Title
Phonological Neighborhood Density in Native and Non-Native Word Production

Permalink
https://escholarship.org/uc/item/34w8w6mp

Journal

ISSN
1069-7977

Authors
Blumenfeld, Henrike K.
Marian, Viorica

Publication Date
2006

Peer reviewed
Phonological Neighborhood Density in Native and Non-Native Word Production

Henrike K. Blumenfeld (k-blumenfeld@northwestern.edu)
Department of Communication Sciences and Disorders, 2240 Campus Drive
Evanston, IL 60208 USA

Viorica Marian (v-marian@northwestern.edu)
Department of Communication Sciences and Disorders, 2240 Campus Drive
Evanston, IL 60208 USA

Abstract
The role of phonological neighborhood density in lexical access was examined during native- and non-native language production. German-English and English-German bilinguals named pictures of German words with high- and low-density neighborhoods. Results revealed that accuracy of picture naming was influenced by phonological neighborhood, with high-density neighborhoods facilitating lexical access in both the native and the non-native languages. However, latency of picture naming was facilitated by high-density phonological neighborhoods only in the non-native, but not the native, language. This suggests that native / non-native language status mediates the effect of phonological neighborhood on the speed of lexical access. Implications for native and non-native language processing dynamics are discussed.

Keywords: Language Processing; Bilingualism

Introduction
How alike are the mechanisms underlying native and non-native language processing? This question was examined by manipulating word similarity during native and non-native naming. Similarity and difference among words can be examined at the phonological, orthographic, lexical, and semantic levels. For example, the size of a word’s semantic network influences how fast the word can be accessed, with words that have larger semantic similarity neighborhoods processed faster than words with smaller semantic neighborhoods (e.g., Locker, Simpson, & Yates, 2003). In the same way, the size of a word’s phonological or orthographic neighborhood influences word access. A phonological neighbor is a word that differs from the target word by a single phoneme (Grainger, Muneaux, Farioli, & Ziegler, 2005; Yates, Locker, & Simpson, 2004). A word’s neighborhood size, also referred to as its neighborhood density, is the number of items that are highly similar to it. While phonology, orthography, and semantics are all subject to neighborhood density effects, with implications for lexical activation, the objective of the present study was to examine the role of phonological neighborhood density in lexical access during language production in native and non-native speakers.

Neighborhood Density in Word Production
In monolingual language production, activation of phonological representations similar to the target has been found to facilitate word access. For example, phonologically similar words have been found to facilitate naming during picture-word interference tasks (e.g., Costa & Sebastian-Galles, 1998). In studies focusing on tip-of-the-tongue (TOT) states, Meyer and Bock (1992) showed that priming with a phonologically similar word facilitated correct retrieval. Correspondingly, when phonological similarity was manipulated by varying neighborhood density, high-neighborhood targets have been found to produce fewer TOTs than low-neighborhood targets (e.g., Vitevitch & Sommers, 2003). Similar error patterns have also been observed in naturally-produced speech (Vitevitch, 1997). In picture naming tasks, targets with dense phonological neighborhoods have been found to be retrieved faster than targets with sparse phonological neighborhoods (e.g., Vitevitch, 2002). In sum, phonological similarity has consistently been found to facilitate monolingual lexical access across language production tasks (e.g., picture naming, picture-word interference, TOT-elicitation, naturally-produced speech).

Neighborhood facilitation during word production has been ascribed to interactive feedback between lexical and phonological processing levels. Gordon and Dell (2001) simulated behavioral findings of neighborhood effects within the framework of an Interactive Spreading Activation Model of production (Dell, 1986). The model was based on three stages of language production: a semantic stage where word-meaning is chosen, a lemma stage where other lexical characteristics are identified, and a phonological stage where the word form is accessed. Simulations suggested that neighborhood facilitation in production is due to feedback between the lemma level and the phonological level during word selection. The lemma activates phonological representations, which in turn activate similar-sounding lemmas (i.e., the phonological neighborhood) that feed back onto the target’s phonological representations and increase their activation levels, facilitating their selection.
For bilingual language production, a study looking at the role of orthographic neighborhoods in native and non-native naming (De Groot, Borgwaldt, Bos, & van den Eijnde, 2002) found that high-density neighborhoods facilitated naming and resulted in shorter response latencies in both languages. This neighborhood effect was facilitative across languages as well as within languages. Moreover, when delayed naming was partialled-out from immediate naming latencies (in order to detect the recognition component of naming) a difference between native and non-native languages was found. Non-native latencies were found to be more dependent on target-language orthographic neighborhood than native latencies, suggesting that production in a non-native language may be more sensitive to orthographic neighborhood effects.

Further insight into bilingual neighborhood effects can be gained from cross-linguistic research with monolinguals. Initial evidence from cross-linguistic research suggests that neighborhood effects in production may depend on structural characteristics of the tested language (Vitevitch & Rodriguez, 2005; Vitevitch & Stamer, in press). For instance, while high-density neighborhoods typically facilitate production in English, high-density neighborhoods have been found to inhibit production in Spanish (Vitevitch & Stamer, in press). These differences were explained in terms of Spanish language characteristics. Spanish is a language that is morphologically richer than English, so that clusters of phonological neighbors also contain morphological neighbors (e.g., niño, niña). These morphological clusters inhibit production, since one word-form needs to be chosen from many semantically-consistent candidates. To avoid any differences across linguistic structures, the target language was kept constant in the present study.

**Developmental Patterns in Neighborhood Density**

One way to understand differences between neighborhood effects in native and non-native language processing is by examining the developmental path of neighborhood density effects. It has been found that while toddlers prefer to listen to high-neighborhood words (Jusczyk, Luce, & Charles-Luce, 1994), children in fact do worse at naming high-neighborhood targets compared to low-neighborhood targets (Arnold, Conture, & Ohde, 2005; Newman & German, 2002). This suggests that facilitated naming of high-neighborhood targets may be the end of a developmental path that requires maturation of the language system. The presence of a developmental path raises questions about the role of language status in sensitivity to neighborhood density. On the one hand, non-native speakers may show patterns similar to those of children who are native speakers, due to lower proficiency and a less entrenched language network. If that were the case, then bilingual speakers would be more sensitive to phonological neighborhood density in the native language compared to the non-native language. On the other hand, findings from TOT studies with adult native speakers suggest that facilitation of dense neighborhoods is more pronounced for low-frequency words than for high-frequency words (Vitevitch & Sommers, 2003). Thus, high-frequency words may be easier to access overall, and therefore may be less susceptible to neighborhood effects. In a native language, more extensive practice with and previous exposure to a language may give a word an “often used” status and produce effects similar to those of high-frequency words. In a non-native language, limited exposure to a word and rare instances of previous use may give a word a “rarely used” status and produce effects similar to those of low-frequency words. As a result, a non-native language may be more susceptible to neighborhood effects than a native language, a hypothesis supported by studies of orthographic neighborhood density (De Groot et al., 2002). If that were the case, then bilingual speakers should be more sensitive to phonological neighborhood effects in the non-native language compared to the native language.

**The Present Study**

The question underlying the current research is how phonological neighborhood influences lexical access in bilingual language production. Do effects of phonological neighborhood parallel those of orthographic neighborhood, and are they similar across native and non-native languages? To answer these questions, the present study examined the role of phonological neighborhood density during native and non-native picture naming. We manipulated whether the target language was native or non-native by testing two groups of bilinguals, one where the target language was the native language and one where the target language was the non-native language. In other words, two groups of bilinguals were tested in the same language, as opposed to testing one group of bilinguals in both of their languages. As a result, any differences observed could be attributed to native/non-native language status rather than to differences in language structure. During a German picture naming task, German-English and English-German bilinguals were asked to produce targets with either high-density or low-density phonological neighborhoods. Thus, German was always the target language. The present study followed a two-by-two design, with two independent variables, neighborhood size and language status. Neighborhood size was a within-group variable and consisted of two levels, high-density phonological neighborhood words and low-density phonological neighborhood words. Language status was a between-group variable and also consisted of two levels, native German speakers and non-native German speakers.

We expected that results in the native language would replicate those of previous studies with monolinguals, and produce higher accuracy and shorter latency rates for words with dense phonological
neighborhoods than for words with sparse phonological neighborhoods. In addition, we aimed to extend the paradigm to production in a non-native language. The lower proficiency levels in the non-native language were predicted to influence the pattern of results. On the one hand, if sensitivity to phonological neighborhood density emerged with language proficiency, then neighborhood effects should be more apparent in native naming than in non-native naming. On the other hand, if lower proficiency levels conferred ‘low frequency status’ onto all words in that language, then sensitivity to phonological neighborhood density should be more apparent in non-native naming than in native naming. In general, participants were predicted to be faster and more accurate when the neighborhood was large than when it was small, and the magnitude of the effect was predicted to differ across native and non-native languages.

Methods

Participants Twenty-nine bilingual speakers of German and English were tested. Of these, 14 were English-German bilinguals (native language = English; 5 females), and 15 were German-English bilinguals (native language = German; 7 females). All bilinguals reported being native-language dominant. English-German bilinguals started learning German at the age of 11.8 years ($SD = 8.6$) and became fluent at 17.4 years ($SD = 10.0$). German-English bilinguals started learning English at the age of 10.7 years ($SD = 3.3$) and became fluent at 18.8 years ($SD = 7.6$). The mean age at the time of testing was 25.6 years ($SD = 8.9$) for the English-German bilinguals and 28.7 years ($SD = 12.9$) for the German-English bilinguals, with no significant difference between the two, $t(27) = 0.8, p > .1$. At the time of study, German-English bilinguals had more exposure (in terms of % time) to German ($M = 23.1\%$, $SD = 16.3$) than English-German bilinguals ($M = 11.1\%$, $SD = 6.8$), $t(27) = 2.6, p < .05$. All participants were administered a German translation of the English Peabody Picture Vocabulary Test (PPVT-III, Dunn & Dunn, 1997), where German-English bilinguals ($M = 193.9$, $SD = 7.6$) performed better than English-German bilinguals ($M = 178.6$, $SD = 18.2$), $t(27) = 2.9, p < .01$.

Materials Fifty-seven pictures corresponding to target German words were used. Picture stimuli were black line drawings with gray shadings, and were selected from the IMSI Master Clips electronic database and the Alta Vista search engine or were hand-drawn. To identify phonological neighbors of each target word, the German corpus of the CELEX lexical database (Baayen, Piepenbrock, & Van Rijn, 1995) was used, with an item coded as a phonological neighbor if it differed from the target by only one phoneme, and had the same number of phonemes in the same positions (Grainger et al., 2005; Yates et al., 2004). For example, the phonological neighborhood of the German word Hase, [həzə], includes such words as Vase, [vəzə], Hose, [hozə], and Habe, [habə]. (Note that the ideal scenario would be to also manipulate the phonological neighborhood density of English, in order to gauge the separate effect of non-target language phonological neighborhood, as well as the cumulative effect of phonological neighborhoods across both languages. However, that was not possible because differences in phonetic features between German and English precluded meaningful computations of cross-linguistic phonological neighborhoods in English.)

Stimuli were grouped into two conditions, where one condition included words with large phonological neighborhoods (3 or more phonological neighbors in German), and the other condition included words with small phonological neighborhoods (2 or fewer phonological neighbors in German). The large-neighborhood condition consisted of 31 German words, with a mean neighborhood size of 5.8 words ($SE = 0.4$). The small-neighborhood condition consisted of 26 German words, with a mean neighborhood size of 1.2 words ($SE = 0.2$). The neighborhood sizes for the two conditions were significantly different from each other ($t(55) = 8.8, p < .001$). The rationale for choosing a small (albeit significant) difference between dense and sparse neighborhood conditions was to specifically examine the sensitivity to relatively small changes in neighborhood density across native and non-native languages. Words in the two conditions were balanced for word length (in phonemes), spoken word frequencies of German and of English translation equivalents, orthographic neighborhood size in German and English, and number of synonyms available in German. There were no significant differences for these measures between the low-density and high-density neighborhood conditions ($p > .05$).

Procedure The experimenter was a native speaker of German, and conversed with participants in German prior to the experimental session to ensure a German language mode during testing. Participants sat in front of a computer and named pictures that appeared on the screen. Responses were recorded using a Logitech microphone. The experiment was self-paced, and each response triggered a 500 msec inter-stimulus-interval. Pictures were presented in a random sequence generated by SuperLab experimental software. Finally, a Language Experience and Proficiency Questionnaire (Marian, Blumenfeld, & Kaushanskaya, submitted) was administered to assess participants’ linguistic profiles.

Coding and Analyses Accuracy and latency of responses were measured. For accuracy, the percentage of pictures named correctly was computed. For latency, the duration of time from onset of picture presentation to onset of word production was measured in milliseconds using the experimental software. Naming accuracy and naming latency (for correct responses only) were analyzed for large-neighborhood and small-neighborhood conditions across the two bilingual groups, using two-way analyses
Results

Of the 1,653 responses produced, 62.5% (1033 cases) were coded as correct and 37.5% (620 cases) were coded as incorrect.

Naming Accuracy Two-way analyses of variance, with phonological neighborhood size (large, small) as a within-subject variable and group (native, non-native) as a between-subject variable in both by-subject (F1) and by-item (F2) analyses, and with number of years of education as a covariate in by-subject analyses, were performed.

Results revealed a main effect of neighborhood size, with faster naming when the phonological neighborhood was larger (M = 2,608.9 msec, SE = 214.6 msec) than when it was smaller (M = 3,719.9 msec, SE = 388.1 msec), F1 (1, 27) = 8.4, p < .01, F2 (1, 55) = 5.2, p < .05. Moreover, a main effect of group revealed that native speakers named pictures faster (M = 2,285.9 msec, SE = 271.3 msec) than non-native speakers (M = 4,042.9 msec, SE = 289.3 msec), F1 (1, 27) = 9.2, p < .01, F2 (1, 55) = 19.6, p < .001. The interaction between neighborhood size and group was significant by items, F2 (1, 55) = 4.6, p < .05, but the by-subject analysis did not reach significance, p > .05. Follow-up t-tests revealed that non-native speakers named pictures faster with large-neighborhood targets (M = 2,964.3 msec, SE = 313.0 msec) than with small-neighborhood targets (M = 5,121.4 msec, SE = 566.1 msec), t (55) = 3.1, p < .005. For native speakers, the difference was not significant (large neighborhoods: M = 2253.4 msec, SE = 548.6; small neighborhoods: M = 2318.4 msec, SE = 303.3), p > .05. When neighborhood size was regressed on naming latency across items, the relationship between the two was not significant for either native speakers or non-native speakers, p > .05.

Since neighborhood size is a continuous variable, follow-up regression analyses were conducted, where neighborhood size (independent variable) was regressed on naming accuracy. A significant relationship between neighborhood size and accuracy was found across both groups (R = 0.2, p < .05).

Naming Latency Two-way analyses of variance were performed, with phonological neighborhood size (large, small) and group (native, non-native) as independent variables in both by-subject and by-item analyses, and with participants’ number of years of education as a covariate in the by-subject analyses. Results (see Figure 2) revealed a main effect of neighborhood size, with faster naming when the phonological neighborhood was larger (M = 2,608.9 msec, SE = 214.6 msec) than when it was smaller (M = 3,719.9 msec, SE = 388.1 msec), F1 (1, 27) = 8.4, p < .01, F2 (1, 55) = 5.2, p < .05. Moreover, a main effect of group revealed that native speakers named pictures faster (M = 2,285.9 msec, SE = 271.3 msec) than non-native speakers (M = 4,042.9 msec, SE = 289.3 msec), F1 (1, 27) = 9.2, p < .01, F2 (1, 55) = 19.6, p < .001. The interaction between neighborhood size and group was significant by items, F2 (1, 55) = 4.6, p < .05, but the by-subject analysis did not reach significance, p > .05. Follow-up t-tests revealed that non-native speakers named pictures faster with large-neighborhood targets (M = 2,964.3 msec, SE = 313.0 msec) than with small-neighborhood targets (M = 5,121.4 msec, SE = 566.1 msec), t (55) = 3.1, p < .005. For native speakers, the difference was not significant (large neighborhoods: M = 2253.4 msec, SE = 548.6; small neighborhoods: M = 2318.4 msec, SE = 303.3), p > .05. When neighborhood size was regressed on naming latency across items, the relationship between the two was not significant for either native speakers or non-native speakers, p > .05.

Figure 1: Naming accuracy for high-neighborhood and low-neighborhood words in English-German bilinguals and German-English bilinguals.

Figure 2: Naming latencies for high-neighborhood and low-neighborhood words in English-German bilinguals and German-English bilinguals.
Discussion

The present experiment extended the study of phonological neighborhood effects during native language production to non-native language production. Similar to accuracy rates in native language naming, accuracy rates in non-native language naming were better with high-density neighborhoods than with low-density neighborhoods. However, latency results varied across native and non-native languages. While high-density neighborhoods facilitated naming latency in the non-native language, no differences between high- and low-density neighborhoods were found in the native language. This pattern suggests that retrieval difficulties for sparse-neighborhood words may be more marked in non-native speakers, supporting the prediction that language status influences sensitivity to neighborhood density. These findings have implications for mechanisms underlying non-native language processing. For native language processing, previous simulations within an Interactive Spreading Activation Model suggest that neighborhood effects are driven by bi-directional feedback between phonological and lexical levels (Gordon & Dell, 2001). Since current findings for non-native naming closely mirror findings for native naming, it is likely that non-native naming also relies on bi-directional feedback between processing levels. Therefore, late acquisition of a language may not influence the nature of processing dynamics during naming.

Findings that efficiency of naming sparse-neighborhood words was lower in the non-native language than in the native language may be due to differences in proficiency levels. Lower proficiency in a non-native language may have conferred overall ‘low-frequency status’ onto non-native words, rendering non-native naming more sensitive to phonological neighborhood density. Within connectionist models, word frequency has been linked to resting activation levels of word representations (Dell, 1988). Lower resting levels of non-native words may require more activation to accrue before a word can be accessed, therefore reducing efficiency of word retrieval. This reduced efficiency in non-native word retrieval may be especially apparent for words with sparse neighborhoods, where facilitative feedback from neighbors is absent. In sum, results suggest that the role of neighborhood density in processing of a late-acquired language can be explained by mechanisms within the framework of adult native language processing. This account indicates that neighborhood effects are more marked during learning of a non-native language than during learning of a native language. Recall that in children, neighborhood effects are negligible during language learning and increase as language development progresses (Arnold et al., 2005; Newman & German, 2002). This might suggest that neighborhood effects are stronger in a native and highly proficient language than in a non-native and less-proficient language. However, that was not the case in the current study, as the differences between dense- and sparse phonological neighborhoods in non-native speakers were more robust than in native speakers. Therefore, neighborhood density effects appear to manifest themselves differently during the course of first and second language learning. Once a native language has been acquired, neighborhood density effects may follow different patterns with the development of subsequent languages.

In the current study, the absence of latency differences between high- and low-density neighborhoods in the native language may be explained by the small contrast between high- and low-density conditions compared to other studies (e.g., in Garlock et al. 2001, Vitevitch, 2002, Yates et al., 2004). The fact that we found a phonological neighborhood effect on naming accuracy in the current dataset speaks to the robustness of the phenomenon. The fact that we did not find a phonological neighborhood effect on naming latency in a native language suggests that speed of access may be less sensitive to small variations in neighborhood density. It is possible then, that accuracy is more sensitive to even slight variations in neighborhood density, while latency differences are triggered by more dramatic changes, at least in a highly proficient language.

In conclusion, the present study confirmed that phonological neighborhood density influences bilingual lexical access, and extended this finding to non-native language production. Cross-linguistic orthographic neighbors have been found to facilitate lexical access (e.g., De Groot et al., 2002) and our results report the same pattern for phonological neighbors. Although phonological neighborhood density effects were apparent in both the native and the non-native languages, latency results indicated that they were more marked in the non-native language. This confirms the hypothesis that native/non-native language status modulates the effect of neighborhood density on language production. The facilitative effects of phonological neighborhood density on lexical access during language production have applied implications. For instance, in second language instruction, the knowledge that dense neighborhood words are associated with better performance might guide the choice of words in vocabulary learning activities to provide additional support for low-neighborhood items. The present research suggests that the native and non-native languages are subject to similar processing dynamics. Proficiency-based differences between native and non-native language processing can be accounted for by similar mechanisms within an Interactive Activation Model (McClelland & Rumelhart, 1981). Further, preliminary results suggest that the cognitive representations underlying neighborhood effects may develop differently in native and non-native languages, with further experimental and computational work necessary to explore this hypothesis.
Acknowledgments

This work was supported in part by Grants NICHD 1R03HD046952-01A1 and NSF BCS-0418495 to the second author and a Northwestern Graduate Research Grant to the first author. The authors thank Laura Guengerich for assistance with data coding and Margarita Kaushanskaya, Roberto Heredia, Ludmilla Isurin, Olga Boukrina, Nadia Cone, Avital Rabin, as well as three anonymous reviewers, for helpful comments on earlier drafts of this paper.

References


