which suggests that our participants did not accurately perceive some of the list items to begin with – in particular, they may not have perceived intended [h] in final position.

To determine if this may have been the case, we counted the number of times that participants gave a recall response which omitted the expected orthographic *s* corresponding to spoken [h]. In the medial /s/ condition, there were 22 such instances (e.g., participants provided the response *gata* instead of the expected *gasta* ‘he/she/you spends’) and in the final /s/ condition, there were 212 such instances (e.g., they provided the response *rezo* instead of the expected *rezos* ‘prayers’).

It is unlikely that the large number of misperceptions in final position is due to excessive reduction (i.e., to Ø) in our stimuli, because our speaker consistently produced audible correlates of [h]. Figure 2 shows a typical pronunciation, where frication noise associated with [h] is clearly present.

Despite the presence of [h] in the acoustic signal, it is nevertheless not entirely surprising that our participants sometimes misperceived it. Previous experiments have examined

**Table 14:** Proportion of “yes” responses to critical items in the recognition task for Experiment 2. Abbreviations indicate participants’ origins: CO = Colombia, DO = Dominican Republic, GT = Guatemala, HN = Honduras, MX = Mexico, PE = Peru, PR = Puerto Rico, ES = Spain, VE = Venezuela, Unspecified = participant did not state origin.

	CO	DO	GT	HN	MX	PE	PR	ES	VE	Unspecified	Mean
Initial /s/	0.67	0.67	1.00	1.00	0.52	0.33	0.25	0.64	0.22	0.67	0.59
Medial /s/	0.67	0.89	0.67	1.00	0.71	1.00	0.67	0.67	0.78	0.89	0.75
Final /s/	0.52	0.78	1.00	0.67	0.52	0.67	0.83	0.33	0.44	0.67	0.56

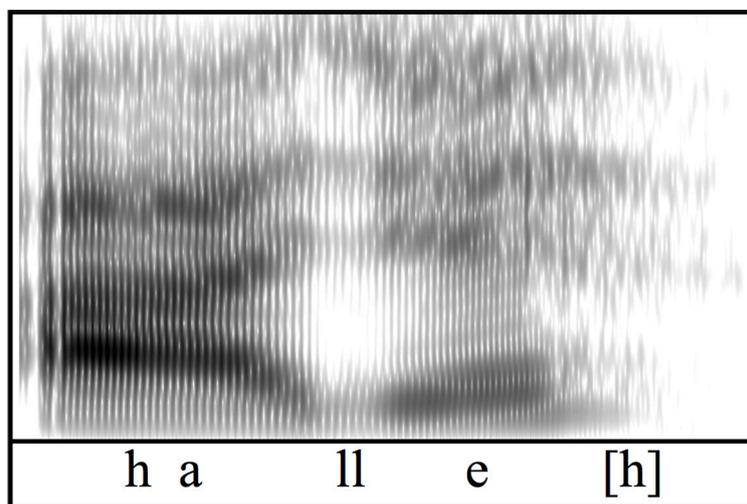
whether Spanish listeners can distinguish between minimal pairs such as *pe[h]cado* ‘fish’ vs. *pecado* ‘sin’ or *pueblo[h]* ‘towns’ vs. *pueblo* ‘town’. Results showed that they can, but only when target [h] occurred word-medially; when target [h] occurred word-finally, listeners’ responses were at chance (the same pattern occurs across three different varieties: for Miami Cuban, Hammond 1978; for Puerto Rican, Figueroa 2000; for Andalusian, Carlson 2012). Thus, even when [h] is present, listeners do not always detect it.

In our experiment, if participants did not perceive list words as intended, this would have an obvious impact on false memories, which, as described in Section 1, arise from converging activation produced by phonological neighbors. That is, a false memory for *remos* depends crucially upon the accurate perception and activation of words like *demos*, *ramos*, *retos*, etc. To further pursue this idea, we performed an alternative coding of the recall data, in which the omitted-s responses such as *gata* and *reto* were counted as “Veridical” instead of “Intrusion”. The resulting proportions are displayed in Table 15.

Importantly, the aberrant pattern is no longer evident: in the final condition, rates of veridical recall (0.45) are now higher than the intrusion rate (0.16), as expected. This strongly suggests that the low rate of false memories for final /s/ words – in both recall and recognition tasks for Experiment 2 – does not arise from a fundamental difference in lexical representations between words like *busto* and *remos*, but rather from difficulties in perceiving the list words. Future work could address this issue, potentially by using cross-splicing to insure robust and uniform realization of [h] in word-final positions.

### 3.4.2 Recognition vs. Recall

We turn now to the second question, namely, why do we see the predicted pattern of higher false memories for *busto* only in recognition, not recall? One possible answer is a test effect: participants completed the recall task first, which may have affected their subsequent performance on the recognition task. Indeed, many previous studies have reported



**Figure 2:** Spectrogram for *halle[h]* ‘you find’.

**Table 15:** Proportions of response types (means, standard deviations) in recall task for Experiment 2, using an alternative coding of responses (see text).

	Veridical	Intrusion	Critical Item
Initial /s/	0.49 (0.18)	0.14 (0.17)	0.28 (0.45)
Medial /s/	0.43 (0.16)	0.12 (0.16)	0.29 (0.45)
Final /s/	0.45 (0.18)	0.16 (0.19)	0.28 (0.45)

significantly higher rates of false recognition for critical items when a recognition task occurs immediately after recall, compared to when it occurs immediately after an unrelated task (such as math problems), (summarized in Gallo 2006: 148–151). In the current study, then, the presence of a prior recall task may have increased false recognition rates somewhat, such that differences between conditions which were not detectable in recall became observable in recognition.

Yet this answer is not completely satisfactory, particularly in light of the many previous studies which report patterns of false recall and recognition that are highly similar to one another (see Gallo 2006: 23–30). Most tellingly, Roediger and colleagues ran a regression analysis on data collected for 55 DRM lists of semantic associates and concluded that “[...] within the limits of this study, the factors responsible for false recognition across lists seem to be the same as those producing false recall” (Roediger et al. 2001: 392). This is clearly not the case for Experiment 2 of the current study, and so future work will need to more fully explore the differences between recall versus recognition tasks, and link these differences to varying versus non-varying lists (one possible starting point is Roediger et al. 2004). Successful recall, for example, takes place in the absence of any cue, and therefore requires a process of explicit recollection. Successful recognition, on the other hand, takes place in the presence of an explicit cue, and may therefore require only a feeling of implicit familiarity (see Yonelinas 2002). What characteristics of the current stimuli tap into this distinction, and are these characteristics unique to lists of words that contain phonological variants? Answering this question will not only reveal more about the difference in lexical representations between varying words like *busto* versus non-varying words like *sopa*, but also help us to reach bigger-picture conclusions about when and why people form false memories.

#### 4 Discussion

The current study investigated two predictions of exemplar theory, which is unique among models of variant word recognition because it proposes that listeners store multiple instances of heard words, rather than a single abstract representation. As outlined in Section 1, a great deal of previous work in exemplar theory has focused on voices. Yet the predictions that the theory makes for words with multiple variants are particularly interesting in their own right, because such words will – by definition – be stored differently in the minds of listeners than words with only a single variant. We focused specifically on [s] ~ [h] variation in Spanish, which is an important pattern because it is influenced by a variety of well-documented factors that have already justified an exemplar approach to production. One goal of the current study was to bolster this approach for perception as well.

Our results provide partial support for our predictions. In Experiment 1, where all heard words were pronounced with [s], the predicted difference between varying and non-varying words was evident only in a non-significant trend in the recall task. That is, although varying words like *busto* and *remos* showed somewhat lower rates of false recall – reflecting lower levels of lexical activation – compared to non-varying words like *sopa*, this difference was not significant. In Experiment 2, where varying words were pronounced with either [s] or [h] but non-varying words were pronounced only with [s], the predicted difference was evident in a significant effect in the recognition task. That is, varying words like *busto* showed significantly higher rates of false recognition – reflecting higher levels of lexical activation – compared to non-varying words like *sopa*. As discussed in Section 3.3.3, the lack of effect for varying words like *remos* is probably due to unrelated perceptual issues.

An important issue for future research is to understand why we obtained significant results for Experiment 2, but not Experiment 1. One possibility is the participant groups,

which differed across the experiments in terms of country of origin and familiarity with [s] ~ [h] variation. Another possibility concerns the nature of the task itself. Experiment 2 made explicit reference to [s] ~ [h] variation (by presenting variants such as *gu[s]to* and *va[h]to* within a single list), while Experiment 1 did not, and it is possible that this difference affected participants' responding strategies in the memory tasks. In general, the trend in the predicted direction in Experiment 1, at least for recall, suggests that in future work with increased number of participants, a significant result could emerge.

#### 4.1 Predictions based on varieties of Spanish and word position

The predictions that we outlined in Section 1 relied upon a somewhat crude distinction between non-varying versus varying words. We worked with the assumption that, for "lowlands" varieties of Spanish, words like *sopa* are realized with [s] "all of the time", whereas words like *busto* and *remos* are realized with [s] "some of the time." Although this crude distinction was sufficient for testing our basic hypothesis, and also appropriate for experiments with participants living in a Spanish-variety contact situation in the United States, the literature on production of Spanish [s] ~ [h] provides sufficient data to make more refined hypotheses that could, in the future, delve more deeply into the predictions of exemplar theory for perception.

One such distinction arises from considering the varieties of Spanish spoken in different parts of the world. As we mentioned in Section 1, even among those varieties that clearly exhibit [s] ~ [h] variation, there are large differences in rates of [h] realization. Table 16, for example, shows the proportion of times that words like *busto* and *remos* were realized with [h] in four different varieties.

With numbers like these, exemplar theory makes some clear predictions. Broadly speaking, for example, differences in activation between varying and non-varying words should be smallest for listeners from Cali, where words like *sopa* are realized with [s] about 100% of the time, but words like *busto* and *remos* are realized with [s] 64% of the time. The difference should be greatest for listeners from San Juan, where words like *sopa* are realized with [s] about 100% of the time, but words like *busto* and *remos* are realized with [s] 8% of the time. These predictions could be tested by, for example, employing the current stimuli and experimental paradigm with homogenous participant groups from Cali versus San Juan. (Note that we would need to adjust these predictions for certain Spanish varieties, such as those found in New Mexico and central Colombia, which are somewhat exceptional in exhibiting [s] ~ [h] variation in initial position.) (Lipski 1994; E. L. Brown & Cacoulios 2003).

Another distinction arises from considering the differences between words like *busto*, where /s/ is word-medial, versus *remos*, where /s/ is word-final. As we mentioned in Section 1, even within a given variety of Spanish that exhibits [s] ~ [h] variation, there are sometimes large differences in rates of [h] realization between these two types of words, as shown in Table 17.

Again, with this type of data, exemplar theory makes some clear predictions that distinguish it from abstractionist theories. For listeners from Cali, for example, the difference

**Table 16:** Proportion of words with /s/ that were realized with [h] in spoken corpora from four different varieties of Spanish (adapted from E. K. Brown 2009: 20).

Variety	% realized as [s]	% realized as [h]
Cali, Colombia	64%	36%
Northern New Mexico, United States	50%	50%
Merida, Venezuela	11%	89%
San Juan, Puerto Rico	8%	92%

**Table 17:** Proportion of words with /s/ that were realized with [h] in spoken corpora from four different varieties of Spanish (adapted from E. K. Brown 2009).

Variety	Position	% realized as [s]	% realized as [h]
Cali, Colombia	Medial /s/	87%	13%
	Final /s/	56%	44%
Northern New Mexico, United States	Medial /s/	74%	26%
	Final /s/	41%	59%
Merida, Venezuela	Medial /s/	10%	90%
	Final /s/	12%	88%
San Juan, Puerto Rico	Medial /s/	10%	90%
	Final /s/	7%	93%

between words with initial versus final /s/ should be relatively large, because words like *sopa* are realized with [s] 100% of the time, while words like *remos* are realized with [s] 56% of the time. By comparison, the difference between words with initial versus medial /s/ should be relatively small, because words like *busto* are realized with [s] 87% of the time. Again, these predictions could be tested by, for example, employing a homogenous participant group from Cali. (As before, we would need to adjust these predictions for the New Mexico variety, but also for the Merida and San Juan varieties, where the differences between word positions are much smaller.) The possibility of such investigations highlights the richness of exemplar theory and the opportunities that it affords for making connections between production and perception patterns.

#### 4.2 Exemplar theories versus other theories of variant word recognition

A fundamental tenet of exemplar theory is that surface phonetic variation does not reduce to a single lexical representation, but is instead stored in the lexicon as multiple episodes, which cluster together according to their acoustic similarity. As such, the theory predicts fundamental differences between words that have multiple surface variants versus those that have only one, and can explain the differences between *sopa* and *busto* that we reported in Experiment 2. It is important to see that other theories of variant recognition cannot account for such data.

In inference models, listeners essentially “undo” phonological rules that produce surface forms such as [buhto], and access a representation containing an underlying form such as /busto/ (Gaskell & Marslen-Wilson 1996, 1998; Pitt 2009). In other words, upon hearing [buhto], the listener undoes the rule stating that /s/ becomes [h] at the end of syllables. With speech input containing [s], then, this model makes no prediction for a difference in activation levels among words like *sopa*, *busto*, and *remos*, because the word is already realized with its underlying form and there is no rule to undo. With speech input containing [h], the model does make a prediction, namely that activation levels for words like *busto* and *remos* should be lower than for words like *sopa*, because the presence of [h] requires the undoing of a rule and therefore exacts a cost in processing. But as we have seen, the data from Experiment 2 indicate the opposite finding.

In underspecification models, listeners map speech inputs onto representations that are not fully specified for all features (Lahiri & Marslen-Wilson 1991; Lahiri & Reetz 2010). Thus, a surface input [busto] or [buhto] would activate an underlying form /buSto/, where /S/ lacks a specification for place of articulation and is therefore compatible with either [s] or [h]. Underspecification models predict no difference between words like *sopa*, *busto* and *remos*, whether they are activated by input with [s] or [h]. This is because in both cases, the speech input maps directly to /S/. The data from Experiment 2 do not pattern in this manner, and underspecification models would be hard-pressed to explain why activation levels for *busto* should exceed those of *sopa*.

“Hybrid” models, such as those suggested by Connine and colleagues, are somewhat different because they adopt certain elements of exemplar theory (Rambom & Connine 2007; Connine et al. 2008; Pinnow & Connine 2014). The basic idea is that all variant forms have their own representations, which are abstract but crucially accompanied by frequency information. Such a model could make some predictions in a manner similar to exemplar theory, because a word like *sopa* is “frequently” realized with the [s] variant, while words like *busto* and *remos* are “infrequently” realized with the [s] variant. However, because the hybrid model rejects the idea that all individual exemplars are stored in memory, it has no mechanism to create blended representations and therefore no way to account for the generic pattern for which we found evidence in Experiment 2.

#### **4.3 Phonological variation and the DRM false memory paradigm**

Most word recognition studies activate lexical representations directly, by presenting listeners with matching information in the speech input: that is, they activate a word like *sopa* by presenting [sopa]. The current study obviously differs from this approach, because we activated representations indirectly, by presenting listeners with phonological neighbors: that is, we activated *sopa* by presenting [popa], [sepa], [soxa], etc. Our use of the DRM false memory paradigm has several implications, which we discuss here.

To begin, we did not ultimately evaluate the idea that speech input with [s] directly activates *sopa* more than *busto* and *remos*, which would have required us to present inputs like [sopa], [busto], and [remos] and measure the resulting activation. As discussed in the Section 1, the disproportionate role that word-initial segments (here, [s]) play in lexical activation ruled out such a paradigm. Instead, we employed indirect activation to investigate a more general idea, namely that comparable information in the speech input – here, in the form of converging phonological neighbors – activates varying words less strongly than it activates non-varying words. A resulting limitation is that we cannot make complete comparisons between our findings and those that use direct activation methods. Another limitation is that we must assume a transitive relationship whereby weaker (or stronger) activation for a target also produces weaker (or stronger) activation for its neighbors, although several previous studies provide evidence for such a relationship (Underwood 1965; Hall & Kozloff 1970; Arndt & Hirshman 1998; Benjamin 2001; Zeelenberg, Plomp, & Raaijmakers 2003; Kawasaki & Yama 2006).

Furthermore, the DRM false memory paradigm differs from more common word recognition paradigms because it includes a monitoring component. That is, even under the assumption that lexical activation for unheard critical items does occur, the recall and recognition tasks still require participants to decide whether that activation originated from an event that actually happened (i.e., from a word that they actually heard) or not. In other words, they must “monitor” the source of their memories before providing a response (M. K. Johnson & Raye 1981; M. K. Johnson, Hashtroudi & Lindsay 1993; Roediger & McDermott 1995; M. K. Johnson 2006). Previous work has demonstrated that manipulating monitoring conditions can alter false memory rates (Israel & Schacter 1997; Schacter, Israel & Racine 1999; Dodson & Schacter 2001), so we must remain open to the idea that differences in monitoring, rather than in activation, could account for our results. Given that no obvious monitoring manipulation occurred in the current study, however, this scenario seems unlikely.

Finally, although we have interpreted our results in terms of activation levels, they also have more general implications for false remembering. It is not unusual for people to experience a memory for an event that did not occur, and false memories have been documented in a range of different contexts (e.g., Loftus 2005; Loftus & Bernstein 2005). False memories triggered the DRM paradigm are particularly robust and highly replicable

(Gallo 2006, 2010); they occur, for example, even when the list of neighbors is small (Robinson & Roediger 1997), or when (orthographic) neighbors are presented for very brief durations (Seamon, Luo & Gallo 1998). Given this, the results of the current study are notable for identifying an instance in which the false memory effect is not uniform. If these results eventually generalize, we would have the basis to argue that varying versus non-varying words are crucially stratified according to their relative susceptibility to memory distortions.

### Acknowledgements

For their crucial assistance with conducting this study, I thank Faviola Valadez, Laura Mendoza, and Nancy Bird-Soto. For their valuable comments on previous versions, I thank Sonia Barnes, Eleni Pinnow, two anonymous reviewers and associate editor Anne-Michelle Tessier, as well as audiences from the Chicago Linguistics Society, Penn Linguistics Conference, and the University at Buffalo. Flaws are mine.

### Competing Interests

The author has no competing interests to declare.

### References

- Amberg, Martha, Jill Yamashita & William Wallace. 2004. Implicit word activation during pre-recognition processing influences correct recognition and estimates of presentation frequency. *Memory* 12(2). 129–139. DOI: <https://doi.org/10.1080/09658210244000298>
- Arndt, Jason & Elliot Hirshman. 1998. True and false recognition in MINERVA2: Explanations from a global matching perspective. *Journal of Memory and Language* 39(3). 371–391. DOI: <https://doi.org/10.1006/jmla.1998.2581>
- Ballardini, Nicole, Jill A. Yamashita & William P. Wallace. 2008. Presentation duration and false recall for semantic and phonological associates. *Consciousness and Cognition* 17(1). 64–71. DOI: <https://doi.org/10.1016/j.concog.2007.01.008>
- Ballou, McKenzie R. & Mitchell S. Sommers. 2008. Similar phenomena, different mechanisms: Semantic and phonological false memories are produced by independent mechanisms. *Memory & Cognition* 36(8). 1450–1459. DOI: <https://doi.org/10.3758/MC.36.8.1450>
- Benjamin, Aaron S. 2001. On the dual effects of repetition on false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27(4). 941–947. DOI: <https://doi.org/10.1037/0278-7393.27.4.941>
- Boersma, Paul & David Weenink. 2014. Praat: Doing phonetics by computer. (Version 5.4.04). Retrieved from: [www.praat.org](http://www.praat.org).
- Boomershine, Amanda. 2006. Perceiving and processing dialectal variation in Spanish: An exemplar theory approach. In *Selected proceedings of the 8th Hispanic Linguistics Symposium*, 58–72. Somerville, MA: Cascadilla Press.
- Bradlow, Ann R., Lynne C. Nygaard & David B. Pisoni. 1999. Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics* 61(2). 206–219. DOI: <https://doi.org/10.3758/BF03206883>
- Brown, Earl K. 2009. *A usage-based account of syllable-and word-final /s/ reduction in four dialects of Spanish*. Munich: LINCOM Studies in Romance Linguistics.
- Brown, Earl K. & Esther Brown. 2012. Syllable-final and syllable-initial /s/ reduction in Cali, Colombia: One variable or two? In Richard J. File-Muriel & Rafael Orozco (eds.), *Colombian varieties of Spanish*, 89–106. Madrid: Iboamericana.
- Brown, Esther L. & Rena Torres Cacoullos. 2003. Spanish /s/: A different story from beginning (initial) to end (final). In Rafel Núñez-Cedeño, Luis López & Richard Cameron

- (eds.), *A Romance perspective on language knowledge and use: Selected papers from the 31st Linguistic Symposium on Romance Languages (LRS�)*. Philadelphia, PA: John Benjamins Publishing.
- Bybee, Joan. 2000. Lexicalization of sound change and alternating environments. *Papers in Laboratory Phonology V: Acquisition and the Lexicon* 5. 250–268.
- Bybee, Joan. 2001. Frequency effects in French liaison. In Joan Bybee & Paul J. Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*, 327–359. Philadelphia, PA: John Benjamins Publishing.
- Bybee, Joan. 2002. Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition* 24(2). 215–221. DOI: <https://doi.org/10.1017/S0272263102002061>
- Carlson, Kristin M. 2012. An acoustic and perceptual analysis of compensatory processes in vowels preceding deleted post-nuclear /s/ in Andalusian Spanish. *Concentric: Studies in Linguistics* 38(1). 39–67. DOI: <https://doi.org/10.6241/concentric.ling>
- Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.
- Church, Barbara A. & Daniel L. Schacter. 1994. Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20(3). 521–533. DOI: <https://doi.org/10.1037/0278-7393.20.3.521>
- Collins, Allan M. & Elizabeth F. Loftus. 1975. A spreading-activation theory of semantic processing. *Psychological Review* 82(6). 407–428. DOI: <https://doi.org/10.1037/0033-295X.82.6.407>
- Connine, Cynthia M., Larissa J. Ranbom & David J. Patterson. 2008. Processing variant forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics* 70(3). 403–411. DOI: <https://doi.org/10.3758/PP.70.3.403>
- Dahan, Delphine, Sarah J. Drucker & Rebecca A. Scarborough. 2008. Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition* 108(3). 710–718. DOI: <https://doi.org/10.1016/j.cognition.2008.06.003>
- Deelman, Thomas & Cynthia M. Connine. 2001. Missing information in spoken word recognition: Nonreleased stop consonants. *Journal of Experimental Psychology: Human Perception and Performance* 27(3). 656–663. DOI: <https://doi.org/10.1037/0096-1523.27.3.656>
- Deese, James. 1959. On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology* 58(1). 17–22. DOI: <https://doi.org/10.1037/h0046671>
- Dodson, Chad S. & Daniel L. Schacter. 2001. “If I had said it I would have remembered it”: Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review* 8(1). 155–161. DOI: <https://doi.org/10.3758/BF03196152>
- Figueroa, Neysa. 2000. An acoustic and perceptual study of vowels preceding deleted post-nuclear /s/ in Puerto Rican Spanish. In Héctor Campos, Elena Herburger, Alfonso Morales-Font & Thomas J. Walsh (eds.), *Hispanic linguistics at the turn of the millennium: papers from the 3rd Hispanic linguistics symposium*, 66–79. Somerville, MA: Cascadilla Press.
- File-Muriel, Richard J. & Earl K. Brown. 2011. The gradient nature of s-lenition in Caleño Spanish. *Language Variation and Change* 23(2). 223–243. DOI: <https://doi.org/10.1017/S0954394511000056>
- Gallo, David A. 2006. *Associative illusions of memory: False memory research in DRM and related tasks*. New York, NY: Psychology Press.
- Gallo, David A. 2010. False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition* 38(7). 833–848. DOI: <https://doi.org/10.3758/MC.38.7.833>

- Garoff-Eaton, Rachel J., Elizabeth A. Kensinger & Daniel L. Schacter. 2007. The neural correlates of conceptual and perceptual false recognition. *Learning & Memory* 14(10). 684–692. DOI: <https://doi.org/10.1101/lm.695707>
- Gaskell, M. Gareth & William D. Marslen-Wilson. 1996. Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 22(1). 144–158. DOI: <https://doi.org/10.1037/0096-1523.22.1.144>
- Gaskell, M. Gareth & William D. Marslen-Wilson. 1997. Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes* 12(5–6). 613–656. DOI: <https://doi.org/10.1080/016909697386646>
- Gaskell, M. Gareth & William D. Marslen-Wilson. 1998. Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 24(2). 380–396. DOI: <https://doi.org/10.1037/0096-1523.24.2.380>
- Goldinger, Stephen D. 1996. Auditory lexical decision. *Language and Cognitive Processes* 11(6). 559–568. DOI: <https://doi.org/10.1080/016909696386944>
- Goldinger, Stephen D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105(2). 251–279. DOI: <https://doi.org/10.1037/0033-295X.105.2.251>
- Goldinger, Stephen D., Paul A. Luce & David B. Pisoni. 1989. Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language* 28(5). 501–518. DOI: [https://doi.org/10.1016/0749-596X\(89\)90009-0](https://doi.org/10.1016/0749-596X(89)90009-0)
- Gow, David W. 2001. Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language* 45(1). 133–159. DOI: <https://doi.org/10.1006/jmla.2000.2764>
- Hall, James W. & Edward E. Kozloff. 1970. False recognitions as a function of number of presentations. *The American Journal of Psychology* 83(2). 272–279. DOI: <https://doi.org/10.2307/1421331>
- Hammond, Robert M. 1978. An experimental verification of the phonemic status of open and closed vowels in Spanish. In Humberto López Morales (ed.), *Corrientes actuales en la dialectología del caribe hispánico: Actas de un simposio*, 93–143. Puerto Rico: Editorial Universitaria, Universidad de Puerto Rico.
- Hintzman, Douglas L. 1986. Schema abstraction in a multiple-trace memory model. *Psychological Review* 93(4). 411–428. DOI: <https://doi.org/10.1037/0033-295X.93.4.411>
- Hualde, José Ignacio. 2005. *The sounds of Spanish*. New York: Cambridge University Press.
- Hundley, James E. 1987. Functional constraints on plural marker deletion in Peruvian Spanish. *Hispania* 70(4). 891–894. DOI: <https://doi.org/10.2307/342562>
- Israel, Lana & Daniel L. Schacter. 1997. Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review* 4(4). 577–581. DOI: <https://doi.org/10.3758/BF03214352>
- Jaeger, T. Florian. 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: <https://doi.org/10.1016/j.jml.2007.11.007>
- Johnson, Keith. 1997. Speech perception without speaker normalization: An exemplar model. In Keith Johnson & John W. Mullenix (eds.), *Talker variability in speech processing*, 145–165. Cambridge, MA: Academic Press.
- Johnson, Keith. 2006. Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics* 34(4). 485–499. DOI: <https://doi.org/10.1016/j.wocn.2005.08.004>
- Johnson, Marcia K. 2006. Memory and reality. *American Psychologist* 61(8). 760–776. DOI: <https://doi.org/10.1037/0003-066X.61.8.760>
- Johnson, Marcia K. & Carol L. Raye. 1981. Reality monitoring. *Psychological Review* 88(1). 67. DOI: <https://doi.org/10.1037/0033-295X.88.1.67>

- Johnson, Marcia K., Shahin Hashtroudi & D. Stephen Lindsay. 1993. Source monitoring. *Psychological Bulletin* 114(1). 3. DOI: <https://doi.org/10.1037/0033-2909.114.1.3>
- Kawasaki, Yayoi & Hiroshi Yama. 2006. The difference between implicit and explicit associative processes at study in creating false memory in the DRM paradigm. *Memory* 14(1). 68–78. DOI: <https://doi.org/10.1080/09658210444000520>
- Kroll, Neal E. A., Robert T. Knight, Janet Metcalfe, Elizabeth S. Wolf & Endel Tulving. 1996. Cohesion failure as a source of memory illusions. *Journal of Memory and Language* 35(2). 176–196. DOI: <https://doi.org/10.1006/jmla.1996.0010>
- Lafford, Barbara. 1989. Is functionalism a fact? *Hispanic Linguistics* 3(1–2). 49–74.
- Lahiri, Aditi & Henning Reetz. 2010. Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics* 38(1). 44–59. DOI: <https://doi.org/10.1016/j.wocn.2010.01.002>
- Lahiri, Aditi & William D. Marslen-Wilson. 1991. The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition* 38(3). 245–294. DOI: [https://doi.org/10.1016/0010-0277\(91\)90008-R](https://doi.org/10.1016/0010-0277(91)90008-R)
- Lipski, John M. 1984. On the weakening of /s/ in Latin American Spanish. *Zeitschrift Für Dialektologie Und Linguistik*, 31–43.
- Lipski, John M. 1994. *Latin American Spanish*. London: Longman.
- Loftus, Elizabeth F. 2005. Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory* 12(4). 361–366. DOI: <https://doi.org/10.1101/lm.94705>
- Loftus, Elizabeth F. & Daniel M. Bernstein. 2005. Rich false memories: The royal road to success. In Alice F. Healy (ed.), *Experimental cognitive psychology and its applications*, 101–113. Washington, D.C.: American Psychological Association Press.
- Luce, Paul A. & David B. Pisoni. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19(1). 1–36. DOI: <https://doi.org/10.1097/00003446-199802000-00001>
- Luce, Paul A., Stephen D. Goldinger, Edward T. Auer & Michael S. Vitevitch. 2000. Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics* 62(3). 615–625. DOI: <https://doi.org/10.3758/BF03212113>
- Mack, Sara. 2011. A sociophonetic analysis of /s/ variation in Puerto Rican Spanish. In Luis A. Ortiz-López (ed.), *Selected Proceedings of the 11th Hispanic Linguistics Symposium*, 81–93. Somerville, MA: Cascadilla Press.
- Magnuson, James S., James A. Dixon, Michael K. Tanenhaus & Richard N. Aslin. 2007. The dynamics of lexical competition during spoken word recognition. *Cognitive Science* 31(1). 133–156. DOI: <https://doi.org/10.1080/03640210709336987>
- Marian, Viorica, James Bartolotti, Sarah Chabal & Anthony Shook. 2012. CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE* 7(8). e43230. DOI: <https://doi.org/10.1371/journal.pone.0043230>
- Ma, Roxana & Eleanor Herasimchuk. 1971. The linguistic dimensions of a bilingual neighborhood. In Joshua A. Fishman, Robert L. Cooper & Roxana Ma (eds.), *Bilingualism in the Barrio*, 349–464. Bloomington, IN: Indiana University.
- Marslen-Wilson, William D. & Alan Welsh. 1978. Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology* 10(1). 29–63. DOI: [https://doi.org/10.1016/0010-0285\(78\)90018-X](https://doi.org/10.1016/0010-0285(78)90018-X)
- McDermott, Kathleen B. & Jason M. Watson. 2001. The rise and fall of false recall: The impact of presentation duration. *Journal of Memory and Language* 45(1). 160–176. DOI: <https://doi.org/10.1006/jmla.2000.2771>
- Minnette Fox, Michelle. 2006. *Usage-based effects in Latin American Spanish syllable-final /s/ lenition*. Philadelphia, PA: University of Pennsylvania dissertation.

- Nooteboom, Sieb G. 1981. Lexical retrieval from fragments of spoken words: Beginnings vs. endings. *Journal of Phonetics* 9(4). 407–424.
- Nygaard, Lynne C., Mitchell S. Sommers & David B. Pisoni. 1994. Speech perception as a talker-contingent process. *Psychological Science* 5(1). 42–46. DOI: <https://doi.org/10.1111/j.1467-9280.1994.tb00612.x>
- Patterson, David & Cynthia M. Connine. 2001. Variant frequency in flap production. *Phonetica* 58(4). 254–275. DOI: <https://doi.org/10.1159/000046178>
- Patterson, David, Paul C. LoCasto & Cynthia M. Connine. 2003. Corpora analyses of frequency of schwa deletion in conversational American English. *Phonetica* 60(1). 45–69. DOI: <https://doi.org/10.1159/000070453>
- Pierrehumbert, Janet B. 2016. Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics* 2. 33–52. DOI: <https://doi.org/10.1146/annurev-linguistics-030514-125050>
- Pinnow, Eleni & Cynthia M. Connine. 2014. Phonological variant recognition: Representations and rules. *Language and Speech* 57(1). 42–67. DOI: <https://doi.org/10.1177/0023830913479105>
- Pitt, Mark A. 2009. How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *Journal of Memory and Language* 61(1). 19–36. DOI: <https://doi.org/10.1016/j.jml.2009.02.005>
- Poplack, Shana. 1980. Deletion and disambiguation in Puerto Rican Spanish. *Language*. 371–385. DOI: <https://doi.org/10.2307/413761>
- Pycha, Anne. 2017. False memories for morphologically simple versus complex words in English. *The Mental Lexicon* 12(1). 71–106. DOI: <https://doi.org/10.1075/ml.12.1.04pyc>
- Ranbom, Larissa J. & Cynthia M. Connine. 2007. Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language* 57(2). 273–298. DOI: <https://doi.org/10.1016/j.jml.2007.04.001>
- Reinartz, Mark Tippens. 2001. Illusions of memory. *Comments on Theoretical Biology* 6. 411–430.
- Robinson, Kerry J. & Henry L. Roediger. 1997. Associative processes in false recall and false recognition. *Psychological Science* 8(3). 231–237. DOI: <https://doi.org/10.1111/j.1467-9280.1997.tb00417.x>
- Roediger, Henry L., David A. Balota & Jason M. Watson. 2001. Spreading activation and arousal of false memories. In Henry L. Roediger, James S. Nairne, Ian Neath & Aimée M. Surprenant (eds.), *The nature of remembering: Essays in honor of Robert G. Crowder*, 95–115. Washington, DC: American Psychological Association Press. DOI: <https://doi.org/10.1037/10394-006>
- Roediger, Henry L., Jason M. Watson, Kathleen B. McDermott & David A. Gallo. 2001. Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review* 8(3). 385–407. DOI: <https://doi.org/10.3758/BF03196177>
- Roediger, Henry L. & Kathleen B. McDermott. 1995. Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21(4). 803–814. DOI: <https://doi.org/10.1037/0278-7393.21.4.803>
- Roediger, Henry L., Kathleen B. McDermott, David B. Pisoni & David A. Gallo. 2004. Illusory recollection of voices. *Memory* 12(5). 586–602. DOI: <https://doi.org/10.1080/09658210344000125>
- Schacter, Daniel L., Lana Israel & Carrie Racine. 1999. Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language* 40(1). 1–24. DOI: <https://doi.org/10.1006/jmla.1998.2611>
- Schacter, Daniel L., Mieke Verfaellie & Michael D. Anes. 1997. Illusory memories in amnesic patients: conceptual and perceptual false recognition. *Neuropsychology* 11(3). 331. DOI: <https://doi.org/10.1037/0894-4105.11.3.331>

- Schmidt, Lauren B. 2013. Regional variation in the perception of sociophonetic variants of Spanish /s/. In Ana M. Carvalho & Sara Beaudrie (eds.), *Selected Proceedings of the 6th Workshop on Spanish Sociolinguistics*, 189–202. Somerville, MA: Cascadilla Press.
- Schmidt, Lauren B. 2015. Effects of regional media, travel, and social contacts on the perception of Spanish aspirated-/s/. *Borealis – An International Journal of Hispanic Linguistics* 4(1). 99–120.
- Seamon, John G., Chun R. Luo & David A. Gallo. 1998. Creating false memories of words with or without recognition of list items: Evidence for nonconscious processes. *Psychological Science* 9(1). 20–26. DOI: <https://doi.org/10.1111/1467-9280.00004>
- Sommers, Mitchell S. & Bryan P. Lewis. 1999. Who really lives next door: Creating false memories with phonological neighbors. *Journal of Memory and Language* 40(1). 83–108. DOI: <https://doi.org/10.1006/jmla.1998.2614>
- Stockall, Linnaea, Andrew Stringfellow & Alec Marantz. 2004. The precise time course of lexical activation: MEG measurements of the effects of frequency, probability, and density in lexical decision. *Brain and Language* 90(1). 88–94. DOI: [https://doi.org/10.1016/S0093-934X\(03\)00422-X](https://doi.org/10.1016/S0093-934X(03)00422-X)
- Sumner, Meghan & Arthur G. Samuel. 2005. Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language* 52(3). 322–338. DOI: <https://doi.org/10.1016/j.jml.2004.11.004>
- Sumner, Meghan & Arthur G. Samuel. 2009. The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language* 60(4). 487–501. DOI: <https://doi.org/10.1016/j.jml.2009.01.001>
- Sumner, Meghan, Seung Kyung Kim, Ed King & Kevin B. McGowan. 2014. The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology* 4. 1–13. DOI: <https://doi.org/10.3389/fpsyg.2013.01015>
- Terrell, Tracy D. 1979. Final /s/ in Cuban Spanish. *Hispania* 62(4). 599–612. DOI: <https://doi.org/10.2307/340142>
- Underwood, Benton J. 1965. False recognition produced by implicit verbal responses. *Journal of Experimental Psychology* 70(1). 122–129. DOI: <https://doi.org/10.1037/h0022014>
- Wallace, William P., Mark T. Stewart & Christine P. Malone. 1995. Recognition memory errors produced by implicit activation of word candidates during the processing of spoken words. *Journal of Memory and Language* 34(4). 417–439. DOI: <https://doi.org/10.1006/jmla.1995.1019>
- Wallace, William P., Mark T. Stewart, Heather L. Sherman & Michael D. Mellor. 1995. False positives in recognition memory produced by cohort activation. *Cognition* 55(1). 85–113. DOI: [https://doi.org/10.1016/0010-0277\(94\)00646-3](https://doi.org/10.1016/0010-0277(94)00646-3)
- Wallace, William P., Mark T. Stewart, Thomas R. Shaffer & John A. Wilson. 1998. Are false recognitions influenced by prerecognition processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24(2). 299. DOI: <https://doi.org/10.1037/0278-7393.24.2.299>
- Wallace, William P., Thomas R. Shaffer, Martha D. Amberg & Vicki L. Silvers. 2001. Divided attention and prerecognition processing of spoken words and nonwords. *Memory & Cognition* 29(8). 1102–1110. DOI: <https://doi.org/10.3758/BF03206379>
- Watson, Jason M., David A. Balota & Henry L. Roediger. 2003. Creating false memories with hybrid lists of semantic and phonological associates: Over-additive false memories produced by converging associative networks. *Journal of Memory and Language* 49(1). 95–118. DOI: [https://doi.org/10.1016/S0749-596X\(03\)00019-6](https://doi.org/10.1016/S0749-596X(03)00019-6)
- Westbury, Chris, Lori Buchanan & Norman R. Brown. 2002. Sounds of the neighborhood: False memories and the structure of the phonological lexicon. *Journal of Memory and Language* 46(3). 622–651. DOI: <https://doi.org/10.1006/jmla.2001.2821>

- Widdison, Kirk A. 1995. An acoustic and perceptual study of the Spanish sound change  $s > h$ . *Rivista Di Linguistica* 7. 175–190.
- Yonelinas, Andrew P. 2002. The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language* 46(3). 441–517. DOI: <https://doi.org/10.1006/jmla.2002.2864>
- Zeelenberg, René, Gijs Plomp & Jeroen G. W. Raaijmakers. 2003. Can false memories be created through nonconscious processes? *Consciousness and Cognition* 12(3). 403–412. DOI: [https://doi.org/10.1016/S1053-8100\(03\)00021-7](https://doi.org/10.1016/S1053-8100(03)00021-7)

**How to cite this article:** Pycha, Anne. 2017. A new test for exemplar theory: Varying versus non-varying words in Spanish. *Glossa: a journal of general linguistics* X(X): X.1–31, DOI: <https://doi.org/10.5334/gjgl.240>

**Submitted:** 25 August 2016    **Accepted:** 04 May 2017    **Published:** XX Month 201X

**Copyright:** © 2017 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

]u[ *Glossa: a journal of general linguistics* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 