Speech Perception: Phonological Neighborhood Effects on Word Recognition Persist Despite Semantic Sentence Context

Teresa Cervera-Crespo¹ and Julio González-Álvarez¹,²

Abstract
This study tested the hypothesis that two lexical properties, both phonological neighborhood density (ND) and neighborhood frequency (NF), influence the recognition of target words when preceded by either a semantically congruent or semantically neutral context. Our study is the first to test this hypothesis using a language other than English (i.e., Spanish). We used highly familiar bisyllabic nouns with medium-frequency occurrence as target words, and we expected recognition accuracy to increase as ND and NF decreased in both semantically congruent and semantically neutral sentences. We presented 48 undergraduate listeners with a set of 80 words, differing in ND and NF, within these two sentence contexts (i.e., 160 sentences). We then tested the relationships between ND, NF, and variations in semantic sentence context within a linear logistic model and found that words with a low frequency of neighbors were more likely to be correctly recognized in both sentence contexts. Thus, during word recognition, the influence of phonological competition outweighed semantic sentence context even when words were presented in Spanish.

¹Department of Basic Psychology, University of Valencia, Spain
²Department of Basic and Clinical Psychology and Psychobiology, Universitat Jaume I, Castellón, Spain

Corresponding Author:
Teresa Cervera-Crespo, Departamento de Psicología Básica, Universitat de Valencia, Blasco Ibañez 21, 46010 Valencia, Spain.
Email: Teresa.Cervera@uv.es
Keywords
auditory word recognition, phonological proximity, sentence context

Introduction

It has been well established that the number of similar-sounding words in a mental lexicon significantly influences word recognition in speech perception. Recognition of a spoken word requires the listener to match the acoustical input corresponding to the speech signal with the correct lexical entry in the mental lexicon. The acoustical information corresponding to a given word is quite variable due to different speaking characteristics, such as speaking accent, speech rate, and dialect, and different environmental listening conditions, such as background noise and reverberation. Therefore, it is unlikely that the acoustical information corresponding to the speech signal would be sufficient to activate a single word’s entry into the mental lexicon consisting of thousands of words stored in long-term memory (Luce & Pisoni, 1998; Vitevitch & Luce, 2016). Most models of spoken word recognition assume that word recognition involves the activation of similar-sounding words in the mental lexicon, but there are some differences among the models regarding the details of this process. The TRACE (McClelland & Elman, 1986), Shortlist (Norris, 1994), and PARSYN models (Luce, Goldinger, Auer, & Vitevitch, 2000) are all interactive, localist connectionist models that have various processing units such as features, phonemes, and words, with excitatory and inhibitory connections that raise or lower the perceiver’s activation of word candidates in the recognition process. Dynamic causal modeling (DCM; Gaskell & Marslen-Wilson, 1997) is also an interactive model, but unlike TRACE, Shortlist, and PARSYN, in DCM, information is distributed so that intermediate levels of representation are not needed. Nonetheless, all of these models postulate that phonological competition is a function of the similarity of competing words to the auditory input. These sets of similar-sounding words compete for recognition. One of the most well-known phonological competition sets is the phonological “neighborhood,” defined as a set of words that can be formed by deleting, adding, or substituting a single phoneme (Luce & Pisoni, 1998).

The neighborhood activation model (NAM; Luce & Pisoni, 1998) provides a theoretical framework for explaining how the lexical neighborhood influences spoken word recognition. This model examines not only the effects of similar-sounding words but also their frequency of occurrence. According to NAM, words with a small number of similar-sounding neighbors or “neighborhood density” (ND) are recognized faster and more accurately than words with a high number of neighbors. Similarly, when the “neighborhood frequency” (NF) is low, words are more easily recognized. Thus, words with both high numbers of neighbors and high frequency of their occurrence are hard words.
that are more difficult to recognize than words with low ND and low NF, or easy words. These effects have been demonstrated in several studies using a variety of experimental tasks such as the perceptual identification task (Goldinger, Luce, & Pisoni, 1989; Luce & Pisoni, 1998; Sommers & Danielson, 1999; Taler, Aaron, Steinmetz, & Pisoni, 2010), auditory lexical decision task (Luce & Pisoni, 1998; Vitevitch & Luce, 1999), naming task (Luce & Pisoni, 1998), and the “slips of the ear” task involving naturally occurring perceptual errors in which it is common for listeners to misperceive a word the speaker pronounces correctly (Vitevitch, 2002). These studies showed that the effects of ND and NF on spoken word recognition depended on the experimental task and the language spoken. For instance, in the Spanish language, high ND has been found to facilitate lexical decision tasks, while, in English, high ND inhibits lexical decision-making (Vitevitch & Rodríguez, 2005). In addition, using a perceptual identification task, the effects of high ND were inhibitory in French (Dufour & Fraunfeldner, 2006) and Japanese (Amano & Kondo, 2000), as first observed in English by Luce and Pisoni (1998).

Although the influence of the phonological neighborhood on word recognition when words are presented out of context has been investigated with various tasks and in various languages, the influence of phonological neighborhood on word recognition when words are embedded in varied sentence contexts has rarely been studied; and so this circumstance is poorly understood. On the one hand, Marslen-Wilson and Tyler (1980) demonstrated that a constraining linguistic context produced a dramatic reduction in the number of possible word candidates that share the same initial sounds as the target word (the “cohort”), while, on the other hand, Zwitserlood (1989) presented a semantically constrained sentence with a final target word followed by either a prime word related to the target word or a neighbor of that word and found some word recognition facilitation for both conditions.

The small number of studies that have investigated the role of ND and NF in word recognition within varied sentence contexts (Sommers & Danielson, 1999; Taler et al., 2010) found that the semantic sentence context did not totally eliminate the influence of ND and NF. In these studies, the ND and NF inhibitory effects persisted, although these effects were lower in semantically congruent versus semantically neutral or less predictable sentence contexts. These results are more consistent with interactive models of language processing than with modular models that postulate a discrete separation of different levels of language representations and processing (Forster, 1978) and would predict that when the individual listens to a word preceded by semantic linguistic information, phonological neighbors are inevitably activated, regardless of the presence of a meaningful sentence context, because the flow of information across the different processing stages is ordered bottom-up from lower to higher stages. But, interactive models such as TRACE, Shortlist, DCM, and NAM assume that information flows in both directions, from lower to higher levels and
top-down; and thus, these models would postulate that context restricts lexical alternatives in the earliest stages (e.g., Brock & Nation, 2013).

The studies by Sommers and Danielson (1999) and Taler et al. (2010) employed experimental tasks to identify a target word within a sentence using English speech stimuli. To our knowledge, there have been no studies in other languages. While it is clear that a meaningful sentence context facilitates word recognition and these beneficial effects were shown quite early using sentences in a noisy environment (e.g., Kalikow, Stevens, & Elliot, 1977), the question of whether the phonological neighborhood plays a significant role in word recognition when a target word is heard in a meaningful sentence context versus in an out-of-context sentence has not been explored in Spanish. This is the objective of this study, and its practical applications include the provision of guidance to researchers and clinicians in developing stimuli for evaluating speech perception in individuals with auditory or cognitive deficits for whom it is important to attend to both acoustical and phonetic characteristics and any other important lexical properties or semantic information conveyed by the sentence heard. Based on the findings from Sommers and Danielson (1999) and Taler et al. (2010), we expected that (a) word recognition accuracy would be higher in congruent than in neutral sentence contexts; (b) recognition accuracy would increase as ND and NF decrease and that ND and NF characteristics would interact, producing the classical difference between easy and hard-to-recognize words; and (c) there would be ND and NF effects in both congruent and neutral sentence contexts.

Method

Participants

Participants in this experiment were 48 undergraduate students (31 females and 17 males), aged 22-28 years (M = 23.7, SD = 2.6 years). These students participated voluntarily and received partial credit for a course requirement. No participants reported having any hearing or language problems, and all were native speakers of Castilian Spanish. The research conformed to the American Psychological Association’s Ethical Principles of Psychologists and Code of Conduct (APA, 2010). All participants signed informed consent forms.

Materials

As described earlier, speech stimuli in the present experiment were 160 spoken sentences extracted from the Spanish Sentence Lists (Cervera & González-Alvarez, 2010). Half of the sentences had a final target word that was predictable from the preceding sentence context (congruent sentences; e.g., Para leer necesito gafas [I have to put my glasses on to read]). The final target word in the other half of the sentences could not be predicted by the preceding words (neutral sentences;
e.g., *Ella estaba hablando sobre gafas,* [She was talking about glasses]). Each congruent sentence had a corresponding neutral sentence so that the same final word appeared in both types of sentences. The predictability of the target word in the congruent sentences was determined by the cloze procedure in which the final word of the sentence was omitted and the listeners were required to fill in the blank, ranging from 45-75%, thus avoiding very low or high probabilities.

The target words were similar with regard to frequency of occurrence, familiarity, contextual diversity, duration (2,000 milliseconds), and syllabic stress with emphasis on the first syllable. In addition, all the words were nouns and had two syllables. Thereby, verbs, which have a rich morphology in Spanish, were avoided. Word morphology and length influence word recognition and seem to explain some of the differences between the results obtained when using English versus Spanish speech stimuli (Vitevitch & Rodríguez, 2005; Vitevitch & Stammer, 2009). In addition, the words differed in ND and NF.

The lexical characteristics of these words were obtained from the EsPal database (Duchon, Perea, Sebastian-Gallés, Martí, & Carreiras, 2013). The target words were mid-frequency but highly familiar words. Words with more than 20 neighbors were considered high ND words, whereas words with fewer neighbors were considered low ND words. Likewise, words whose neighbors had a mean frequency greater than nine occurrences per million were considered high NF, and words below this value were considered low NF words. In both cases, we used the median number to divide words into ND and NF groups (Table 1). The target words are presented in Online Appendix I.

**Table 1.** Lexical Properties of the Target Words and Significant Group Differences.

|                | High ND–high NF Mean (SD) | High ND–low NF Mean (SD) | Low ND–high NF Mean (SD) | Low ND–low NF Mean (SD) | F (df) | p     |
|----------------|----------------------------|--------------------------|--------------------------|-------------------------|--------|-------|
| Log frequency  | 1.37 (0.31)                | 1.94 (0.27)              | 1.46 (0.37)              | 1.45 (0.23)             | 0.29 (3) | ns    |
| Contextual diversity | 12,372 (11,012)       | 11,651 (12,020)          | 12,594 (13,683)          | 10,591 (8,334)          | 0.11 (3) | ns    |
| Familiarity    | 6 (0.37)                   | 6 (2)                    | 6 (1)                    | 6 (0.3)                 | 0.13 (3) | ns    |
| ND             | 40 (11)                    | 31 (9)                   | 14 (7)                   | 11 (4)                  | 50.56 (3) | <.01  |
| NF             | 40 (28)                    | 5 (3)                    | 24 (19)                  | 5 (3)                   | 20.81 (3) | <.01  |

Note. ND = neighborhood density (number of neighbors); NF = neighborhood frequency (mean average frequency of the neighbors); ns = no significant; SD = standard deviation.
sentence was equated on the root-mean-square across the entire sentence and stored in different digital files. Babble noise, at a $+10\,\text{dB}$ signal-to-noise ratio, was added to the speech signal to avoid ceiling effects on the scores of the perceptual identification task. Previous data suggested that the lexical neighborhood emerges in the auditory domain when the speech signal is degraded (e.g., Goldinger et al., 1989). The babble noise was generated by mixing 12 voices: 6 males and 6 females. The $+10\,\text{dB}$ signal-to-noise condition was created by manipulating the overall root-mean-square of both the signal and the noise. These manipulations were performed using Adobe Audition Pro software.

**Procedure**

We used the perceptual identification task because it seems to be the most consistent task cross-linguistically in studies of neighborhood effects of isolated word recognition and because it was used in past studies by Sommers and Danielson (1999) and Taler et al. (2010). The perceptual identification task took place in a sound-attenuated room. The listeners were presented with the stimuli through Sennheiser headphones connected to a Pentium PC. A computer program especially designed for this task administered and registered the listeners’ responses. The listeners were instructed to listen to each sentence and type the last word of the sentence they heard using the computer keyboard and then press the space bar. After two seconds, the next stimulus was presented. Prior to the experiment, we presented eight different practice sentences. Each listener was presented with half of the words in a congruent sentence and the other half in neutral sentence conditions so that the listener did not hear the same word twice. The ND and NF characteristics of word presentation were counterbalanced in each condition. The 48 participants were randomly assigned to either the congruent sentence or neutral sentence condition, and the order of stimuli presentation within each condition was random.

**Results**

Recognition accuracy was measured as the number of target words correctly identified later, converted to percentages. For descriptive purposes, the mean recognition percentages in each experimental condition are presented in Table 2. The dependent variable was the binary value of correct versus incorrect word recognition accuracy for each word; and analyses were by mixed-effects binary logistic regression. Participants and items were random effects, whereas Context (congruent or neutral sentences), ND (high or low), and NF (high or low) were fixed effects. In addition, $\text{Exp}(B)$ (the exponentiation $B$ coefficients) values had an odds ratio interpretation and could be used to assess the predicted magnitude of the effects of each independent variable. The statistical analyses were performed using the SPSS software package.
An omnibus test of model coefficients was significant ($\chi^2 = 428.55$, $p < .001$). The three fixed effects (Context, ND, and NF) were significant. A significant fixed effect of Context was obtained; as expected, participants provided significantly more correct responses in semantically congruent sentences than in neutral sentences, Wald $Z = 33.88$, $p < .001$, $\text{Exp}(B) = 3.41$. Likewise, we found significant effects for ND, Wald $Z = 36.07$, $p < .05$, $\text{Exp}(B) = 0.41$, and NF, Wald $Z = 34.58$, $p < .001$, $\text{Exp}(B) = 0.42$, and as expected, correct responses were greater for low ND words than for high ND words and for low NF words than for high NF words. From the $\text{Exp}(B)$ values, it can be observed that the magnitude of the effect was greater for Context than for the ND and NF variables. Regarding the interactions, the Context $\times$ NF interaction was significant, Wald $Z = 13.27$, $p < .001$, $\text{Exp}(B) = 2.97$, indicating that the recognition was higher for low NF than for high NF in the neutral sentence condition but not in the congruent sentence condition. However, the Context $\times$ ND interaction was not significant, Wald $Z = 3.27$, $p = .07$, $\text{Exp}(B) = 0.63$. The ND $\times$ NF interaction was significant, Wald $Z = 13.18$, $p < .001$, $\text{Exp}(B) = 2.04$, indicating that the effects of ND differed, depending on whether NF was high or low. In other words, as expected, low ND and low NF words (easy words) were identified more accurately than were high ND and high NF words (hard words). The most interesting result was the absence of a significant Context $\times$ ND $\times$ NF interaction, Wald $Z = 0.31$, $p = .58$, $\text{Exp}(B) = 0.81$, indicating that the effects of lexical characteristics (ND and NF) on word recognition were evident in both congruent and neutral sentence contexts.

**Discussion**

The objective of this study was to assess lexical neighborhood effects (ND and NF) on the recognition of spoken words embedded in varied sentence contexts.
Although the lexical neighborhood has had well-known factor effects on phonological categorization in word recognition in the past research, there have been fewer studies on lexical neighborhood effects on word recognition when target words are presented within a sentence, rather than individually, and when they are preceded by semantically meaningful or neutral information.

Our results showed that not only sentence context but also lexical phonological properties, ND and NF, affected word recognition. As expected, a semantically congruent sentence context increased target word recognition compared with a semantically neutral context. As in previous studies using English sentences (Sommers & Danielson, 1999; Taler et al., 2010), our results showed inhibitory effects of the lexical neighborhood on target word recognition, even in semantically constrained sentences. In other words, when participants listened to an acoustic signal, a congruent sentence context biased a given meaning but did not keep some similar-sounding words from being activated in memory competing for recognition and producing inhibitory word recognition effects.

Our prediction of an interactive influence of ND and NF on word identification was confirmed in both the meaningful and neutral sentence contexts. Similar to the results of prior studies using isolated words as stimuli (Luce & Pisoni, 1998), not only the number of phonological neighbors but also their frequency of occurrence influenced word recognition. In both meaningful and neutral sentence contexts predicted, word recognition differences between words that were easy and hard in their phonological characteristics were evident. There was a nonstatistically significant trend toward a greater difference for the neutral sentence condition (21% recognition difference of easy vs. hard words) versus the semantically congruent sentence condition (5% recognition difference for easy vs. hard words). In the neutral sentence condition, hard words were correctly perceived with 58% frequency and easy words were perceived with 79% frequency. In congruent sentences, hard words were recognized with 88% frequency and easy words were recognized with 93% frequency. These results are consistent with those from Sommers and Danielson (1999) who compared easy and hard words in isolation and in a preceding sentence context. These results also coincide with Taler et al. (2010), who used sentences consisting of three target words that differed in ND and NF and in meaningful or neutral sentence contexts, as in this study. In general, we showed inhibitory effects of the lexical neighborhood on word recognition performance, in spite of a semantically biasing sentence context. Thus, semantic context seems insufficient to restrict lexical competition.

It should be noted that that these results were obtained with one particular task (word identification), two types of sentences, and specific target words (nouns, disyllabic, and mid-frequency), restricting the generalization of these findings to other experimental tasks, types of sentences, and target words. For instance, the congruent sentences used in this study were quite predictable, but they did not constrain responses to a unique word candidate. Our semantically presented in Spanish neutral sentences were not semantically inappropriate or
implausible. In addition, this study controlled the target words for length and morphology, addressing suggestions that these two lexical characteristics explained some of the differences between Spanish and English in prior studies with isolated words. Thus, the limitation of this study is that the conclusions are restricted to words with the characteristics employed in this study. Future researchers should consider different types of target words and experimental tasks.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical Approval
The research conformed to the American Psychological Association’s (2010) “Ethical Principles of Psychologists and Code of Conduct”.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by the Spanish Ministry of Economy and Competitiveness (Research Grant FFI2017-84951-P).

ORCID iD
Teresa Cervera-Crespo https://orcid.org/0000-0001-8957-2483
Julio González-Alvarez https://orcid.org/0000-0002-0389-5263

References
Amano, S., & Kondo, T. (2000, May). Neighborhood and cohort in lexical processing of Japanese spoken words. Nijmegen, the Netherlands: Spoken Word Access Processes. Retrieved from https://www.isca-speech.org/archive_open/archive_papers/swap/swap_091.pdf

American Psychological Association. (2010). Ethical principles of psychologist and code of conduct. American Psychologist, 65, 493. Retrieved from http://www.apa.org/ethics/code

Brock, J., & Nation, K. (2013). The hardest butter to button: Immediate context effects in spoken word recognition. Quarterly Journal of Experimental Psychology, 67, 114–123. doi:10.1080/17470218.2013.791331

Cervera, T. & Gonzalez-Alvarez, J. (2010). Lists of Spanish sentences with equivalent predictability, phonetic content, length and frequency of the last word. Perceptual and Motor Skills, 111, 1–13.

Duchon, A., Perea, M., Sebastian-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. Behavior Research Methods, 4, 1246–1258. doi:10.3758/s13428-013-0326-1

Dufour, S., & Fraunfeldner, U. H. (2006). Phonological neighborhood effects in French spoken word recognition. Quarterly Journal of Experimental Psychology, 63, 226–238. doi:10.1080/17470210903308336
Forster, K. I. (1978). Accessing the mental lexicon. In E. Walker (Ed.), *Explorations in the biology of language* (pp. 139–174). Montgomery, VT: Bradford Books. doi:10.1017/S00222670000684

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12*, 613–656. doi:10.1080/0169096973866646

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

Kalikow, D. N., Stevens, K. N., & Elliot, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61*, 1337–1360. doi:10.1121/1.381436

Luce, P. A., Goldinger, S., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics, 62*, 615–625. doi:10.3758/BF03212113

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear & Hearing, 19*(1), 1–36. doi:10.1097/00003446-199802000-00001

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8*(1), 1–71. doi:10.1016/0010-0278(80)90015-3

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

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

Sommers, M. S., & Danielson, S. M. (1999). Inhibitory processes and spoken word recognition in young and old adults: The interaction of lexical competition and semantic context. *Psychology and Aging, 14*, 458–472. doi:10.1037/0882-7974.14.3.458

Taler, V., Aaron, G. P., Steinmetz, L. G., & Pisoni, D. B. (2010). Lexical neighborhood density effects on spoken word recognition and production in healthy aging. *Journal of Gerontology: Psychological Sciences, 65B*(5), 551–560. doi:10.1093/geronb/gbq039

Vitevitch, M. S. (2002). Naturalistic and experimental analyses of word frequency and neighborhood density effects in slips of the ear. *Language and Speech, 45*, 407–434. doi:10.1177/2F00238309020450040501

Vitevitch, M. S., & Luce, P. A. (2016). Phonological neighborhood effects in spoken word perception and production. *Annual Review of Linguistics, 2*, 75–94. doi:10.1146/annurev-linguistics-030514-124832

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

Vitevitch, M. S., & Stammer, M. K. (2009). *The influence of neighborhood density (and neighborhood frequency) in Spanish speech production: a follow-up report* (Spoken Language Technical Report). Lawrence: University of Kansas. Retrieved from http://hdl.handle.net/1808/5500

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition, 32*, 25–64. doi:10.1016/0010-0277(89)90013-9
Author Biographies

**Teresa Cervera-Crespo**, PhD, is an associate professor on psychology of language at the University of Valencia, Spain. Her interest includes issues related to speech perception and production in different challenging situations. Recently her interest is also focused on the inter-relationship of language and other cognitive systems in adults with sensory or cognitive impairments.

**Julio González-Álvarez**, PhD, is an associate professor on psychology of language and speech processes at the University Jaume I, Spain. His research focuses on cognitive and neural bases of language and perception processes, and lately on gender studies.