

Reconciling phonological neighborhood effects in speech production through single trial analysis

Jasmin Sadat ^{a, b, 1}, Clara D. Martin ^{b, c, d}, Albert Costa ^{b, e}, F.-Xavier Alario ^a

^a Laboratoire de Psychologie Cognitive, CNRS & Aix-Marseille Université, France

^b Departament de Tecnologies de la Informació i les Comunicacions, Universitat Pompeu Fabra,
Barcelona, Spain

^c Basque Center on Cognition, Brain Lang, Donostia-San Sebastian, Spain

^d IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

^e Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Running head: Phonological neighborhood effects in speech production

Address correspondence to:

Jasmin Sadat

¹ Now at Department of Psychology,
Royal Holloway, University of London,
Egham, Surrey TW20 0EX,
United Kingdom

Email: jasmin.sadat@upf.edu

Abstract

A crucial step for understanding how lexical knowledge is represented is to describe the relative similarity of lexical items, and how it influences language processing. Previous studies of the effects of form similarity on word production have reported conflicting results, notably within and across languages. The aim of the present study was to clarify this empirical issue to provide specific constraints for theoretical models of language production. We investigated the role of phonological neighborhood density in a large-scale picture naming experiment using fine-grained statistical models. The results showed that increasing phonological neighborhood density has a detrimental effect on naming latencies, and re-analyses of independently obtained data sets provide supplementary evidence for this effect. Finally, we reviewed a large body of evidence concerning phonological neighborhood density effects in word production, and discussed the occurrence of facilitatory and inhibitory effects in accuracy measures. The overall pattern shows that phonological neighborhood generates two opposite forces, one facilitatory and one inhibitory. In cases where speech production is disrupted (e.g. certain aphasic symptoms), the facilitatory component may emerge, but inhibitory processes dominate in efficient naming by healthy speakers. These findings are difficult to accommodate in terms of monitoring processes, but can be explained within interactive activation accounts combining phonological facilitation and lexical competition.

Reconciling phonological neighborhood effects in speech production through single trial analysis

Native speakers of a language know a myriad of different words. This so-called mental lexicon is often described as an interconnected network in which representation distance may depend on meaning or form similarities among the words. One crucial step for understanding this network is to describe which kind of similarity influences language processing and how it modulates performance. In this context, the current research is concerned with the role of phonological similarity in word retrieval and speech production. This issue has been addressed in earlier theoretical work (e.g. Chen & Mirman, 2012; Dell & Gordon, 2003), but the empirical evidence on which this work is grounded remains controversial. Our goal in this article is to clarify the empirical facts regarding form similarity effects in speech production, and to integrate them within a single account. Clearing the empirical constraints will allow further refinement of theoretical models to advance our understanding of the cognitive processes underlying speech production.

An approximation of how similar or interconnected a word is within the lexical network can be obtained by computing its phonological neighborhood density (PhND). PhND refers to the number of words that can be formed from a given word by substituting, adding or deleting one phoneme (Luce, 1986). For example, the word “bat” sounds similar to many other words (e.g., “cat”, “fat”, “rat”, “mat”, “bad”, “sat”, etc.; i.e., dense neighborhood), whereas the word “elk” shares its phonological form with only a few words (“ilk”; “elm”, “elf”, “else”; i.e., sparse neighborhood).

Studies on the role of form similarity in word *recognition* have reported that the identification of a word is influenced by the existence of orthographic or phonological neighbors. Most word recognition theories assume that similar words are co-activated together with the target word and compete for selection, thereby slowing down the recognition process (e.g., Luce & Pisoni, 1998; McClelland & Elman, 1986; Norris, 1994). There is a large number of studies concerning form similarity in visual (e.g., Andrews, 1992, 1989; Carreiras et al., 1997; Forster & Shen, 1996; Grainger, 1990; Grainger et al., 1992, 1989; Grainger & Seguí, 1990; Pollatsek et

al., 1999; Sears et al., 1995) and auditory word recognition (e.g., Dufour & Frauenfelder, 2010; Goldinger et al., 1989; Landauer & Streeter, 1973; Luce, 1986; Luce & Pisoni, 1998; Vitevitch & Luce, 1999, 1998; Vitevitch & Rodríguez, 2005; Ziegler et al., 2003; Ziegler & Muneaux, 2007). In comparison, research on PhND effects in the field of speech production is relatively scarce, and the evidence is quite unclear. It is often thought that, in contrast to word recognition, PhND similarity has a *facilitative* effect in word production (i.e., better performance for words with many similar sounding words; e.g., Chen & Mirman, 2012). Importantly, however, there is also evidence for interference effects (i.e., worse performance for words with many similar sounding words) on naming latencies, (e.g. Vitevitch & Stamer, 2006) and on accuracy measures (e.g. Newman & German, 2005; the evidence will be carefully reviewed later on).

In the remainder of this Introduction, we will first present previous research on the effect of phonological similarity in speech production. We will then describe how speech production models account for the effect of phonological similarity. Finally, we will illustrate the inconsistent pattern of results reported for PhND in speech production.

Phonological influences on word retrieval during language production

Language production research is concerned with how speakers retrieve words from the mental lexicon to express their communicative intentions. The main theories of speech production distinguish between three major levels of processing and mental representations involved in word retrieval (Caramazza, 1997; Dell, 1986; Levelt, 1989). The first step involves the retrieval of a concept that matches the idea to communicate (e.g., the idea of a feline pet). This level of processing is referred to as the semantic level. In addition to the semantic information of the intended concept, semantically similar information is also activated (e.g., concept of related domestic animals like dog, the action to meow, etc.; further distinctions entail taxonomic versus associative semantic relations that will not be considered here). Subsequently, the word corresponding to the highest activated concept has to be retrieved from the mental lexicon, along with its grammatical properties (e.g., “cat”, noun, neuter). This stage is usually referred to as the lexical (or “lemma”) level. Most importantly for our purposes here, there is a consensual distinction between accessing those lexical representations, and accessing the

corresponding phonological make-up of the word (in our example, the sequence /k/, /ae/, /t/).¹ Accomplishing these steps ensures that the relevant information is sent to the motor programming stages and subsequently the articulatory organs, such that the speech signal corresponding to the intended idea is produced eventually.

Given the description above, in which language production proceeds from the semantic to the lexical and then the phonological level, it is not immediately evident that phonologically similar words should affect word retrieval. Yet there are studies indicating that phonological properties do indeed influence word retrieval. A case in point is the mixed error effect (e.g., Dell & Reich, 1981; Harley, 1984; Martin et al., 1996). Mixed errors share both semantic and phonological similarity with the intended word (e.g., “rat” for “cat”). The fact that they occur more often than predicted based on rates of purely semantic (e.g. “cat” for “dog”) or phonological (e.g., “cab” for “cat”) errors signals that phonological information can interact with semantic information to constrain lexical selection. Relatedly, Miceli et al. (1999) showed that semantic errors of aphasic patients could be influenced by recently produced words in the same (oral) or different (written) modality. This finding indicates that word retrieval processes may be constrained by previously activated form information (phonological or orthographical; see also Alario et al., 2003). Further evidence for the influence of phonological properties on word retrieval comes from priming studies (Ferreira & Griffin, 2003; see also Jaeger et al., 2012, and Rapp & Samuel, 2002). Ferreira and Griffin asked participants to name pictures (e.g. of a priest) right after they read sentences that primed a semantic or homophone competitor (e.g., “nun” or “none”). As expected, pictures were more often misnamed after presentation of a semantic priming sentence (i.e., saying “nun” instead of “priest”) than after an unrelated sentence. Notably, the homophone priming sentences had the same effect as the semantic ones. For the word “none” to be a successful intruder, its phonological similarity had to combine with the semantics of its homophone “nun”. The authors suggested that phonological similarity can promote errors at the lexical level. There are important differences between the intrinsic activation of phonological information generated within the language production system and that induced in priming paradigms, where activation stems from prime and target sources. What is

¹ The distinction between the retrieval of a word's lexical representation and its phonological properties captures well the everyday phenomenon known as tip-of-the-tongue state. A speaker in this kind of state has typically retrieved the word's lexical representation, but failed to retrieve its phonology (e.g., Badecker et al., 1995).

common to the observations reviewed above, however, is that they suggest that word retrieval can be influenced by phonological properties.

To account for the fact that phonological properties influence word retrieval, it is often assumed that there are feedback links between the phonological and lexical representational levels in speech production models (i.e., interactivity; Dell, 1986; Harley, 1993; Rapp & Goldrick, 2000). According to such interactive models, phonological relatedness generates increased feedback from shared phonological segments to the lexical layer. As an alternative to feedback links, many of these observations have been explained within strictly feed-forward models of lexical access complemented with monitoring processes (Levelt, 1989; Levelt et al., 1999). These are usually assumed to employ the comprehension system to compare a to-be produced word with its intended form. Phonological similarity of an error to the target word, or of an intended word to other known words, increases the ease and likelihood with which the word may slip through monitoring processes. Summarizing, then, the fact that phonological similarity influences word retrieval, whether it is through interactivity or monitoring mechanisms, motivates a detailed investigation of the role of PhND in speech production.

Phonological neighborhood density effects on naming latencies

We will start by considering the effect of PhND on naming latencies and defer a more thorough discussion of the effect on errors (e.g., in aphasic speakers) to the General Discussion. A seminal study conducted by Vitevitch (2002) examined the role of PhND in speech production of younger adults. He observed that naming latencies were shorter for English picture names drawn from dense neighborhoods (e.g. “bat”, see above) than for those drawn from sparse neighborhoods (e.g. “elk”). Since that study, however, several studies have reported conflicting results in English. In these, either no effect of PhND was observed (e.g., in children: Bernstein Ratner et al., 2009; in adults: Gordon & Kurczek, in press; Newman & Bernstein Ratner, 2007; Vitevitch et al., 2004), or inhibitory effects were recorded on naming latencies (e.g., in children: Arnold et al., 2005; in older adults: Gordon & Kurczek, in press). Importantly, to our knowledge there is no study of healthy speakers that has reported a replication of Vitevitch’s (2002) finding of a facilitative effect of PhND on English naming latencies. Thus at present the effect of PhND on the most studied language (English) is still unclear.

Unfortunately, evidence from other languages is also controversial. Vitevitch and Stamer (2009, 2006) observed an inhibitory effect of PhND in Spanish picture naming, the denser the neighborhood, the longer the picture naming latencies (contrary to the facilitative effect reported in English; Vitevitch, 2002). These contrasting cross-linguistic effects were attributed by Vitevitch and Stamer to differences between the richness of the morphological system of Spanish and that of English (see Ziegler & Perry, 1998, for a similar argument in visual word recognition). However, Baus et al. (2008) challenged the observations of Vitevitch and Stamer. Baus et al. reported a facilitative effect of PhND in Spanish (i.e. faster naming for words from dense than sparse neighborhoods). The authors concluded that the PhND effect is facilitative in Spanish, as had been previously observed by Vitevitch (2002) in English, and they argued for a language-independent interactive spreading activation model of speech production. A similar facilitatory effect of PhND in Spanish has been reported by Pérez (2007), although the focus of his study was on frequency effects (see also Rodriguez-Gonzalez, 2012, for an apparent facilitatory effect in certain cases of verb production). Lastly, in a study conducted with Dutch native speakers, Bien et al. (2005) reported a non-linear effect of PhND, whereby naming responses were fastest for very sparse and very dense PhND of the initial constituents of compound words (although see Tabak et al., 2010, for no PhND effect at all in Dutch). In summary, the available evidence does not provide a consistent pattern for the effect of PhND on naming latencies across various languages (see Table 3 for an overview of studies on PhND effects on speech production latencies, including the novel evidence reported in the current article). Furthermore, the explanations given for the discrepancies between studies (i.e., cross-linguistic properties determining the effect of PhND as argued in Vitevitch and Stamer, versus an interactive speech production system that should always show a facilitatory effect for PhND as argued in Baus et al.) do not seem to provide a satisfactory account to the inconsistent pattern of results that has been observed within and across languages.

The present study

In the present article, our goal is to clarify the effect of PhND in language production by reporting new data, re-analyzing data from previously reported studies and establishing a

detailed summary of the available evidence. This should allow evaluating the theoretical accounts put forward for the effect of this variable.

Below we report a large-scale experiment on word production in Spanish in which we estimated the effect of PhND on naming latencies to test whether co-activation of similar sounding words facilitates or hinders retrieval of the desired word. We did so by measuring naming latencies in a standard picture naming paradigm. In addition, we will report re-analyses of several published data sets on Spanish, French and Dutch.

Our current study incorporates several improvements compared to previous studies on PhND in speech production. First, previous research has been limited to relatively small sets of items in factorial designs with PhND as a between-item manipulation (i.e., comparing performance for words from dense versus sparse neighborhoods; although see Baus et al., 2008, for a cross-linguistic control within “items”). The data set we collected and analyzed is an order of magnitude larger than that of the studies conducted previously (31,980 trials here vs. 1,482 trials in Baus et al., or 1,152 trials in Vitevitch & Stamer, 2006; see however Newman & German, 2005, for testing 1075 individuals on 44 words for a total of 47,300 trials). Second, the data were analyzed with regression models performed at the single trial level, providing a fine-grained approach in which the properties of each individual word and participant are considered explicitly. This approach offers detailed insights beyond the typical condition level afforded by conventional general linear model approaches (e.g., dichotomous ANOVA), for example, to reveal the variability or stability of the effect of PhND at the individual participant level. Third, we included several additional variables in our analyses to control for possible confounds for the effect of PhND. This is important because the previously conflictive findings have often been attributed to poor control of confounding variables. Given the intermingled relationships between these confounding variables, the present study considered them all as potential regression predictors to ensure a reliable and independent effect of PhND. We took special care to address issues of collinearity between the large numbers of variables included, for example by comparing results of alternative analysis and using residualization methods. In doing so, we could test the effect of PhND on speech production while controlling for possible confounding variables and maintaining a large and representative set of stimuli.

Method

Participants

Thirty native Spanish monolinguals participated in the experiment. They were all students from the University of Murcia in Spain. They grew up in Spanish speaking families and used only Spanish for daily communication. All participants had normal or corrected-to-normal vision, and received a monetary reward (20 Euros) for participating in the experiment.

Stimuli

533 black-and-white line drawings of common objects were selected from various picture databases (Bonin et al., 2003; Pérez & Navalón, 2003; Snodgrass & Vanderwart, 1980; Székely et al., 2004). Spanish picture names met the following criteria: (a) they consisted of a single word; (b) they were present in the Spanish database BuscaPalabras (Davis & Perea, 2005); (c) they had no other meanings with higher frequency usage to be confused with (e.g. “tienda”, meaning “tent” or “shop” in English, or “sobre”, meaning “envelope” or “on”/“over”/“about”); and (d) they had relatively high name agreement in Spanish (we only used pictures with name agreement values higher than 60%; this value was taken from Cuetos et al., 1999, if available, or assessed through offline pretests with four Spanish monolinguals from among the university staff). Pictures had black outlines and white surfaces and were presented 300 pixels wide x 300 pixels high on a white rectangle with a monitor resolution of 800 x 600 pixels.

Predictor variables

Previous studies on speech production identified some of the most important predictors of naming performance (e.g., Alario et al., 2004; Barry et al, 1997; Cycowicz et al., 1997; Ellis & Morrison, 1998; Lachman et al., 1974; Severens et al., 2005; Snodgrass & Yuditsky, 1996; Vitkovitch & Tyrrell, 1995). Based on these studies, we included several influential predictors such as name agreement, lexical frequency and age of acquisition to investigate whether PhND would impact speech performance beyond the effects of these variables. In addition, other

variables such as neighborhood frequency (i.e., the average lexical frequency of all neighbors), phonotactic probability (i.e., frequency with which phonological segments occur in a given position in a word), onset density (i.e., number of neighbors that share the onset with the given word), syllable frequency or word length have been proposed as potential confounds for PhND (see Storkel, 2004; Vitevitch et al., 2004; Vitevitch & Sommers, 2003). For example, shorter words tend to have more neighbors than longer words, and words consisting of high frequency syllables or segments with higher phonotactic probabilities tend to have more phonological neighbors than words consisting of low frequency syllables or segments with lower phonotactic probabilities. These potentially confounding variables were also included in our multiple regression analysis.

Altogether, the following item-related predictors were collected from the Spanish lexical database BuscaPalabras (Davis & Perea, 2005) if not specified otherwise:

- Phonological neighborhood density (PhND; range: 0-37 number of neighbor words, $M = 5$, $SD = 7$), as defined in the second paragraph of the Introduction.
- Neighborhood frequency (range: 0-3.03 log occurrences, $M = 0.70$, $SD = 0.69$). It refers to the average lexical frequency of all neighbors. A logarithmic transformation was applied to avoid the undue influence of extreme values in the regression.
- Onset density (range: 0-35 number of onset neighbor words, $M = 4$, $SD = 5$). It refers to the number of neighbors from among the total number of phonological neighbors sharing the first phoneme.
- Word length measured in phonemes (range: 2-11 number of phonemes, $M = 6$, $SD = 2$) and syllables (range: 1-5 number of syllables, $M = 3$, $SD = 1$).
- First syllable frequency (range: 0.08-4.61 log occurrences, $M = 3.33$, $SD = 0.91$). Values were taken from Alario et al. (2010). A logarithmic transformation was applied to avoid the undue influence of extreme values in the regression.
- Phonotactic probability (range: 0-0.03 log occurrences, $M = 0.009$, $SD = 0.005$). Values were taken from the online database CLEARPOND providing phonological

neighborhood information in Spanish (Marian et al., 2012). To standardize, the summed biphone probabilities were divided by word length measured as number of phonemes.

- Word form frequency with values for written (range: 0.07-2.80 log occurrences, $M = 0.98$, $SD = 0.54$) and spoken frequency (range: 0.00-3.12 log occurrences, $M = 1.02$, $SD = 0.60$). A logarithmic transformation was applied to avoid the undue influence of extreme values in the regression.

- Objective visual complexity (range: 4,048-48,874 bytes, $M = 14,715$, $SD = 7,557$). Values were calculated based on the compressed JPEG file size (Székely & Bates, 2000).

- Concreteness (range: 2.7-6.9, $M = 5.8$, $SD = 2.7$), familiarity (range: 2.7-7.0, $M = 5.8$, $SD = 2.7$), and imageability (range: 2.8-6.9, $M = 6.1$, $SD = 2.8$) ratings were taken from Cuetos et al. (1999). The ratings for these three variables were obtained on a one to seven scale, with one indicating low, and seven indicating high concreteness/ familiarity/ imageability. These values were available only for a subset of the 533 words (concreteness and familiarity: 385 words; imageability: 377 words).

- Name agreement (range: 10-100 percentage, $M = 85$, $SD = 18$), number of alternative names (range: 0-8 number of alternative responses, $M = 1$, $SD = 1$), and H statistics which accounts for the proportion and spread of alternative responses (range: 0-2.6 unit of response agreement, $M = 0.4$, $SD = 0.5$; for details see Snodgrass & Vanderwart, 1980). Values were established on the basis of the analysis of the data we collected (see below for response scoring).

- Subjective estimates of age of acquisition from adult ratings (range: 2.3-10.3 years; $M = 4.6$, $SD = 1.4$). Questionnaires consisted of four randomized lists of all picture names together with a 1 to 7 point rating scale (Barbón Gutiérrez & Cuetos Vega, 2006), where 1 corresponded to an age of acquisition before 2 years old and 7 to an age of acquisition after 12 years old. They were filled in by 50 graduate Spanish monolinguals from different universities.

Design and procedure

Six experimental lists were created with each of the 533 pictures being presented twice. The order of presentation of the pictures was pseudo-randomized with the following restrictions: (a) the 1066 pictures appeared in two successive sets of 533, with each picture presented only once per set; (b) picture names in two successive trials were neither semantically (i.e., objects were not from the same semantic category) nor phonologically related (i.e., picture names did not share the initial phoneme nor did they rhyme). Participants were randomly assigned to one of the six lists.

Participants were tested in a sound-proof room. Stimulus presentation and the software voice-key were controlled via DMDX (Forster & Forster, 2003). The sensitivity of the voice-key was adjusted for each participant. Each trial started with a fixation cross displayed at the centre of the computer screen for 500 ms. After a 300 ms blank screen, the picture of the object to name was displayed. The picture remained on the screen until either the voice key detected the response or a 2500 ms deadline was reached without any overt response detected. The next trial began 700 ms after the recording period finished.

The experiment consisted of a short training followed by two sessions that were separated by a break of 15 minutes. First, participants were asked to name eight practice black-and-white pictures similar to the materials used in the experiment. They were instructed to name the pictures as fast and as accurately as possible using single nouns. After that, in the first session, they had to name the first set of 533 object pictures divided into eight short blocks. The responses were monitored by the experimenter and if participants gave another name for the picture than the intended one, they were corrected by the experimenter at the end of the first session.² In the second session, the same 533 pictures were presented in the same way as in the first session, but in a different order. Participants' responses were automatically recorded by the computer as digitized sound files, and errors were noted online by the experimenter. Each session lasted about 45 minutes. In total, the experiment lasted about two hours.

²This follows common practice in the picture naming literature where participants are familiarised with the task and the materials in a first naming block (e.g., see “specified elicitation” in Bock, 1996). Note that the first session is often considered for familiarization purposes only, and performance is not analyzed. As we will show below, session did not interact with the variables of interest (see also Alario et al. 2004) and thus we report the analysis collapsing the two sessions.

Response scoring

All 31,980 vocal responses and onset markers (533 pictures x 2 presentations x 30 participants) were visually checked offline with the software Check-Vocal (Protopapas, 2007) and corrected if necessary. Responses other than the intended one were classified as non-target responses (i.e., categories 2 to 4 below) or errors (i.e., categories 5 to 7 below) and excluded from naming latency analyses. To establish name agreement values from the present responses and to properly characterize the naming behavior of the participants, responses were coded into seven categories:

1. The produced name was the target name.
2. The response was a morphological variant of the target name, defined as a variation that shares the word root without changing the word's core meaning. Examples are clippings (e.g., "tele" for "television"- television) or plural/singular alternations (e.g., "ojos" for "ojo"- eyes / eye).
3. The response was a synonym for the target name, not sharing its word root (e.g., "frigorifico" for "nevera" - fridge).
4. The response included hyponyms (e.g. "fruta"-fruit for "naranja"- orange), semantic coordinates that share the same class but do not have the target word's core meaning (e.g., "cebolla"- onion for "ajo"- garlic), part-whole relations at the visual-semantic level (e.g., "mano"- hand for "brazo"- arm), and visual errors (e.g., "cacahuete"- peanut for "patata"- potato).
5. The response was the target name, but included a phonological error.
6. The response was the target name, but started with a hesitation (e.g., "aeh") followed by participant's auto-correction.
7. The response was omitted or was a non-word.

Data analysis

Due to the large number of variables known to affect naming performance and their potentially high level of collinearity, three statistical procedures were applied to select the most important variables and to reduce collinearity. We first used a hierarchical clustering analysis to assess the correlations among predictors (using the “Hmisc” package in R and the “varclus” function; Harrell et al., 2010). Secondly, we ran a random forest analysis on naming latencies (e.g., Breiman, 2001) using the package “party” in R and the function “cforest” (Hothorn et al., 2006; Strobl et al., 2007; Strobl et al., 2008). A random forest analysis is a non-parametric type of data analysis, in which a single measure of importance is provided for each predictor through a collection of classification trees (see Tagliamonte & Baayen, 2012, for details and application to linguistic data). All the predictor variables listed above were included in the random forest analysis on naming latencies. When several variables appeared to be highly correlated in the cluster analysis, the one variable having the highest measure of variable importance provided by the random forest analysis was selected to represent the cluster. Before entering the selected predictors in the main linear mixed-effects models described below, we systematically tested for correlations between them. Predictors that were correlated above $r = 0.25$ were orthogonalized by running a linear regression, which allowed isolating the variance specific to one of the predictors in the form of a residual. The residuals of these linear models were entered as fixed effects in the linear mixed-effects model. For example, since our focus of analyses was to explore the influence of PhND, we tried to remove from PhND any source of variance that was shared with other correlated variables. We regressed PhND against phoneme length and neighborhood frequency, and the residuals of this model were used as PhND_{res} variable (i.e., residualized PhND). In addition, phoneme length was residualized against neighborhood frequency (yielding $\text{phoneme length}_{\text{res}}$, and leaving the original neighborhood frequency variable), and onset density was residualized against PhND (i.e., yielding $\text{onset density}_{\text{res}}$; see the General Discussion section for a consideration of the reverse regression). Finally, to reduce the influence of highly related variables inherent to the estimation of age of acquisition ratings, lexical frequency and name agreement were regressed out of age of acquisition (i.e., yielding $\text{age of acquisition}_{\text{res}}$). This way, a minimal number of residualized variables were selected in the final linear mixed-effects model, and all important confounding variables could be included in the

model without introducing unreasonable amounts of collinearity.³ We attempted to use a minimal number of residualized variables in the analysis to preserve the interpretability of our results. For this reason, when a residualized variable was entered into the model together with the variable against which it was residualized and the latter was not significant, the model was recalculated taking out non-significant variables and using the original variable.⁴

Finally, we used linear mixed regression models to analyze the naming latencies and accuracy rates. One important aspect of the applied analyses is that regression models are performed at the single trial level, and not on averages (Baayen et al., 2008; Jaeger, 2008). In addition to fixed predictors considered in simple linear regressions, linear mixed-effects models account for random variation induced by specific words or speakers. Several models were fitted and compared step-wise by means of log likelihood tests, removing non-significant predictors, to identify the optimal linear mixed-effects model. For the analysis, we introduced by-participant random intercepts, by-participant random slopes for all significant fixed effects estimated within participants (except for the control variable trial order), by-item random intercepts, and finally by-item random slopes for session (see Barr et al., 2013).⁵ Additional predictors were always entered in the models before PhND. This procedure ensured that any effect of the theoretically central predictor of PhND was significant over and above the variation explained by other secondary predictors.

All statistical analyses were run with the statistical software R (R Development Core Team, 2011) and linear mixed-effects models were computed with the package “lme4” in R (Bates et al., 2011). P-values were validated by Markov chain Monte Carlo simulations using the

³ We also conducted an analysis in which the predictors were normalized (i.e., centered and scaled), a procedure that is often thought to reduce collinearity (but see Belsley, 1984; Dalal & Zickar, 2012). This analysis yielded the same results as those obtained from non-normalized orthogonalized predictors.

⁴ As a final check, for the main statistical models described below, we computed and report variance inflation factors (VIF) of each of the predictors to ensure reliability of the analyses (using the function “vif” in the package “HH” in R, Heiberger, 2009). VIF indicates how much the variance of an estimated coefficient is increased due to collinearity in the regression model (VIF values around 2 are considered as problematic for a regression coefficient).

⁵ A log-likelihood test indicated significant improvement by including such random effect structure over the simple random intercept effect model. Trial order could not be included as a random effect, because the model did not converge. Although the contribution of some of the fixed effects (e.g., session) was reduced by including random slopes, it never resulted in the disappearance of a main effect.

function “pvals.fnc” in the package “languageR” (Baayen, 2008) whenever possible (i.e., in the case of models without random slopes). The Box-Cox test (Box & Cox, 1964; using the function “boxcox” in the package “MASS” in R, Venables & Ripley, 2002) indicated that the reciprocal transformation of the latencies was the most appropriate transformation for the data to reduce skewness and approximate a normal distribution. The Box-Cox test is commonly used in linear mixed-effect modeling to find the most appropriate transformation function that produce normally distributed residuals (Gurka et al., 2006; Gurka et al., 2007; see also Baayen & Milin, 2010). Instead of the suggested simple reciprocal transformation (i.e., 1/RT), we used -1000/RT to facilitate the reading and interpretation of our results. Note that by applying this transformation, the normality assumption of the data is met and order is preserved. We followed Baayen’s (2008) procedure of model criticism in which trials whose standardized residual value is above 2.5 are removed and the model is recomputed. Only such recomputed models are reported.

Results

Predictor selection

The statistical procedures described above led us to choose the set of predictors for the linear mixed-effects analysis of our data. Figure 1 provides a graphical overview of (a) the correlation structure among the full set of predictors and (b) their importance values in predicting naming latencies. From every cluster of highly correlated variables for the same concept (e.g. word length measured in phonemes and syllables), we selected the one with the highest measure of variable importance provided by the random forest analysis (e.g., phoneme length). Based on these explorations, we excluded word length measured in syllables, number of alternative names, H statistics, and spoken lexical frequency. Finally, the following nine variables were retained for further analysis: PhND, phoneme length_{res}, neighborhood frequency, phonotactic probability, onset density_{res}, first syllable frequency, name agreement, age of acquisition_{res}, written lexical frequency, visual complexity. In addition, variables coding for trial order presentation (from 1 to 533) and session (first or second) were included to capture statistical variance in the linear mixed-effects analysis.

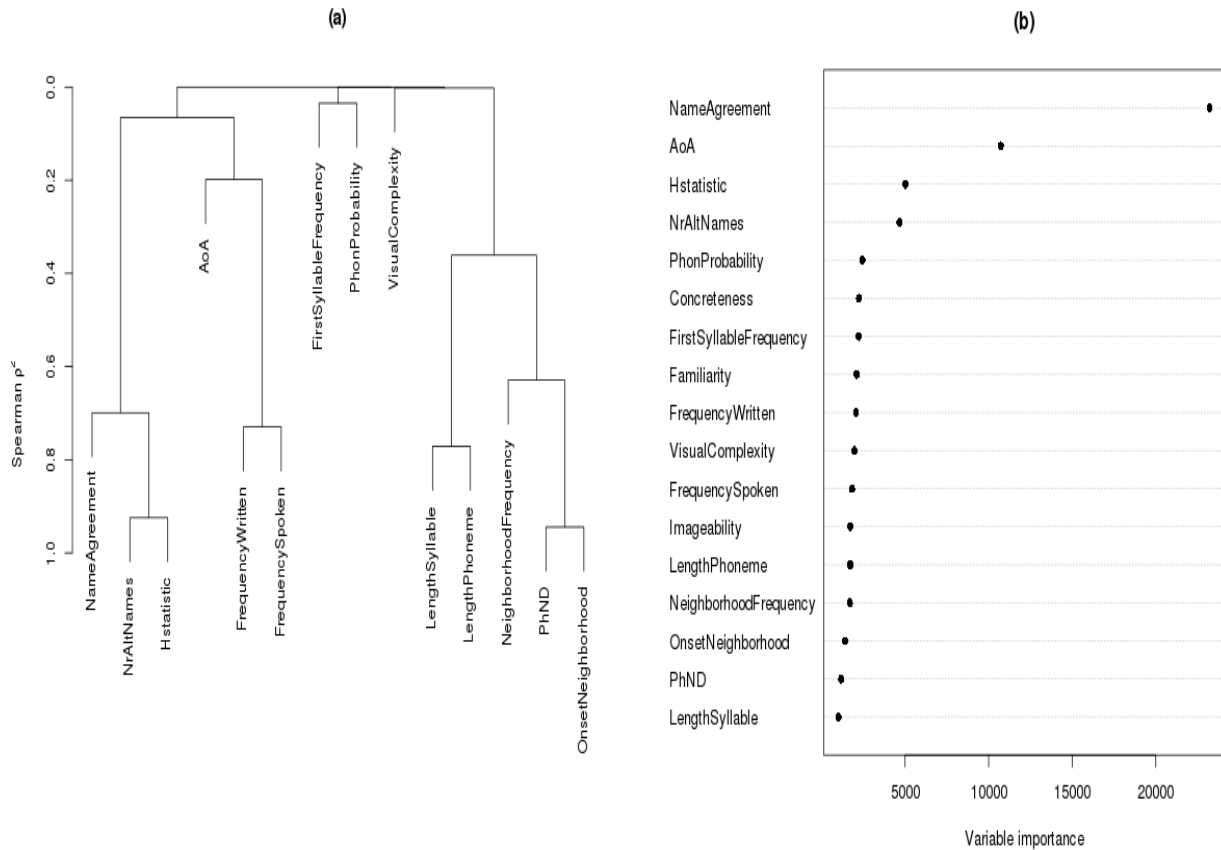


Figure 1: (a) Hierarchical clustering analysis using Spearman’s ρ^2 of confounding variables with PhND and theoretically motivated predictors for naming latencies in word production. (b) Variable importance plot obtained through random forest analysis for the same set of predictors. AoA = age of acquisition; NrAltNames = number of alternative names; PhonProbability = phonotactic probability; PhND = phonological neighborhood density.

Naming latencies

After removing non-target responses (i.e., only category 1 responses were used) and model outliers (see Methods), 27 581 responses (86%) remained for analyses. The average naming latency was 910 ms (SD = 117, range: 704-1146 ms). Statistical values for the fixed and random effects in the final linear mixed-effects model together with the variance inflation factors

of the predictors are reported in Table 1 (see Appendix for further details on the correlation matrix of the random effects structure). Results for the main variable of interest PhND_{res} showed that naming latencies increased significantly with the phonological density of a word's neighborhood (i.e., inhibitory PhND effect; see Figure 2). In addition, Figure 3 presents an estimate of the variation of this effect across individuals. Note that the size of the PhND effect increases with increasing response speed of the participants which is a common observation in psycholinguistic experiments (e.g., Baayen & Milin, 2010; Balota & Ferraro, 1996, 1993; Cerella, 1985; Sadat et al., submitted; Spieler & Balota, 2000). All participants showed an inhibitory effect (i.e. the individual estimates for the PhND effect were all above zero).

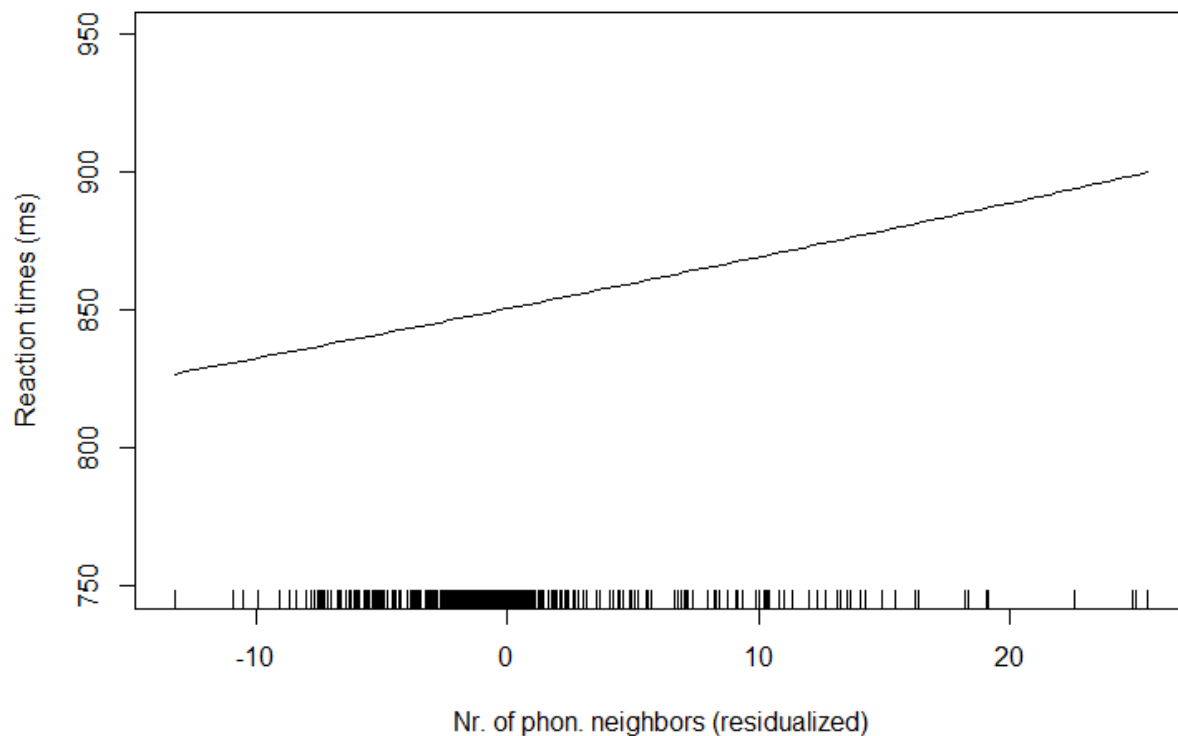


Figure 2: Partial effect of residualized phonological neighborhood density on naming latencies.

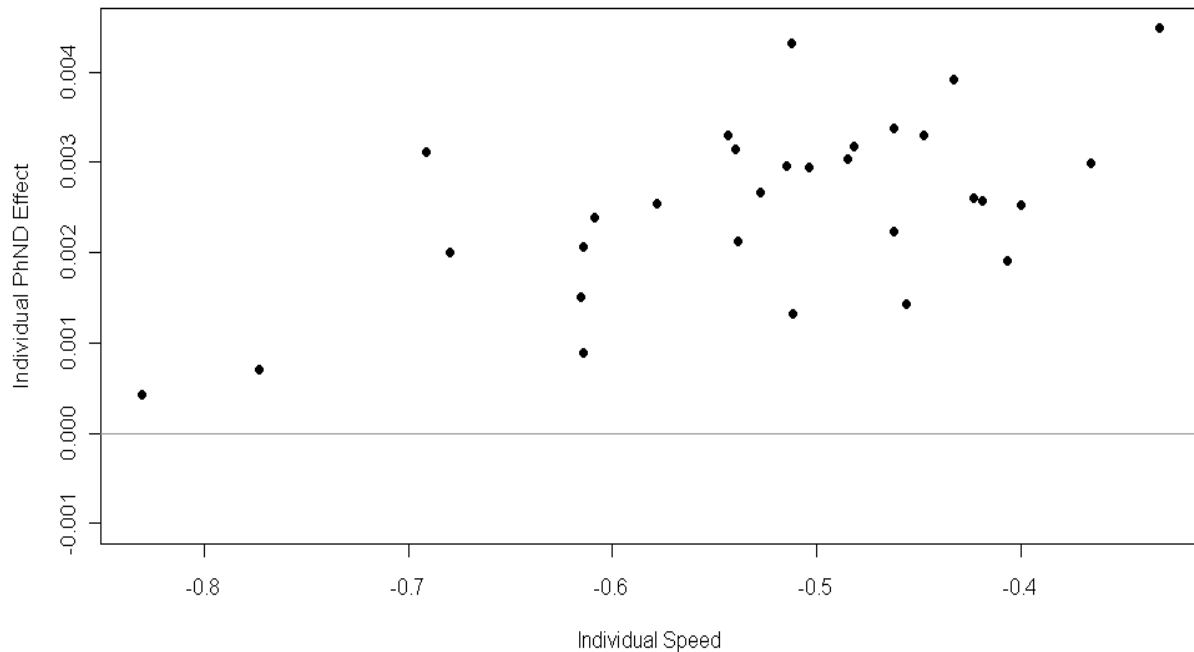


Figure 3: Individual coefficients estimates of the final linear mixed-effects model. Relationship between magnitude of phonological neighborhood density (PhND) effect and overall speed estimated at the individual level. Each point represents a participant. Larger values in the x-axis indicate slower responding participants. Larger values in the y-axis indicate positive coefficients for the PhND effect (i.e. stronger inhibition). Notably, all participants showed an inhibitory effect.

As expected, name agreement, age of acquisition_{res}, and lexical frequency contributed significantly towards predicting naming speed: earlier learned words were named faster than later learned ones, words with high percentages of name agreement were named faster than words with lower percentages, and high frequency words were named faster than low frequency ones. The effects of trial order and session were significant, showing that responses to pictures became slower with increasing trial order and that responses in the second session were faster than in the first. The effect of neighborhood frequency was marginally significant: words with more frequent neighbors tended to be named slower than words with less frequent neighbors. The effects of visual complexity, phoneme length_{res}, first syllable frequency, phonotactic

probability, and onset density_{res} were not significant. Predictors not showing significant effects were removed from the model.

Table 1: Variance and Standard deviation (SD) for each of the random effects, and beta coefficients, effect magnitude in ms, standard errors (SE), t- and p-values together with the variance inflation factor (VIF) for each of the fixed effects in the final linear mixed effect model on naming latencies

Random effects		Variance	SD				
Item	Intercept	2.17×10^{-2}	0.15				
	Session	6.25×10^{-3}	0.08				
Participant	Intercept	1.44×10^{-2}	0.12				
	Session	2.06×10^{-3}	0.05				
	Nagr	1.10×10^{-6}	< 0.01				
	AoA _{res}	2.30×10^{-4}	0.02				
	Frequency	1.08×10^{-3}	0.03				
	NeighFreq	7.99×10^{-5}	0.01				
	PhND _{res}	1.44×10^{-6}	< 0.01				
Residual		4.91×10^{-2}	0.22				
Fixed effects	Predictor range	Effect magnitude	raw β	SE β	t-value	p-value	VIF
Intercept			-5.26×10^{-1}	3.35×10^{-2}	-15.70	< 0.001	
Session		40.93	-1.26×10^{-1}	9.38×10^{-3}	-13.40	< 0.001	1.0
Trial	1 - 533	0.07	9.20×10^{-5}	9.14×10^{-6}	10.06	< 0.001	1.0
Nagr	10 - 100	9.54	-6.97×10^{-3}	3.54×10^{-4}	-19.68	< 0.001	1.1
AoA _{res}	-4 - 4	19.93	2.74×10^{-2}	5.26×10^{-3}	5.21	< 0.001	1.1
Frequency	0 - 3	29.74	-4.26×10^{-2}	1.21×10^{-2}	-3.51	0.001	1.1
NeighFreq	0 - 3	10.37	1.41×10^{-2}	8.12×10^{-3}	1.73	0.045	1.1
PhND _{res}	-13 - 25	1.92	2.53×10^{-3}	1.01×10^{-3}	2.50	0.008	1.0

Note: Nagr = name agreement; AoA_{res} = residualized age-of-acquisition; Frequency = lexical Frequency; NeighFreq = neighborhood frequency; PhND_{res} = residualized phonological neighborhood density; effect magnitude shows the increase in ms per scaling unit for each of the predictors

Since ratings for concreteness, familiarity and imageability were available only for a subset of the 533 words, we conducted an additional analysis on this restricted data set (20 492 data points; 74%). The same random effect structure as in the previous model was included. When entering familiarity, imageability, and concreteness into the model, along with the other predictors, the effect of PhND_{res} remained significant ($\beta = 2.34 \times 10^{-3}$, $\text{SE} = 1.11 \times 10^{-3}$, $t[20091] = 2.10$, $p = 0.021$; $\text{VIF} = 1.0$). The effect of concreteness was marginally significant ($\beta = -1.68 \times 10^{-2}$, $\text{SE} = 1.03 \times 10^{-2}$, $t[20091] = -1.63$, $p = 0.056$; $\text{VIF} = 1.3$). The effects of familiarity and imageability were not significant.

Analysis of accuracy rates

Responses other than the intended one were classified as non-target responses (i.e., categories 2 to 4 below) or errors (i.e., categories 5 to 7 below) and excluded from naming latency analyses.

There were 88% of correct responses (category 1; 28,046 trials out of 31,980), 5% of responses containing speech errors (categories 5, 6, and 7; a total of 1,681 trials), and 7% of non-target responses (categories 2, 3, and 4; a total of 2,253 trials). Most of the non-target responses (5%) were of category 4 (i.e., semantic variations around the target, see Methods). For the error analysis, we excluded non-target responses, and correct responses were contrasted with error responses to accurately predict the probability of an error-free response. Thus the total amount of responses considered in the error analysis was 29,727 trials (correct: 28,046 and errors: 1681). The average error rate across participants was 6% ($SD = 3$, range: 2-14%; 1,681 out of 29,727 trials). The logistic mixed-effects model, computed over 28,930 trials after model outlier removal (3%), estimated how the predictors modulate the odds of error.⁶ Statistical values for the final mixed-effects model together with the variance inflation factors of the predictors are reported in Table 2. The odds of error were smaller for words that have higher name agreement, higher lexical frequency, and are learned earlier. The effect of session was significant, showing that the odds of error were less in the second than in the first session. There was no significant

⁶The control variable trial order could not be included as fixed or random effect, and session could only be included as fixed effect, because the model would not converge.

effect of phoneme length_{res}, first syllable frequency, phonotactic probability, neighborhood frequency, onset density_{res}, or PhND.

Table 2: Variance and Standard deviation (SD) for each of the random effects, and beta coefficients, standard errors (SE), z- and p-values together with the variance inflation factor (VIF) for each of the fixed effects in the final linear mixed effect model on accuracy rates

Random effects		Variance	SD			
Item	Intercept	1.58	1.26			
Participant	Intercept	0.99	1.00			
	Nagr	< 0.01	0.02			
	AoA _{res}	0.07	0.26			
	Frequency	1.23	1.11			
	PhND _{res}	< 0.01	0.04			
Fixed effects		raw β	SE β	z-value	p-value	VIF
	Intercept	5.39	3.35	12.88	< 0.001	
	Session	-2.43	0.11	-21.81	< 0.001	1.0
	Nagr	-0.13	0.01	-20.52	< 0.001	1.2
	AoA _{res}	0.52	0.10	5.06	< 0.001	1.1
	Frequency	-1.39	0.33	-4.27	< 0.001	1.2
	PhND _{res}	0.03	0.03	1.06	0.29	1.0

Note: Nagr = name agreement; AoA_{res} = residualized age-of-acquisition; Frequency = lexical frequency; NeighFreq = neighborhood frequency; PhND_{res} = residualized phonological neighborhood density

Discussion

The results of this large scale experiment show that increasing PhND has a detrimental effect on naming latencies, i.e. an inhibitory effect. This is the case even when many other important variables affecting the speed of speech production are taken into account. In addition, our study incidentally confirms other findings, showing that name agreement, age of acquisition,

and lexical frequency were the three most influential variables for predicting naming behavior (for a review see Alario et al., 2004).

Further evidence from chronometric studies

Our main finding is in line with Vitevitch and Stamer's (2009, 2006) results, and supports the view of increased competition during the production of dense PhND words in Spanish. Our results are at odds with those of Baus et al. (2008) and Pérez (2007) who reported a facilitative effect of PhND in Spanish naming performance - the same language of response we used in our experiment. Given the discrepancy of the present results with these two findings, we conducted several re-analyses of our data, the data of Baus et al. and that of Pérez. Moreover, we conducted analyses of the effect of PhND on three sets of chronometric data from previously published papers that were available to us: a naming study in Spanish (Sadat et al., 2012), a large-scale naming study in French (Alario et al., 2004), and an additional large-scale naming study in Dutch (Severens et al., 2005; see Table 3 for an overview of re-analyses).

Reanalysis of Baus et al. (2008)

Our set of 533 object pictures comprised most of the pictures used in the study by Baus et al. (2008). Thus, we selected from our dataset the trials that corresponded to those pictures and applied (1) the same analyses as in the original study (Student t-test). We also applied (2) the analyses used in the current study to this subset of trials from our dataset and (3) to the original data from Baus et al. The Box-Cox test (Box & Cox, 1964) indicated a logarithmic transformation of the data of Baus et al. as the most appropriate to approximate a normal distribution. In the following re-analyses, linear mixed model analyses were conducted with PhND and all relevant control variables (phoneme length, neighborhood frequency, first syllable frequency, onset density, name agreement, age of acquisition, lexical frequency, visual

complexity, session and trial order) and in case of multicollinearity the same residualization procedures were used as in the main analyses.⁷

Four stimuli of the set used by Baus et al. (2008) were not part of our stimuli (two of each of their PhND categories, high versus low). Applying the same statistical analysis (Student t-test) used by Baus et al. to our data led to the same pattern of results that they reported: words coming from dense neighborhoods were named faster than words coming from sparse neighborhoods ($t[29] = 2.8$, $p < 0.01$; magnitude of the effect in our subset: 32 ms and in Baus et al.: 33 ms). This means that for an almost identical set of stimuli as used in Baus et al. (and even with more observations: 1,920 trials in our subset versus 1,238 trials in Baus et al.), our data also showed a facilitative effect of PhND. However, when using linear mixed model analysis the previously observed facilitative effect of PhND disappeared in the original data of Baus et al. ($\beta = 1.41 \times 10^{-3}$, $SE = 2.25 \times 10^{-3}$, $t[1199] = 0.63$, $p = 0.27$; $VIF = 1.0$) and showed a trend towards inhibition in our selected subset ($\beta = -6.61 \times 10^{-3}$, $SE = 4.11 \times 10^{-3}$, $t[1881] = 1.61$, $p = 0.08$; $VIF = 1.1$; number of observations after model outlier removal: 1892 trials; 99%). Note that in the re-analyses of both sets, the effect of word length was significant, indicating increasing naming latencies with increasing word length. Thus, the disappearance of the facilitatory effect could be due to a better control of confounding variables by including them in the model explicitly as well as a better fit of individual variations and residual noise often present in such relatively small datasets.⁸ In support of this reasoning, Baayen (2010) showed that the dichotomization of continuous predictors leads to a loss of power. Consequently, results of small scale studies are difficult to generalize. On the basis of further examples (see below), we will argue that the effect of PhND is often small and difficult to detect. More powerful designs with a large number of trials and more fine-grained analysis methods like in the present study provide better tools to reveal such subtle effects. For now we note that the facilitatory effect reported by Baus et al. did not survive a fine-grained analysis.

⁷ Note that random by-participant and by-item slopes were introduced only when log-likelihood tests indicated significant improvement in the model's fit. Including these individual slopes never resulted in the disappearance of a fixed effect.

⁸ A reviewer suggested that the disappearance of the facilitatory effect might be due to the transformation of the data (i.e., normalizing the data distribution in regression but not Student t-tests). We conducted the same analysis as in Baus et al. (2008) on the log-transformed data and still obtained a significant facilitative effect. This means that the applied data transformation is not responsible for the disappearance of the effect when analyzing the data of Baus et al. by means of linear mixed-effects models.

Reanalysis of Pérez' subset (2007)

Additional previous evidence for a facilitative effect of PhND in Spanish naming was reported by Pérez (2007). Although the focus of his study was on frequency effects, the author also reported a facilitative effect of PhND in Spanish speech production. We selected the subset of our data ($N_{\text{items}}=146$) that overlapped with the stimuli by Pérez ($N_{\text{items}}= 161$). Applying the same statistical analysis (linear regression) as used by Pérez to our data did not yield the same results as he reported: there was no effect of PhND on naming latencies ($\beta = 1.5$, $SE = 1.2$, $t[133] = 1.2$, $p = 0.23$; $VIF = 1.1$). This means that for a very similar set of stimuli and same analyses as used in Pérez (and even with more observations: 8,092 trials in our subset versus 5,937 trials in Pérez), our data did not replicate the facilitatory effect of PhND reported by Pérez. However, when introducing the predictor variables used by Pérez (H statistics, age of acquisition, written lexical frequency, image variability, familiarity, visual complexity, and PhND) together with the variables trial order and session into a linear mixed-effects model, the re-analyses of our subset showed a marginally significant inhibitory PhND effect ($\beta = 2.70 \times 10^{-3}$, $SE = 1.5 \times 10^{-3}$, $t[7742] = 1.77$, $p = 0.04$; $VIF = 1.0$; number of observations after model outlier removal: 7,766 trials; 98%). In short, our subset of data similar to Pérez' stimuli showed an inhibitory effect when analyzed with fine-grained measures on more observations.

Reanalysis of Sadat et al. (2012)

To gain further insights, we considered a third set of Spanish picture naming data that had not been tested previously for an effect of PhND (Sadat et al., 2012). In addition to the above mentioned control predictors, we entered the variables speaker group and task from the original study into the linear mixed-effects model.⁹ We used $-1000/RT$ for transforming the latencies as indicated by the Box-Cox test (Box & Cox, 1964). The results on 21,264 trials (98 % after model outlier removal) showed an inhibitory effect of $PhND_{\text{res}}$ on naming latencies (effect magnitude: 2.1 ms per unit, predictor range: 0-26; $\beta = 6.01 \times 10^{-3}$, $SE = 2.71 \times 10^{-3}$, $t[21252] = 2.22$, $p = 0.02$; $VIF = 1.0$).

⁹ We used the data from the two tasks (bare noun and noun phrase production) and the two speaker groups (mono- and bilingual) because there were no significant interactions with PhND.

Reanalysis of Alario et al. (2004)

We explored speech production data from another Romance language besides Spanish using a large set of French naming data (Alario et al., 2004). We introduced the variables of the original study (H statistics, image agreement, lexical frequency, syllable length, visual complexity) together with session (first or second), first syllable frequency (taken from Alario et al., 2010) and PhND (from the French database LEXIQUE; New et al., 2001) into a linear mixed-effects model. The Box-Cox test (Box & Cox, 1964) indicated a logarithmic transformation of the data of Alario et al. as most appropriate. Since word length and PhND were highly correlated in this dataset ($r = -0.78$), we chose to include the original PhND variable and to enter word length as residualized fixed effect. The model showed a significant inhibitory effect of PhND (effect magnitude: 1.2 ms per unit, range: 0–62; $\beta = 1.41 \times 10^{-3}$, $SE = 6.12 \times 10^{-4}$, $t[16248] = 2.30$, $p = 0.01$; VIF = 1.2) and no effect of word length_{res} on naming latencies of 16,258 trials (98% after model outlier removal). However, when the opposite residualization was performed (i.e., PhND was residualized against word length), there was no effect of PhND_{res} ($\beta = 7.74 \times 10^{-4}$, $SE = 7.77 \times 10^{-4}$, $t[16248] = 1.00$, $p = 0.25$; VIF = 1.0), whereas the effect of word length was significant, indicating decreasing naming latencies with increasing word length (as reported in the original study). This means that the effect of PhND in this dataset can be explained by the variance common to both word length and PhND. However, since the direction of the observed length effect is opposite to what would be given most theoretical accounts (i.e., word length was facilitatory), we are inclined to attribute the effect to PhND.

Reanalysis of Severens et al. (2005)

Finally, we explored the effect of PhND_{res} in a non-Romance language, using a large-scale naming dataset in Dutch (Severens et al., 2005). We introduced the variables of the original study (H statistics, age of acquisition, lexical frequency, phoneme length) together with PhND_{res}, summed neighborhood frequency (from the Dutch database CELEX; Baayen et al., 1993), and trial order into a linear mixed-effects model. We used $-1000/RT$ for transforming the latencies as indicated by The Box-Cox test (Box & Cox, 1964). The results on 14,277 trials (98% after model outlier removal) showed no effect of PhND_{res} at all on naming latencies ($\beta = 5.87 \times 10^{-4}$,

SE=1.1x10⁻³, t[14247] = 0.55, p = 0.29; VIF = 1.0). As suggested by the study by Bien et al. (2005) in Dutch, we also explored the data for non-linear effects of PhND_{res}. The results indicated a non-significant linear component and a marginally significant quadratic term with a positive coefficient (i.e., a U-shaped PhND_{res} effect; $\beta = 2.00 \times 10^{-4}$, SE=1.3x10⁻³, t[14524] = 1.56, p = 0.07; VIF = 1.2).¹⁰ This result hints to an inhibitory component of the PhND effect that is present even in a non-Romance language like Dutch. This result is relevant with respect to Vitevitch and Stamer's (2006) argument that the effect of PhND may vary depending on specific language properties as mentioned in the Introduction. Part of the General Discussion is dedicated to discuss this possibility (cf. "Are there cross-linguistic differences in the direction of the PhND effect?").

Summary of re-analyses results

We re-analyzed the data from five previously published studies to test for the influence of PhND on naming latencies. In the case of Baus et al. (2008), we replicated their findings with their statistical method, but showed that with finer-grained analysis, the previously reported facilitatory effect disappeared. Regarding the data of Pérez (2007), we could not replicate the facilitatory effect of PhND, but observed an inhibitory effect with more detailed analysis. Likewise, when exploring the data of Sadat et al. (2012) for PhND we observed an inhibitory effect. The reanalysis of naming data in French (Alario et al., 2004) provided suggestive evidence to support the claim of an inhibitory effect of PhND on speech production. The results from Dutch did not show any clear effect, but indicated a trend towards inhibition related to a non-linear component of the PhND effect. Overall, our results together with the re-analyses reported above support the view proposed by Vitevitch and Stamer (2009, 2006) that PhND shows an inhibitory effect in morphologically rich languages like Spanish and French. The evidence previously reported in favor of a facilitative effect in Spanish was shown to be inconclusive. Possible explanations of these discrepancies could be due to a better control of confounding variables in the present study and a better fit of individual variations and residual noise as compared to small datasets (as argued in Baayen, 2010).

¹⁰Note that none of the PhND effects reported in the previous analyses showed significant non-linear effects.

General Discussion

In the present study we performed statistical analysis at the single trial level (Baayen et al., 2008) to assess the impact of PhND on word production performance. The result of a picture naming paradigm with a very large set of materials revealed an inhibitory effect of PhND on naming latencies. In addition, our results on individual participant effects show that PhND influences speech production in a systematic way. Five previous studies were reanalyzed using these statistical methods, which, for a variable that is intrinsically defined as “between-items”, are arguably more fine-grained than those used in the original reports. The new analyses challenged previous conflicting findings from two studies and provided new results from three published datasets. These analyses converged on showing PhND inhibition, while significant facilitation was never observed. Thus, we conclude that the effect of PhND on word production latencies has an inhibitory component in the languages considered (Spanish, French, and Dutch). This broad-based analysis of PhND effects on naming latencies together with the evidence at the individual participant level sets important constraints for theoretical models of speech production. Below we will elaborate on how current models can accommodate the empirical findings.

Regarding accuracy rates, our results showed no significant effect of PhND. While such effects are usually reported in naming performance of patients, the absence of an effect in healthy speakers like in the current study is not uncommon (e.g., Vitvitch & Stamer, 2006). Given our findings of an inhibitory PhND effect on naming latencies, we can speculate about a similar influence of PhND on accuracy measures of healthy speakers (e.g., Newman & German, 2005, 2002, for behavioral evidence) under certain circumstances (see below for a discussion of specific task effects) .

Are there cross-linguistic differences in the direction of the PhND effect?

Our findings are in line with those of other studies assessing the influence of PhND on naming latencies of word production in Spanish (Vitevitch & Stamer, 2009, 2006). In English, however, Vitevitch (2002) showed a facilitatory effect of PhND. On the contrary, following that seminal article, several studies in the speech production literature have since then reported an

inhibitory effect in this language (see Table 3 for an overview of studies on PhND effects on naming latencies). For example, when assessing children's speech production latencies in English, naming was slower for words from dense phonological neighborhoods compared to sparse ones (e.g., Arnold et al., 2005; Munson et al., 2005). The authors appealed to developmental change considerations to explain the discrepancy between their inhibitory effect in children and the facilitatory result reported in adults by Vitevitch. The idea was that the impact of PhND changes with vocabulary growth during development. Further studies reported an inhibitory effect of PhND in English speech production latencies (Luce & Pisoni, 1998; Munson et al., 2005; Taler et al., 2010; Vitevitch & Luce, 1998). In these studies, participants were asked to repeat as fast as possible a word or sentence presented acoustically (i.e., shadowing task). The shadowing task involves both speech perception and speech production, and hence may reflect contributions of either process (Bates & Liu, 1996). For example, shadowing experiments are used to investigate the relationship between input and output phonological representations (e.g., Mitterer & Ernestus, 2008; Taler et al., 2010). Therefore, the inhibitory effect may stem in part from production processes, yet attributing the results from shadowing studies to perception versus production processes should be done cautiously.

To our knowledge, there is no study besides Vitevitch (2002) that has reported a facilitative effect of PhND on English naming latencies (for error rates see below). Further studies in this language would be useful to test the robustness of these findings. In view of the relatively small data set sizes of previous studies, we highlight that large-scale studies will certainly aid in clarifying such effects. We speculate on the basis of admittedly partial evidence that in an English large-scale naming study the effect of PhND should turn out to be inhibitory. In fact, an inhibitory effect of PhND has very recently been observed in older adults naming pictures in English (Gordon & Kurczek, in press). In our view, then, cross-linguistic differences in the direction of the PhND effect are not to be expected.

Table 3: Overview of speech production studies of the effects of phonological neighborhood density on naming latencies in healthy speakers

Study	Task	Population	Language	Design	Trials	Direction of PhND effect
Bernstein Ratner et al. (2009)	Picture naming	children (control group)	English	Factorial	660	ns
Arnold et al. (2005)	Picture naming	children (control group)	English	Factorial	180	Inhibition
Frank et al. (2007)	Picture naming	young adults	Artificial	Factorial	3840	Inhibition
Tabak et al. (2010)	Picture naming	young adults	Dutch	Continuous	2890	ns
Reanalysis of Severens et al. (2005)*	Picture naming	young adults	Dutch	Continuous	22191	ns
Jescheniak & Levelt (1994)	Picture naming	young adults	Dutch	Continuous	nr	ns
Vitevitch (2002) - Exp. 4	Picture naming	young adults	English	Factorial	1200	Facilitation
Vitevitch (2002) - Exp. 3	Picture naming	young adults	English	Factorial	1632	Facilitation
Vitevitch et al. (2004) - Exp. 3	Picture naming	young adults	English	Factorial	1100	ns
Gordon & Kurczek (2013)	Picture naming	younger and older adults	English	Continuous	14600	Inhibition, but ns for youngest adults
Newman & Bernstein Ratner (2007)	Picture naming	adults (control group)	English	Factorial	1100	ns
Reanalysis of Alario et al. (2004)*	Picture naming	young adults	French	Continuous	20720	Inhibition
Marian et al. (2008) - Exp. 1	Picture naming	young adult bilinguals in first and dominant language	German	Factorial	855	ns
Pérez (2007)*	Picture naming	young adults	Spanish	Continuous	6440	Facilitation, but see reanalysis in current study
Baus et al. (2008) - Exp. 2*	Picture naming	young adults	Spanish	Factorial	1482	Facilitation, but see reanalysis in current study
Vitevitch & Stamer (2009)	Picture naming	young adults	Spanish	Factorial	22400	ns (reanalysis of Bates et al., 2003), but trend for inhibition
Vitevitch & Stamer (2006)	Picture naming	young adults	Spanish	Factorial	1152	Inhibition
Reanalysis of Sadat et al. (2012)*	Picture naming	young adults	Spanish	Continuous	11200	Inhibition
Current study*	Picture naming	young adults	Spanish	Continuous	31980	Inhibition
Vitevitch (2002) - Exp. 5	Button-press followed by naming	young adults	English	Factorial	1200	Facilitation
Munson et al. (2005)	Word repetition	younger children	English	Factorial	512	ns
Munson et al. (2005)	Word repetition	older children	English	Factorial	480	Inhibition
Vitevitch & Luce (1998)	Shadowing	young adults	English	Factorial	4500	Inhibition
Luce & Pisoni (1998) - Exp. 3	Shadowing	young adults	English	Factorial	7200	Inhibition
Taler et al. (2010)	Sentence repetition	younger and older adults	English	Factorial	4440	Inhibition

Note: PhND = phonological neighborhood density; Exp. = experiment; ns = not significant; nr = not reported; asterisk indicates studies assessed in the current paper

Considering finer grained modulations of neighborhood density effects

PhND may not be detailed enough to characterize the phenomena under scrutiny. Vitevitch and Stamer (2006) suggested that the difference between language production performance in Spanish and in English lies in the location within the word where neighbors can be formed in either language (e.g., frequently the rhyme in English, but often the onset in Spanish: rhyme vs. onset density). A study by Vitevitch et al. (2004) showed that when PhND is controlled, English words with a high number of onset neighbors are produced more slowly than words with few onset neighbors. This suggests that phonological neighborhood effects in English may have several facets: overall neighborhood has a facilitatory contribution (Vitevitch, 2002) whereas onset neighborhood has an inhibitory one. Given the highly inflectional nature of Spanish, the inhibitory effect of onset density could overpower an overall facilitation of PhND.

Our analyses did not reveal independent effects of onset density and PhND. Admittedly, the relative contributions of PhND and onset density are difficult to tease apart, because the latter count is a subset of the former (see definition in Methods), and the two are highly correlated in Spanish and in our materials ($r = 0.97$; 82 % of all neighbors were onset neighbors). In all the models we reported, onset density was residualized against PhND (i.e. $\text{onset density}_{\text{res}}$). This means that the variance common to both variables was attributed to PhND. We also ran a series of models in which the residualization was performed the opposite way (i.e., PhND was residualized against onset density). In these models, onset density had a significant inhibitory effect, whereas PhND_{res} was no longer significant.¹¹ This indicates that the reported inhibitory effect can be attributed to the variance common to PhND and onset density. The results of the random forest analysis performed with unresidualized variables showed that onset density was ranked as being of only slightly higher importance towards predicting naming latencies than PhND (see Figure 1b). Relatedly, note that Bien et al.'s (2011) study showed independent effects for onset cohort (facilitatory) and rhyme neighbors (inhibitory) on naming latencies in Dutch. This is in conflict with the pattern of effects described above for English, where onset neighborhood had an inhibitory effect. Note however that the paradigm used by Bien et al. (i.e., a position-response association task) differs largely from standard picture naming; it involves

¹¹ In particular, PhND was not facilitatory in this analysis, as could have been suggested by the dual-faceted phenomenon present in English (see above).

recently learned arbitrary stimulus-response associations and a large number of item repetitions. In short, the interactions of PhND effects with position (e.g. onset, rhyme, and overall neighbors) still deserve further empirical clarifications.

Another property that could modulate the effect of PhND is semantic and morphological similarity. Due to the inflectional properties of Spanish, semantically similar words often bear high phonological resemblance, with differences in word affixes being smaller (e.g., “mejor” [better] -“mejorar” [improve]; see Arbesman et al. 2010, for a comparison of Spanish and English morphological networks in relation to PhND). Phonologically similar words are thus likely to receive additional activation through semantic similarity. Depending on further assumptions, this may increase competition at stages of lexical selection. Thus, one general caveat regarding an independent effect of phonological similarity is to consider the morphological structure of the words. To better characterize how the PhND effect relates to semantic similarity, future studies will need to take into account measures like morphological family size (again see Bien et al., 2011, for some steps in this direction).

Opposing effects of neighborhood density

Based on the present results and the review of previous studies suggesting an inhibitory effect of PhND on onset latencies in naming tasks, one may wonder why the common assumption in language production research is that increasing PhND facilitates word production. This is for example the assumption made by Chen and Mirman (2012; see also Dell & Gordon, 2003) when they modeled PhND effects across modalities. It appears that many previous studies of PhND effects in speech production assessed accuracy measures like word finding rates or speech errors, rather than naming latencies. We now briefly review this evidence, first considering healthy speakers and then aphasic patients.

With healthy speakers, words from dense neighborhoods tend to be less amenable to tip-of-the-tongue (ToT) states than words from sparse neighborhoods (e.g., Brennen et al., 1990; Brown & McNeill, 1966; Burke et al., 1991; Harley & Bown, 1998; James & Burke, 2000; Vitevitch & Sommers, 2003). This observation is consistent with Meyer and Bock's (1992) findings that phonologically related cues help the resolution of ToT states, compared to unrelated

ones (semantically related cues do not show any benefit). Somewhat in contrast, several studies suggest that higher PhND is associated with poorer accuracy in healthy children and adults (German & Newman, 2004; Newman & German, 2005, 2002; but see Mirman et al., 2010, for a facilitatory effect of PhND in accuracy measures of the control group).¹²

Regarding aphasic speakers, it has been repeatedly reported that PhND has a facilitatory effect on their spoken productions. Overall, aphasic speakers tend to make fewer phonological errors in words from dense than sparse PhND (e.g., Best, 1995; Goldrick et al., 2010; Gordon, 2002; Kittredge et al., 2008; Middleton & Schwartz, 2010; Mirman et al., 2010). However, it seems that an important distinction can be made by considering the effect of PhND across error types. When errors are classified into semantic, non-word, and formal errors based on the relationship between the outcome and the intended utterance, it appears that only semantic and non-word rates show a facilitatory effect of PhND (e.g., Mirman et al., 2010, in the case of non-words). For example, Gordon (2002) reported that patients produced more formal errors on words from dense PhND than on words resulting in non-word errors. Similarly, Laganaro et al. (in press) distinguished formal errors in their analysis and observed an inhibitory effect of PhND (see also Table 1 in Dell & Gordon, 2003, for a simulation showing that with increasing PhND density, semantic and non-word errors decreased, while formal errors slightly increased).

In short, healthy speakers are less likely to enter in ToT states for words from higher PhND, yet a few studies suggest overall decreased accuracy for higher PhND. In aphasic speakers, semantic and non-word errors decrease with PhND whereas, conversely, formal errors increase with PhND. We now attempt to provide an integrative view of these findings and our own in the context of the theoretical models of speech production.

¹² Using a task purposely designed to elicit phonological errors from healthy speakers (the SLIP paradigm), Vitevitch (2002) reported a facilitative effect of phonological neighborhood. However, Stemmer (2004) used the same paradigm and found no overall effect of neighborhood size (see also Vitevitch, 1997, for analysis of spontaneous speech errors with opposing results). Since the effect of PhND in error elicitation paradigms appears to be inconclusive, we will not include it in our interpretations.

Theoretical interpretation of neighborhood density effect(s) in language production models

The first general kind of account we will consider relies on the monitoring processes commonly included in feed-forward models of speech production (Levelt, 1989; Levelt et al., 1999). Nootboom's (2005) description of these processes explicitly postulated that the monitor carries out comparisons based on phonological similarity between the error and the target word. The more an error sounds like its intended counterpart, the more likely it is to be mistaken and to slip through the monitor (see also Slevc & Ferreira, 2006). When a formal error is committed for a word from a dense neighborhood, it is more likely to result in another real word. Accordingly, errors stemming from dense neighborhood words would be less likely to be filtered out by the monitoring system. This prediction is in line with the inhibitory PhND effect showing that aphasics produce more formal errors on words from dense than sparse neighborhoods (e.g., Laganaro et al., in press). Thus, it is possible that the monitoring account can accommodate for some of the findings of PhND effects in aphasic speech error patterns.

Considering now error-free production, words from dense neighborhoods are overall more word-like within the lexical network, and thus they might pass the monitor more easily (or more quickly) compared to words from sparse neighborhoods. This prediction is at odds with the results we report. Alternatively, a monitoring account of neighborhood effects could build on the general hypothesis that monitoring processes rely on the speech comprehension system. This could imply that the PhND effect in speech production mirrors the effects observed in comprehension processes. However, in studies of speech comprehension, PhND manipulations have shown cross-linguistic differences, with facilitatory PhND effects in auditory word recognition in Spanish (Vitevitch & Rodriguez, 2005) and inhibitory PhND effects in English (e.g., Vitevitch & Luce, 1998). Thus, such an account could explain the inhibitory effect of PhND in English speech production, but could not explain the inhibitory effect in Spanish that is reported in the current article. In short, without further explicit assumptions, the monitoring account does not seem to provide a valid account for PhND effects in word production.

An alternative are the lexical network accounts proposed by Dell and Gordon (2003) or Chen and Mirman (2012), which are based on the basic principles of interactive activation applied to speech production. PhND effects are assumed to result from feedback between shared phonological representations and the intended lexical representation. For a word with many

phonological neighbors, many shared phonological representations will send feedback that contributes to activating the word to be produced. This results in facilitation. Note however that the phonological segments of the target word send activation to all lexical items with similar phonology; in some cases these co-activated items can enhance lexical competition. Our results showed that there is a clear inhibitory component that arises through phonological activation in speech production. This is important given that, up to now, computational models of speech production only concentrated on the facilitatory effect of PhND. In what follows we elaborate on how the prevalent interactive activation models can fit the empirical grounds that were delineated in the current study.

Dell and Gordon (2003) investigated how the opposing dynamics at the phonological and lexical levels were modulated by certain model parameters intended to reflect functional impairments (i.e. activation decay and weights). They showed that, in general, the model accounts for facilitatory PhND effects in accuracy rates. With certain model parameters, however, the neighbors reached an activation level sufficiently close to that of the target word. Such activation of the phonological neighbors hampered lexical selection, hence yielding an inhibitory PhND effect on accuracy (see Table 3 in Dell & Gordon, 2003). As Dell and Gordon suggested, in a parenthesized comment, “when the facilitation provided by neighbors is reduced, as in the weight-lesioned model, the formal neighbors do become serious contenders.”

The opposing facilitatory and inhibitory forces present in this model can account for the complex pattern of error type modulation by PhND, which contributes to a decrease in semantic and non-word formal errors while increasing formal errors (see the simulations based on “lesioned” networks in Dell & Gordon, 2003). The overlapping phonemes boost the activation of the target word, compared to semantically related candidates that are not phonologically related. On average, this helps preventing semantic errors and substitutions of non-target phonemes. Regarding formal errors, the same mechanism also increases the likelihood of mis-selecting a highly active phonological similar competitor during lexical access. This opposing effect of PhND across error types has been reported in aphasic speakers (e.g., Gordon, 2002; Laganaro et al., in press; Middleton & Schwartz, 2010).

The same model can also account for the “protective effect” of PhND on TOT occurrence. A TOT state is classically described as a situation in which lexical access is

disrupted mainly at the stage of phonological retrieval (Badecker et al., 1995; for a review, see Chapter 9 in Brown, 2012). This means that while a lexical representation has been successfully singled-out, the retrieval of its full phonological content is impaired. Such a situation can be described as one of efficient lexico-semantic processing but inefficient phonological retrieval. Words from dense neighborhoods should benefit from increased feedback from a more interconnected phonology, while being preserved from additional lexical competition precisely because of the efficient lexico-semantic processing. In addition, while feedback from the phonological layer may be weaker during a TOT, the lexical level should not suffer from lexical competition given that the target has been successfully identified. Increased PhND can only result in facilitation of word retrieval and easier ToT resolution.

In our results on naming latencies, formal neighbors induce an inhibitory effect and thus appear to be “serious contenders” in Dell and Gordon's (2003) terms. Yet their simulations based on lesioned models would constitute an implausible account for the data we report, given the high accuracy rates of our healthy young participants. In addition, their unlesioned model predicts facilitatory effects on error rates. This being said, given the opposing facilitatory and inhibitory forces present in the model, it is not unlikely that a quantitative adaptation of the model to naming latencies would account for the findings we report.

Chen and Mirman (2012) further developed the above described interactive activation model of language production to account for inhibitory as well as facilitatory effects of neighbors in naming latencies. One major difference between this model and Dell and Gordon's (2003) is the inclusion of lateral inhibition connections at the lexical level, the strength of which is scaled by a sigmoid function. In their simulations of error-free word production, Chen and Mirman's model showed a facilitatory effect of phonological neighbors on naming latencies. This is because, in their implementation, the effect of the feedback connections from the phonological to the lexical level is damped by the sigmoid function. Importantly, under other circumstances (e.g. reading), the dynamic properties of the model *do* generate inhibitory neighborhood effects, i.e. if there is activation feeding back from the phonological to the lexical level. Chen and Mirman attributed the difference between facilitatory and inhibitory effects to a contrast between highly and weakly activated neighbors, amplified by the sigmoid function. Altogether, this means that through a more extensive exploration of the model's parameter space, the inhibitory effect of

PhND on naming latencies of the present study could be accounted for. This would be the case if, for example, neighbors in word production were modeled as highly active neighbors (as in reading).

In sum, the interactive activation model(s) with the appropriate parameter settings can likely account for the full range of findings. This is because the two forces that underlie PhND effects are inherent to the model: facilitatory feedback connections between the phonological and lexical level and a selection mechanism that is slowed by increased co-activation of lexical competitors.¹³ Overall, the extent to which the competition mechanism surfaces in word production performance appears to be determined, more broadly, by the parameter settings of the lexical network. These may depend on the similarity distance among the neighbor words (e.g., Chen & Mirman, 2012), the communication context (e.g. Jaeger et al., 2012), developmental changes (e.g., German & Newman, 2005; Gordon & Kurczek, in press), and possibly the particular language (although see above the discussion of cross-linguistic differences). Given the evidence presented in the current study and the review of the accuracy literature, we conclude that the interactive activation model provides an appropriate account of the full range of variations in the direction of PhND effects in language production.

Relationship to other phonological effects in word production tasks

There are two additional findings that have to be discussed in the light of an inhibitory PhND effect. Both findings have been described as facilitatory effects driven by phonological similarity, and therefore they are in apparent contrast with the inhibitory effect observed in the present study. First, bilingual speakers name cognate words (translations having a high phonological overlap like “tomate” [Spanish] -“tomato” [English]) faster than non-cognates (e.g., “mesa” [Spanish] – “table” [English]; Costa et al., 2000). This means that the high phonological overlap across translations facilitates naming performance. The current PhND findings could only be considered consistent with the facilitative cognate effect if one assumes

¹³ In addition to the above, a mechanism of competition has sometimes been postulated at the phonological level (Dell, 1988). The requirement to select phonemes in a “left-to-right” sequential fashion from among the activated phonemes may trigger increased competition for words sharing onsets than those sharing endings (O’Seaghdha & Marin, 2000; see also Sevald & Dell, 1994; Sullivan & Riffel, 1999; Wheeldon, 2003).

that the lexical representations of a bilingual's two languages do not engage in cross-linguistic lexical competition (see Runnqvist et al., 2012, and Sadat et al., submitted, for discussion).

The second observation refers to a phonological facilitation effect reported in studies using the picture-picture paradigm (e.g., Morsella & Miozzo, 2002; Navarrete & Costa, 2005). In this task participants are asked to name a picture while ignoring a superimposed distractor picture whose name is phonologically-related to the target. Naming latencies were shorter for pictures that bear a phonological relationship with the distractor picture than for pictures that have no relation. This finding indicates that speakers access the phonological information of unattended stimuli and that this overlap in phonology facilitates the production of the intended word. One caveat here is that the phonological relationship between the tested words is not as precisely defined as in standard picture naming studies. Results of these picture-picture studies would be more comparable to the present results if the picture words would be either phonological neighbors (i.e., differing by only one phoneme) or controlled for positional overlap (i.e., only differing at the onsets or endings). Moreover, although participants were selectively attending to the target picture, the representation of the distractor was nevertheless processed at early perceptual stages (Bles & Jansma, 2008). As already noted in the Introduction, one may need to differentiate between paradigms in which phonologically similar words are externally pre-activated or primed, from paradigms in which phonological activation spreads naturally within the network (i.e., without explicit distractors or primes as in the present study). One way in which this distinction might be captured in word production models is by keeping track of the origin of the activation within the lexical network, through flagging and verification procedures (e.g. Roelofs, 2003).

Conclusion

While it is commonly acknowledged that form similarity influences word production, the nature of the effect remained disputed. In the present study, we showed that phonological neighborhood density has a significant inhibitory effect on naming latencies, and that this finding is reliable across participants and studies. The effect of form similarity on accuracy measures is more complex. Phonological neighborhood density promotes form-related errors, but it helps

reducing semantic and non-word errors as well as tip-of-the-tongue states. This pattern is consistent with the idea that phonological neighborhood generates two opposite forces, one facilitatory and one inhibitory. In cases where speech production is disrupted (e.g. certain aphasic symptoms), the facilitatory component may emerge. By contrast, in efficient naming by healthy speakers, inhibitory processes dominate. These findings are difficult to accommodate in terms of monitoring processes, but can be explained within interactive activation accounts combining phonological facilitation and lexical competition. While this interpretation awaits explicit quantitative modeling our research establishes the empirical ground for PhND effects in speech production.

Acknowledgements

This research was supported by a grant from the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013 Grant agreement n° 263575), three grants from the Spanish government (PSI2008-01191, PSI2011-23033, Consolider Ingenio 2010 CSD2007-00012), and the Catalan government (Consolidado SGR 2009-1521). We thank the Brain and Language Research Institute and the Fédération de Recherche 3C (both at Aix-Marseille Université) for institutional support. Jasmin Sadat was supported by a pre-doctoral fellowship from the Spanish Government (FPU-2008). Clara Martin was supported by the Spanish Government (Grant Juan de la Cierva) and is now supported by the Basque Foundation for Science (IKERBASQUE) and the BCBL institution. We thank Laure Minier and Elena Guerra for checking a portion of the vocal responses, and Luis Fuentes for providing access to the laboratory.

References

- Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U.H., Seguí, J., 2004. Predictors of picture naming speed. *Behav Res Methods Instrum Comput*, 36, 1, 140-155.
- Alario, F.-X., Goslin, J., Michel, V., Laganaro, M., 2010. The functional origin of the foreign accent evidence from the syllable-frequency effect in bilingual speakers. *Psychol Sci*, 21, 1, 15-20.
- Alario, F.-X., Schiller, N.O., Domoto-Reilly, K., Caramazza, A., 2003. The role of phonological and orthographic information in lexical selection. *Brain Lang*, 84, 3, 372-398.
- Andrews, S., 1989. Frequency and neighborhood effects on lexical access: Activation or search? *J Exp Psychol Learn Mem Cogn*, 15, 5, 802.
- Andrews, S., 1992. Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *J Exp Psychol Learn Mem Cogn*, 18, 2, 234.
- Arbesman, S., Strogatz, S.H., Vitevitch, M.S., 2010. Comparative analysis of networks of phonologically similar words in English and Spanish. *Entropy*, 12, 3, 327-337.
- Arnold, H.S., Conture, E.G., Ohde, R.N., 2005. Phonological neighborhood density in the picture naming of young children who stutter: Preliminary study. *J Fluency Disord*, 30, 2, 125-148.
- Baayen, R. H. (2010). A real experiment is a factorial experiment? *Ment Lex*, 5, 149-157.
- Baayen, R.H., Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3(2), 12-28.
- Baayen, R.H., 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.
- Baayen, R.H., Davidson, D.J., Bates, D.M., 2008. Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang*, 59, 4, 390-412.
- Baayen, R.H., Piepenbrock, R., Van Rijn, H., 1993. *The CELEX lexical data base on CD-ROM*. Linguistic Data Consortium, Philadelphia, PA.
- Badecker, W., Miozzo, M., Zanuttini, R., 1995. The two-stage model of lexical retrieval: Evidence from a case of anomia with selective preservation of grammatical gender. *Cognition*, 57, 2, 193-216.

- Balota, D.A., Ferraro, F.R. (1993). A dissociation of frequency and regularity effects in pronunciation performance across young adults, older adults, and individuals with senile dementia of the Alzheimer type. *J Mem Lang*, 32, 573-573.
- Balota, D.A., Ferraro, F.R. (1996). Lexical, sublexical, and implicit memory processes in healthy young and healthy older adults and in individuals with dementia of the Alzheimer type. *Neuropsychology*, 10(1), 82.
- Barbón Gutiérrez, A., Cuetos Vega, F., 2006. Efectos de la edad de adquisición en tareas de categorización semántica. *Psicológica: Revista De Metodología y Psicología Experimental*, 27, 2, 207-224.
- Barr, D.J., Levy, R., Scheepers, C., Tily, H.J., 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J Mem Lang*, 68(3), 255–278.
- Barry, C., Morrison, C.M., Ellis, A.W., 1997. Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *Q J Exp Psychol A*, 50, 3, 560-585.
- Bates, E., D'Amico, S., Jacobsen, T., Szekely, A., Andonova, E., Devescovi, A., et al. (2003). Timed picture naming in seven languages. *Psychon Bull Rev*, 10, 2, 344–380.
- Bates, D., Maechler, M., Bolker, B., 2011. lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-39. <http://CRAN.R-project.org/package=lme4>
- Bates, E., Liu, H., 1996. Cued shadowing. *Lang Cogn Process*, 11(6), 577–581.
- Baus, C., Costa, A., Carreiras, M., 2008. Neighbourhood density and frequency effects in speech production: A case for interactivity. *Lang Cogn Process*, 23, 6, 866-888.
- Belsley, D. A. (1984). Demeaning conditioning diagnostics through centering. *Am Stat*, 38, 73–82.
- Bernstein Ratner, N., Newman, R., Streckas, A., 2009. Effects of word frequency and phonological neighborhood characteristics on confrontation naming in children who stutter and normally fluent peers. *J Fluency Disord*, 34, 4, 225-241.
- Best, W., 1995. A reverse length effect in dysphasic naming: When elephant is easier than ant. *Cortex*, 31(4), 637-652.
- Bien, H., Baayen, R.H., Levelt, W.J.M., 2011. Frequency effects in the production of Dutch deverbal adjectives and inflected verbs. *Lang Cogn Process*, 26, 4-6, 683-715.

- Bien, H., Levelt, W.J.M., Baayen, R.H., 2005. Frequency effects in compound production. *PNAS*, 102, 49, 17876-17881.
- Bles, M., Jansma, B., 2008. Phonological processing of ignored distractor pictures, an fMRI investigation. *BMC Neurosci*, 9, 1, 20.
- Bock, J. K. (1996) Language production: Methods and methodologies. *Psychon Bull Rev*, 3, 395-421.
- Bonin, P., Peereman, R., Malardier, N., Méot, A., Chalard, M., 2003. A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behav Res Methods*, 35, 1, 158-167.
- Box, G.E.P., Cox, D.R. (1964). An analysis of transformations. *J R Stat Soc Series B Stat Methodol*, 26(2), 211–252.
- Breiman, L., 2001. Random Forests. *Mach Learn*, 45, 5-32.
- Brennen, T., Baguley, T., Bright, J., Bruce, V., 1990. Resolving semantically induced tip-of-the-tongue states for proper nouns. *Mem Cognit*, 18, 4, 339-347.
- Brown, A.S., 2012. *The tip of the tongue state*, Psychology Press, New York.
- Brown, R., McNeill, D., 1966. The “tip of the tongue” phenomenon. *J Verbal Learning Verbal Behav*, 5, 4, 325-337.
- Burke, D. M., MacKay, D. G., Worthley, J. S., Wade, E., 1991. On the tip of the tongue: What causes word finding failures in young and older adults? *J Mem Lang*, 30, 5, 542-579.
- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cogn Neuropsychol*, 14, 1, 177-208.
- Carreiras, M., Perea, M., Grainger, J., 1997. Effects of orthographic neighborhood in visual word recognition: Cross-task comparisons. *J Exp Psychol Learn Mem Cogn*, 23, 4, 857–871.
- Cerella, J. (1985). Information processing rates in the elderly. *Psychol Bull*, 98(1), 67.
- Chen, Q., Mirman, D., 2012. Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychol Rev*, 119, 2, 417-430.

- Costa, A., Caramazza, A., Sebastián-Gallés, N., 2000. The cognate facilitation effect: Implications for models of lexical access. *J Exp Psychol Learn Mem Cogn*, 26, 5, 1283-1296.
- Cuetos, F., Ellis, A.W., Alvarez, B., 1999. Naming times for the Snodgrass and Vanderwart pictures in Spanish. *Behav Res Methods*, 31, 4, 650-658.
- Cycowicz, Y.M., Friedman, D., Rothstein, M., Snodgrass, J. G., 1997. Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *J Exp Child Psychol*, 65, 2, 171-237.
- Dalal, D.K., Zickar, M.J., 2012. Some common myths about centering predictor variables in moderated multiple regression and polynomial regression. *ORM*, 15, 339–362.
- Davis, C.J., Perea, M., 2005. BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behav Res Methods*, 37, 4, 665-671.
- Dell, G.S., Reich, P.A., 1981. Stages in sentence production: An analysis of speech error data. *J Verbal Learning Verbal Behav*, 20, 6, 611-629.
- Dell, G.S., 1986. A spreading-activation theory of retrieval in sentence production. *Psychol Rev*, 93, 3, 283-321.
- Dell, G.S., 1988. The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *J Mem Lang*, 27, 2, 124-142.
- Dell, G.S., Gordon, J.K., 2003. Neighbors in the lexicon: Friends or foes? In Schiller, N., & Meyer, A. (Eds.). *Phonetics and phonology in language comprehension and production: Differences and similarities*. Germany: Mouton de Gruyter.
- Dufour, S., Frauenfelder, U.H., 2010. Phonological neighbourhood effects in French spoken-word recognition. *Q J Exp Psychol*, 63, 2, 226-238.
- Ellis, A.W., Morrison, C.M., 1998. Real age-of-acquisition effects in lexical retrieval. *J Exp Psychol Learn Mem Cogn*, 24, 2, 515-523.
- Ferreira, V.S., Griffin, Z.M., 2003. Phonological influences on lexical (mis) selection. *Psychol Sci*, 14, 1, 86-90.
- Forster, K.I., Forster, J.C., 2003. DMDX: A windows display program with millisecond accuracy. *Behav Res Methods Instrum Comput*, 35, 1, 116-124.

- Forster, K.I., Shen, D., 1996. No enemies in the neighborhood: Absence of inhibitory neighborhood effects in lexical decision and semantic categorization. *J Exp Psychol Learn Mem Cogn*, 22, 3, 696-713.
- Frank, A.F., Tanenhaus, M.K., Aslin, R.N., Salverda, A.P., 2007. Frequency, Neighborhood Density, and Phonological Similarity Effects in Picture Naming: An Artificial Lexicon Study. *Proc 29 Annu Conf Cogn Sci Soc*, 995-1000.
- German, D.J., Newman, R.S., 2004. The impact of lexical factors on children's word-finding errors. *J Speech Lang Hear Res*, 47, 3, 624.
- Goldinger, S.D., Luce, P.A., Pisoni, D.B., 1989. Priming lexical neighbors of spoken words: Effects of competition and inhibition. *J Mem Lang*, 28, 5, 501-518.
- Goldrick, M., Folk, J.R., Rapp, B., 2010. Mrs. Malaprop's neighborhood: Using word errors to reveal neighborhood structure. *J Mem Lang*, 62, 2, 113-134.
- Gordon, J.K., Kurczek, J. (2013). The aging neighborhood: Phonological density in naming. *Lang Cogn Process*.
- Gordon, J.K., 2002. Phonological neighborhood effects in aphasic speech errors: Spontaneous and structured contexts. *Brain Lang*, 82, 2, 113-145.
- Grainger, J., 1990. Word frequency and neighborhood frequency effects in lexical decision and naming. *J Mem Lang*, 29, 2, 228-244.
- Grainger, J., O'Regan, J.K., Jacobs, A.M., Seguí, J., 1989. On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Atten Percept Psychophys*, 45, 3, 189-195.
- Grainger, J., O'Regan, J.K., Jacobs, A.M., Seguí, J., 1992. Neighborhood frequency effects and letter visibility in visual word recognition. *Atten Percept Psychophys*, 51, 1, 49-56.
- Grainger, J., Seguí, J., 1990. Neighborhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies. *Atten Percept Psychophys*, 47, 2, 191-198.
- Gurka MJ, Edwards LJ, Muller KE, Kupper LL. 2006. Extending the Box-Cox transformation to the linear mixed model. *J Roy Stat Soc A*. 2:273-288.
- Gurka, M. J., Edwards, L. J., & Nylander-French, L. (2007). Testing transformations for the linear mixed model. *Comput Stat Data Anal*, 51, 4297-4307.

- Harley, T.A., 1984. A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cogn Sci*, 8, 3, 191-219.
- Harley, T.A., 1993. Phonological activation of semantic competitors during lexical access in speech production. *Lang Cogn Process*, 8, 3, 291-309.
- Harley, T.A., Bown, H.E., 1998. What causes a tip-of-the-tongue state? Evidence for lexical neighbourhood effects in speech production. *Br J Psychol*, 89, 1, 151-174.
- Harrell, F.E.Jr., with contributions from many other users, 2010. Hmisc: Harrell Miscellaneous. R package version 3.8-3. <http://CRAN.R-project.org/package=Hmisc>
- Heiberger, R.M., 2009. HH: Statistical Analysis and Data Display: Heiberger and Holland. R package version 2.1-32. <http://CRAN.R-project.org/package=HH>
- Hothorn, T., Buehlmann, P., Dudoit, S., Molinaro, A., Van Der Laan, M., 2006. Survival Ensembles. *Biostatistics*, 7, 3, 355-373.
- Jaeger, T.F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *J Mem Lang*, 59, 4, 434-446.
- Jaeger, T.F., Furth, K., Hilliard, C., 2012. Phonological overlap affects lexical selection during sentence production. *J Exp Psychol Learn Mem Cogn*, 38, 5, 1439-1449.
- James, L.E., Burke, D.M., 2000. Phonological priming effects on word retrieval and tip-of-the-tongue experiences in young and older adults. *J Exp Psychol Learn Mem Cogn*, 26, 6, 1378.
- Jescheniak, J.D., Levelt, W.J.M., 1994. Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *J Exp Psychol Learn Mem Cogn*, 20(4), 824-843.
- Kittredge, A.K., Dell, G.S., Verkuilen, J., Schwartz, M.F., 2008. Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cogn Neuropsychol*, 25, 4, 463-492.
- Lachman, R., Shaffer, J.P., Hennrikus, D., 1974. Language and cognition: Effects of stimulus codability, name-word frequency, and age of acquisition on lexical reaction time1. *J Verbal Learning Verbal Behav*, 13, 6, 613-625.
- Laganaro, M., Chetelat-Mabillard, D., Frauenfelder, U.H. (2013). Facilitatory and interfering effects of neighborhood density on speech production: evidence from aphasic errors. *Cogn Neuropsychol*.

- Landauer, T.K., Streeter, L.A., 1973. Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *J Verbal Learning Verbal Behav*, 12, 2, 119-131.
- Levelt, W.J.M., 1989. *Speaking: From intention to articulation*. Cambridge, Mass.: MIT Press.
- Levelt, W.J.M., Roelofs, A., Meyer, A.S., 1999. A theory of lexical access in speech production. *Behav Brain Sci*, 22, 1, 1-38.
- Luce, P.A., 1986. *Neighborhoods of words in the mental lexicon*. Ph.D. dissertation, Indiana University.
- Luce, P.A., Pisoni, D.B., 1998. Recognizing spoken words: The neighborhood activation model. *Ear Hear*, 19, 1, 1-36.
- Marian, V., Bartolotti, J., Chabal, S., Shook, A., 2012. CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE*, 7, 8: e43230.
- Marian, V., Blumenfeld, H.K., Boukrina, O.V., 2008. Sensitivity to phonological similarity within and across languages. *J Psycholinguist Res*, 37(3), 141–170.
- Martin, N., Gagnon, D.A., Schwartz, M.F., Dell, G.S., Saffran, E.M., 1996. Phonological facilitation of semantic errors in normal and aphasic speakers. *Lang Cogn Process*, 11(3), 257-282.
- McClelland, J.L., Elman, J.L., 1986. The TRACE model of speech perception. *Cogn Psychol*, 18(1), 1-86.
- Meyer, A.S., Bock, K., 1992. The tip-of-the-tongue phenomenon: Blocking or partial activation? *Mem Cognit*, 20, 6, 715-726.
- Miceli, G., Capasso, R., Caramazza, A., 1999. Sublexical conversion procedures and the interaction of phonological and orthographic forms. *Cogn Neuropsychol*, 16(6), 557–572.
- Middleton, E.L., Schwartz, M.F., 2010. Density pervades: An analysis of phonological neighbourhood density effects in aphasic speakers with different types of naming impairment. *Cogn Neuropsychol*, 27, 5, 401-427.
- Mirman, D., Kittredge, A.K., Dell, G.S., 2010. Effects of near and distant phonological neighbors on picture naming. *Proc 32 Annu Conf Cogn Sci Soc*, 1447-1452.
- Mitterer, H., Ernestus, M., 2008. The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, 109(1), 168-173.

- Morsella, E., Miozzo, M., 2002. Evidence for a cascade model of lexical access in speech production. *J Exp Psychol Learn Mem Cogn*, 28, 3, 555-563.
- Munson, B., Swenson, C.L., Manthei, S.C., 2005. Lexical and phonological organization in children: Evidence from repetition tasks. *J Speech Lang Hear Res*, 48, 1, 108.
- Navarrete, E., Costa, A., 2005. Phonological activation of ignored pictures: further evidence for a cascade model of lexical access. *J Mem Lang*, 53 (3), 359–377
- New, B., Pallier, C., Ferrand, L., Matos, R., 2001. Une base de données lexicales du français contemporain sur internet: LEXIQUE™//A lexical database for contemporary French: LEXIQUE™. *Annee Psychol*, 101, 3, 447-462.
- Newman, R.S., Bernstein Ratner, N., 2007. The role of selected lexical factors on confrontation naming accuracy, speed, and fluency in adults who do and do not stutter. *J Speech Lang Hear Res*, 50, 1, 196.
- Newman, R.S., German, D.J., 2002. Effects of lexical factors on lexical access among typical language-learning children and children with word-finding difficulties. *Lang Speech*, 45, 3, 285-317.
- Newman, R.S., German, D.J., 2005. Life span effects of lexical factors on oral naming. *Lang Speech*, 48, 2, 123-156.
- Nooteboom, S.G., 2005. Listening to oneself: Monitoring speech production. In R. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.). *Phonological encoding and monitoring in normal and pathological speech*. Hove, UK: Psychology Press.
- Norris, D., 1994. Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52(3), 189-234.
- O'Seaghdha, P.G., Marin, J.W., 2000. Phonological competition and cooperation in form-related priming: Sequential and nonsequential processes in word production. *J Exp Psychol Hum Percept Perform*, 26, 1, 57-73.
- Pérez, M.Á., Navalón, C., 2003. Normas españolas de 290 nuevos dibujos: Acuerdo en la denominación, concordancia de la imagen, familiaridad, complejidad visual y variabilidad de la imagen. *Psicológica*, 002, 215-241.
- Pérez, M.Á., 2007. Age of acquisition persists as the main factor in picture naming when cumulative word frequency and frequency trajectory are controlled. *Q J Exp Psychol*, 60, 1, 32-42.

- Pollatsek, A., Perea, M., Binder, K.S., 1999. The effects of neighborhood size in reading and lexical decision. *J Exp Psychol Hum Percept Perform*, 25(4), 1142– 58.
- Protopapas, A., 2007. CheckVocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behav Res Methods*, 39, 4, 859.
- R Development Core Team, 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org/>.
- Rapp, B., Goldrick, M., 2000. Discreteness and interactivity in spoken word production. *Psychol Rev*, 107, 3, 460.
- Rapp, D.N., Samuel, A.G., 2002. A reason to rhyme: Phonological and semantic influences on lexical access. *J Exp Psychol Learn Mem Cogn*, 28, 3, 564.
- Rodriguez-Gonzalez, E., 2012. Processing of Spanish preterite regular and irregular verbs: The role of neighborhood density. *Span J Psychol*, 15, 1, 35-47.
- Roelofs, A., 2003. Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychol Rev*, 110, 88 – 125.
- Runnqvist, E., Strijkers, K., Alario, F.-X., Costa, A., 2012. Cumulative semantic interference is blind to language: Implications for models of bilingual speech production. *J Mem Lang*, 66(4), 850-869
- Sadat, J., Martin, C.D., Alario, F.-X., Costa, A., 2012. Characterizing the bilingual disadvantage in noun phrase production. *J Psycholinguist Res*, 41(3), 159-179.
- Sadat, J., Martin, C.D., Magnuson, J.S., Alario, F.-X., Costa, A., submitted. Breaking down the bilingual cost in speech production.
- Sears, C.R., Hino, Y., Lupker, S.J., 1995. Neighborhood size and neighborhood frequency effects in word recognition. *J Exp Psychol Hum Percept Perform*, 21(4), 876-900.
- Sevold, C.A., Dell, G.S., 1994. The sequential cuing effect in speech production. *Cognition*, 53, 2, 91-127.
- Severens, E., Van Lommel, S., Ratinckx, E., Hartsuiker, R.J., 2005. Timed picture naming norms for 590 pictures in Dutch. *Acta Psychol*, 119(2), 159–187.
- Slevc, L.R., Ferreira, V.S., 2006. Halting in single word production: A test of the perceptual loop theory of speech monitoring. *J Mem Lang*, 54, 4, 515-540.

- Snodgrass, J.G., Vanderwart, M., 1980. A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *J Exp Psychol Hum Learn Mem*, 6, 2, 174-215.
- Snodgrass, J.G., Yuditsky, T., 1996. Naming times for the Snodgrass and Vanderwart pictures. *Behav Res Methods*, 28, 4, 516-536.
- Spieler, D. H., Balota, D. A. (2000). Factors influencing word naming in younger and older adults. *Psychol Aging*, 15(2), 225.
- Stemberger, J.P., 2004. Neighbourhood effects on error rates in speech production. *Brain Lang*, 90, 1-3, 413-422.
- Storkel, H.L., 2004. Methods for minimizing the confounding effects of word length in the analysis of phonotactic probability and neighborhood density. *J Speech Lang Hear Res*, 47, 6, 1454.
- Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional Variable Importance for Random Forests. *BMC Bioinformatics*, 9(307).
- Strobl, C., Boulesteix, A.-L., Zeileis, A., Hothorn, T., 2007. Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution. *BMC Bioinformatics*, 8(25).
- Sullivan, M.P., Riffel, B., 1999. The nature of phonological encoding during spoken word retrieval. *Lang Cogn Process*, 14, 1, 15-45.
- Székely, A., Bates, E., 2000. Objective visual complexity as a variable in studies of picture naming. *Center for Research in Language Newsletter*, 12, 2
- Székely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., . . . Wicha, N., 2004. A new on-line resource for psycholinguistic studies. *J Mem Lang*, 51, 2, 247-250.
- Tabak, W., Schreuder, R., Baayen, R.H., 2010. Producing inflected verbs: A picture naming study. *Ment Lex*. 5(1), 22-46.
- Tagliamonte, S., Baayen, R.H. (2012). Models, forests and trees of York English: Was/were variation as a case study for statistical practice. *Lang Var Change*, 24(2): 135-178.
- 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. *J Gerontol B Psychol Sci Soc Sci*, 65B(5), 551-560.

- Venables, W.N., Ripley, B.D., 2002. *Modern Applied Statistics with S*. Fourth Edition. Springer, New York.
- Vitevitch, M.S., 1997. The neighborhood characteristics of malapropisms. *Lang Speech*, 40, 3, 211-228.
- Vitevitch, M.S., 2002. The influence of phonological similarity neighborhoods on speech production. *J Exp Psychol Learn Mem Cogn*, 28, 4, 735-747.
- Vitevitch, M.S., Armbruster, J., Chu, S., 2004. Sublexical and lexical representations in speech production: Effects of phonotactic probability and onset density. *J Exp Psychol Learn Mem Cogn*, 30, 2, 514-529.
- Vitevitch, M.S., Luce, P. A., 1998. When words compete: Levels of processing in perception of spoken words. *Psychol Sci*, 9, 4, 325-329.
- Vitevitch, M.S., Luce, P.A., 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *J Mem Lang*, 40(3), 374-408.
- Vitevitch, M.S., Rodríguez, E., 2005. Neighborhood density effects in spoken word recognition in Spanish. *J Multiling Commun Disord*, 3(1), 64-73.
- Vitevitch, M.S., Sommers, M.S., 2003. The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Mem Cognit*, 31, 4, 491-504.
- Vitevitch, M.S., Stamer, M.K., 2006. The curious case of competition in Spanish speech production. *Lang Cogn Process*, 21, 6, 760-770.
- Vitevitch, M.S., Stamer, M.K., 2009. The influence of neighborhood density (and neighborhood frequency) in Spanish speech production: A follow-up report. *Spoken Language Laboratory Technical Report (University of Kansas)*, 1, 1-6.
- Vitkovitch, M., Tyrrell, L., 1995. Sources of disagreement in object naming. *Q J Exp Psychol*, 48, 4, 822-848.
- Wheeldon, L., 2003. Inhibitory form priming of spoken word production. *Lang Cogn Process*, 18, 1, 81-109.
- Ziegler, J., Perry, C., 1998. No more problems in Coltheart's neighborhood: Resolving neighborhood conflicts in the lexical decision task. *Cognition*, 68, 2, B53-B62.
- Ziegler, J.C., Muneaux, M., 2007. Orthographic facilitation and phonological inhibition in spoken word recognition: a developmental study. *Psychon Bull Rev*, 14(1), 75-80.

Ziegler, J.C., Muneaux, M., Grainger, J., 2003. Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *J Mem Lang*, 48(4), 779–793

Appendix A:

Correlation matrix for the random effects in the final linear mixed effect model on naming latencies

Random effects							
Item	Intercept						
	Session	-0.69					
Participant	Intercept						
	Session	0.00					
	Nagr	0.02	0.10				
	AoA _{res}	0.37	0.20	-0.62			
	Frequency	-0.34	-0.27	0.35	-0.37		
	NeighFreq	-0.59	-0.11	-0.40	0.17	0.24	
	PhND _{res}	0.55	0.33	-0.07	0.04	0.09	-0.36

Note: Nagr = name agreement; AoA_{res} =residualized age-of-acquisition; Frequency = lexical Frequency; NeighFreq = neighborhood frequency; PhND_{res} = residualized phonological neighborhood density