Neighbourhood density effects in spoken word production

Solène Hameau

A thesis submitted in fulfillment of the requirements for thedegree of

Doctor of Philosophy

ARC Centre of Excellence in Cognition and its Disorders (CCD)

Department of Cognitive Science

Macquarie University, Sydney, Australia

September 2016

Declaration

The work in this thesis is my own original work. It has not been submitted for a higher degree in any other university of institution. All of the work reported in this thesis was undertaken during the time I was enrolled as a PhD student at Macquarie University, under the supervision of Prof. Lyndsey Nickels, Dr. Britta Biedermann and Dr. Nora Fieder. Ethics approval for the studies reported in this thesis was obtained from Macquarie University's Human Research Ethics Committee, Reference No. 5201200905.

Signed:

Solène Hameau, 20th September, 2016

Table of contents

General Abstract

Most theories of speech production postulate that in the process of producing words, representations that are similar to the target word with respect to its meaning and, for some theories, in terms of form, also become active. However, in previous speech production research, findings are still inconclusive with regards to the influence of the number of these neighbours, with data supporting the full range of possibilities - no effect, inhibitory and facilitatory effects. This thesis investigates what constitutes a word's meaning and form neighbourhood in the lexicon and whether and how this neighbourhood might affect the production of the intended target word.

Paper 1 is a comparison of several measures of semantic neighbourhood density, and an investigation of their influence on the picture naming performance of English monolinguals and individuals with aphasia. While no effect was observed in monolinguals, a facilitation effect of semantic neighbours was observed in aphasic speakers. Paper 2 follows this up with a case study investigating the effects of different semantic neighbourhood measures on a facilitated naming paradigm involving two aphasic participants with different levels of impairment. Although inhibitory effects of semantic neighbourhood variables were observed for both participants at baseline, the change in each participant's performance at post-test was affected differently and by distinct measures in each participant.

In Paper 3, the picture naming performance of English monolinguals and late French-English bilinguals is investigated with respect different types of phonological neighbourhood density and frequency measures , including a novel phonological neighbourhood density metric adapted to these bilingual speakers. Results showed, for both speaker groups, an interaction between phonological neighbourhood measures and the target word's frequency, with inhibitory effects on low frequency targets and facilitatory effects on the most frequent targets, frequency being affected, for bilinguals, by language experience and the overlap between the target and its translation equivalent. The complex patterns of facilitation and inhibition coming from neighbours that were found across this thesis are discussed in light of the main theories of lexical access in speech production, and are best explained within theories that assume interactivity between the different levels of representation.

Acknowledgements

I would like to express my deepest gratitude to my primary supervisor Lyndsey Nickels. It was quite a challenge to transition from over decade of clinical practice as a speech pathologist to the world of research, but from the beginning I knew I was in good hands to start this fascinating journey. I have learnt a lot and it has always been enjoyable thanks to your warmth, your patience, your expertise, your simplicity and your very rigorous feedback: Thank you, Lyndsey, for being such a wonderful and dedicated supervisor!

I would also like to thank my other supervisors: Britta Biedermann and Nora Fieder. Britta, thank you for sharing your enthusiasm for research, for including me in very interesting projects during this PhD, and for your constant support in PhD matters, even after you had to move from Sydney. Thank you also for offering me great emotional support in times when I needed it. And Nora, thank you for your support during my PhD, and for allowing me to take part in very nice research projects.

This thesis would not have been possible without the contribution of all participants who generously accepted to give me some of their time and efforts. A warm thank you to all of them.

I am very grateful I was part of such a friendly and supportive environment at CCD. I would like to thank Lesley, Teri, Marcus and Craig, Steve, for their great help along the way. Thanks to all the members of the Aphasia and the CogNeuro group for being such a stimulating team of language enthusiasts!

If my PhD journey was so smooth, it is also thanks to the fact I had such wonderful friends to laugh and enjoy Sydney life with. A special thank you to Fanou, Alia and Lou, my friends and housemates, but also Carmen, Elaine, Ivan, Vincent, Lisa, Anastasiia, Anna, Anne, Vishnu, Loes, Rimke, Polly, Soline, Camille, George, and many more.

Et le merci de la fin revient à ma merveilleuse famille… la distance m'a fait réaliser ma chance d'avoir des parents et des frères et soeurs si uniques. Je suis tellement fière de vous, les darons, Lupin et Roby, Alice, Martin et Laura, et Alex, vous êtes mon énergie, et resterez dans mon coeur pour toujours!

CHAPTER 1

Introduction

How do we speak? Language is one of the most important aspects of life, allowing us to express our inner thoughts and emotions, to make sense of complex and abstract thought, to communicate with others and to be able to fulfil our needs. Indeed, producing words appears to be a simple thing humans 'just do', and most of us barely give a conscious thought to the process of speaking. But, in fact, saying the word that matches what we want to express is one of the most complex actions we perform. We have to translate the meaning that we want to express into the form of a word, which involves picking of the right word from amongst tens of thousands of available words in memory. We also have to use of the voice and articulatory systems to accurately perform the right movements corresponding to each sound in that word. In addition, for individuals who speak more than one language, two or more different language systems have to be negotiated. Yet this amazing procedure allows us, almost always successfully, to produce the intended word within a fraction of a second.

The question of how humans achieve the task of producing words has been addressed for several decades in the field of psycholinguistics. Several models of lexical access in speech production have been implemented to describe the structure and dynamics necessary to explain how we produce words (e.g., Caramazza, 1997; Dell, Schwartz, Martin, Saffran & Gagnon, 1997; Levelt, Roelofs & Meyer, 1999). These models have addressed several broad questions, relating to the structure of representations (how many different levels of representation exist and what type of information is included in those representations), the general dynamics of the system (how information flows between levels of representation and whether there is interaction between levels, which other representations become active during the process, and how a word is selected from among all activated words), and, for bilinguals, to what extent representations from the other language are also activated. In this thesis, I am particularly interested in the role, in the course of speech production, of words that are similar in meaning (semantic neighbours) or similar in form (phonological neighbours), in the target language but also, in the case of bilinguals, in the non-target language. This chapter will briefly describe the main processes proposed in the most influential theories of speech production, and show how they broadly conceptualise the potential role of neighbours.

Activation of neighbours in speech production

Most theories assume the representation of three types of information for spoken word production: the meaning of the word (semantics), its grammatical characteristics (lexical-syntax), and its phonological form (phonemes). All theories agree that there are distinct levels representing word meaning (semantic nodes) and the set of sounds (phoneme nodes), and at least one abstract representation of the word intervening between semantics and phonemes (lexical nodes). However, there is debate regarding the number of these intervening levels and their role in the representation and retrieval of lexical-syntax (cf, e.g., Caramazza (1997) and Levelt et al. (1999)). The description here follows Rapp and Goldrick (2000) in remaining neutral in this debate and referring to a single intervening lexical level (rather than specifying this as a lemma or lexeme (word form) level).

When trying to express a concept, for example, the concept of a cat, whether from an internally generated idea or memory or an externally perceived object or picture, one first has to retrieve its semantic representation. Next, according to the spreading activation principle, whereby activation spreads between levels of representation, activated semantic nodes send proportional activation to their corresponding lexical nodes. Note that theories differ in whether they represent semantics as a series of semantic features (e.g., Dell et al., 1997) or as a network of interconnected holistic semantic nodes (e.g., Levelt et al., 1999). However, in both theories it is not just the target that is active at the lexical level, semantically related concepts are also activated. For the word ‘cat’, then, there is activation of ‘cat’, ‘dog’, ‘mouse’, etc..., at the lexical level (semantic neighbours). In featural theories, this occurs because the target shares features with these concepts and every semantic feature activates its corresponding lexical node. In theories with holistic semantic representation, this occurs because, at the semantic level, each semantic node will activate a network of related concepts and then, in turn, each of these concepts activate their corresponding lexical nodes. In addition, in bilinguals, it is generally agreed that activation at the level of semantics (common to both languages) spreads to lexical nodes in both languages. Hence, in a French-English bilingual, the nodes for both ‘cat’ (English) and ‘chat’ (French) will be activated.

Since several representations are activated at the lexical level, it is necessary to provide a mechanism responsible for selection of the relevant lexical representation. Models of speech production agree that the selection of a lexical node depends on its level of activation. The lexical node with the highest level of activation is selected (usually, the target: here, ‘cat’) (e.g., Dell et al., 1997). However, in some theories there may also be *competition* occurring between lexical nodes such that the relative activation of nodes influences target selection (e.g., using the Luce Choice Rule, Roelofs, 1992). This competition may be implemented as lateral inhibition where each activated lexical node inhibits the activation of other lexical nodes (e.g., Harley, 1984; Schade & Berg, 1992). At this stage, if the lexical node that receives the highest activation is not the target, the speaker produces an error. Hence, a word with many semantic neighbours might be more difficult to select compared to a word with few semantic neighbours resulting in slower and/or more error prone responses. In bilinguals, the extent to which the non-target language is activated could also influence selection, and there is still debate regarding the mechanisms that can account for the selection of the right lexical node in the right language (e.g., de Bot, 1992, Costa & Caramazza, 1999; Green, 1998; Poulisse & Bongaerts, 1994; Roelofs, 1998).

Once the target lexical node is selected, the word's phonological segments (phoneme nodes) must also be selected. Here, theories vary widely in terms of the dynamics of activation and selection. In discrete serial models of lexical access (e.g. Levelt et al., 1999), the phonological segments of only the selected lexical node are accessed (so here, only the phoneme nodes of 'cat', /k/, /æ/, /t/), and activation of the phoneme nodes happens only once the target lexical node has been selected. Alternatively, cascaded models of lexical access (e.g. Caramazza, 1997; Dell et al., 1997) assume that every lexical node activated at the lexical level ('cat', 'dog', 'mouse', etc…) activates its phonological segments, and that this occurs before lexical selection is completed. Furthermore, interactive theories (Dell et al., 1997) allow for all activated lexical nodes to activate their phonological representations, but critically, also postulate *interactivity* between the lexical level and the level of phonemes. Consequently, lexical and phonological levels continually cycle activation between one another. In this framework, phonological neighbours of a word will also become activated. For example, as the lexical node 'cat' activates the phonemes /k/, /æ/, /t/, each of these phonemes in turn activates all the lexical nodes that comprise these phonemes. Hence, /k/ will also activate 'car' 'key' 'kilt' 'coin' etc., and /æ/ , 'hand' 'sack' 'dam' and so forth. Consequently, interaction between lexical and phoneme levels results in phonologically related words (phonological neighbours) being activated at the level of the lexicon. While this can apply for monolinguals, in the case of bilinguals, there is still no agreement on the extent to which phonological representations of non-target languages are active, so it is unclear whether phonological neighbours of the non-target language would be activated.

The last steps in the process of expressing a concept as a word include the assembly of phonemes into syllables, and retrieval of the articulatory plans corresponding to the syllables of the selected word (e.g., the exact position of the muscles involved in the production of speech) enabling the word to be articulated.

As can be seen from this short review, different theories make different predictions about the extent to which neighbours become active and influence processing. Models agree on the fact that *semantic* neighbours are activated, but whether or not they influence processing depends on model assumptions. On the other hand, discrete models do not assume activation of *phonological* neighbours, even within cascaded models , only interactive models predict an influence of phonological neighbours on processing.

In this thesis, I am interested in the processes involved in speech production, with a focus on the influence of words that are likely to become active along with the target in processing: semantic neighbours, and phonological neighbours. This is investigated using a picture naming task, and targeting different populations of speakers, namely unimpaired monolingual and bilingual speakers and individuals with acquired language impairment after stroke (speakers with aphasia).

Experimental approaches for the study of neighbourhoods in speech production

The study of speech production at the single word level commonly uses complex paradigms that are meant to highlight a specific aspect of processing. Paradigms such as priming, error elicitation or picture-word interference have been used to build theoretical models. In these paradigms, a target has to be named but the ease of its correct production is affected, for example, by the presence of a prime or a distractor. In contrast, the 'simple' picture-naming paradigm ('simple' because no potentially distracting information is presented along with the picture) has been used primarily with individuals with language impairments (aphasia), studying error patterns to reveal language structure and processing. More recently, the picture naming paradigm has increasingly been used to study speech production mechanisms at the word level with unimpaired populations, where it has the advantage of being less prone to strategic influences. This simple task requires all of the processes involved in lexical access, although, as only a single word is required, it underrepresents the influence of lexical-syntax (at least in English). In picture naming studies with unimpaired participants, the critical measure that is used for analysis is the time between the presentation of a picture and the onset of the response when the participant produces the (correct) word corresponding to that picture. This task has allowed researchers to identify which factors influence the ease of naming as reflected by shorter latencies (e.g. Alario, Ferrand, Laganaro, New, Frauenfelder, & Segui, 2004). Given that picture naming allows one to assess the influence of a word's characteristics on processing, it follows that it is appropriate for the study of neighbourhood factors, as it is possible to compare the speed and accuracy of words that have many neighbours, compared to few neighbours.

Brain damage (for example, as a consequence of a stroke) can result in aphasia, an impairment that affects the processing of language. Although aphasia can affect every modality of language, the most common symptom of aphasia is the inability to produce known words in a fast and accurate manner. Indeed, individuals with aphasia tend to produce errors at a high rate, and this has enabled researchers to use error data theoretically (e.g., Caramazza & Coltheart, 2006), contributing to the development and evaluation of models of speech production (e.g. Dell et al., 1997). For example, the fact that some aphasic speakers produce mostly phonological errors, but others, mostly semantic errors, is taken as evidence for two distinct stages in word retrieval. Interestingly for our thesis, several studies have used aphasic data to investigate the influence of neighbourhood density factors on the pattern of errors, both for semantic neighbours (e.g., Blanken, Dittmann, & Wallesch, 2002; Bormann, 2011; Bormann, Kulke, Wallesch, & Blanken 2008; Kittredge, Dell, & Schwartz, 2007a; 2007b; Mirman, 2011, Mirman & Graziano, 2013) and phonological neighbours (e.g., Gordon, 2002; Laganaro, Cheletat-Mabillard & Frauenfelder, 2013; Middleton & Schwartz, 2011).

Our choice of picture naming as the main paradigm in this thesis, and of individuals with aphasia as one of the populations under investigations, seems then appropriate to the study of neighbourhood factors in speech production.

Preview of the thesis

This thesis presents three experimental chapters written in journal article format followed by a General Discussion. As noted above, the overall aim is to define what constitutes a word's neighbourhood, with a focus on both semantically related neighbours, and on phonologically related neighbours in monolinguals and bilinguals, to understand how these neighbours affect processing, to further our understanding of the processes involved in the production of words.

In **Chapter 2** (Paper 1), the influence of semantic neighbourhood density is investigated on the picture naming performance of English monolinguals and a large group of individuals with aphasia. In Experiment 1, several different measures that have been previously used in the literature to account for the set of words in the lexicon that are related in meaning to the target (semantic neighbourhood density measures) are compared by means of correlations and a principal component analysis. Following this procedure, the new measures are used as predictors in analysis of picture naming data (Experiment 2) involving 50 English monolinguals. Finally, in Experiment 3, an analysis of the patterns of response on a picture naming task for a large group of individuals with aphasia (n=193) is performed, looking at the influence of the semantic neighbourhood density variables, with a focus on the different level of impairment.

Chapter 3 (Paper 2) is a further investigation of the effect of semantic neighbourhood density variables, with the addition of measures representing the similarity, frequency and strength of association between the target and its neighbours. In this chapter, a cognitive neuropsychological approach is adopted, in a facilitated naming task involving two participants with chronic aphasia. The response to the facilitation task is analysed in view of the effects of semantic neighbourhood and the pattern of impairment of each participant, to determine under what circumstances the semantic neighbourhood properties of words determine the outcomes of the facilitation task.

Chapter 4 (Paper 3) is an investigation of the influence of several phonological neighbourhood density and frequency measures on monolingual and bilingual picture naming. In Experiment 1, the performance of 40 English monolinguals on a picture naming task is analysed with respect to these different measures using regression techniques, and in Experiment 2, a novel set of within and cross-language phonological neighbourhood measures is implemented that endeavours to take into account an estimate of the vocabulary range and of the representation of non-native phonemes in French-English late bilinguals. The performance of a group of 50 French-English late bilinguals is analysed on the same English picture naming task as Experiment 1, investigating the effect of these novel measures on speed and accuracy.

Finally, **Chapter 5**, the General Discussion, provides a summary of every experiment in the thesis, discusses their contribution to our understanding of the influence of neighbours in speech production, and reflects on some of the choices made in this thesis as well as its limitations and potential future directions.

References

Alario, F. X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers*, *36*(1), 140-155.

Blanken, G., Dittmann, J., and Wallesch, C-W. (2002). Parallel or serial activation of word forms in speech production? Neurolinguistic evidence from an aphasic patient. *Neuroscience Letters, 325,* 72-74.

Bormann, T. (2011). The role of lexical-semantic neighbourhood in object naming: implications for models of lexical access. *Frontiers in Psychology, 2(127),* 1-11.

Bormann, T., Kulke, F., Wallesch, C.W., and Blanken, G. (2008). Omissions and semantic errors in aphasic naming: Is there a link? *Brain and Language, 104,* 24-32.

Caramazza, A. (1997). How many levels of processing are there in lexical access?. *Cognitive Neuropsychology*, *14*(1), 177-208.

Caramazza, A., & Coltheart, M. (2006). Cognitive neuropsychology twenty years on. *Cognitive Neuropsychology*, *23*(1), 3-12.

Costa, A., & Caramazza, A. (1999). Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish–English and English–Spanish bilinguals. *Bilingualism: Language and Cognition, 2*(3), 231-244.

De Bot, K. (1992). A bilingual production model: Levelt's' speaking' model adapted. *Applied Linguistics, 13* (1), 1-24.

Dell, G.S., & Gordon, J.K. (2003). Neighbors in the lexicon: Friends or foes? In N.O. Schiller and A.S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities*. New York: Mouton de Gruyter.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*(4), 801.

Gordon, J. K. (2002). Phonological neighborhood effects in aphasic speech errors: Spontaneous and structured contexts. *Brain and Language, 82*(2), 113-145.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*(2), 67-81.

Harley, T. A. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science, 8,* 191-219.

Kittredge, A.K., Dell, G.S., and Schwartz, M.F. (2007a). Omissions in aphasic picture naming: Late age-of-acquisition is the culprit, not low semantic density. *Brain and Language (103),* 132-133.

Kittredge A. K., Dell G. S., Schwartz M. F. (2007b). "*Omissions in aphasic picture naming: AoA is the culprit, as well as low semantic density*", in Poster Presented at the 45th Meeting of the Academy of Aphasia, Washington, DC.

Laganaro, M., Chetelat-Mabillard, D., & Frauenfelder, U. H. (2013). Facilitatory and interfering effects of neighbourhood density on speech production: Evidence from aphasic errors. *Cognitive Neuropsychology, 30*(3), 127-146.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(01), 1-38.

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. *Cognitive Neuropsychology, 27*(5), 401-427.

Mirman, D. (2011). Effects of Near and Distant Semantic Neighbors on Word Production. *Cognitive, Affective, and Behavioral Neuroscience, 11(1),* 32-43.

Mirman, D., & Graziano, K.M. (2013). The Neural Basis of Inhibitory Effects of Semantic and Phonological Neighbors in Spoken Word Production. *Journal of Cognitive Neuroscience, 25(9),* 1504-1516.

Poulisse, N., & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics, 15*(1), 36-57.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological review, 107*(3), 460.Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1-3), 107-142.

Roelofs, A. (1998). Lemma selection without inhibition of languages in bilingual speakers. *Bilingualism: Language and Cognition, 1*(2), 94-95.

Schade, U., & Berg, T. (1992). The role of inhibition in a spreading-activation model of language production. II. The simulational perspective. *Journal of Psycholinguistic Research, 21*(6), 435-462.

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 735.

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

CHAPTER 2

Paper 1

Effects of semantic neighbourhood density on unimpaired and aphasic spoken word production.

Introduction

Among the thousands of words that comprise an adult's vocabulary, the successful speaker constantly has to select the one that matches best his or her communicative goals. Many theories of word processing assume that during speech production, not only the intended lexical representation but also other representations that are similar in terms of meaning (semantics) and/or form (phonology) are activated. In order to characterise which related representations are co-activated, a growing number of studies of word production have investigated the influence of phonological neighbourhood density (e.g., Dell & Gordon, 2003; Vitevitch, 2002), which represents the number of words in an individual's lexicon that sound similar to a given word. However, when it comes to semantically related words, psycholinguistic research is more scarce (but see, for example Bormann, 2011; Kittredge, Dell, & Schwartz, 2007a, 2007b; Mirman, 2011), possibly because of the challenges required to capture semantic relationships.

There is in fact no agreement on the most accurate way of defining and calculating the number of words semantically related to any given word. Some authors have attempted to characterise the set of words that share aspects of conceptual representation with a given word, and to investigate the effects of the size of this set on language performance. The size of this collection of related representations has been referred to either as semantic density (Kittredge et al., 2007a,2007b), semantic neighbourhood size, semantic neighbourhood density (Chen & Mirman, 2012; Mirman, 2011; Mirman & Graziano, 2013; Mirman & Magnuson, 2008) or number of semantic competitors (Blanken, Dittmann, and Wallesch, 2002; Bormann, 2011; Bormann, Kulke, Wallesch, and Blanken, 2008). A further distinction has been made between words that have very similar semantic representations (such as apple and pear) and words with some degree of semantic similarity (apple and eggplant) (near versus distant semantic neighbours: Chen & Mirman, 2012; Mirman, 2011; Mirman & Graziano, 2013; Mirman & Magnuson, 2008).

These different labels reflect the variety of methods employed to obtain measures of semantic cohorts. In this paper, we will refer to the variable under investigation as semantic neighbourhood density, while noting that different studies have used different means of calculation for this measure. We have labelled these different measures as: rated competitors, near and distant feature neighbours, contextual neighbours, and association neighbours.

Rated competitors: Some authors (Blanken et al., 2002; Bormann, 2011; Bormann et al., 2008) have collected subjective semantic neighbourhood density ratings using the following procedure: after being provided with instructions about the type of semantic relationship that was considered, students were asked to estimate the number of members in the target's semantic category (in other words, the number of coordinates) by rating each target item on a scale from having ’hardly any competitors‘ to ’many competitors‘. Hence, words would be considered as items with ’high competition‘ or ’low competition‘.

Near and distant feature neighbours: Semantic features have been generated by asking speakers to define and describe different words (McRae, Cree, Seidenberg, and McNorgan, 2005): the number of features shared between words is then used as a measure of semantic similarity (e.g., Mirman, 2011; Vigliocco, Vinson, Damian, and Levelt, 2002). Mirman and his colleagues (Chen & Mirman, 2012; Mirman, 2011; Mirman & Graziano, 2013; Mirman & Magnuson, 2008) defined ’near‘ and ’distant‘ semantic neighbours based on the number of shared features in the McRae et al. (2005) database: ’near neighbours‘ shared at least 40% of features and therefore were semantically very similar to the target word whereas ’distant neighbours‘ shared more than 0 and fewer than 25% of features.

Contextual neighbours: Other authors have used measures based on words that occur frequently in similar semantic contexts. Kittredge et al. (2007a; 2007b) developed a measure of ’semantic density‘ based on Latent Semantic Analysis (LSA) (Landauer, Foltz, and Laham, 1998). LSA is a method that defines the semantic space of a given word based on its contextual-usage meaning from a large corpus of texts. The primary assumption of this method is that the similarity of the meanings of words can be determined by the contexts in which these words are likely or not to occur. It is possible with LSA to determine a measure of semantic ’similarity‘ between words, and, more importantly for our purposes, to obtain a list of ’near neighbours‘ of target words. ’Density‘ measures were hence gathered by Kittredge and her colleagues by counting, among the words that were the nearest the target in semantic space, those that had an LSA similarity score of at least 0.4, and that were also members of the target word’s category (based on any of six different published category norms).

Association neighbours: Finally, subjective free association norms have been used as a tool to characterize the concepts related to a given word: Mirman and Magnuson (2006) reviewed different approaches to semantic representation, one being based on such association norms. They noted that one view of semantic neighbourhood defined it as the set of associates (words that first come to mind that are meaningfully related or strongly associated) generated by participants for a given word (from the University of South Florida free association norms: Nelson, McEvoy, & Schreiber, 1998).

Importantly, these different measures have been associated with varying effects on language processing, especially on the picture naming performance of both healthy speakers and speakers with aphasia (these findings are reviewed in more detail below). It remains unclear whether naming is facilitated or inhibited by the number of semantic neighbours, and when and how these neighbours play a role in the course of speech production. More evidence is needed in order to characterize the effect of semantic neighbourhood density on speech production. Therefore, in this paper, we aim to further investigate and disentangle effects of the different measures of semantic neighbourhood density that have been used in previous research. We do this by examining the naming performance of a group of unimpaired English speakers, and of speakers with aphasia. By doing so, we expect to gain a better understanding of the semantic representation of words in the lexicon, and how theories of word production are constrained by the effects of semantically related words on word production.

Previous research

There has been a body of research that has investigated the influence of different measures of semantic neighbourhood density using input language tasks. For instance, words with many *contextual neighbours* are extensively associated with shorter reaction times (RTs) in lexical decision tasks (e.g., Buchanan et al., 2001; Hargreaves & Pexman, 2012; Pexman, Hargreaves, Siakaluk, Bodner, & Pope, 2008; Shaoul & Westbury, 2010; Yap, Pexman, Wellsby, Hargreaves, & Huff, 2012; Yap, Tan, Pexman, & Hargreaves, 2011). *Association neighbours* have also been related to faster RTs in lexical decision (Buchanan et al., 2001 (not in all analyses), Duñabeitia, Avilés & Carreiras 2008, Yates et al., 2003, but see Yap et al., 2011, for no effect found on RTs). In tasks that require semantic analysis (e.g., semantic judgment or categorisation tasks), in most cases none of the semantic neighbourhood density measures seem to predict response times (Mirman & Magnuson 2006; Pexman et al., 2008; Yap et al., 2011; Yap et al., 2012). However, Shaoul & Westbury (2010), found longer RTs associated with many *contextual neighbours* in semantic decision and judgment tasks, and Mirman & Magnuson (2008) reported longer RTs on words with many *near (feature) neighbours*, and shorter RTs on words with many *distant (feature) neighbours*). Finally, the influence of both *contextual* and *association* neighbours in word reading aloud tasks remains unclear, with the observation of both facilitation and no significant effect across studies (Buchanan et al., 2001; Dunabeitia et al., 2008; Yap et al., 2011; 2012).

In sum, with respect to input tasks, the main findings are that words that have many semantic neighbours, regardless of the type of semantic relationship between the target word and its neighbours, seem to be faster recognised; however, in tasks that require a semantic judgement, different types of neighbours seem to be associated with different effects in performing the task.

Table 1. Semantic neighbourhood effects in word production: previous research.

Study	Population	Language	Task	Latency	Measure	Accuracy	Type of errors
Mirman (2011): Experiment 2	35 controls	English	Speeded PN (57 items)	no effect	Near feature Distant feature	↘ ↗	↗ semantic errors ↘ semantic errors
Bormann (2011): Experiment 2	18 controls	German	PN (54 items)	no effect	Rated competitors	no effect	NA
Bormann (2011): Experiment 3	24 controls	German	PWI (54 items + distractors)	no effect	Rated competitors	no effect	NA
Mirman (2011): Experiment 1	62 PWA	English	PN (57 items)	NA	Near feature Distant feature	↘ No effect	↗ semantic errors ↘ semantic errors
Mirman & Graziano (2013)	47 PWA	English	PN (95 items)	NA	Near feature	↗↘	NA
Blanken, Dittmann, & Wallesch (2002)	1 PWA	German	PN (138 items)	NA	Rated competitors	no effect	↗ semantic errors ↘ omissions
Bormann, Kulke, Wallesch, & Blanken (2008)	17 PWA	German	PN (66 items)	NA	Rated competitors	no effect	↗ semantic errors ↘ omissions
Bormann, 2011: Experiment 1	7 progressive anomia	German	PN (54 items)	NA	Rated competitors	no effect	↗ semantic errors ↘ omissions
Kittredge, Dell, & Schwartz (2007a)	50 PWA	English	PN (175 items)	NA	Contextual	↗	no effect
Kittredge, Dell, & Schwartz (2007b)	100 PWA	English	PN (175 items)	NA	Contextual	no effect	↗ semantic errors ↘ omissions

PWA= participants with aphasia, PN=picture naming, PWI=picture-word interference, ↘=decrease, ↗=increase

In speech production, the evidence does not seem any more consistent. Across production studies, subjects with unimpaired language have shown different sensitivity to semantic neighbourhood density on picture naming ability depending on the type of neighbours under investigation (see Table 1 for a summary). For instance, in a classic picture naming experiment, Bormann (2011) observed the same rate of correct responses, and no differences in latencies on targets with many versus few *rated competitors*. In the same study, Bormann also reported no difference in accuracy or response time depending on the number of *rated competitors*, this time in a picture-word interference experiment: Although semantically related distractors produced more interference than unrelated distractors, the number of competitors of the target word did not predict response time across conditions. In contrast, with a group of 35 healthy English-speaking older adults, and in a speeded naming task that was designed in order to generate more errors than in a classic picture naming task, Mirman (2011, Study 2) observed an influence of the number of *near feature neighbours*: there was a higher overall rate of errors and more specifically, of semantic errors, on words with many near feature neighbours than on words with few near feature neighbours. An opposite effect of *distant feature neighbours* was also found: words with many distant neighbours were more likely to be correctly named and less likely to result in a semantic error than words with few distant feature neighbours.

In sum, while many studies have looked at the effects of semantically related distractors on picture naming (e.g., Glaser & Düngelhoff, 1984; Rosinski, Golinkoff, & Kukish, 1975), to our knowledge these are the only two studies that have looked at semantic neighbourhood density in unimpaired picture naming performance: these studies featured different languages (German and English), different tasks (speeded versus classic picture naming, and picture-word interference) and different types of neighbours (*rated competitors* versus *feature neighbours*). It is therefore difficult to determine to what extent semantic neighbourhood impacts speech production processes in unimpaired speakers and further evidence is required.

Semantic neighbourhood density has also been shown to have different effects on the naming performance of individuals with aphasia: no effect, inhibitory and facilitative effects on accuracy have been reported in the literature. Moreover, semantic neighbourhood density also seems to influence the probability of different error types (semantic and omission errors) inconsistently. Blanken et al. (2002) report a single case study involving a person with chronic aphasia (MW) who had a marked lexical-semantic impairment with intact conceptual processing, as shown by his ’faultless‘ performance on the Pyramids and Palm Trees test (Howard & Patterson, 1992). Blanken et al. (2002) compared MW’s picture naming performance on two matched sets of items: one with many *rated competitors* and the other with few *rated competitors*. There was no significant difference in accuracy between the two sets. However, the set of words with many *rated competitors* was significantly more prone to semantic errors but less prone to omissions, and the opposite for the set of words with few *rated competitors*. More recently, Bormann et al. (2008) replicated these findings in a picture naming study involving 17 people with aphasia following a stroke, who were selected because they produced very few phonological errors in picture naming. While some of the participants had impaired comprehension skills, it is unclear whether conceptual processing was intact across the group, because the authors do not report the performance of the participants on tasks addressing conceptual knowledge without involving language. Once more, there was no significant difference in accuracy between words with many or few *rated semantic competitors* for the group or any individual. However, once again, there was a higher rate of semantic errors and fewer omissions on the set of words with many *rated competitors* compared to the set of words with few *rated competitors*. Bormann (2011) also found the same results with seven people with progressive aphasia (five with a diagnosis of semantic dementia, one with a suspected behavioural variant of frontotemporal lobar degeneration, and one with probable Alzheimer's disease).

To summarise, these three studies of aphasic picture naming that stem from the same group of researchers, found that participants with naming impairments (predominantly at the lexical-semantic level) showed effects of the number of *rated semantic competitors* on semantic errors and omissions but not on overall accuracy. However, in contrast, Kittredge and her colleagues (2007a), report better overall accuracy on words with many *contextual neighbours* than on words with few contextual neighbours for a group of 50 people with aphasia, as shown using binomial logistic regressions where 'semantic density' (the number of contextual neighbours) was entered as one of the predictors. Despite this effect on accuracy, further analysis showed no effect of the number of *contextual neighbours* on the probability of each error type (semantic, phonological or omission) compared to a correct response. When a subgroup of 15 participants who had more of a semantic impairment were analysed, the overall facilitation effect was now only marginally significant (p<.062). In a subsequent extension of their study which included the 50 participants reported previously and an additional 50 more people with aphasia with a mixture of levels of impairment, Kittredge et al. (2007b) used regression models to investigate the effect of the number of *contextual neighbours* on the probability of each response type. With this larger sample, they found that the effect of *contextual neighbours* on accuracy disappeared for the whole group as well as for separate groups that were defined based on whether individuals made predominantly phonological errors or few phonological errors (58 'phonological patients' versus 42 'non-phonological patients'). In addition, and in contrast to their previous study, they found an effect of *contextual neighbours* on semantic errors and omissions in the non-phonological group: words with many *contextual neighbours* showed decreased omissions relative to correct responses and increased semantic errors relative to omissions. The findings of Kittredge's second study (Kittredge et al., 2007b) are then more similar to those of the three *rated semantic competitors* studies (Blanken et al., 2002; Bormann, 2011; and Bormann et al., 2008): no effect on overall accuracy, but more correct responses than omissions, and more semantic errors than omissions on words in the high semantic density condition.

Mirman (2011, Study 1) analysed the picture naming performance of a group of 62 unselected people with chronic aphasia, and looked at effects of numbers of *feature neighbours* on performance. Four different conditions were defined depending on the number of near feature neighbours and distant feature neighbours of each word, and hence four corresponding groups of words were formed (12 to 18 words each) for a total of 57 words. Using logistic regressions, Mirman found that words in the 'many *near feature* neighbours' conditions were more likely to be incorrectly named than words in the 'few *near feature* neighbours' conditions. In contrast, accuracy was not affected by the number of *distant feature* neighbours. In addition, more semantic errors were observed on targets with many *near feature* neighbours compared to targets with few *near feature* neighbours. By contrast, there were fewer semantic errors on words that had many *distant feature* neighbours compared to words that had few *distant feature* neighbours. In a further analysis of semantic errors that were produced by the speakers with aphasia, Mirman showed that the words that were substituted for the target were more likely to be words from dense *near feature* neighbourhoods than predicted by chance, and less likely to be words from dense *distant feature* neighbourhoods than predicted by chance. In a subsequent analysis of individual participants' performance, the difference in the proportion of semantic errors in the many *near feature* neighbours and in the "few *near feature* neighbours" conditions was compared to each individual's score on independent semantic processing tests. Mirman reported that there was a negative correlation between the difference in proportion of semantic errors across the conditions and the tests of semantic processing, suggesting that individuals with a greater semantic impairment showed larger detrimental effects of the number of *near feature* neighbours in terms of semantic errors.

Finally, Mirman and Graziano (2013) reported both facilitatory and inhibitory effects of *near feature* neighbourhood density in subgroups of a larger group of unselected aphasic individuals. Semantic neighbourhood effect size was calculated for each participant, reflecting the difference in picture naming accuracy between two sets of 36 items each with either few *near feature* neighbours or many *near feature* neighbours. The two sets were matched for log frequency, length, number of semantic features and number of phonological neighbours. Accuracy patterns varied across individuals: some showed better performance for words with many *near feature* neighbours and others the opposite (although the exact number of participants showing each pattern is not specified, nor is the significance of the difference between the two conditions within each individual). The aim of this study was to compare these different patterns of sensitivity to semantic neighbourhood density with the location of the patient's brain lesion: increased inhibitory effects were associated primarily with inferior frontal lesions, while increased facilitatory effects were not associated with any particular lesion location.

Overall, the effect of number of semantic neighbours on naming accuracy and type of error, remains unclear both for unimpaired speakers and speakers with aphasia. Although more evidence tends to point towards an increase in the number of semantic errors and a decrease in omission errors for words with many semantic neighbours, and no overall effect of the number of semantic neighbours on accuracy, it is not always the case. We still cannot conclude whether, and under what circumstances, having many words with similar meanings in the lexicon makes a given word easier or more difficult to retrieve compared to having few semantically related items in the lexicon, nor how semantic neighbourhood density affects the type of errors subjects make. The relationship between the effect of semantic neighbourhood density and the type of impairment (for example, some individuals might show a facilitatory effect for words with many semantic neighbours, and others might show an inhibitory effect for these type of words) also needs further investigation. In addition, it is still unclear whether each measure of semantic neighbourhood density captures the same aspect of behaviour and affects naming similarly because these different measures have not been compared within the same study.

Here, we report experiments on spoken word production in English in which we investigated the effect of several measures of semantic neighbourhood density on (i) naming latencies of a group of 50 unimpaired English speakers, and (ii) naming accuracy and type of errors in a large group of aphasic speakers (n=193). The aim of the investigation was to determine whether the number of semantically related neighbours facilitates or hinders naming performance, and, namely in the case of participants with aphasia, how they affect type of errors. Furthermore, we ask if the nature of the language impairment influences the effects of semantic neighbourhood on word production.

In addition, we wish to investigate the differences and similarities across the different measures of neighbourhood density and investigate the extent to which their effects on picture naming vary. Hence, in Experiment 1, we examine the existing measures of semantic neighbourhood density in more detail, while Experiment 2 explores the effects of these measures on picture naming in unimpaired subjects, and, subsequently, Experiment 3 investigates their effects on picture naming in people with aphasia.

Experiment 1: Comparison between the six semantic neighbourhood density measures.

In this section, we will first detail the stimuli that were used, and examine the similarities and differences in the characteristics of the different measures of semantic neighbourhood density for these items.

Method

Stimuli

The data for Experiments 2 and 3 was taken from databases of naming responses for unimpaired subjects (Székely et al., 2003) and people with aphasia (MAPPD, Mirman et al., 2010). In order to make our comparisons strictly comparable across experiments, we analysed only 86 stimuli that overlapped between two different sets of items. Data in the MAPPD (Mirman et al., 2010) consists of the performance of each individual with aphasia on the Philadelphia Naming Test (PNT) (Roach, Schwartz, Martin, Grewal, & Brecher, 1996). Consequently, all items in our selection were part of the Philadelphia Naming Test in which all items have at least 85% name agreement. Our naming data set was hence of high name agreement. The 86 final items were selected because they had full data available on a set of psycholinguistic variables: visual complexity, imageability, age of acquisition, word frequency, printed familiarity, phonological neighbourhood density, as well as measures of semantic neighbourhood density (see Appendix A). In particular, they had to appear in the McRae et al. (2005) feature norms and appear as targets in the University of Florida Free Association Norms (Nelson et al., 1998).

While these items are represented by black-and-white line drawings in both sources, the pictures themselves do not necessarily overlap between the two sets (20 of the 86 pictures are identical between the two sources). Consequently, most items had two different visual complexity measures, one for the picture from each source. These visual complexity measures were retrieved from Székely et al. (2004) and Mirman et al. (2010). The two databases used the same procedure to define visual complexity based on the file size of the image file (Székely & Bates, 2000). As noted earlier, name agreement measures for the 175 pictures from the PNT (Roach et al., 1996) were 85% and above (individual name agreement for each item is not provided). There are other name agreement measures for pictures from the IPNP (Székely et al., 2003); for the 86 pictures we selected, mean name agreement was 97%. However, we did not include name agreement within our control variables: Although name agreement measures are available for each item on the IPNP pictures, no measures at the item level were available for the pictures from the PNT. Collecting our own name agreement measures for all pictures was not feasible, since the English speakers at our disposal were speaking a different variety of English (Australian English) to the speakers from both data sets we analysed (both American English). Since name agreement was very high for both sets of pictures overall, and since we controlled for many variables that are usually highly correlated with name agreement (such as visual complexity, imageability, and familiarity), we believe this is unlikely to affect our results. Imageability ratings, reflecting the ease with which a word evokes a mental image for each item, were drawn from the Medical Research Council (MRC) Psycholinguistic Database (Wilson, 1988). Familiarity ratings (subjective ratings of how frequently a word had been encountered in the person's experience; Coltheart, 1981) were also obtained from the MRC database. Age of acquisition (AoA) ratings were taken from a database of 30,000 AoA ratings of English words (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012, obtained online: http://crr.ugent.be/archives/806). Lexical frequency measures (base 10 log lemma frequency per million) were taken from the Centre for Lexical Information (CELEX) database (Baayen, Piepenbrock, & Van Rijn, 1993). Phonological neighbourhood density was obtained using the program 'N-Watch' (Davis, 2005).

Measures of Semantic Neighbourhood Density

Values of semantic neighbourhood density were collected using the different approaches described in previous research: number of *rated competitors*, number of *near and distant feature* neighbours, number of *contextual* neighbours, and number of *association* neighbours. Table 2 gives examples of the different semantic neighbours under each measure for two items in our set ('butterfly' and 'lamp').

1) *Rated competitors*

For each of the 86 items, the number of *rated competitors* was obtained following the same procedure as Bormann (2011): 23 undergraduate students from Macquarie University were recruited (19 females, mean age = 21years, range 18-39,) and received course credit for their participation. They were given an explanation about the types of semantic relations that can occur between words (coordinates, associates, superordinates and subordinates), along with examples (e.g., crab and lobster are coordinates, but spider and web are associates). They were then asked to rate the experimental stimuli (the target word was presented with the original picture from the Philadelphia Naming Test on a 7 point scale, according to the estimated number of coordinates (from 1 = 'hardly any coordinates' to 7 = 'many coordinates'). The instructions included examples such as 'oak', a word that should be rated as having 'many coordinates' (pine, eucalyptus, gumtree, cypress, etc.), whereas, for example 'island', is expected to be rated as having 'hardly any coordinates'. The obtained means of the ratings for each item ranged from 2.17 ('anchor') to 6.17 ('dog') (mean number of *rated competitors*: 4.41).

Table 2. Examples of semantic neighbours according to the different semantic neighbourhood density measures.

	Rated competitors	Raw contextual	Categorical contextual	Near feature	Distant feature	Association
Butterfly	Mean rating: 6.13 NA	82 words: (including) larva, nymph, moth, grasshopper, metamorphosis, wasps, insect, hatch, lays, adult, crawls, tadpole, thorax, stages, insecticide, gypsy, eats	1 word: moth	9 words: airplane, blackbird, hornet, housefly, moth, owl, pigeon, raven, wasp	194 words: (including) canary, pony, sardine, tangerine, sweater, stone, pearl, peg, socks, sailboat, pistol, hut	20 words: (including) pretty, moth, caterpillar, fly, free, insect, bird, beautiful, flower, net, wings, cocoon, yellow, bug, delicate
Lamp	Mean rating: 4.30 NA	31 words: (including) bulb, incandescent, switch, flashlight, filament, light, fuses, toaster, ammeter, glows, circuit, socket, flashlights, plugged, amperes	3 words: bulb, flashlight, light	1 word: chandelier	41 words: (including) blender, cabinet, candle, carpet, drill, escalator, orange, toaster, telephone, razor, plate, napkin, mat	7 words: light, shade, bulb, post, desk, table, burn.

2) Raw Contextual neighbours & 3) Categorical Contextual neighbours

We obtained two measures of contextual neighbourhood density: raw contextual neighbourhood density and categorical contextual neighbourhood density. We first obtained lists of 'near neighbours' for each target using the 'Latent Semantic Analysis' (LSA) website of the University of Colorado at Boulder. 'Near neighbours' are defined as words within a chosen semantic space (topic spaces include specific ones as 'biology' or 'French fairy tales', we choose a 'general reading' one), that have a high LSA similarity to the target, based on co-occurence of these words within the semantic space. The number of words obtained following this procedure that had a similarity higher than 0.4 was used as a measure of ’raw contextual neighbourhood density‘. Then, following Kittredge et al.’s (2007a; 2007b) procedure, we selected words from these lists that belonged to the same semantic category as the target word. Word categories were selected from the six different published category norm studies used by Kittredge and her colleagues (Hunt & Hodge, 1971; Loess, Brown, & Campbell, 1969; Rosch, 1975; Shapiro & Palermo, 1970; Uyeda & Mandler, 1980; Van Overschelde, Rawson, & Dunlowsky, 2004). In order to be considered as a member of the same category as the target word, a given word in the LSA output had to be found as an exemplar of the same category as the target word according to any of these six databases. *Raw* measures of the number of *contextual* neighbours for the 98 items ranged from 1 (‘fan‘) to 153 (‘church‘) (mean 31.9), and *categorical contextual* neighbours (following Kittredge and colleagues’ category trimming procedure) ranged from 0 (‘train‘) to 9 (‘corn‘) (mean: 1.33).

4) Near Feature neighbours & 5) Distant Feature neighbours:

Feature neighbours were obtained by following the procedure used initially by Mirman and his colleagues (Chen & Mirman, 2012; Mirman, 2011; Mirman & Graziano, 2013; Mirman & Magnuson, 2008). Feature norms were drawn from the McRae et al. (2005) feature norm database, a corpus of 541 concepts for which features have been generated. This database was created by asking participants to list features that best described the given concept. They were encouraged to give different types of features, such as physical (perceptual) properties, functional properties, the category it belongs to or other encyclopaedic facts. For example, features for the word ’book‘ include ’found in libraries, has a hard cover, has authors, made of paper, used by reading, used for learning, etc…’. Each concept had features generated by 30 participants. A matrix is provided that indexes the number of shared features between two concepts taking into account the number of participants who produced a given feature. Based on Mirman and colleagues’ procedure, we defined near feature neighbourhood density as the number of words that had at least .4 cosine similarity (a ratio measure of the number of features in common) with the target word, and distant feature semantic neighbourhood density as the number of words that had more than 0 and up to .25 cosine similarity with the target word. Therefore, in terms of feature neighbourhood, our stimuli had from 0 (‘key‘) to 24 (‘owl‘) *near feature* neighbours (mean = 4.01) and from 12 (‘book‘) to 366 (‘bridge‘) *distant feature* neighbours (mean = 155.16). For example, the concept of ’sock‘ had 5 near feature neighbours including ’coat‘, ’mittens‘, ’shawl‘, ’slippers‘ and ’sweater‘, and 88 distant feature neighbours, such as ’dress‘, ’belt‘ but also ’trout ’flute‘ and ’ashtray‘.

We note here that words that share between 25 and 40% of features with targets do not belong to any feature neighbour group, although some of these concepts are clearly semantically related: for example, table and bench have 32% of features in common and hence are neither near feature neighbours nor distant feature neighbours.

6) *Association neighbours*:

This measure of neighbours was based on the University of South Florida free association norms (Nelson et al., 1998), where participants were asked to write the first word that came to mind or that was meaningfully related or strongly associated to the target word. An *association neighbour* was defined as a word produced by at least two different participants in this free association task. Consequently, when there was high agreement between participants in the words that were associated to a given word, the number of association neighbours was low. For example, ’cat‘ only had three association neighbours: ’dog‘, ’mouse‘ and ’kitten‘, because all subjects produced one of these three words in the free association task. Words with many association neighbours might then reflect those words that have a wide range of free associations but that these are less strongly associated to the target. For example, the 19 association neighbours of ’bottle‘ include ’beer‘, ’coke‘, ’glass‘, ’cap‘, ’drink‘, ’opener‘, ’fragile‘, ’medicine‘, etc… The number of association neighbours for each of the 86 words ranged from 3 (‘cat‘) to 25 (‘seal‘) (mean 13.63, standard deviation 4.96).

These different approaches all address the question of what the set of semantically related words is for any given word. However, they have different conceptual bases and they use quite different methods to define this neighbourhood. Moreover, as can be seen from the examples in Table 2, there is little overlap between the resulting set of semantic neighbours for a given word. Hence, it is important to examine how these measures are related to each other.

Analyses

We examined the relationships between the different measures of semantic neighbourhood density in three ways: by examining correlations between these measures, and correlations between them and other psycholinguistic variables, and, finally, by performing a principal component analysis.

Results

Correlations between measures of semantic neighbourhood density

The semantic neighbourhood measures showed little intercorrelation (Table 3), with only three significant correlations. The two measures of *contextual* semantic neighbourhood density (*raw* and *categorical*) were weakly but significantly correlated: the more *raw contextual* neighbours, the more *categorical contextual* neighbours. As the *categorical* contextual neighbours are a subset of the *raw* contextual neighbours this correlation is relatively unsurprising. Another significant (moderate) correlation was observed between *rated competitors* and *near feature* neighbours. These two measures are likely to reflect the number of words belonging to the same category: *rated competitors* are an estimate of the number of words within the same semantic category, and *near feature* neighbours share many semantic features and are so also likely to be from the same semantic category.

Table 3. Bivariate correlations (Pearson's r) among the different measures of semantic neighbourhood density and other psycholinguistic predictors (N=86).

	Rated competitors	Contextual neighbours	Contextual neighbours, category-trimmed	Near feature neighbours	Distant feature neighbours	Associates
Contextual neighbours	.184					
Contextual neighbours, category-trimmed	**.250***	**.376****				
Near feature neighbours	**.521****	-.056	-.116			
Distant feature neighbours	-.012	-.020	-.008	.058		
Associates	.117	.107	.028	.016	.040	
Visual complexity	-.044	**.293****	.085	-.026	.208	.176
Imageability	**.336****	.240*	.199	**.265***	-.061	.076
Log lemma frequency	.107	.195	.153	-.169	-.116	.123
Familiarity	.074	.013	.168	-.078	-.144	-.028
Age of acquisition	**-.394****	-.025	-.118	-.065	.115	-.022
Phonological neighbourhood density	.171	.001	.063	-.023	-.181	-.052
Length in phonemes	-.157	-.041	-.032	-.067	.103	.142

**Correlation is significant at the 0.05 level (2-tailed); ** Correlation is significant at the 0.01 level (2-tailed). Significant correlations are in bold*

Similarly, there was a weak correlation between the number *of contextual neighbours from the same category* and the number of *rated competitors*. Again, these two measures are meant to reflect the number of words from the same semantic category, it is hence unsurprising that they show some correlation. The number of *association* neighbours and the number of *distant* feature neighbours showed very low correlations overall and no significant correlations with any other semantic neighbourhood density measures.

Correlations with Psycholinguistic Variables

The semantic neighbourhood density measures also showed four significant but weak correlations with the other psycholinguistic predictors: *rated competitors* were correlated with imageability and age of acquisition: items with more rated competitors were of higher imageability, and acquired earlier. The number of *near feature* neighbours was positively correlated with imageability and finally, the number of *raw contextual* neighbours was also positively correlated with visual complexity and imageability (correlations are shown in Table 3).

In addition, many significant correlations were observed between the psycholinguistic variables that are most often taken into account in picture naming studies, the strongest correlations being between familiarity, log lemma frequency and age of acquisition on the one hand, and length in phonemes and phonological neighbourhood density on the other. These intercorrelations are shown in Appendix B.

Principal Components Analysis

In order to better understand the structure of correlations among the six measures of semantic neighbourhood density and the other psycholinguistic predictors, a principal component analysis was conducted, allowing estimation of the pattern of relations between potential common factors behind these measures, and each of these individual measures. The principal component analysis was conducted on 13 measures (the six measures of semantic neighbourhood density and the seven other psycholinguistic variables) with orthogonal rotation (varimax). The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis, KMO=.603 (acceptable according to Kaiser, 1974). Bartlett's test of sphericity (χ^2 (78) = 286.914, p<.001), indicated that correlations between items were sufficiently large for principal component analysis. An initial analysis was carried out to obtain 'eigenvalues' for each component in the data. Seven factors had eigenvalues over Joliffe's criterion of .7 and in combination explained 80.44% of the variance. We deliberately chose a

Table 4. Principal component analysis: Summary of the factor loadings for each predictor variable.

	1	2	3	4	5	6	7
	Lexical	Post-lexical	Feature-based	Contextual	Visual complexity	Distant feature	Associates
Log lemma frequenc	**0.816**	-0.272	-0.053	0.139	0.207	0.068	0.065
Familiarity	**0.888**	-0.066	-0.039	0.033	-0.204	-0.047	-0.075
AoA	**-0.716**	0.211	-0.244	0.002	0.262	-0.107	0.010
PND	0.141	**-0.893**	0.050	0.036	-0.074	0.008	-0.031
Length	-0.271	**0.863**	-0.103	-0.031	0.036	0.142	-0.025
Near Feature	-0.074	-0.082	**0.841**	-0.209	0.070	-0.030	0.053
Rated Competitors	0.094	-0.133	**0.825**	0.205	-0.132	0.163	0.022
Raw contextual	0.069	-0.017	0.044	**0.798**	0.266	-0.044	0.001
Cat. contextual	0.050	-0.052	-0.014	**0.815**	-0.210	-0.178	0.008
Visual complexity	-0.153	0.089	-0.018	0.039	**0.884**	-0.143	0.008
Distant feature	0.064	-0.099	0.062	0.101	0.139	**0.932**	0.022
Associates	-0.024	0.007	0.040	0.004	0.009	0.022	**0.995**
Imageability	*0.371*	*-0.329*	*0.506*	*0.329*	0.068	-0.174	-0.097

AoA= Age of acquisition, PND=phonological neighbourhood density, Cat. contextual = categorical contextual. Highest loadings on each factor are in bold.

lenient criterion regarding the number of factors to retain, to have as many factors as needed to have each of the measures of semantic neighbourhood density included in a factor. Table 4 above shows the factor loadings after rotation (values over .7 are in bold). The loading of a given variable can be understood as the degree to which each variable “correlates” with the common factor. Therefore, inspection of factor loadings (ranging from -1 to 1) reveals the extent to which each variable contributes to the meaning of the factor.

The seven resulting factors were as detailed below, with the corresponding factor scores being used in subsequent analyses (factor scores for each of the 86 items are presented in Appendix C):

Factor 1 (Lexical Predictor): Log frequency, familiarity and age of acquisition had a high loading on the first factor. This is consistent with previous findings that have shown that familiarity is dependent on age of acquisition and word frequency (e.g., Gilhooly & Logie, 1980), and that word frequency and age of acquisition are highly related (e.g., Carroll & White, 1973). We labelled this factor ’lexical predictor‘. Since lemma frequency and familiarity are positively correlated with this factor, and age of acquisition is negatively correlated with it, a high factor score for a given word’s lexical predictor means this item is of high frequency and familiarity, and low (early) age of acquisition.

Factor 2 (Post-Lexical Predictor): Phonological neighbourhood density and word length in phonemes were the most important measures in a second factor. It has been widely shown that longer words have fewer phonological neighbours and that in general these two measures 'length' and 'phonological neighbourhood density' are highly (negatively) correlated (e.g., Bard & Shillcock, 1993; Pisoni et al., 1985). Length was positively correlated with this factor, and phonological neighbourhood density negatively correlated with it. Hence an item with a high (positive) score on this second factor was longer and with fewer phonological neighbours.

Factor 3 (Feature-based semantic neighbourhood density): Near feature semantic neighbours and number of *rated competitors* were the main loadings on the third factor, consistent with the pattern of correlations. These two measures could be considered to both reflect the shared semantic features (semantic properties) between target and neighbours, consequently, we named the factor that included the largest loadings for these two variables '*feature-based* semantic neighbourhood density'.

Factor 4 (Contextual semantic neighbourhood density): A fourth factor was identified that included both *contextual* semantic neighbourhood density measures (the *raw* and the *categorical* measures), again in line with the results of the correlational analyses. These two measures reflect the similarity of use of given words in a semantic context (semantic properties given by the context). Hence, the fourth factor '*contextual* semantic neighbourhood density' represented the 'combination' of these two variables.

Factors 5-7: Visual complexity, number of association neighbours and the number of distant feature neighbours did not cluster with any other variables, suggesting that each of these measures is quite independent from the others and, hence, they represent independent factors.

Imageability did not have a sufficiently high loading to be included in any factor but only had low loadings on the first four factors (lexical, post-lexical, categorical semantic neighbours and contextual semantic neighbours).

Discussion: **Experiment 1**

We identified six different types of semantic neighbours used in previous studies, and examined how these measures were related to each other (within our set of 86 pictureable nouns). Perhaps surprisingly, there were very few correlations between the different measures. Indeed, the combination of the correlations and the principal components analysis suggests that the semantic neighbourhood density measures can be thought to be of four different types, perhaps reflecting different semantic relationships. This perhaps provides some insight into the different effects of semantic neighbourhood density found in the literature previously and reinforces the importance of examining the effect of each measure with the same stimuli and the same sample population.

Some semantic neighbourhood density measures did not seem to share an underlying factor with other measures. First, a closer look at the *association* neighbours reveals a variety of types of meaning relationship between the responses and the target: some are clearly conceptually associated, sharing no or very few semantic features (like 'saddle-horse'), others are just word associations (like 'banana-republic'), while many share several semantic features (as in 'knife-fork').

We further analysed this by calculating the percentage of association neighbours that were also coordinates or superordinates (words that shared many semantic features with the target, like 'cat' and 'animal' for the target 'dog') for each associated word of each target word (there were very few superordinates overall). This proportion correlated negatively with the number of association neighbours ($r(84)=-.475$, $p<0.001$), meaning that, for words with few association neighbours (strongly associated neighbours), these neighbours were also likely to share many features with the target.

Hence, *association* neighbours include different relationships with the target, some even not being semantic, and this measure reflects more the strength of the association between the target and some of its neighbours, these strongly associated neighbours being likely to share many semantic features with the target, too. In sum, this measure is different from *feature-based* and *context-based* semantic neighbours, as it gives more information about how strongly some words are semantically related to the target rather than information about the number of such words.

The *distant* feature neighbourhood density measure also seems unrelated to the other measures. This measure, unlike the others, does not target the words whose meaning is most "similar" to a given word, but rather the number of words that are not completely semantically unrelated but somehow, albeit sometimes tangentially, related. For example, 'bear' and 'submarine' are distant feature neighbours because both have the feature 'is large'. Hence, the conceptual basis of this type of neighbour is very different from the other semantic neighbours.

Finally, none of the semantic neighbourhood density measures were shown to have an underlying component with any of the other psycholinguistic measures (see Table 4 above), reinforcing that this is an independent factor with potentially important theoretical implications.

Considering how different these semantic neighbourhood measures are, as supported by the principal components analysis, it is not surprising that they have been associated with different effects in speech production. It is important to compare their specific influence in picture naming, to understand better what type of semantic representations play a role in the production of words and in what manner. The potential influence of feature-based measures, for example, would make sense within theories of word production that assume a level where semantic features of the target word are represented, such as in Dell's interactive activation model (e.g., Dell, Schwartz, Martin, Saffran, & Gagon, 1997). Moreover, an influence of contextual neighbours could be explained by theories that state that the similarity of meaning between words in the lexicon is determined by the contexts in which these words appear (e.g., Landauer et al., 1998), while the possible influence of associates in speech production would speak for theories that assume words in the lexicon are organised based on individuals' real-life experience with words (e.g., Nelson et al., 1998), or would underline the importance of the strength of the relationship between a word and its semantic neighbours, rather than the number of semantic neighbours.

In the next two experiments, the influence of the four resulting semantic neighbourhood measures factors will be investigated in spoken picture naming, both with unimpaired speakers and with individuals with aphasia. We want to be able to compare how differently each of these measures impact both types of speakers' picture naming behaviour, what insights this can provide into the discrepancies in the previous literature, and how their influence can be accommodated within theories of spoken word production.

Experiment 2: Influence of semantic neighbourhood density in unimpaired picture naming

In this experiment, we examined the effect of the different measures of semantic neighbourhood density on unimpaired picture naming in order to dissociate the effects of these factors. We used response time (RT) data from Székely et al. (2003) who report mean spoken picture naming latencies from 50 unimpaired English speakers on a large set of items which are accessible online ('International Picture Naming Project' (IPNP) on the Centre for Research in Language of the University of California at San Diego CRL-UCSD website: http://crl.ucsd.edu/experiments/ipnp/index.html).

Method

Participants

The fifty participants from Székely et al. (2003) were right-handed, monolingual native speakers of English, who were students (aged 18-25; 35 female and 15 male) at the University of California at San Diego. **Stimuli**

Stimuli used in Székely et al. (2003) consisted of black and white drawings of 520 common objects taken from different sources of picture materials (including from Snodgrass & Vanderwart, 1980). Among these 520 items, we focused our analyses on the 86 nouns that we selected based on the availability of all of our predictors (see Experiment 1). As mentioned earlier, the 86 pictures had 97% name agreement on average (SD=6%), and their visual complexity ranged between 5156 and 39085 (mean 16213, SD 7866).

Procedure

The procedure of the picture naming task is fully described in Székely et al. (2003). At the start of the naming task, voice trigger sensitivity was calibrated for each of the participants in a repetition task, followed by practice items. Participants were instructed to name the pictures that would appear on the screen as fast as possible without making mistakes. Pictures were displayed on a computer screen, preceded by a fixation cross (200msec) and a blank (500msec). The reaction time (RT) associated with each response was recorded, and the whole word response was audio-recorded.

Scoring

The IPNP database describes two types of coding of the responses produced by participants: first, the response is coded as ’valid‘, ’invalid‘ (when the recording is not usable because of noise, or low quality etc.), or ’no response‘. Only valid responses were further coded into four categories: ‘Lexical code 1’ referred to the target name (a correct response), ‘Lexical code 2’ was used for any morphological or morpho-phonological alteration of the target name (‘bike‘ for bicycle, ’cookies‘ for ’cookie‘), while synonyms that did not share any morphology with the target (‘couch‘ for ’sofa‘ or ’chicken‘ for ’hen‘) were labelled under ‘Lexical code 3’. Finally, any ’other‘ responses were included under ‘Lexical code 4’. The latter category contains superordinates (‘animal‘ for ’dog‘), semantic associates/coordinates (‘cat‘ for ’dog‘), part-whole relationships (‘finger‘ for ’hand‘), visual errors and unrelated errors. Therefore, this ’Lexical code 4’ category is the one most likely to reflect the rate of semantic errors for a given word. For each response type, mean reaction times across trials are available from Székely et al. (2003) database.

Analyses

In order to assess the influence of the four semantic neighbourhood density predictors on the speed of response while controlling for the effect of common predictors of picture naming latencies, multiple regressions were applied to mean response times for correct responses (valid responses coded with 'Lexical code 1'), with the seven factor scores described above entered simultaneously as predictors. Using factor scores has the advantage of overcoming multicollinearity (here, two or more predictor variables are highly correlated), because these scores are not correlated, unlike the initial variables. Next, the pattern of response was analysed by means of multiple regressions assessing whether semantic neighbourhood predictors could predict, first, the rate of correct responses on each item, and second, the rate of semantic errors ('Lexical code 4') for each item as well. Analyses were performed using SPSS (IBM, 2013).

Results

Influence of semantic neighbourhood measures on naming response latency

Table 5. Summary of multiple regression analysis on response times for correct responses (n=86).

Variable	*B*	*SE B*	*β*
(Constant)	860.419	12.731	
Lexical predictors	-61.544	12.806	**-.463****
Post-lexical predictors	30.141	12.806	**.227***
Visual Complexity	9.565	12.806	.072
Semantic neighbourhood factors			
Feature-based	4.821	12.806	.036
Context-based	-7.395	12.806	-.056
Association	.630	12.806	.005
Distant	4.606	12.806	.035

Overall model : R^2=.276, and F= 4.257**

**p<.05. **p<.01 (significant effects in bold).*

The first analysis used response time for correct answers as a dependent variable. As shown in Table 5, no measure of semantic neighbourhood density was found to predict response time. Only the “lexical” predictor (the factor underlying log frequency, familiarity and age of acquisition), and the “post-lexical” predictor (the factor representing length and phonological neighbourhood density) significantly predicted response time.

Influence of semantic neighbourhood measures on naming accuracy

Table 6. Summary of multiple regression analysis on accuracy (n=86).

Variable	*B*	*SE B*	*β*
(Constant)	.971	.006	
Lexical predictors	.013	.006	**.223***
Post-lexical predictors	-.009	.006	-.149
Visual complexity	.005	.006	.076
Semantic neighbourhood factors			
Feature-based	-.001	.006	-.019
Context-based	.003	.006	.052
Association	-.001	.006	-.019
Distant	.004	.006	.065

Overall model : R^2=.086, F=1.044

**p<.05 (significant effect in bold).*

A second analysis assessed the influence of the different semantic neighbourhood density variables on accuracy (using per item mean accuracy- as the dependent variable). The multiple regression model was not significant and there were no significant effects of the semantic neighbourhood density variables on the rate of correct responses (see Table 6). Only the factor underlying lexical predictors predicted a correct response.

Influence of semantic neighbourhood measures on semantic errors

The last analysis focused on the rate of semantic errors that were observed for the targets. This was computed by using ‘Lexical code 4’ as the dependent variable. This code is most likely to reflect the rate of semantic errors that were spontaneously produced by participants, although it does also include visual and unrelated errors. As for accuracy, the model was not significant overall and no measure of semantic neighbourhood density predicted the incidence of ‘code 4’ errors, only the lexical predictors were significant (see Table 7).

Table 7. Summary of multiple regression analysis on ‘Lexical Code 4’ errors (N=86).

Variable	*B*	*SE B*	*β*
(Constant)	.023	.005	
Lexical predictors	-.011	.006	**-.221***
Post-lexical predictors	.010	.006	-.187
Visual complexity	-.007	.006	-.141
Semantic neighbours			
Feature-based	.000	.006	-.006
Context-based	-.003	.006	-.061
Association	.000	.006	.006
Distant	-.005	.006	-.106

Overall model: R^2=0.119, F=1.504

*p<.05. (significant effect in bold)

Discussion: Experiment 2

Consistent with Bormann (2011) and with Mirman (2011), we found no effect of semantic neighbourhood density on picture naming latencies in this sample of unimpaired participants. Moreover, this was true not only for the *feature-based* semantic neighbours (the component that underlies the number of *rated competitors* of Bormann’s analysis and the number of *near feature* neighbours of Mirman’s analysis) but also for all the other types of semantic neighbourhood density. However, the findings were not consistent with Mirman (2011) who found an inhibitory effect of *near feature* neighbours on accuracy, although the paradigms differed, with Mirman using highly speeded naming rather than a standard picture naming paradigm. These are, to our knowledge, the only studies to have looked at the influence of semantic neighbourhood density on unimpaired picture naming and their main findings were not challenged by the present study. It seems clear from this experiment and previous literature that, in unimpaired picture naming, there is no detectable influence of the number of semantically related neighbours (however defined) on the speed of lexical access nor on the production of errors for spoken picture naming. Hence, however these effects arise, it is apparent that under normal processing conditions their effects are too small to be detectable in picture naming latencies under normal conditions.

Experiment 3: Semantic neighbourhood density in aphasic picture naming

In this experiment we studied the influence of the different measures of semantic neighbourhood density on spoken picture naming in people with aphasia using a similar procedure to Experiment 1, except with a focus on response type rather than reaction time

Our aim was to investigate whether, in aphasic speakers, semantic neighbours make a given word more or less accurate. Indeed, if many semantically related representations are activated along with the target, if there is competition occurring between these representations, it would be predicted to be harder to select the relevant word, as observed by Mirman (2011) and Mirman and Graziano (2013). Conversely, activation of the target might benefit from the presence of many semantic neighbours via increased activation of its semantic features in an interactive framework (facilitation), as observed by Kittredge et al. (2007a) and Mirman and Graziano (2013).

We also wished to examine whether having many semantic neighbours specifically affects the likelihood of selecting a semantically related word instead of the correct target word, hence influencing the rate of semantic errors compared to correct responses, as observed by Mirman (2011): in individuals with aphasia, it might be more difficult to select the right word in the presence of highly activated competitors who might be selected instead. On the other hand, having many co-activated semantically related items might make it less likely for an omission to occur. If this is the case, then the likelihood of an omission compared to a correct response would be affected by semantic neighbourhood as found by Kittredge et al. (2007b).

We also investigated the relative probability of producing a semantic error over an omission (Blanken et al., 2002; Bormann, 2011; Bormann et al., 2008; Kittredge et al., 2007b): if there is increased activation of a set of semantically related words, then it may be more likely to observe relatively more semantic errors and fewer omissions on targets with many semantic neighbours. Blanken, Bormann, Kittredge and colleagues (Blanken et al. , 2002; Bormann, 2011; Bormann et al. , 2008; Kittredge et al. , 2007b) argued that semantic errors and omission errors resulted from non-independent processes, hence the relevance of this particular comparison.

Finally, as it is believed that, in speech production, activation spreads from the semantic level down to the phonological level, it might be the case that semantic neighbourhood density impacts the possibility of (not) producing a phonological error: if the target is activated more strongly at the lexical level by the presence of many semantic neighbours, it might help overcoming potentially reduced activation at subsequent levels of processing (e.g. phoneme level) resulting in relatively fewer phonological errors. We therefore decided to examine the probability of a correct response compared to a phonological error in our analyses.

Therefore, below, we will report the results of the investigation of the effects of semantic neighbourhood on the picture naming performance of aphasic participants, looking at correct versus incorrect responses, semantic errors versus correct responses, omissions versus correct responses, semantic errors versus omissions, and phonological errors versus correct responses. We hope to be able to differentiate possible effects of distinct semantic neighbourhood measures within the same study, and to further distinguish the effects at different levels of impairment.

Method

Participants

Data from aphasic speakers was taken from the Moss Aphasia Psycholinguistic Project (MAPPD), an online database of aphasic picture naming (Mirman et al., 2010, available upon registration at www.mappd.org). The naming performance of 265 patients with aphasia on the 175 items included in the Philadelphia Naming Test (PNT) (Roach et al., 1996) was available on the website at the time when the data was downloaded, along with relevant subject- and item-related variables.

The participants with aphasia included in the MAPPD database were 265 individuals from which we selected 193 who were neither at ceiling nor at floor on the Philadelphia Naming Test (Roach et al., 1996; accuracy between 10 and 90%). These 193 selected participants (mean age: 59 years, SD=13, range 22-86; mean number of years of education: 14, SD=3, range 7-22) were tested at a subacute or chronic stage of aphasia (average number of months post-onset 32, SD=47, range 1-381). Based on their Western Aphasia Battery (WAB; Kertesz, 1982) profile 70 participants were diagnosed as anomic, 48 Broca, 43 conduction, 29 Wernicke and 3 transcortical sensory, and covered a wide range of severity (Western Aphasia Battery Aphasia Quotient mean 72, SD=16, range 27- 95). English was the first and primary language for all individuals (less than 5% reported the regular use of another language), and all were recruited at the Moss Rehabilitation Research Institute. In addition to Philadelphia Naming Test (Roach et al., 1996) data, the MAPPD includes patient performance on other linguistic and neuropsychological tests. For instance, patient performance on the Philadelphia Repetition Test using the 175 items of the Philadelphia Naming Test is available for most participants. Performance on other tests in the database include aphasia diagnostic tests such as the Western

Aphasia Battery (Kertesz, 1982), speech perception and recognition tests (e.g., lexical decision), semantic tests like the Pyramids and Palm Trees Test (Howard & Patterson, 1992) and other tests of cognitive function (e.g., measures of short-term memory).

In order to be able to determine the nature of the impairment in word production of the participants with aphasia (that is, whether they had more difficulties at the semantic or at the phonological level), we computed two different measures based on the participant's performance on several tests. We chose, as much as possible, to use measures that would be independent from their performance on the Philadelphia Naming Test (Roach et al., 1996).

First, to obtain a measure of the degree of phonological impairment, we computed z-scores for measures of phonological processing: performance on the Philadelphia Repetition Test, a non-word repetition test, as well as the percentage of phonological errors that the patient made on the Philadelphia Naming Test. The Philadelphia Repetition Test (PRT) is a repetition test that uses the 175 items of the Philadelphia Naming Test, and the non-word repetition tests (Dell, Martin, & Schwartz, 2007) are tests that were created based on 60 words from the PNT: each word was turned into a phonologically legal non-word by replacing two of its phonemes. The phonological z-score was the mean of these three measures, and hence provided an indication of how severe the individual's phonological impairment was relative to the overall sample.

Second, a semantic processing measure was created based on the participant's performance on the following tests (all converted into z-scores): the Synonymy Judgments with Nouns and Verbs Test (Martin, Schwartz, & Kohen, 2006), the Pyramids and Palm Trees test (Howard & Patterson, 1992), and the Camel and Cactus test (Bozeat, Lambon Ralph, Patterson, Garrard, & Hodges, 2000). The Synonymy Judgments with Nouns and Verbs Test (Martin et al., 2006) requires participants to determine which two words are most similar in meaning from a choice of three words (either three nouns or three verbs) that are presented in both visual and auditory modalities. In the 52 trials of the Pyramids and Palm Trees test (Howard & Patterson, 1992), participants are presented with a pictured item that has to be matched to the picture representing the closest associate from two possibilities. The Camel and Cactus test is very similar, except that it has more trials (64) and the participant has four different pictures to choose from.

Of the 193 patients in our sample, 50 were excluded from these further analyses as they did not have data on the 'semantic' and 'phonological' measures. For the 143 remaining participants we used median splits of the semantic z-scores, and the phonological z-scores, to create four subgroups manipulating the degree of phonological and semantic impairment as summarised in Table 8.

Individuals with a lower z score (negative) are those who have scored relatively more poorly on that measure and so have a relatively greater impairment.

Table 8. Distribution of patients with aphasia as a function of the type of impairment: Mean semantic and phonological z-scores for each group.

Impairment	n	Mean z-score	
		Phonological	Semantic
Phon↘Sem↘ = Mixed impairment Poorer phonological & poorer semantic abilities	40	-1.39	-2.07
Phon↗Sem↗ = Mild impairment Better phonological & better semantic abilities	39	1.44	1.86
Phon↘Sem↗ = Phonological impairment Poorer phonological & better semantic abilities	32	-1.35	1.87
Phon↗Sem↘ = Semantic impairment Better phonological & poorer semantic abilities	32	1.33	-1.33

Stimuli

The MAPPD database reports naming performance on the Philadelphia Naming Test. The Philadelphia Naming Test is a standardized set of 175 black-and-white line drawings of limited visual complexity representing single nouns covering a relatively large range of semantic categories (animal, body parts, clothing, food, furniture, tools, vehicles, etc.). We used a subset of 86 items chosen according to the criteria described in Experiment 1. As we noted earlier, although the stimuli are identical, some of the pictures were different to those of Experiment 2, and hence, had different objective visual complexity ratings and name agreement. The rest of the control predictors (imageability, age of acquisition, familiarity, log lemma frequency, phonological neighbourhood density, length, and the different measures of semantic neighbourhood density) were identical to those used in in Experiment 1. The same procedure of a principal component analysis was applied, in order to generate new factor scores that took into account the different measure of visual complexity. This analysis led to the same seven factors as those obtained in Experiment 1 with the visual complexity measures corresponding to the stimuli of Experiment 2, with very similar factor scores.

Procedure

As described in Mirman et al. (2010), the aphasic participants were tested according to the Philadelphia Naming Test administration guidelines (available online, along with test materials: http://www.mrri.org/index.php/philadelphia-naming-test). Pictures were displayed using a Powerpoint presentation on a computer screen, and the examiner advanced the pictures manually during presentation, using a 30 second deadline for naming attempts. Ten practice trials were displayed before the 175 naming trial items. Participants were instructed to name the picture using only one word. The examiner avoided providing any verbal or non-verbal cue prior to the participant's response, but would give him/her feedback following his/her response (saying if the response is accurate, and giving the correct word if the patient could not produce it). Responses were audio-recorded and transcribed during test administration.

Scoring

The MAPPD database provides transcriptions (phonetic and orthographic) of the first complete response produced and an accuracy measure. In addition, two different categorizations of the response types are provided: the 'conventional' coding, and the 'model' coding that was implemented to test the interactive two-step computational model of speech production of Dell and his colleagues (Dell & O'Seaghdha, 1991). Here, we used the conventional coding. This coding refers to the most commonly used designation for aphasic and non-aphasic naming performance (correct, semantic error, formal error (a real word sharing phonemes with the target), mixed error (both semantically and phonologically related), non-word error (phonologically related to the target), description, omission, picture part, perseveration, abstruse neologism (non-word not sharing phonemes with the target, etc.). It follows the guidelines recommended for the Philadelphia Naming Test administration: phonological relatedness is defined as sharing a minimum of one phoneme in the correct position (excluding schwa; more details are provided in the Philadelphia Naming Test scoring guide). This coding of formal errors differs, for example, from the criterion used by Goodglass and Kaplan (1983), which requires an overlap of half the target's phonemes. Each response receives a Level 1 code: for example Level 1 code 1 is correct, code 2 is 'target attempt' and encompasses all phonologically related responses, including words and non-words, 3 is semantic, 4 is mixed, etc.. 'Level 2' codes are used to specify the type of error in the case of a Level 1 code 2 (phonological errors). For our analyses, in addition to accuracy measures, we were interested in several types of errors, and we used the Level 1 coding to select these error types. We combined semantic and mixed errors to represent all those responses with a semantic component: these errors were labelled 'semantic errors'. We used the 'no response' coding for omissions, and the 'target attempt' coding as phonological errors. Hence, phonological errors include both word and non-word errors that are phonologically related to the target word. Table 9 shows the breakdown of the different response types (correct, semantic errors, phonological errors, and omissions) for the whole group as well as for each subgroup of participants.

Table 9. Mean percentage (and standard deviation) of each response type when naming 86 pictures for participants of the whole group (N=193), the sum of the four subgroups (n=143) and of each of the subgroups.

	Correct		Semantic errors		Phonological errors		Omissions	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
N=193	52.48	19.91	7.45	4.93	11.66	10.64	6.77	10.26
n=143	52.09	19.58	7.01	4.74	12.39	11.15	6.89	6.89
Mixed impairment	36.45	17.50	8.73	5.52	16.88	12.52	10.83	13.15
Mild impairment	63.46	16.55	5.00	3.86	5.67	4.65	3.18	4.52
Phonological impairment	50.84	18.98	4.38	3.16	21.13	10.56	2.81	4.50
Semantic impairment	52.48	19.91	8.72	4.58	4.88	2.97	8.84	10.67

Analyses at the group level

For the analyses in Experiment 3, we took the same predictors as used in Experiment 1. We used generalised linear mixed effects models for binomial outcomes to analyse the factors influencing accuracy and the occurrence of different types of errors. Statistical analyses were performed using the package lme4 (Bates, Maechler, Nolker, & Walker, 2014) in the statistical software R (R core team, 2014). Analyses were run for the whole group of 193 patients, as well as on the subgroup of 143 patients for whom degree of semantic and phonological impairment could be established, then on each of the four subgroups with different patterns of impairment.

To account for random variation caused by specific items or participants, participants and items were entered as random factors (with random intercepts). Predictors that were entered as fixed effects were the seven factor scores that were previously computed by means of the factor analysis. Separate models were fitted with different dependent variables as described above: accuracy (correct vs incorrect responses), number of semantic errors versus correct responses, number of omissions versus correct responses, number of semantic errors versus omissions, and number of phonological errors versus correct responses.

All models included the lexical, post-lexical, and visual complexity factor scores in order to control for these usual predictors of picture naming behaviour, as well as the four semantic neighbourhood factor scores[1]. Given that our interest here is whether semantic neighbourhood density measures significantly change the odds of obtaining one type of response over the other, we only report the coefficients for the four semantic neighbourhood density predictors.

Results

Table 10 summarises the results of analyses for the whole (n=193) and reduced (n=143) groups, and Table 11 for the impairment-based subgroups divided by the level of semantic and phonological impairment.

Correct responses versus any other response type: In the analyses examining which semantic neighbourhood measures predicted correct responses rather than any other response type, there was a significant effect of feature-based neighbourhood density for the whole group as well as for the reduced group: words with many feature-based semantic neighbours were more likely to be correctly named than words with few feature-based semantic neighbours. Within the impairment-based subgroups, the 'mild' impairment group (and the 'mixed' impairment group, but the effect was marginally significant) also showed this beneficial effect of having many feature-based neighbours.

None of the other semantic neighbourhood density measures were significant predictors of accuracy at the full group level, but two subgroups showed some effect of other semantic neighbourhood density types on accuracy. First, in the mixed impairment group, words with many association neighbours (few strongly associated neighbours) were less likely to result in a correct response. Second, in the group with more of a phonological impairment, words with many contextual neighbours were more likely to be correctly named compared to words with few contextual neighbours. The interaction between the distant semantic neighbourhood factor and the lexical factor was marginally significant for the mixed impairment group: a beneficial effect of distant neighbours on accuracy was present on targets with a low lexical factor score, but this effect decreased with increasing lexical factor score (no other interaction was found to be significant in the accuracy analyses).

[1] Hence, each model was run on a different set of data: there were five response type comparisons, and six different groups / subgroups of participants, resulting in 30 different data sets / models.

Table 10. Summary of the fixed effects in the generalised linear mixed models fitted for different error types, in the total group (N=193) and the reduced total group (n=143), along with all significant or marginally significant interactions (all significant effects are in bold).

		193				143			
	Neighbour type	*β*	*SD β*	*z*-value	*p*-value	*β*	*SD β*	*z*-value	*p*-value
correct vs incorrect	Feature	**0.13**	**0.06**	**2.18**	**0.03**	**0.12**	**0.06**	**2.02**	**0.04**
	Contextual	0.08	0.06	1.36	0.17	0.09	0.06	1.53	0.13
	Distant	0.04	0.06	0.66	0.51	0.03	0.06	0.52	0.60
	Association	-0.08	0.06	-1.43	0.15	-0.09	0.06	-1.50	0.13
correct vs semantic	Feature	-0.09	0.10	-0.93	0.35	-0.06	0.10	-0.59	0.56
	Contextual	0.14	0.10	1.38	0.17	**0.21**	**0.11**	**1.96**	**0.05**
	Distant	0.05	0.10	0.45	0.65	0.06	0.10	0.58	0.56
	Association	-0.05	0.10	-0.48	0.63	-0.05	0.10	-0.44	0.66
	Association*Lexical	*-0.17*	*0.10*	*-1.70*	*0.09*				
correct vs omissions	Feature	**0.18**	**0.08**	**2.21**	**0.03**	**0.18**	**0.08**	**2.13**	**0.03**
	Contextual	0.05	0.08	0.58	0.56	0.05	0.09	0.64	0.52
	Distant	0.13	0.08	1.59	0.11	0.14	0.09	1.62	0.11
	Association	**-0.27**	**0.08**	**-3.30**	**<0.001**	**-0.27**	**0.09**	**-3.09**	**<0.001**
	Distant*Lexical	**-0.21**	**0.08**	**-2.54**	**0.01**	**-0.22**	**0.09**	**-2.55**	**0.01**
correct vs phonological	Feature	**0.18**	**0.05**	**3.71**	**<0.001**	**0.16**	**0.05**	**3.13**	**<0.001**
	Contextual	0.04	0.05	0.89	0.38	0.04	0.05	0.77	0.44
	Distant	0.01	0.05	0.27	0.79	0.02	0.05	0.36	0.72
	Association	-0.04	0.05	-0.91	0.36	-0.04	0.05	-0.85	0.40
	Feature*Lexical					*0.09*	*0.05*	*1.67*	*0.09*
semantic vs omissions	Feature	**0.24**	**0.08**	**3.05**	**<0.001**	**0.23**	**0.08**	**2.74**	**0.01**
	Contextual	*-0.14*	*0.08*	*-1.78*	*0.07*	**-0.21**	**0.09**	**-2.29**	**0.02**
	Distant	0.06	0.08	0.82	0.41	0.05	0.09	0.58	0.56
	Association	**-0.22**	**0.08**	**-2.79**	**0.01**	**-0.23**	**0.09**	**-2.71**	**0.01**
	Association*Lexical	**0.17**	**0.08**	**2.16**	**0.03**				

Table 11. Summary of the fixed effects of semantic neighbourhood variables and significant or marginally significant interactions in the generalised linear mixed models fitted for different error types, for each subgroup of participants with aphasia (significant effects in bold, marginally significant in italics).

		Mixed				mild				semantic				phonological			
	Neighbour type	*β*	*SD β*	*z*-value	*p*-value	*β*	*SD β*	*z*-value	*p*-value	*β*	*SD β*	*z*-value	*p*-value	*β*	*SD β*	*z*-value	*p*-value
correct vs incorrect	Feature	*0.11*	*0.06*	*1.81*	*0.07*	**0.23**	**0.09**	**2.65**	**0.01**	0.08	0.07	1.07	0.29	0.06	0.07	0.84	0.40
	Contextual	0.08	0.06	1.34	0.18	0.07	0.09	0.82	0.41	0.08	0.07	1.06	0.29	0.13	0.07	1.73	0.08
	Distant	0.08	0.07	1.22	0.22	-0.03	0.09	-0.40	0.69	0.03	0.07	0.47	0.64	0.08	0.07	1.07	0.29
	Association	**-0.18**	**0.06**	**-2.79**	**0.01**	-0.13	0.09	-1.45	0.15	-0.03	0.07	-0.38	0.71	0.01	0.07	0.08	0.94
	Distant*Lexical	*-0.12*	*0.06*	*-1.83*	*0.07*	-	-	-	-	-	-	-	-	-	-	-	-
correct vs semantic	Feature	0.00	0.11	0.01	0.99	0.05	0.15	0.31	0.76	-0.08	0.11	-0.69	0.49	*-0.28*	*0.15*	*-1.84*	*0.07*
	Contextual	**0.27**	**0.12**	**2.21**	**0.03**	0.17	0.15	1.09	0.28	0.19	0.12	1.58	0.11	0.19	0.16	1.18	0.24
	Distant	0.09	0.11	0.83	0.41	0.07	0.15	0.50	0.62	0.04	0.11	0.36	0.72	0.02	0.16	0.13	0.90
	Association	-0.02	0.11	-0.16	0.87	-0.17	0.14	-1.19	0.23	0.03	0.11	0.23	0.82	-0.22	0.15	-1.45	0.15
	Association*Lexical	-	-	-	-	-	-	-	-	-	-	-	-	**-0.33**	**0.16**	**-2.11**	**0.03**
correct vs omissions	Feature	0.09	0.10	0.87	0.38	**0.29**	**0.15**	**1.97**	**0.05**	**0.24**	**0.10**	**2.38**	**0.02**	0.05	0.18	0.30	0.77
	Contextual	0.09	0.10	0.96	0.34	0.06	0.16	0.39	0.70	0.00	0.10	0.05	0.96	0.12	0.20	0.63	0.53
	Distant	0.16	0.10	1.57	0.12	0.02	0.15	0.14	0.89	0.13	0.10	1.24	0.22	0.30	0.19	1.57	0.12
	Association	**-0.33**	**0.10**	**-3.28**	**<0.001**	-0.22	0.15	-1.53	0.13	**-0.27**	**0.10**	**-2.61**	**0.01**	0.04	0.18	0.22	0.82
	Distant*Lexical	**-0.27**	**0.10**	**-2.67**	**0.01**	-	-	-	-	**-0.24**	**0.10**	**-2.33**	**0.02**	-	-	-	-
correct vs phonological	Feature	**0.14**	**0.06**	**2.21**	**0.03**	**0.40**	**0.10**	**3.85**	**<0.001**	0.07	0.11	0.62	0.53	0.10	0.07	1.33	0.19
	Contextual	0.05	0.06	0.83	0.40	-0.02	0.09	-0.21	0.84	0.00	0.11	0.00	1.00	0.07	0.08	0.98	0.33
	Distant	0.01	0.06	0.23	0.82	*-0.14*	*0.08*	*-1.68*	*0.09*	0.03	0.11	0.23	0.82	0.09	0.07	1.23	0.22
	Association	**-0.15**	**0.06**	**-2.48**	**0.01**	-0.06	0.08	-0.65	0.51	0.00	0.11	0.01	0.99	0.06	0.08	0.82	0.41
	Feature*Lexical	-	-	-	-	**0.30**	**0.09**	**3.23**	**<0.001**	-	-	-	-	-	-	-	-
semantic vs omissions	Feature	0.16	0.11	1.47	0.14	0.28	0.20	1.42	0.16	**0.31**	**0.12**	**2.61**	**0.01**	**0.54**	**0.24**	**2.25**	**0.02**
	Contextual	**-0.24**	**0.12**	**-1.97**	**0.05**	-0.05	0.21	-0.24	0.81	-0.21	0.13	-1.57	0.12	-0.34	0.30	-1.11	0.27
	Distant	0.10	0.11	0.89	0.37	0.09	0.19	0.45	0.65	0.04	0.11	0.35	0.72	0.31	0.26	1.18	0.24
	Association	**-0.28**	**0.11**	**-2.49**	**0.01**	-0.17	0.18	-0.92	0.36	**-0.28**	**0.12**	**-2.38**	**0.02**	0.28	0.24	1.18	0.24
	Association*Lexical	**0.22**	**0.11**	**1.95**	**0.05**	-	-	-	-	-	-	-	-	**0.53**	**0.25**	**2.16**	**0.03**
	Distant*Lexical	-	-	-	-	-	-	-	-	*-0.21*	*0.11*	*-1.81*	*0.07*	-	-	-	-

To summarise the accuracy analyses, a high number of feature-based semantic neighbours seems to facilitate the accuracy of word retrieval in aphasia in general, and this is more specifically the case in people that have comparable levels of impairments in semantic and phonological processing. These findings are inconsistent with some previous studies of aphasic picture naming that used similar measures of semantic neighbourhood density, where either no effect on accuracy (Blanken et al., 2002; Bormann, 2011; Bormann et al., 2008) or a detrimental effect on accuracy (Mirman, 2011) were found. However, Mirman and Graziano (2013) reported both facilitatory and inhibitory trends from near feature neighbours on accuracy across different individuals with aphasia, although it is unclear whether these effects were significant. Finally, other types of neighbourhood density impacted accuracy in the mixed impairment group and in the phonological impairment group, and these effects were also in the direction of facilitation: the more strongly associated neighbours (fewer association neighbours; mixed impairment group), or, the more contextual neighbours (phonological impairment group) the more correct responses. Remarkably, only the group with more of a semantic impairment showed no effect of any semantic neighbourhood factor on naming accuracy.

Correct responses versus semantic errors: None of the semantic neighbourhood density measures significantly influenced the probability of semantic errors compared to correct responses for the whole group. However, for the reduced group, we found a significant effect of contextual neighbours: words with many contextual neighbours were more likely to be correctly named than to result in a semantic error. This effect for the reduced group was only present in the subgroup of people with a mixed impairment. This finding also differs from most studies that found that words with many neighbours increased semantic errors, however, the group with a phonological impairment did show the effect in the direction of previous studies: people in this group were more likely to make a semantic error compared to getting a correct response on targets with many feature-based neighbours, but this effect was only marginally significant. Finally, there was a marginally significant interaction between association neighbours and the lexical factor in the full group as well as in the "phonological impairment" subgroup. Interestingly again, the semantic impairment group did not show any effect of a semantic neighbourhood factor on the probability of making a semantic error compared to a correct response.

In sum, there were rather few effects of semantic neighbourhood factors on the probability of a semantic error compared to a correct response, and most of these effects were in the direction of facilitating correct spoken production of the target rather than making a semantic error (except from a marginally significant effect of feature-based neighbours in the other direction, in the phonological impairment group). This is at odds with most of the findings of the semantic neighbourhood density studies we reviewed (except for Kittredge et al., 2007a, who did not observe any effect of their semantic neighbourhood measures on the probability of each error type) where a high number of semantic neighbours was associated with more semantic errors compared to words with few semantic neighbours.

Correct responses versus omissions: Feature-based neighbours also seemed to influence the likelihood of omissions: the probability of producing a correct response compared to an omission was higher for words with many feature neighbours. This was true for the subgroups with less phonological impairment (mild and semantic subgroups), but not for those with phonological difficulties (mixed and phonological impairment subgroups). Association neighbours were also found to affect the probability of producing a correct response over an omission: words with few association neighbours/some strongly associated neighbours were more likely to result in a correct response than in an omission. Once again, this effect of association neighbours differed across subgroups, being present for those with semantic impairment (mixed and semantic) but not for those with less semantic impairment (mild and phonological subgroups). Finally, a significant interaction was found between the distant neighbourhood measure and the lexical predictor, again this was the case for those with semantic impairment but not for those with few semantic difficulties (mild and phonological impairment): while there was no effect of distant neighbours overall, words with many distant neighbours were more likely to be correctly named than result in an omission if the target was low in the lexical factor, but this effect decreased with increasing values of the lexical variable, and even turned into the opposite effect for words with highest lexical values.

To sum up, both feature-based (for the whole group and people with phonological difficulties) and association neighbours (for the whole group and people with semantic difficulties) increased the probability of a correct response over an omission while there was an effect of distant neighbours in a cross-over interaction with the lexical factor, for all groups except those with no semantic difficulties. The effect of feature-based neighbours on omissions is in line with Blanken et al. (2002), Bormann (2011) and Bormann et al. (2008) although the participants in their samples had more semantic than phonological difficulties, which appears to contrast with the results here.

Correct responses versus phonological errors: Feature-based neighbourhood density significantly predicted the production of phonological errors for every group, except for the subgroups with “unbalanced” impairments (the phonological and the semantic impairment groups):

words with many feature-based neighbours were more likely to result in a correct response than in a phonological error. For the reduced group and the subgroup with a mild impairment, there was also a marginally significant interaction between this neighbourhood measure and the lexical variable: the effect increased with an increased lexical factor score. Association neighbours also had an influence on the probability of a phonological error compared to a correct response: in the mixed impairment group, words with fewer association neighbours / some highly associated neighbours were more likely to result in a correct response than in a phonological error. There was also a marginally significant effect of distant neighbours in people with a mild impairment.

In this analysis, findings are mostly in the direction of many semantic neighbours favouring a correct response instead of a phonological error (except for distant neighbours although the effect is marginal and is only present in the mild impairment subgroup). The effect of semantic neighbours on phonological errors has not been reported before. Indeed, Mirman (2011) found no difference between high and low semantic neighbourhood conditions on the rate of phonological (formal) errors, and Kittredge et al. (2007a) included analyses that compared phonological errors to correct responses but did not report the results of these analyses.

Semantic errors versus omissions: participants in group as a whole and the reduced group were more likely to produce semantic errors than omissions on targets with many feature-based neighbours. This effect was also found in the semantic impairment and in the phonological impairment subgroups. A similar effect of the number of associates was observed in both overall groups, as well as in the two subgroups with semantic difficulties (semantic and mixed subgroups): words with fewer association neighbours (or some highly associated words in the lexicon) were more likely to result in a semantic error than in an omission. Moreover, in the whole group and in the phonological impairment subgroup, there was a significant interaction between the lexical predictor and the number of association neighbours: the effect of association neighbours on the probability of a semantic error over an omission decreased with increased values of the target's lexical factor in such a way that the effect was no longer present on targets with high lexical values. In addition, an effect of contextual neighbours in the other direction was observed for the whole group (although only marginally significant), for the reduced whole group and for people with a mixed impairment: words with many contextual neighbours were more likely to result in an omission than in a semantic error. Finally, a marginally significant cross-over interaction between the lexical factor and the number of distant neighbours was observed in people with more of a semantic impairment.

In a nutshell, a high number of semantic neighbours or presence of highly associated neighbours was shown to increase the probability of a semantic error compared to an omission, but the opposite pattern was found with contextual neighbours for the whole group and for the group with a mixed impairment. Here, we replicate the findings of Blanken et al. (2002), Bormann (2011) and Bormann et al. (2008), who found that words with high semantic neighbourhood density (a measure close to our feature-based measure) were associated with more semantic errors compared to omissions, but results for contextual neighbours are at odds with Kittredge et al. (2007b), who observed the opposite pattern for semantic errors versus omissions. However, both patterns of results suggest non-independent processes underlying semantic errors and omissions, as suggested by the Blanken and Bormann studies.

Additional individual logistic regressions on accuracy were performed at the individual level: 73 of the 193 participants showed a significant effect of at least one semantic neighbourhood density measure: 65 facilitatory and 33 inhibitory effects.

Experiment 3: Discussion

Unlike in Experiment 2 with unimpaired speakers, many different effects of several semantic measures were found in participants with aphasia. Every different group of aphasic speakers was affected in some way by some semantic neighbourhood measures. So how do the different measures of semantic neighbourhood density influence aphasic picture naming?

It seems that feature-based semantic neighbourhood density is the measure that best predicts aphasic picture naming behaviour overall. It is the only measure that significantly predicts accuracy for the whole group, and it has a significant influence on the probability of each response combination for at least some aphasic speaker subgroups. The facilitatory effect of this type of semantic neighbours on accuracy is inconsistent with previous studies where no effect (Blanken et al., 2002, Bormann, 2011, Bormann et al., 2008) or an inhibitory effect (Mirman, 2011) was found in aphasic picture naming. While Kittredge at al. (2007a) found a facilitatory effect of semantic neighbourhood density on accuracy, their measure was similar to our contextual neighbourhood measure, which was not significant in our analyses. Our study is therefore mostly inconsistent with previous research with respect to accuracy and semantic neighbourhood density. However, we included a larger participant sample than any of these previous studies, resulting in a higher number of data points. In addition, the study design allows fine grained analyses that take into account both item- and participant-specific variation (we acknowledge that Kittredge et al. (2007a) also used mixed models) and that, rather than comparing between different sets of items of low versus high semantic neighbourhood density, the critical semantic neighbourhood density measure was a continuous variable in our experiment, and its influence was assessed controlling for the influence of common predictors of picture naming for each item rather than for each subset. We believe then that our study allows for more control and has better power than the other studies in which different effects were found.

In addition, feature-based neighbourhood density was shown not only to increase the likelihood of a correct response over all error types, but also to predict better a semantic error than an omission. The influence of semantic neighbourhood density on the probability of semantic errors compared to omissions is consistent with all studies investigating this particular comparison (Blanken et al., 2002, Bormann, 2011, Bormann et al., 2008, Kittredge et al., 2007b), showing the robustness of this effect across semantic neighbourhood measures and experimental paradigms.

The second measure to show significant effects across a number of analyses and groups is the association-based measure. Although it only influenced accuracy in the subgroup with a mixed impairment, it was shown to predict a correct response over an omission or a phonological error for several groups of participants with aphasia, and, like the feature-based measure, it predicted a semantic error over an omission, for most groups of speakers. Importantly, both feature-based and association-based measures had effects on accuracy and errors in the same direction, even if the type of semantic representations that they are based on is different and even if the effect of both variables was not always present in the same groups of speakers. This study is the first, to our knowledge, to investigate effects of the number and/or the strength of associates on spoken picture naming, and these effects are similar to those of feature-based neighbours.

Finally, contextual neighbours only showed negligible effects, inconsistent with Kittredge et al. (2007a, 2007b), and distant semantic neighbours had no reliable main effect (unlike in Mirman, 2011), but were involved in significant interactions with the lexical variable in predicting some response patterns.

We turn to the potential mechanisms underlying these effects in the General Discussion.

General Discussion

This study focused on the effects of semantic neighbourhood in spoken word production. It used a systematic analysis of different measures of neighbourhood density and relatively large samples from two different speaker populations (unimpaired participants and speakers with aphasia) in an attempt to shed light on the inconsistencies in the literature, tease apart the nature of these effects and inform our understanding of the cognitive mechanisms underpinning them.

In Experiment 1, we identified six different types of semantic neighbourhoods that had been investigated in previous studies (i.e., number of rated competitors, feature-based near and distant semantic neighbourhoods, contextual neighbourhoods, number of association neighbours), and detailed the conceptual differences and similarities between them. An analysis of correlations between these measures as well as a Principal Component Analysis allowed us to distinguish four broad types of semantic neighbours: feature-based, context-based, distant and association-based neighbours. Each of these measures illustrates an attempt to define the set of words in the lexicon that are related in meaning to a given word, but these "neighbourhoods" differ quite importantly, not only in size, but also qualitatively. The feature-based measure reflects an organisation of the lexicon based on featural overlap between concepts, while the contextual measure gives a different view of the semantic network, organised according to how often words occur in similar contexts. Distant neighbours represent the set of words that only share limited semantic characteristics with the target while not being completely unrelated. Finally, the association-based neighbourhood reflects the strength of association between a given word and words it is associated to, rather than the number of words that are related to it.

Experiment 2 examined the picture naming latencies and response types of 50 unimpaired speaker and we found no significant effects of any semantic neighbourhood density measure. These results are consistent with those of Bormann (2011) in a similar experiment. It may be that the effects of semantic neighbours on processing might not be strong enough to influence naming behaviour in the unimpaired language system. However, it could also be the case that analyses at the single trial level may provide additional sensitivity and hence facilitate detection of these effects. This possibility was not available to us as the data we used only included responses averaged across subjects, and thus did not allow us to take into account potential individual variation.

Despite the absence of an effect in unimpaired speakers, the influence of some of the semantic neighbourhood density factors was apparent in the picture naming of people with aphasia: The reduced target activation in the impaired language system amplifies the effects of these factors and enables us to observe them. The most important and robust results were that words with many feature-based neighbours were more likely to be correctly named than to result in any type of error, and that words with many feature-based neighbours and with strongly associated neighbours were more likely to result in a semantic error than to be omitted. It is important to remember that although we described the semantic neighbourhood measure that had the greatest effect as feature-based, that this does not necessarily mean that representations at the conceptual level are decomposed into features. Even the measures based on direct featural counts (e.g., Mirman, 2011) could reflect semantic similarity as captured in non-decomposed theories of semantics. In addition, the fact that an effect of the number of association neighbours (or rather, the strength of association between a word and its associates) was also observed could suggest that it matters how strongly words are associated to each other in the lexicon. However, our analyses have shown that these strongly associated neighbours were also likely to share many features with the target (i.e., they were coordinates) as well as being associated, hence the influence of association neighbours may also be a result of feature overlap/semantic similarity: The difference between these two measures may not be features versus association but rather number of neighbours (feature-based neighbourhood density) versus similarity of the neighbours (association-based). Effects on accuracy for these two types of neighbourhood factor were in the direction of facilitation, words with many feature neighbours or with some strong associates were more likely to be correctly named than to result in different types of errors. How can these findings be explained within models of speech production?

Serial discrete lexical access theories (e.g., Levelt, Roelofs & Meyer, 1999; Roelofs, 1992) describe a speech production model in which, first, during conceptual preparation, a concept which consists of a non-decomposed representation of the meaning of the intended target is activated and then this activation spreads to the lemma level. Activated lemmas compete for selection with the highest activated lemma being selected (using the Luce Choice rule). There are no semantic feature representations is these models. It is nevertheless assumed here that activation from the target's concept node spreads to other concepts that are related in meaning (the type of relationship between the target concept and other semantically related concepts is not fully specified but includes hypernyms, coordinates, parts, functions) via links within the conceptual level, and these, in turn, send (weak) activation to their corresponding lemma nodes. It is difficult to accommodate a *facilitation* effect of the number of semantic neighbours in this framework, as it is unclear how co-activated semantic representations could further facilitate word retrieval. In fact, it would seem to predict the opposite because of the competitive selection mechanism: activation of semantic neighbours at the lemma level would create inhibition because the sum of the activation of these lemmas would compete with the activation of the target.

In contrast, interactive activation theories (e.g., Dell, 1986; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997) postulate a level of conceptual representations that consists of a set of semantic features relevant to the target. Activation spreads from this level to the lexical level, activating both the target lexical representation and other lexical representations whose semantic features overlap with the target. Activation is bidirectional, which entails that both the target word and its semantic neighbours will send activation back to the level of semantic features. Features of the target then receive supplementary activation that in turn increases the activation of the target and its neighbours. Importantly, there is no competition: the word with the highest level of activation will be selected irrespective of the level of activation of other neighbours. In this framework, it is thus possible that access to words with many semantic neighbours (based on features) is *facilitated.* Given the evidence that in unimpaired subjects naming at 'normal' rates there is no detectable influence of neighbours, the parameters of the model would need to be set such that the effect of feedback is limited. This would be in line with Goldrick and Rapp (2002) who propose restricted interaction between semantic features and the lexical level. When there is damage to the language system, whereby target lemma node activation is reduced, the activation from neighbours could provide additional support for retrieval, benefitting those items with relatively more neighbours.

Howard, Nickels, Coltheart, and Cole-Virtue (2006) simulate lexical access with a featural representation (as well as with a non-decomposed representation) but, in contrast to Dell and colleagues (1997), also include a competitive mechanism. In this model, therefore, there should be inhibition from neighbours as is the case in Levelt et al.'s (1999) model, described above. Of course, once again, simulation parameters would need to be set such that, in the unimpaired model, these effects would not be apparent on behaviour. However, when the language system is impaired so that the lemma has reduced activation, it is possible that the feedback to shared features from neighbours may provide sufficient additional activation to the target for it to be successfully retrieved. Once again, however, it seems that one might expect that these effects may not be strong, as reflected in the 'mild' subgroup being the group that showed the most benefit from neighbourhood on accuracy,

The effect of association neighbours on accuracy was not observed overall for the combined group of participants with aphasia, but only in the group with the most severe naming impairment (the 'mixed' subgroup; which is also the group with the most severe *semantic* impairment). This effect was in the direction of facilitation for words that have strongly associated neighbours. If there is noise or dysfunction at the conceptual level or between this level and the lemma level because of a severe impairment, maybe then only those semantic neighbours that are strongly associated can still have an influence, but the mechanism for this is unclear. As proposed earlier, facilitation effects of semantic neighbours can most easily be explained by models that postulate interactivity between levels of representation and no competition, but these models do not address associative relationships. One possibility is that associative relationships are encoded differently to featural relationships (e.g., Plaut, 1995), and some authors have suggested this could be by way of lateral activation at the lexical level (Weigl & Bierwisch, 1970, cited in Nickels, 1997) whereby an activated node sends activation to all those nodes with which it is associated. However, the extent to which this would influence processing would likely depend on whether these links were interactive (every pair of associates mutually activates the other) or unidirectional (an item could activate another but not vice versa). Here, we examined the number of words with which a target was associated, but not the number that were bi-directionally associated. An alternative account is found in the fact that when there were few, strongly associated neighbours, these neighbours were also coordinates which shared many features with the targets. Hence, as noted above, the apparent effect of associations, may in fact be an effect of *similarity* of neighbours – neighbours that share the most features with the target will have the strongest effects (e.g., Rabovsky, Schad, & Abdel Rahman, 2016).

Finally, facilitatory effects of both feature-based and association neighbours were only found in the group with the mixed impairment (the one with the most severe semantic impairment). This finding, along with the absence of an effect in unimpaired speakers, suggests that large effects of semantic neighbourhood factors are less likely to be observed in the absence of semantic impairments and/or severe difficulties in successfully activating the lexical form.

Another (rather unexpected) finding was that words with many feature-based neighbours were more likely to be correctly named than to result in a phonological error, a finding that has not been reported in previous studies. A potential explanation can be provided within the interactive activation framework (e.g., Dell et al., 1997): a larger number of semantic neighbours creates stronger activation of the target lexical representation, this in turn would result in stronger activation propagation from this level to the phonemes. As a result, there is a greater chance that the phonemes of the target word will be successfully activated rather than a phonologically related error occurring.

Finally, at the whole group level, both feature-based and association neighbourhood predicted the probability of a semantic error compared to an omission. This seems most likely to be because semantic neighbours receive enough activation to be selected in cases when the target is not available. These effects of feature-based and associate neighbours were found in the whole group and in the semantic impairment group (consistent with Blanken et al. (2002), Bormann (2011), Bormann et al. (2008) and Kittredge et al. (2007b), with their respective measures of semantic neighbourhood density), suggesting again that people with semantic difficulties are those most likely to be sensitive to both measures with respect to semantic errors versus omissions. This view is particularly consistent with Kittredge et al. (2007b), who found decreased probability of omissions and increased probability of semantic errors on high semantic neighbourhood density targets, but only in a group of participants with more of a semantic impairment.

Conclusion

While every influential theory of lexical access in spoken word production agrees that in the course of naming, semantically related words are activated at the lexical level, the nature of these semantic representations and their effect on processing were still unclear. We compared different semantic neighbourhood measures, and investigated the influence of four conceptually different measures on unimpaired and aphasic picture naming performance. No effect of semantic neighbourhood variables was observed in unimpaired participants, showing that these effects may be too small to be detected in unimpaired speech production. In contrast, two different types of measures were shown to affect aphasic performance: while contextual and distant neighbourhood density measures showed negligible effects, feature-based semantic neighbourhood density showed a surprising facilitation effect, but also predicted a higher rate of semantic errors versus omissions. This was also the case for association-based neighbours, but essentially in participants with a semantic impairment. We suggested that these effects can best be explained within interactive activation theories (e.g., Dell, Martin, Schwartz, & Gagnon, 1997), with the addition of within level links related to the strength of associations between words. Hence this study improves our understanding of the type of semantic representations that matter in the course of spoken word production (at least in aphasic individuals), and of the dynamics involved (namely, a strengthening of the activation of features relevant to the target via interactive mechanisms). It challenges the idea that there is substantial competition between semantically related words, and suggests that, in aphasia, stronger activation of the common features of the target and its neighbours via feedback mechanisms minimizes the influence of competition.

References

Alario, F. X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers*, *36*(1), 140-155.

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). The CELEX lexical data base on CD-ROM. Philadephia, PA: Linguistic Data Consortium.

Bard, E. G., & Shillcock, R. (1993). Competitor effects during lexical access: Chasing Zipf's tail. In G. T. M. Altmann, & R. Shillcock (Eds.), *Cognitive models of speech processing: The second Sperlonga meeting* (pp. 235–275), Hillsdale, NJ: Erlbaum.

Bates, D., Maechler, M., Bolker, B. and Walke,r S. (2014). _lme4: Linear mixed-effects models using Eigen and S4_. R package version 1.1-7, <URL: http://CRAN.R-project.org/package=lme4>.

Blanken, G., Dittmann, J., and Wallesch, C-W. (2002). Parallel or serial activation of word forms in speech production? Neurolinguistic evidence from an aphasic patient. *Neuroscience Letters, 325,* 72-74.

Bormann, T. (2011). The role of lexical-semantic neighbourhood in object naming: implications for models of lexical access. *Frontiers in Psychology, 2(127),* 1-11.

Bormann, T., Kulke, F., Wallesch, C.W., and Blanken, G. (2008). Omissions and semantic errors in aphasic naming: Is there a link? *Brain and Language, 104,* 24-32.

Bozeat, S., Lambon Ralph, M.A., Patterson, K., Garrard, P., & Hodges, J.R. (2000). Non-verbal semantic impairment in semantic dementia. *Neuropsychologia, 38*, 1207-1215.

Buchanan, L., Westbury, C., & Burgess (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review, 8(3),* 531-544.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*(1), 177-208.

Carroll, J.B., White, M.N. (1973). Age of acquisition norms for 220 pictureable nouns. *Journal of Verbal Learning* & *Verbal Behavior,12*, 563-576

Chen, Q., & Mirman, D. (2012). Competition and Cooperation Among Similar Representations: Towards a Unified Account of Facilitative and Inhibitory Effects of Lexical Neighbors. *Psychological Review, 119(2),* 417-430.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 33(4)*, 497-505, DOI: 10.1080/14640748108400805

Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods, 37(1),* 65-70.

Dell, G.S., & O'Seaghdha, P.G. (1991), Mediated and convergent lexical priming in language production: A comment on Levelt et al. (1991). *Psychological Review, 98,* 604-614.

Dell, G.S., Martin, N., & Schwartz, M.F. (2007). A case-series test of the interactive two-step model of lexical access: Predicting word repetition from picture naming. *Journal of Memory and Language, 56*(4), 490-520.

Dell, G.S., Schwartz, M.F., Martin, N., Saffran, E.M., & Gagnon, D.A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104(4),* 801-838.

Fenson, L., Dale, P.S., Reznick, J.S., Bates, E., Thal, D.J., Pethick, S.J., Tomasello, M., Mervis, C.B., & Stiles, J. (1994). Variability in Early Communicative Development. *Monographs of the Society for Research in Child Development, 59(5),* 1-185.

Gilhooly, K.J., Logie, R.H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation, 12,* 395-427.

Glaser,W.R., & Düngelhoff, F.-J. (1984). The time course of picture–word interference. *Journal of Experimental Psychology: Human Perception & Performance*, *10*, 640-654.

Goldrick, M., & Rapp, B. (2002). A restricted interaction account (RIA) of spoken word production: The best of both worlds. *Aphasiology*, *16*(1-2), 20-55.

Goodglass, H., & Kaplan, E. (1983). *The assessment of aphasia and related disorders*, (2nd ed.). Philadelphia: Lea & Febiger

Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., Tzeng, A., Tzeng,g O., Arévalo, A., Vargha, A., Butler, A.C., Buffington, R., & Bates, E. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language, 51,* 247–250

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, *100*(3), 464-482.

Howard, D. and Patterson, K. (1992). *The Pyramids and Palm Trees Test*. Thames Valley Test Company.

Hunt, K. P., & Hodge, M. H. (1971). Category-item frequency and category-name meaningfulness (m'): Taxonomic norms for 84 categories. *Psychonomic Monograph Supplements, 4(6),* 97-121.

IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

Kertesz, A. (1982). The Western Aphasia Battery. New York: Grune and Stratton.

Kittredge, A.K., Dell, G.S., and Schwartz, M.F. (2007a). Omissions in aphasic picture naming: Late age-of-acquisition is the culprit, not low semantic density. *Brain and Language (103),* 132-133.

Kittredge A. K., Dell G. S., Schwartz M. F. (2007b). "*Omissions in aphasic picture naming: AoA is the culprit, as well as low semantic density*," in Poster Presented at the 45th Meeting of the Academy of Aphasia, Washington, DC.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44(4),* 978-990.

Landauer, T.K., Foltz, P.W., and Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes, 25,* 259-284.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(01), 1-38.

Loess, H., Brown, A., & Campbell, J. (1969). Cultural norms for items in 30 taxonomic categories. *Psychonomic Monograph Supplements, 3(7),* 69-86.

Martin, N., Schwartz, M.F., & Kohen, F.P. (2006). Assessment of the ability to process semantic and phonological aspects of words in aphasia: A multi-measurement approach. *Aphasiology, 20(2/3/4),* 154-166.

McRae, K., Cree, G.S., Seidenberg, M.S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, 37*, 547-559.

McRae, K., de Sa, V., and Seidenberg, M.S. (1997). On the Nature and Scope of Featural Representations of Word Meaning. *Journal of Experimental Psychology: General, 126(2),* 99-130.

Mirman, D. (2011). Effects of Near and Distant Semantic Neighbors on Word Production. *Cognitive, Affective, and Behavioral Neuroscience, 11(1),* 32-43.

Mirman, D., & Graziano, K.M. (2013). The Neural Basis of Inhibitory Effects of Semantic and Phonological Neighbors in Spoken Word Production. *Journal of Cognitive Neuroscience, 25(9),* 1504-1516.

Mirman, D. & Magnuson, J. S. (2006). The impact of semantic neighborhood density on semantic access. *Proceedings of the Annual Meeting of the Cognitive Science Society,* 1823-1828.

Mirman, D., & Magnuson, J.S. (2008). Attractor Dynamics and Semantic Neighbourhood Density: Processing is Slowed by Near Neighbors and Speeded by Distant Neighbors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(1),* 65-79.

Mirman, D., Strauss, T.J., Brecher, A., Walker, G.M., Sobel, P., Dell, G.S., and Schwartz, M.F. (2010). A large, searchable, web-based database of aphasic performance on picture naming and other tests of cognitive function. *Cognitive Neuropsychology, 27(6),* 495-504. www.mappd.org

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. http://w3.usf.edu/FreeAssociation/

Nickels, L. A. 1997. *Spoken word production and its breakdown in aphasia*, Hove, UK: Psychology Press

Pisoni, D.B., Nusbaum, H.C., Luce, P.A., & Slowiaczek, L.M. (1985). Speech perception, word recognition and the structure of the lexicon. *Speech Communication, 4*, 75–95.

Plaut, D. C. (1995). Semantic and associative priming in a distributed attractor network. *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society, 17,* 37-42.

Rabovsky, M., Schad, D. J., & Abdel Rahman, R. (2016). Language production is facilitated by semantic richness but inhibited by semantic density: Evidence from picture naming. *Cognition*, *146*, 240-244.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Roach, A., Schwartz, M., Martin, N., Grewal, R.S., & Brecher, A. (1996). The Philadelphia Naming Test: Scoring and Rationale. *Clinical Aphasiology, 24,* 121-133.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1-3), 107-142.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104(3),* 192-233.

Rosinski,R.R.,Golinkoff,R.M., & Kukish,K. S. (1975).Automatic semantic processing in a picture–word interference task. *Child Development*, *46*, 247-253.

Shapiro, S.I., & Palermo, D.S. (1970). Conceptual organization and class-membership: Normative data for representatives of 100 categories. *Psychonomic Monograph Supplements, 3(11),* 107-127.

Snodgrass, J.G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174-215.

Székely, A., & Bates, E. (2002). Objective Visual Complexity as a Variable in Studies of Picture Naming. *CRL Newsletter, 12(2),* 1-33.

Székely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., Jacobsen, T., & Bates, E. (2003). Timed picture naming: Extended norms and validation against previous studies. *Behavior Research Methods, Instruments, & Computers, 35(4),* 621-633.

Uyeda, K.M., & Mandler, G. (1980). Prototypicality norms for 28 semantic categories. *Behavior, Research Methods & Instrumentation, 12(6),* 587-595.

Van Overschelde, J.P., Rawson, K.A., & Dunloski, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language, 50,* 289-335

Vigliocco, G., Vinson, D.P., Damian, M.F., Levelt, W. (2002). Semantic distance effects on object and action naming. *Cognition, 85,* B61-B69.

Vinson, D.P., & Vigliocco, G. (2002). A semantic analysis of grammatical class impairments: semantic representations of object nouns, action nouns and action verbs. *Journal of Neurolinguistics, 15,* 317-351.

Weigl, E., & Bierwisch, M. (1970). Neuropsychology and linguistics: Topics of common research. *Foundations of language*, 1-18.

Wilson, M.D. (1988) The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. *Behavioural Research Methods, Instruments and Computers, 20(1),* 6-11.

Yates, M., Locker, L., & Simpson, G. (2003). Semantic and phonological influences on the processing of words and pseudohomophones. *Memory & Cognition, 31(6),* 856-866.

Appendix A

Psycholinguistic properties of the 86 experimental items

AoA=age of acquisition, VC=visual complexity, Imag=Imageability, Freq=Log10 lemma frequency, Fam=familiarity, PND=phonological neighbourhood density, Near=number of near feature neighbours, Distant=number of distant feature neighbours, Assoc=number of association neighbours, Comp=number of rated competitors, Raw context=number of contextual neighbours, Cat context=number of contextual neighbours, category-trimmed.

item	length	AoA	VC	Imag	Freq	Fam	PND	Near	Distant	Assoc	Comp	Raw context	Cat context
apple	3	4.15	120	637	1.48	598	4	22	121	17	6.09	6	1
ball	3	2.9	226	622	2.05	575	28	1	139	17	5.17	71	7
balloon	5	4.37	87	583	0.78	520	4	0	268	21	2.57	20	0
banana	6	3.78	175	644	0.9	576	1	7	218	13	5.52	3	1
basket	6	5.67	300	560	1.38	485	4	1	111	19	3.35	10	0
bat	3	4.85	180	586	1.15	514	34	9	194	17	5.43	69	0
bed	3	2.89	192	635	2.43	636	33	0	63	11	4.43	46	0
Belt	4	4.62	211	494	1.43	550	18	0	115	11	3.87	3	1
bench	4	4.21	290	555	1.42	488	6	6	294	14	4.04	10	0
book	3	3.68	230	591	2.64	643	22	0	12	15	4.61	20	0
boot	3	3.89	138	604	1.59	566	26	2	183	13	5.57	3	0
bottle	4	3.56	100	619	2.06	591	4	2	116	20	3.3	5	1
bowl	3	4.26	157	579	1.52	557	27	2	161	22	4.48	71	1
bread	4	3.58	188	619	1.87	611	15	1	161	15	5.83	107	9
bridge	4	5.58	526	608	1.82	561	6	0	366	10	3.22	45	4
broom	4	5.5	181	608	0.9	547	10	2	179	11	3.65	9	0
butterfly	7	3.67	421	624	1	481	0	6	184	20	6.13	82	1
Cake	3	3.26	185	624	1.53	594	24	0	87	19	5.78	50	8
camel	4	5.11	184	561	1.4	421	5	0	209	11	5.57	30	0
candle	5	5.37	129	594	1.2	544	7	0	76	8	2.78	14	1
cannon	5	7.9	183	588	0.78	498	5	4	267	16	3.43	39	4
carrot	5	2.74	149	577	0.9	539	6	4	108	16	5.83	2	0

item	length	AoA	VC	Imag	Freq	Fam	PND	Near	Distant	Assoc	Comp	Raw context	Cat context
cat	3	3.68	171	617	1.83	582	32	12	83	3	5.83	38	1
chair	3	3.43	191	610	2.13	617	18	6	145	14	4.3	21	3
church	3	5.15	340	616	2.26	560	4	2	107	15	4.26	153	3
clock	4	4.42	283	614	1.59	608	17	0	105	7	3.39	14	1
closet	6	5	247	525	1.04	540	1	5	129	14	3.17	19	3
corn	3	4.61	235	601	1.38	548	29	0	102	14	5.57	70	9
cow	2	3.94	193	632	1.6	529	17	9	213	10	5.57	29	2
crown	4	7.8	318	602	1.38	531	10	0	112	8	2.96	41	0
desk	4	5.56	238	574	1.96	583	3	3	225	18	4.13	5	0
dog	3	2.8	160	636	2.06	598	16	5	92	5	6.17	82	0
door	3	3.05	266	599	2.59	630	21	4	163	23	3.7	48	0
drum	4	4.63	345	599	1.2	506	9	10	192	16	5.43	42	2
duck	3	3.5	265	632	1.15	529	25	11	137	16	5.65	16	1
elephant	7	4.8	233	616	1.38	459	3	11	163	19	5.3	22	2
fan	3	5.63	288	582	1.23	520	21	0	154	17	3	1	0
football	6	4.84	174	597	1.52	565	1	0	140	23	5.13	110	4
fork	3	3.63	118	598	1.18	584	20	1	161	8	4.04	34	1
frog	4	4.32	179	617	0.95	507	7	2	143	20	5.78	43	2
glove	4	4.3	146	596	1.28	575	5	3	33	19	3.48	58	1
goat	3	5.21	202	585	1.45	469	19	7	120	16	5.48	3	0
hammer	5	5.42	133	618	1.04	515	5	2	201	7	5.13	16	6
harp	3	7.55	223	621	0.48	430	18	13	153	17	5.7	41	0
horse	3	4.15	232	624	2.12	560	13	8	184	16	5.65	97	3
hose	3	5.33	229	572	0.6	449	33	0	183	13	2.39	69	0
house	3	3.16	179	606	2.78	600	21	4	291	13	4.78	8	2
Key	2	3.58	160	618	1.93	603	20	0	208	9	2.3	14	0

Item	length	AoA	VC	Imag	Freq	Fam	PND	Near	Distant	Assoc	Comp	Raw context	Cat context
knife	3	4.15	112	633	1.64	573	9	9	164	13	4.09	18	3
lamp	4	4	100	575	1.54	578	11	1	41	7	4.3	31	3
Lion	4	4.42	255	626	1.4	511	7	5	171	15	5.52	42	4
microscope	9	9.16	178	617	0.9	493	0	0	151	19	3	27	1
necklace	6	5	136	606	0.6	536	2	2	107	12	4.61	2	0
Owl	2	6.21	235	595	0.85	477	13	24	108	9	5.74	20	0
pear	3	4	84	590	0.78	567	22	22	141	9	5.96	2	0
Pen	3	5.11	98	576	1.42	554	22	0	283	4	4.52	24	0
pencil	5	4.06	136	607	1.28	598	3	0	169	9	4.65	34	2
piano	5	5.5	305	630	1.43	545	1	11	219	14	5.48	110	0
Pig	3	3.84	128	635	1.63	509	18	2	154	22	5.65	38	0
pillow	4	3.47	115	624	1.28	602	4	1	80	14	3.22	36	0
pyramid	7	7.61	298	613	0.85	386	1	0	112	10	2.39	38	0
Rake	3	5.32	148	550	0.3	476	26	4	214	9	3.83	2	0
Ring	3	4.53	113	601	1.69	589	16	2	201	13	4.48	12	0
rope	3	5.44	303	596	1.62	539	26	0	90	21	3	10	0
ruler	5	5.94	88	543	1.26	571	5	0	179	16	3.61	37	0
saddle	4	6.42	201	578	1	436	5	1	89	4	2.35	93	0
scissors	5	4.5	138	609	0.6	559	3	8	167	4	3.35	10	2
Seal	3	5.42	171	563	1.15	482	23	7	197	25	5.78	4	0
shoe	2	2.6	150	601	1.9	569	21	1	58	12	4.91	4	0
Skis	4	6.68	216	615	0.9	551	12	2	146	14	3	17	0
slippers	6	5.26	123	595	0.95	494	6	1	52	8	5.09	14	1
snail	4	5.79	154	577	0.6	489	8	3	168	8	5.22	8	0

Item	length	AoA	VC	Imag	Freq	Fam	PND	Near	Distant	Assoc	Comp	Raw context	Cat context
Sock	3	2.94	110	553	1.26	578	20	5	88	8	4.35	19	0
spider	6	3.43	290	597	0.85	526	3	1	261	12	5.52	70	2
spoon	4	2.5	142	584	1.18	612	11	2	195	12	4.04	17	1
squirrel	6	4.44	236	642	0.78	511	1	14	192	11	5.39	37	1
strawberries	9	4.21	199	631	0.78	539	0	19	174	21	5.7	23	0
table	4	4.39	173	582	2.37	599	10	6	202	10	4.7	24	2
Tent	4	5.16	168	593	1.64	521	20	0	21	15	3.35	11	0
toilet	5	3.54	182	603	1.45	567	2	0	69	15	3.17	4	0
tractor	7	5.5	206	585	1.04	518	1	0	268	10	3.43	2	0
train	4	4	368	593	1.91	548	16	1	174	19	5	19	0
typewriter	8	6.74	289	615	1.04	524	0	0	157	17	3.13	7	1
Vest	4	5.83	185	581	0.85	472	20	1	68	9	4.35	19	6
whistle	4	5.42	158	574	0.95	505	5	1	236	16	2.35	7	0

Appendix B

Bivariate correlations (Pearson's r) among the (control) psychoinguistic predictors:

	Visual complexity	Imageability	Log lemma frequency	Familiarity	Age of Acquisition	Phonological neighbourhood density
Imageability	.082					
Log lemma frequency	.119	.187				
Familiarity	**-.227***	**.230***	**.638****			
Age of Acquisition	**.215***	-.182	**-.466****	**-.621****		
Phonological neighbourhood density	-.107	-.066	**.305****	.202	**-.290****	
Length in phonemes	.156	.033	**-.417****	**-.297****	**.358****	**-.724****

**. Correlation is significant at the 0.05 level (2-tailed).*

***. Correlation is significant at the 0.01 level (2-tailed).*

(Significant correlations are in bold)

Appendix C

Psycholinguistic properties of the 86 experimental items: Factors scores.

F1=lexical factor, F2=post-lexical factor, F3=Feature-based neighbours, F4=contextual neighbours, F5=Visual complexity, F6=Distant neighbours, F7=Association neighbours.

item	F1	F2	F3	F4	F5	F6	F7
apple	0.80	0.47	2.79	-1.00	-0.63	-0.47	0.23
ball	0.93	-1.05	-0.04	2.47	0.18	-0.10	0.65
balloon	-0.11	0.89	-1.01	-0.45	-1.33	1.38	1.25
banana	0.80	1.79	1.33	-0.32	-0.64	0.76	-0.38
basket	-0.76	0.58	-1.13	-0.87	0.55	-0.41	1.44
bat	-0.96	-1.90	0.85	0.10	0.17	0.33	0.80
bed	2.00	-1.06	-0.18	-0.48	0.83	-1.57	-0.56
belt	-0.58	-1.24	-1.79	-0.27	-1.23	0.46	0.47
bench	-0.32	-0.27	-0.26	-0.88	0.36	2.50	0.38
book	1.81	-0.86	-0.74	-0.89	0.54	-1.76	0.58
boot	0.57	-1.03	0.33	-0.54	-0.72	0.46	0.09
bottle	1.64	0.90	-0.51	-0.74	-0.69	-0.68	1.00
bowl	-0.02	-1.34	-0.42	0.45	-0.17	-0.06	1.74
bread	0.98	0.05	0.04	3.83	-0.24	0.19	0.14
bridge	0.65	0.26	-0.97	0.53	3.33	3.25	-1.17
broom	-0.16	0.32	-0.27	-0.70	-0.13	0.16	-0.78
butterfly	-0.39	1.36	1.27	0.57	1.97	0.30	1.37
cake	0.67	-0.66	0.12	2.79	-0.85	-0.73	1.04
camel	-1.22	-0.40	-0.19	0.08	-0.34	1.00	0.09
candle	-0.08	0.75	-1.06	-0.30	-0.62	-1.24	-1.16
cannon	-1.45	0.63	-0.58	1.05	-0.12	1.31	0.07
carrot	0.08	0.46	0.52	-0.39	-1.70	-0.18	1.04
cat	0.53	-1.41	1.46	-0.17	0.15	-0.97	-1.93
chair	1.50	-0.41	0.05	0.08	-0.04	0.08	-0.04
church	0.52	-0.06	-0.37	1.68	2.95	-1.09	-0.16
clock	0.98	0.02	-0.71	-0.58	0.90	-0.64	-1.44
closet	-0.68	0.66	-1.24	0.31	-0.80	0.39	0.56
corn	-0.48	-1.33	-0.24	3.44	-0.29	-0.46	0.19
cow	0.34	-0.81	1.34	0.11	0.30	0.80	-0.95
crown	-0.72	0.03	-1.03	-0.62	2.04	-0.98	-1.44
desk	0.74	0.33	-0.66	-1.05	0.35	1.24	0.88
dog	1.43	-0.36	1.13	0.28	0.58	-1.12	-1.71
door	1.99	-0.74	-0.50	-0.90	1.25	0.14	1.75
drum	-0.53	-0.17	1.02	0.14	1.14	0.76	0.49
duck	-0.05	-0.94	1.67	-0.50	0.48	-0.24	0.39
elephant	-0.59	1.28	1.23	0.01	0.16	0.00	1.19
fan	-0.53	-0.98	-1.06	-1.06	0.76	0.05	0.64
football	0.21	1.20	-0.38	2.17	-0.06	-0.49	1.82
fork	0.50	-0.57	-0.34	0.12	-0.83	0.10	-1.15
frog	-0.38	0.31	0.63	0.75	-0.46	-0.30	1.24
glove	0.30	0.51	-0.49	0.07	-0.33	-1.89	0.89

item	F1	F2	F3	F4	F5	F6	F7
goat	-0.95	-1.16	0.63	-0.77	-0.10	-0.35	0.81
hammer	-0.40	0.90	0.20	1.86	-1.17	0.72	-1.40
harp	-2.46	-0.92	1.82	-0.35	0.82	-0.59	0.35
horse	0.71	-0.41	0.94	1.17	1.22	0.20	0.22
hose	-1.86	-1.89	-1.13	0.09	0.68	0.06	-0.13
house	2.05	-0.72	-0.04	-0.31	0.10	2.17	-0.08
key	1.60	-0.62	-0.97	-1.05	0.34	0.59	-1.36
knife	0.88	0.29	0.67	0.10	-0.52	0.00	-0.60
lamp	0.37	-0.01	-0.71	0.80	-1.38	-1.26	-0.98
lion	-0.06	0.30	0.77	1.10	0.43	0.25	0.15
microscope	-1.34	2.48	-0.91	-0.08	0.38	-1.00	0.71
necklace	-0.35	1.38	0.12	-0.45	-1.14	-0.78	-0.26
owl	-1.76	-1.18	2.61	-1.13	0.54	-0.51	-1.03
pear	-0.45	-0.96	2.47	-1.00	-1.60	0.10	-0.82
pen	0.04	-1.01	-0.54	-0.14	-0.81	1.84	-1.75
pencil	0.87	1.11	-0.25	0.57	-0.85	0.23	-0.96
piano	0.01	0.87	1.41	0.24	2.20	0.37	-0.38
pig	0.31	-0.49	0.76	-0.16	-0.15	-0.46	1.56
pillow	1.16	0.98	-0.39	-0.49	-0.50	-1.35	-0.30
pyramid	-2.07	1.33	-0.85	-0.42	1.71	-1.29	-0.88
rake	-1.81	-1.56	-0.40	-0.47	-1.25	1.10	-0.60
ring	0.86	-0.36	-0.18	-0.64	-0.62	0.60	-0.20
rope	-0.10	-1.14	-1.03	-1.13	1.24	-1.02	1.38
ruler	-0.23	0.42	-1.37	-0.14	-1.28	0.41	0.69
saddle	-1.73	-0.09	-1.11	0.32	1.09	-1.37	-2.00
scissors	-0.05	1.15	0.27	-0.08	-0.97	0.22	-2.19
seal	-1.23	-1.48	0.52	-0.62	-0.83	0.79	2.64
shoe	1.05	-1.10	-0.04	-0.74	-0.60	-1.09	-0.03
skis	-0.43	0.28	-0.53	-0.82	0.58	-0.56	-0.41
slippers	-0.81	0.86	0.00	0.20	-1.13	-1.49	-0.72
snail	-1.28	-0.09	0.19	-0.26	-0.94	0.32	-0.85
sock	0.24	-1.04	-0.29	-0.47	-1.53	-0.40	-0.60
spider	-0.06	0.96	0.26	1.15	0.32	1.65	-0.14
spoon	1.13	0.24	-0.38	-0.17	-1.27	0.94	-0.22
squirrel	-0.29	1.45	2.01	-0.17	0.43	0.26	-0.82
strawberries	-0.05	2.48	2.43	-0.79	-0.37	-0.02	1.38
table	1.24	-0.01	-0.15	-0.07	-0.13	1.00	-0.56
tent	-0.18	-0.50	-0.90	-0.80	-0.05	-2.04	0.44
toilet	0.90	1.11	-0.76	-0.92	-0.40	-1.16	0.28
tractor	-0.17	1.53	-0.82	-0.63	-0.28	1.58	-0.62
train	0.61	-0.48	-0.21	-0.79	1.44	0.55	1.33
typewriter	-0.36	2.15	-0.79	-0.52	0.79	-0.34	0.47
vest	-1.64	-0.71	-0.57	1.81	-0.94	-0.95	-0.64
whistle	-0.50	0.31	-1.16	-0.82	-0.52	1.11	0.30

CHAPTER 3

Paper 2

Investigation of the effects of semantic neighbours in aphasia: a facilitated naming study.

Introduction

What makes a given word easier or harder to retrieve from memory? Addressing this question is relevant, both when one wants to conceptualise word retrieval processes in unimpaired speakers and in individuals with aphasia, and when developing therapies that can benefit word retrieval in people with aphasia.

The production of a spoken word is often viewed as a series of steps where selection between relevant representations is required. In particular, it is assumed that before selection of the right word form, one has to ’choose‘ between semantically similar alternatives that are concurrently activated. However, the exact nature and impact of these semantic representations during the (unimpaired and aphasic) production of words remains unclear. For instance, the influence, for a given word, of the number of words similar in meaning (or ’semantic neighbourhood density‘), needs further investigation. Moreover, the potential interaction of this factor on the effectiveness of word retrieval treatment in aphasia has, to our knowledge, not yet been investigated.

In fact, there has been little attention so far as to the impact of specific properties of the words used in aphasia treatment on the effectiveness of that treatment: Are some words in the lexicon more likely to benefit from naming therapy than others? This knowledge would have implications both for treatment and for theories of word production. For example, Romanova (2015) examined the effects of a facilitation task (repetition in the presence of a picture) on the later naming performance of two individuals with aphasia, contrasting proper and common nouns. She found no difference in the sensitivity to facilitation between common versus proper nouns for either participant. Consequently, there was no evidence supporting distinct mechanisms for processing of proper and common nouns. In a different domain, Kohnert (2004) provided a treatment based on cognate and non-cognate words to a bilingual individual with aphasia (cognates are words of similar meaning and form in a given pair of languages, as in the English-French pair ’tomato-tomate‘). The participant showed a transfer of the benefits of therapy from the treated to the untreated language, but only for cognate words. Kohnert discussed this effect of cognates on aphasia treatment in relation to theories of bilingual speech production (e.g. Costa, Santesteban, & Caño, 2005). Finally, and of relevance to this study, there is some evidence that semantic properties of words can influence the success of treatment: Kiran and Thompson (2003) provided a semantic treatment to four individuals with aphasia targeting either typical or atypical members of semantic categories, and argued that, following treatment, only atypical items showed generalisation to other words within the category. Kiran and Thompson suggested that exposure to atypical items, that share some features of the prototype as well as disparate features, results in activation of both typical and atypical items, whereas exposure to typical items results in activation only of a limited set of items with shared features.

Here we aim to extend the literature by examining how semantic neighbourhood affects the outcomes of a naming facilitation task in aphasia in order to both inform the theoretical debate and our understanding of treatment effectiveness in aphasia. Facilitation, in this context refers to the effect of performing one task on the accuracy of another task is examined a short time later. It can be thought to be analogous to a long lag priming technique in healthy speakers and it has been suggested that the results of facilitation can predict the results of treatment (e.g. Hickin, Best, Herbert, Howard & Osborne, 2002).

There is some evidence that semantic neighbourhood density has an influence on spoken word production in people with aphasia. This has been observed in standard spoken picture naming tasks (when there is no particular manipulation of the presence of semantically related words), but the effects reported on response are inconsistent, with either facilitation, interference or no effect of semantic neighbourhood density on response (Blanken, Dittman, & Wallesch, 2002; Bormann, 2011; Bormann, Kulke, Wallesch, & Blanken, 2008; Hameau, Biedermann, & Nickels, present thesis, Chapter 2; Kittredge, Dell, & Schwartz, 2007a; 2007b, Mirman, 2011; Mirman & Graziano, 2013; See Hameau et al., present thesis, Chapter 2, for a review).

An effect of the number of semantic neighbours on spoken word production implies that the co-activated semantic neighbours all contribute to increased activation resulting in either facilitation or interference, and that this increased activation varies as a function of the number of semantic neighbours in the lexicon. What exactly a semantic neighbourhood comprises, is, however, still unclear.

Semantic neighbourhood density: what metric?

Although semantic relatedness between two words is relatively easy to establish, it is harder to determine what constitutes the set of semantically related items that are activated in the course of spoken word production. What is included in the semantic neighbourhood of the word “rabbit” for example? Possible words could include coordinates such as mouse and guinea pig, but also associates like carrot, burrow, and many more. In a previous study, we (Hameau et al., present thesis, Chapter 2) submitted different measures of semantic neighbourhood density (feature-based, co-occurrence-based, ratings-based and associate-based) to a principal component analysis, and examined the influence of the different components underlying these measures on the picture naming performance of controls and individuals with aphasia. The component that had the best predictive power for aphasic performance was that which weighted highly on both the number of feature-based neighbours (based on McRae, Cree, Seidenberg, & McNorgan's (2005) feature norms) and on rated competitors (a subjective measure of the number of coordinates of a given target). We suggested that this component is most likely to reflect the words that share many features with the target and/or that belong to the same semantic category. This measure predicted accuracy in individuals with aphasia. A second measure had predictive power on aphasic response type: Words that had few different associates. For these words, associates also tended to share many features with the target, hence this measure also reflected the influence of feature based neighbours, but expressing the similarity of these neighbours to target. These words with few, strongly associated and semantically similar, associates were more likely to be correctly named than result in an error in participants with the most severe impairment, and more likely to be correctly named than to result in an omission for the whole group. Hence, it seems that the set of words that share many features with the target, and/or those that are strongly associated to the target both have some influence on picture naming in aphasia. Thus, while the *number* of semantically related words seems to matter, there is some evidence that some semantic neighbours have a stronger influence than others, if they are strongly associated. Further evidence that the degree of semantic similarity of these neighbours to the target may be important is provided by Rabovsky, Schad, and Abdel Rahman (2016), and by Britt, Ferrara, and Mirman (2016).

Rabovsky and colleagues (2016) investigated the influence of 'intercorrelational feature density' (IFD) on the picture naming performance of healthy speakers. IFD values are available in McRae et al's feature norms (2005), and represent the extent to which each pair of features of the target is also present in other concepts. The higher the IFD of a given concept, the denser the region it inhabits in semantic space. Hence, although it does give an indication about the presence of semantically related words in the lexicon, this measure also indicates how similar the target is to its semantic neighbours. Moreover, Rabovsky and colleagues found that targets with high IFD were slower named than targets with low IFD, demonstrating that the degree of similarity between a target and its semantic neighbours influences lexical retrieval.

In another recent study, Britt et al. (2016) examined picture naming outcomes in controls, healthy older participants and individuals with aphasia, on three different types of items: words with high name agreement, and words in two low name agreement conditions: items that had an alternative acceptable name (such as sofa and couch), and items that had at least one 'near semantic neighbour' (such as jam and jelly). Items in the three conditions were matched for length, word frequency and objective visual complexity. The item set was constructed based on the results of a previous norming study, and the two low name agreement conditions were defined based on the responses made by participants. For instance, near neighbours were words that were sometimes produced in the place of the target word in picture naming and considered as appropriate names for the picture, while being judged as having a different meaning from the target (e.g., jam and jelly). In addition to the expected faster responses in the high name agreement condition, results showed that healthy controls as well as older individuals were significantly slower in the ’near semantic neighbour‘ condition compared to the ’acceptable alternative‘ condition , even within items of similarly low name agreement. However, individuals with aphasia did not show a significant difference in accuracy across the two low name agreement conditions. Once again, this study draws our attention to the fact that the presence of a very similar semantic neighbour may be critical in yielding interference effects in picture naming.

Finally, there is also evidence in the visual word recognition literature, that orthographic neighbours that are higher in *frequency* than the target have a stronger influence on visual lexical decision time (e.g., Ferraro & Hansen, 2002; Grainger, O’Regan, Jacobs, & Segui, 1989; Grainger & Segui, 1990). It could be the case, then, that, as for orthographic neighbours of higher frequency in visual word recognition, semantic neighbours that are higher in frequency than the target may affect accuracy and latency more strongly in spoken word production than those that are lower in frequency.

Based on this evidence, when thinking about potential influences in aphasia treatment, it seems important to consider, not only measures of the number of semantically related words, but also measures taking into account how some semantic neighbours might have a stronger influence than others. This may be because of their high semantic similarity with the target word, or because of their high frequency relative to the target.

Influence of semantic neighbours on the effects of facilitation: predictions

First, it is worth noting that predictions with respect to the *direction* of the influence of semantic neighbourhood variables in speech production depend on the processing assumptions of the theories under consideration. For example, if competition is present between similar representations at the lexical level (as in WEAVER ++: Levelt, Roelofs, & Meyer, 1999), semantic neighbourhood predictors would be expected to result in inhibition (but see Hameau et al., present thesis, Chapter 1, for an account of facilitatory effects of semantic neighbours in a model with competition but also interactivity). Alternatively, in a model where there is no competition but only different activation levels of related representations and, crucially, interactivity between the semantic and the lexical level (e.g., Dell, 1986; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997), production of the target would be predicted to be facilitated by its semantic neighbours.

Predictions of the influence of semantic neighbours on the effects of facilitation on naming have to draw on the mechanism by which facilitation is effective. In aphasia studies, facilitation differs from treatment in that it consists of using a specific task on one occasion only and evaluating the effect of that task on later word finding. As noted above, it has been considered as a test for the effectiveness of aphasia treatment, because the participant's response to facilitation has the potential to determine future response to therapy (e.g., Hickin et al, 2002).

Howard, Hickin, Redmond, Clark, and Best (2006) list several possible processes that could account for the benefits that are obtained from a facilitation task (such as a word to picture matching, or repetition in the presence of the picture) in aphasia due to priming at different levels of speech production: There could be priming at the semantic level, since the semantic characteristics of the target item are accessed. Alternatively, priming could happen at the lemma level (or access to the lemma), because the lemma has to be accessed for the correct word form to be retrieved. Otherwise, facilitation could occur because of a strengthened mapping from the lemma level to the phonological output lexicon. Based on a series of facilitation tasks in individuals with aphasia who had either impaired or spared semantic processing, Howard et al. (2006a) demonstrated that facilitation was not likely to be due to priming at the semantic level, but was due to either priming of lemma retrieval or because of a stronger mapping from the lemma level to the phonological word form level. They made this claim based on the different patterns of response to the facilitation task depending on the level of impairment of the aphasic participants. The basic assumption was that facilitation would be expected only at a point in the process of naming where the individuals 'were operating with inadequate information' (Howard et al., 2006a, p.948). Since they found that individuals with no semantic impairment benefitted to a greater extent from the facilitation task, they concluded that priming was not likely to occur at the level of semantics but rather at later levels.

Taking these possible mechanisms underlying facilitation into account, what role could the semantic neighbours of a word play in modulating the outcomes of facilitation? The effects of facilitation result from the priming of the target at the time of facilitation, which means that this target is easier to access later. If there is an effect of semantic neighbours influencing the outcomes of the facilitation task, this could be the result of either, a) semantic neighbours also being primed during the facilitation task because they are activated concurrently, or, b) priming is restricted to the target but consequently neighbours are more highly activated in subsequent naming. Alternatively, a) and b) could both contribute to a potential effect of semantic neighbours in facilitation.

Whether or not semantic neighbours are primed during the facilitation task depends on exactly how priming is thought to occur. For instance, no priming of semantic neighbours would be expected within the logogen theory (Morton, 1980). Morton (1980) explains priming by a temporary lowering of the threshold of a lexical entry (logogen) for retrieval, but importantly, this only occurs after lexical selection. In other words, as only the target is selected during facilitation, only the target and not the neighbour is primed. Hence under the logogen theory the only effect of semantic neighbours because of the facilitation task, would be the result of target priming. On the other hand, some theories would predict weaker effects of semantic neighbours upon priming: Oppenheim, Dell and Schwartz (2010) implement priming as an error-based implicit learning process that adapts the language production system to recent experience. Every time a word is retrieved, the connection weights from active semantic features to the target word are increased and those to all other active words (that share semantic features with the target) are decreased. In this theory, production of the target in the facilitation task would lead to decreasing the connection weights to semantic neighbours, hence less effect of semantic neighbours would be expected compared to before priming. A third possibility would be that semantic neighbours are also primed along with the target during the facilitation phase. Howard, Nickels, Coltheart, and Cole-Virtue (2006) suggest that priming consists of incrementing the connection weights between semantics and lemmas when both are active. Hence, when the target lemma is active, connections from the semantic features shared with semantic neighbours to the lemmas of those semantic neighbours will also be active. Consequently, these connection weights would also be strengthened and hence the lemmas of the neighbours would also benefit, to a lesser extent, from this priming at later naming.

Hence, it is possible to envisage a full spectrum of effects on the semantic neighbours as a result of the facilitation phase. In addition, at the point of naming, post-facilitation, the priming of the target is also likely amplify any effects of neighbourhood: Greater activation of the target will result in correspondingly greater activation of the neighbours.

The combination of the two sources of potential priming (of target and neighbours), and the variety of potential effects of (different measures of) neighbourhood on naming therefore lead to many potential patterns of effects, particularly in combination with different levels of spoken word retrieval impairment. This study aims to further constrain theories by investigating these patterns. It will answer the question: Does the semantic neighbourhood of a given word affect the efficacy of a facilitation task in improving naming? We compare the effects of facilitation in two participants with aphasia and two different levels of impairment, and analyse the effects of several different measures of semantic neighbourhood. This will enable specification of the role of semantic neighbours in naming facilitation, and contribute to a better understanding of both the dynamics of speech production and mechanisms of language recovery with facilitation in aphasia.

Methods

Participants

Two monolingual English-speaking men with chronic aphasia took part in this study, DEH and SJS. Both participants were right-handed, with corrected-to-normal vision and normal hearing. Both had been involved in several aphasia-related research projects through Macquarie University (e.g., Biedermann, Beyersmann, Mason, & Nickels, 2013; Fieder, Nickels, Biedermann, and Best, 2015; Mason et al., 2011). An aphasia friendly consent form (approved by the Macquarie University Ethics Review Committee) was signed by both participants prior to the study. Participants were selected because reports from previous studies showed they produced predominantly semantic errors and few phonological errors in naming consistent with impaired word retrieval.

DEH was 71 years old at the time of the study, 12 years after the onset of aphasia following an infarct in the territory of the left Middle Cerebral Artery that was secondary to infective endocarditis. DEH worked as a typesetter prior to his stroke and was retired at the time of the study, living at home with his wife. His language difficulties were predominantly in production, with non-fluent speech and anomia, while his comprehension was largely preserved. Several months before the study, he suffered a heart attack and he and his wife report a worsening of his aphasia since then.

SJS was 53 years old when the study began. He had suffered a left CVA (left Middle Cerebral Artery territory and frontal lobe) 16 years prior to the study. SJS was living by himself and working part time in an adapted work setting (factory). Before his stroke, he worked as project manager in a large telecommunication company. SJS presented with a non-fluent aphasia (communicating mostly with single words), deep dyslexia and comprehension difficulties. In the context of conversation these comprehension difficulties were attenuated.

In order to determine the type and extent of the language impairment for both participants, a number of background assessment tasks were carried out. The results of these tests are summarised in Table 1 and Table 2. While both participants show at least some evidence of comprehension difficulties, only SJS appears to have impaired conceptual-semantic processing, as shown by his score below the normal range in the Pyramids and Palm Trees test (Howard & Patterson, 1992).

Table 1. Background assessment for DEH and SJS: performance on published tests.

Task	No. of items	% Cut-off[1]	DEH % correct	SJS % correct
Conceptual semantic processing				
PPT 3 pictures[2]	52	94	94	85*
CAT semantic recognition[3]	10	80	90	100
CAT semantic memory[3]	10	80	90	100
Comprehension				
Spoken comprehension				
CAT spoken word comprehension[3]	15	83	93	77*
PALPA auditory synonym judgement[4]	60	-	93	68
High imageability	30	-	100	70
Low imageability	30	-	87	67
Written comprehension				
CAT written word comprehension[3]	15	90	100	100
PALPA written synonym judgement[4]	60	87	87	58*
High imageability	30	91	97	67*
Low imageability	30	82	77*	50*
Sentence comprehension				
CAT spoken sentence comprehension[3]	16	84	87	84*
CAT written sentence comprehension[3]	16	72	69*	84
CAT spoken paragraph comprehension[3]	4	50	100	87
Production				
Spoken picture naming				
CAT object naming[3]	24	90	62*	60*
CAT action naming[3]	5	80	40*	0*
Word fluency				
CAT word fluency[3]	na	13 words	8 words*	7 words*
Reading				
CAT word reading[3]	24	94	54*	12*
CAT complex word reading[3]	3	67	0*	0*
CAT function word reading[3]	3	50	33*	0*
CAT nonword reading[3]	5	60	0*	0*
Repetition				
CAT word repetition[3]	16	91	100	62*
CAT complex word repetition[3]	3	83	33*	0*
CAT sentence repetition[3]	12	83	67*	42*
CAT nonword repetition[3]	5	50	40*	40*
PALPA nonword repetition[4]	30	-	57	73
Writing				
CAT spoken picture description[3]	na	33	9*	1*

[1]*The cut-off is the score two standard deviations below the mean of the performance of healthy controls. The normal range when available is taken from the instruction manuals of the respective tests, or from the Nickels and Cole-Virtue (2004) norms. Scores with an asterisk are those below the normal range.*

[2] *PPT: Pyramids and Palms Trees test (Howard & Patterson, 1992). The percentage at cut-off represents the lowest boundary of normal range.*

[3] *CAT: Comprehensive Aphasia Test (Porter & Howard, 2004). Cut-off values (percentages) represent the lowest boundary of normal range.*

[4]*PALPA: Psycholinguistic Assessments of Language Processing in Aphasia (Kay, Lesser, Coltheart, 1996).*

Table 2. Background assessment for DEH and SJS: performance on the 24 item naming test[1]

Task	DEH % correct				SJS % correct			
	Total	1 syll	2 syll	3 syll	Total	1 syll	2 syll	3 syll
Spoken picture naming	50	62	62	25	58	50	37	37
Written picture naming	50	78	50	0	4	12	0	0
Reading	83	87	87	75	62	62	75	50
Repetition	79	87	87	62	96	100	100	87
Writing to dictation	67	100	62	25	8	12	12	0

[1]A set of 24 pictured items consisting of 8 words of each of 1, 2, and 3 syllables matched for name agreement, age of acquisition, word form frequency and visual complexity, to be assessed across five different modalities (spoken and written picture naming, repetition, reading aloud, and writing to dictation). All items differed from the experimental items for the current study, and assessment in each modality was made on a different day. Norming is not available, but since the same items are used in different tasks, the relative impairment can be compared across modalities.

DEH shows good auditory comprehension, which indicates intact lexical-semantic processing, but his understanding of written language is not perfect, especially when dealing with syntactically complex sentences or abstract words, suggesting impaired access from the written modality. On the other hand, SJS shows impaired comprehension in both the auditory and written modalities, with better comprehension in the context of paragraphs or sentences.

Both participants' speech is non-fluent (see Appendix A for samples of spoken picture description for each participant), and both show difficulties in spoken and written picture naming. DEH shows an effect of length in picture naming. His incorrect responses were mostly omissions, and semantic errors, in addition to some phonological errors, especially on consonant clusters. He made no semantic errors in written picture naming. His repetition and his reading aloud performance was impaired on some words with consonant clusters, long words, and on nonwords. The type of errors made, the length effect, the difficulties in repetition together with the absence of a conceptual or semantic processing deficit show that DEH, in addition to lexical syntactic difficulties, most likely has an impairment of the phonological output lexicon and/or the link between the lexical-syntactic (lemma) level and phonological output lexicon, as well as an impairment of the phonological output buffer.

SJS made predominantly semantic errors and descriptions in naming, but made some phonological errors on longer words. He showed no clear length effect in production tasks, which, together with the fact that his semantic processing seems impaired, leads us to think his impairment is predominantly at the conceptual/lexical-semantic level. However he also performed outside the normal range in several repetition tasks (phonological errors on long words and on nonwords), which speaks in favour of an additional mild impairment of the phonological output buffer. In addition, SJS is severely impaired in every task involving written language, in both reading and writing modalities. We will not detail further the respective impairments in written language of our participants, given the focus of this study is on spoken word production.

Materials

197 black and white drawings from the International Picture Naming Project (IPNP: Székely et al., 2004) were selected. These pictures were chosen because they depicted single words and appeared both in McRae et al.'s feature norm database (McRae et al., 2005), and in the Edinburgh Associative Thesaurus (Kiss, Armstrong, Milroy, and Piper, 1973). The McRae et al. feature norm database (McRae et al., 2005) is a corpus of 541 concepts for which features have been generated by participants asked to provide those that best described each concept. A matrix indexes the number of shared features between two concepts taking into account the number of participants who produced a given feature. The Edinburgh Associative Thesaurus is a set of word association norms showing the counts of word association (Kiss et al., 1973): each stimulus word was presented to 100 different subjects, who, for each word, had to write down the first word that came to mind, as quickly as possible.

From these 197 items, 196 were included in the final analysis, as the word "bag" was named "paper bag" and was therefore removed.

Item related properties:

In this section we describe the different variables that were used as predictors in response time and response type analyses. Several variables have been shown to influence latencies in unimpaired picture naming (e.g., Alario, Ferrand, Laganaro, New, Frauenfelder, & Segui, 2004; Baayen & Milin, 2015) and performance in aphasic naming (e.g., Nickels & Howard, 1994). In order to be able to identify an effect of our variables of interest (the semantic neighbourhood variables) over and above the effect of other common predictors of picture naming performance, we chose to include as many as possible of these variables, that we labelled "control predictors". We first describe these control predictors, followed by the semantic neighbourhood variables that were included in the analyses.

Control predictors

- **Trial number:** each item in each instance when it was presented, appeared in a given order. There is evidence of temporal dependencies between successive trials in many experiments for an individual, or 'autocorrelation' (e.g., Baayen & Milin, 2015): participants might get faster and faster because they are getting used to the experiment, or they might get slower because of fatigue. Including trial number as a predictor allows a better control of these interdependencies.
- Objective **visual complexity** (IPNP, Székely et al., 2004), range 4325-62243, average= 16061, SD=8038. This is the size of the digitized stimuli picture files in Kbytes. This measure seems to be preferable to subjective ratings of visual complexity that have been shown to be confounded with familiarity (Székely and Bates, 2000).
- **Word form frequency** (CELEX: Baayen, Piepenbrock & Gulikers, 1995), log transformed: range 0.95-4, average=2.37, SD=0.54. This measure is the sum of spoken and written word frequency; word form written frequencies represent the number of times the word occurs in a 16,600,000 word corpus, and word form spoken frequencies represent the number of times the word occurs in a 1,300,000 word corpus.
- **Lemma frequency** (CELEX, Baayen et al., 1995), log transformed: range 1-4.10, average=2.57, SD=0.53, the sum of spoken and written lemma frequencies: lemma frequency represents the occurrence of the headword for a given word in the same corpus as word form frequency. For example, the word 'apple' has a lemma frequency of 546, which is the sum of the word-form frequency of 'apple' (315), and 'apples' (231).
  There is still no agreement regarding where the frequency effect arises in the process of producing words. Some evidence points to the word form level (e.g., Jescheniak & Levelt, 1994), but also to the lemma level (Gahl, 2008). Having both options of word form and lemma frequency seems reasonable as they might both be good predictors of response time and accuracy (Brysbaert & New, 2009). Note however, that CELEX lemma frequency does not necessarily map directly onto the lemma level, it could equally well reflect the frequency of the word form in a theory where word forms are stored decomposed (e.g. Nickels, Biedermann, Fieder, & Schiller, 2015)
- **Name agreement**: range 0.38-1, average 0.9, SD=0.13. Name agreement refers to the degree to which participants agree on the name of the picture. Name agreement values were taken from our own Australian English name agreement data when available (n=89) and were otherwise drawn from the IPNP (Székely et al., 2004; IPNP is an American database).

- **Age of Acquisition** (Johnston, Dent, Humphreys, & Barry, 2010): range 1.37-4.97, average=2.85, SD0.73. Age of Acquisition ratings were drawn from a British English norming study (Johnston et al., 2010): participants were presented with pictures and were asked to estimate the age at which they thought they had first learned the name of the depicted object, choosing between seven age bands.
- **Familiarity** (Johnston et al., 2010): range: 2.61-6.84, average=5.09, SD=1.10. Familiarity values were drawn from the same source as Age of Acquisition ratings. Values were obtained by asking participants to rate how usual or unusual the depicted object was in their realm of experience, on a scale ranging from very unfamiliar to highly familiar.
- **Imageability** (MRC database: Coltheart, 1981): range 506-668, average=601.3, SD=29.4. Imageability represents the ease with which a word gives rise to a sensory mental image. Values were only available for 168 of the 196 final items.
- **Phonological neighbourhood density** (N-Watch: Davis, 2005): range 0-34, average=11, SD 9.18. Phonological neighbourhood density (PND) is the number of words in the lexicon that differ from the target word by only one phoneme, either substituted, added, or deleted.
- **Length in syllables**: range 1-3, average=1.53, SD=0.67.
- **Length in phonemes**: range 2-9, average=4.28, SD=1.39.

Semantic neighbourhood predictors

- *Feature-based*

Feature-based neighbourhood measures were obtained based on McRae and colleagues' (2005) database. When applicable, frequency of the semantic neighbours was drawn from CELEX (Baayen et al., 1995). We evaluated the effects of both measures of the number of semantically similar words (density), and also measures of how similar the most similar neighbours of a target were. In addition, we included the frequency of these very similar neighbours as another predictor, as high frequency words in the lexicon are likely to be more highly activated than low frequency words in the spoken word production process.

- **Semantic neighbourhood density** (SND) was defined as the number of feature-based close neighbours (words sharing 40% of more of their semantic features with the target, following Mirman's definition (2011): range 0-36, average=5.00, SD=6.10.
- **Proportion of shared features between the target and its closest neighbour** (CloseSim): this measure indicates how "similar" the closest neighbour in the lexicon is: range 0.00 -0.93, average=0.47, SD=0.27.
- **Mean similarity** (MeanSim): this measure is the average proportion of shared features of all the feature based neighbours: range 0.402-0.93, average=0.50, SD=0.07.

CloseSim and MeanSim are strongly positively correlated in our dataset (*r*(152)=.734, p<0.001), which is unsurprising as both measure the featural similarity of semantic neighbours. However, MeanSim shows no significant correlation with the measures of the number of feature-based neighbours (all rs< 0.1; *p*s>0.1).

- **Number of feature neighbours that are higher in frequency than the target word**, **based on word form frequency** (WFSNDh): range 0-12, average=1.38, SD=2.32.
- **Number of feature neighbours that are higher in frequency than the target word, based on lemma frequency** (LSNDh): range 0-14, average=1.43, SD=2.46.
- **Log word form frequency of the closest neighbour** (CloseWFF): range 0-3.69, average=1.95, SD=0.70.
- **Log lemma frequency of the closest neighbour** (CloseLF): range 0.30-3.80, average=2.13, SD=0.71.

- *Associate-based*

Associate-based semantic neighbourhood values were defined using the Edinburgh Association Thesaurus database (Kiss et al., 1973). Although association is argued to reflect a different type of semantic relationship to feature-based neighbours, and associates may be processed differently to words that share features with the target (e.g. Plaut, 1995), about 18% of the first associates (and 17% of all associates) for our experimental set were coordinates of the target word (such as fork-knife), and hence also shared many semantic features. When applicable, frequency of the associates was drawn from CELEX (Baayen et al., 1995). Similar to feature-based neighbours, we included both measures of the count of associates, but also the possible strength of association between the target and its closest associate, and the frequency of this closest associate.

- **Number of different associates** given by at least two participants (NumAss): range 4-22, average-12.88, SD=3.50. Since the free association task requires participants to give only one associate per target word, this measure gives an idea of the agreement between subjects as to which words are most associated with the target word (low NumAss means that most participants produced the same associate, whereas high NumAss indicates that there is less agreement amongst participants on which words are strongly associated to the target word).
- **Percentage of participants choosing the first associate** (StrengthAss): range 7-65, average=23.93, SD=13.41. This measure represents the strength of the association between the target word and its most frequently given associate.

  NumAss and StrengthAss work in opposite directions: in general words with a strong associate (high StengthAss) have fewer different associates (low NumAss), and words with many different associates (high NumAss) have a less strongly associated first associate (low

StrengthAss).These two measures are significantly negatively correlated in our dataset ($r(194)=-.689$, $p<.001$).

- **Log word form frequency of the first associate** (WFFAss): range 0.70-4.79, average=2.78, SD=0.70.
- **Log lemma frequency of the first associate** (LFAss): range 1-4.81, average=2.96, SD=0.67.

Procedure

The facilitation task used was repetition of the target word in the presence of its corresponding picture. While Howard et al. (2006a) used word-to-picture matching, similar effects have been found following application of a range of different tasks in treatment and/or facilitation of word retrieval in aphasia (e.g., Howard, 2000; Nickels, 2002). Specifically, tasks that require activation of both semantics (e.g., from presentation of the picture) and word form (e.g., from repeating the word) have all been argued to engage similar mechanisms and hence tasks that focus on ’semantic‘ processing (e.g. word-to-picture matching) have similar outcomes compared to treatments with a ‘phonological’ focus (e.g., phonemic cueing, repetition in the presence of the picture; Howard, 2000; Nickels, 2002).

The experimental stimuli were randomly allocated to two sets and for each set the facilitated naming treatment procedure was run over two experimental sessions:

Session 1

1a. Picture naming of the first 99 items (baseline, set 1),

1b. Facilitation of these 99 items in blocks of 20 followed by the final timed naming of each block of items (facilitated naming and post-test, set 1).

Session 2

2a. Picture naming of the remaining 98 items (baseline, set 2),

2b. Facilitation of these 98 items in blocks of around 20 followed by timed naming of those 20 items (facilitated naming and post-test, set 2).

The baseline and post-test picture naming sessions followed the same procedure: pictures were presented centred on a laptop screen, using the software DMDX (Forster & Forster, 2003), and each picture appeared concurrently with a beep. The participant was asked to name the picture with a single word. The picture remained on the screen until the participant gave a response or manifested his failure to respond (there was no time out), and the next picture was presented upon the examiner’s button press. The whole naming session was audio recorded using a digital voice recorder. Naming responses and response time (time between the onset of the beep and start of the participant’s response) were transcribed and calculated using Audacity (Mazzoni, Brubeck, and Haberman, 2005).

Facilitated naming sessions were each organised in five blocks of approximately 20 items, as follows: the first 20 pictures were displayed one after another on a laptop screen, along with the written name represented by the picture, while the examiner said the target word out loud. Presentation of the target was hence multimodal, providing many opportunities for facilitation (Best, Howard, Bruce, & Gatehouse, 1997). The participant was instructed to repeat/read the target word. Each word had to be produced (successfully) by the participant, and to make sure the word was correctly produced, the participant was asked to repeat it again immediately. Participants were aware that they would have to name these pictures again subsequently and were encouraged to try to remember the names (although awareness of the next task does not seem to affect effectiveness, see Howard et al., 2006).

After this facilitated naming task, and a short intervening chat with the examiner (about 3-5 minutes), the same 20 pictures were presented again for naming (with no cues. using DMDX, with a beep and audio-recorded). Pictures were presented in a different order in the facilitation task and in the final naming task, but pseudo-randomization in both facilitation and post-naming ensured that the two words that preceded or followed the target were not from the same semantic category. Following this first block, the remaining pictures were presented for facilitated naming and post-test using the same procedure, in four blocks of 19 or 20 items. Participants were free to take longer breaks between each block if they wished.

Response coding

The first complete response was coded, excluding false starts, and subsequent (correct or incorrect) responses. A response was considered as **correct** if it consisted of a correctly pronounced target word, with no phonological error. **Acceptable alternatives** were words that were alternative responses to the target which are also acceptable responses for the stimuli (e.g., ’trousers‘ for ’pants‘). **Semantic errors** were responses that consisted of a single noun, and that were clearly semantically related to the target. Semantic errors were further split in four subcategories: **superordinate** (e.g., ’toy‘ for ’ball‘), **subordinate** (e.g., ’grizzly bear‘ for ’bear‘), **coordinate** (another member of the same semantic category, like “skirt” for “dress”), and **associate** (a word with no taxonomic relationship with the target, but related in meaning through associations, like ’snow‘ for

’skis‘). **Formal errors** were responses that shared at least 50% of their phonemes with the target or vice versa (if the target shared 50% of its phonemes with the response). **Mixed errors** were responses that shared a semantic relationship with the target and met the criteria for a formal error. **Unrelated** word errors were real words bearing no semantic or less that 50% of phonemes with the target. **Nonwords** were responses that did not correspond to a real word but that shared at least 50% of their phonemes in common with the target, or vice versa (e.g., ’beaner‘ for ’beaver‘). **Descriptions** were an attempt to describe the target (its appearance, use, etc...), descriptions could also be a single word, provided it was not a noun (e.g., ’sweep‘ for ’broom‘). **Omissions** were instances were no response was made, or when there was a comment expressing failure to respond, such as “I don’t know”. Errors that did not fall into any of these categories (unrelated nonwords with no similarity to the target, fragments) were labelled as **other errors.**

Analyses

We performed a number of different analyses to examine the effects of semantic neighbourhood on the participants’ performance. Separate analyses were performed for each participant.

One set of analyses targeted naming behaviour (response time and accuracy) for each participant at baseline, in order to determine the effect of semantic neighbourhood variables on picture naming before the facilitation task, and another set of analyses was performed to analyse patterns of change between the two time points. Response time data for correct responses at baseline was analysed by means of linear regressions, and accuracy with logistic regressions, taking into account control factors and examining the effects of each critical semantic neighbourhood predictor when added to the model independently (i.e. never more than one semantic neighbourhood predictor in the model at any one time). Then, for each individual participant, linear mixed effect models were applied to response time for items that were correct both at baseline and at post-test, and generalised linear mixed effects models for binomial outcomes were applied to each participant’s complete set of responses, to analyse the influence of each SND factor on changes in response between baseline and post-test. In these linear mixed effects models, target item was entered with a random intercept, and control predictors were the same as in baseline analyses, with the addition of the factor time (baseline vs post-test). Response type and latency were analysed for each participant using R (R Development Core Team, 2011).

Results

Overall, performance was better at post-test than at baseline for both participants, showing the beneficial effect of the facilitation task. Consistent with their respective pattern of impairment that was described following background testing, DEH produced more omission errors than other error types, whereas SJS produced more semantic errors than any other type of error. A summary of all error types for both participants at baseline and at post test is provided in Table 3.

Table 3. Distribution of response types at baseline (BL) and post-test (PT) for both participants, with per item mean and standard deviation of response time (RT) (number of targets: 196).

	DEH		SJS	
	BL	PT	BL	PT
correct	91	125	100	140
Mean RT	2.54	1.64	2.09	2.62
(SD)	*(1.96)*	*(0.77)*	*(1.99)*	*(3.18)*
acceptable alternatives	6	2	1	3
semantic	22	7	54	26
superordinate	*1*	*0*	*1*	*1*
subordinate	*1*	*1*	*1*	*1*
coordinate	*13*	*4*	*34*	*19*
associate	*7*	*2*	*18*	*5*
mixed	1	2	1	3
formal	8	7	1	5
unrelated	0	0	2	0
nonword	7	14	7	4
description	3	0	14	4
omissions	43	26	13	5
other errors	15	13	3	6

We first analyse the effects on performance at baseline and then the influence of semantic neighbourhood variables on the effects of facilitation

Baseline analyses

DEH: Response time at baseline

Latencies for correct responses (including acceptable alternatives) were transformed to approximate normal distribution: the Boxcox test indicated a reciprocal transformation (power -1.3) to be the most appropriate for DEH's baseline response time. Linear regressions were conducted to assess the potential influence of each SND predictor on baseline naming latency for correct responses while controlling for other relevant psycholinguistic predictors. A base model was defined that best-fitted the data (i.e., had the lowest Akaike Information Criterion (AIC)) by entering every control predictor (item-related predictors and trial number) in a stepwise procedure (both backwards and forward), using the function stepAIC from the package MASS in R (Venables & Ripley, 2002). This procedure led to only trial number and lemma frequency being retained as control predictors in the base model: responses were slower the later in the session the item occurred, and faster for targets of higher frequency (lemma frequency was only marginally significant but improved the model's fit). This base model was significant ($F(2,94)=4.69$, $p=0.011$). When each semantic neighbourhood predictor of interest was added, one at a time, to this base model, three (SND, MeanSim, WFSNDh) were significant or marginally significant. A summary of the coefficients for the base model and for each semantic neighbourhood variable, as well as the Akaike Information Criterion of each model is provided in Table 4.

Table 4. Summary of the coefficients for each control predictor in the base multiple regression model and each model including a critical predictor for DEH's response time on correct responses at baseline):

Variable	*β*	*SE β*	*t*-value	*p*-value	AIC
Base model					-37.688
(Intercept)	0.345	0.103	3.347	0.001	
Trial	-0.001	<0.001	-1.997	0.049	
Log lemma frequency	0.065	0.034	1.888	0.062	
Models including semantic neighbourhood variables					
SND	*-0.005*	*0.003*	*-1.724*	*0.088*	-38.738
CloseSim	0.004	0.073	0.051	0.960	-35.690
MeanSim	**0.808**	**0.367**	**2.204**	**0.031**	-19.752
WFSNDh	**-0.019**	**0.009**	**-2.121**	**0.037**	-40.271
LSNDh	-0.013	0.009	-1.427	0.157	-37.789
CloseWFF	-0.033	0.042	-0.790	0.433	-15.441
CloseLF	-0.041	0.040	-1.032	0.306	-15.903
NumAss	0.000	0.005	0.012	0.990	-35.688
StrengthAss	0.000	0.001	0.185	0.854	-35.723
WFFAss	0.017	0.028	0.626	0.533	-36.096
LFAss	0.012	0.029	0.422	0.674	-35.873

Significant effects of the critical predictors are in bold (p<.05), marginally significant in italics (p<.1). Response times are inverted: the higher the coefficient value, the faster the response.

Semantic Neighbourhood Density (the number of items sharing 40% of their semantic features or more with the target) had a significant effect on response time, with words with many semantic neighbours being slower named that words with few semantic neighbours. Similarly, the number of feature-based semantic neighbours that were higher in (word-form) frequency than the target (WFSNDh) predicted longer latencies. On the other hand, a high mean similarity of the feature-based neighbours of the target (MeanSim) predicted faster responses. Amongst the three models, the best fit to the data (with the lowest AIC) was the one including the number of neighbours of higher word form frequency than the target (WFSNDh).

DEH: Accuracy at baseline

Table 5. Summary of the coefficients for each control predictor in the base model and each critical predictor in logistic regressions on accuracy at baseline for DEH:

Variables	*β*	*SE β*	*z*-value	*p*-value	AIC
Base model					266.590
(Intercept)	-2.610	1.292	-2.020	0.043	
Name agreement	3.028	1.269	2.385	0.017	
Length (syllables)	1.106	0.448	2.469	0.014	
Length (phonemes)	-0.43	0.21	-2.045	0.041	
Models including the semantic neighbourhood variables					
SND	0.039	0.026	1.471	0.141	266.332
CloseSim	-0.666	0.567	-1.176	0.240	267.195
MeanSim	-2.268	2.425	-0.935	0.350	213.638
WFSNDh	0.039	0.069	0.563	0.574	268.270
LSNDh	0.019	0.065	0.289	0.773	268.510
CloseWFF	-0.021	0.242	-0.088	0.930	214.540
CloseLF	-0.038	0.238	-0.161	0.872	214.520
NumAss	0.042	0.043	0.973	0.330	267.630
StrengthAss	-0.001	0.011	-0.108	0.914	268.578
WFFAss	-0.125	0.215	-0.580	0.562	268.250
LFAss	-0.137	0.223	-0.617	0.537	268.210

Logistic regressions were performed with accuracy (correct responses including acceptable alternatives versus incorrect responses) as a dependent variable. Similar to RT analyses, the optimal base model was defined using a stepwise procedure, which resulted in name agreement, length in syllables, and length in phonemes being included as significant predictors of accuracy (a test of the full model against a constant only model was statistically significant, p=0.004). Performance was hence more accurate on words that had high name agreement, and that had fewer phonemes and (unexpectedly) more syllables[2]. Each critical predictor was then added to that base model, but none of the measures of semantic neighbourhood had a significant effect on DEH's accuracy at baseline (all *ps*>0.05). A summary of the coefficients for each predictor in the base model and each semantic neighbourhood variable is provided in Table 5.

SJS: Response time at baseline

For SJS, latencies for correct (and acceptable alternatives) responses at baseline were inverse transformed to approximate normal distribution, as indicated by the Boxcox test. Stepwise regressions indicated that the only significant predictor in the base model should be imageability : words that were more imageable were faster named. This base model was significant ($F(1,84)=7.96$, $p=0.006$). However, when added to this base model, none of the semantic neighbourhood variables significantly predicted response latency (all $p>0.05$). Coefficients for each control variable in the base model and each semantic neighbourhood variable are displayed in Table 6.

Table 6. Summary of the coefficients for each control predictor in the base multiple regression model and each model including a critical predictor for SJS's response time on correct responses at baseline:

Variable	*β*	*SE β*	*t*-value	*p*-value	AIC
Base model					81.898
(Intercept)	-1.664	0.886	-1.879	0.064	
Imageability	0.004	0.001	2.821	0.006	
Models including the semantic neighbourhood variables					
SND	0.003	0.010	0.290	0.773	83.811
CloseSim	-0.066	0.141	-0.470	0.639	83.669
MeanSim	0.012	0.913	0.013	0.990	58.097
WFSNDh	-0.016	0.032	-0.498	0.620	83.642
LSNDh	-0.011	0.031	-0.357	0.722	83.766
CloseWFF	-0.031	0.066	-0.468	0.642	57.867
CloseLF	-0.032	0.066	-0.480	0.633	57.855
NumAss	0.004	0.012	0.302	0.763	83.803
StrengthAss	-0.004	0.004	-1.033	0.305	82.800
WFFAss	-0.031	0.054	-0.574	0.568	83.557
LFAss	-0.032	0.057	-0.567	0.572	83.565

Response times are inverted: the higher the coefficient value, the faster the response

[2] This could be related to his apparent difficulty with words with more clusters as multisyllabic words often have a simpler CV structure. However this effect of length in syllables and in phonemes is not reliable, as neither of these variables is significant when entered as a single predictor in the model.

SJS: Accuracy at baseline

The base logistic regression model for accuracy for SJS included log lemma frequency, phonological neighbourhood density, name agreement and visual complexity as control predictors. Words that were more frequent, that had more phonological neighbours and better name agreement were more accurately named. Whole model significance was calculated comparing the full model against a constant only model: $p<0.001$). Visual Complexity was not significant but improved the model. Of all the semantic neighbourhood variables, only the strength of association (StrengthAss: percentage of people choosing the closest associate) even marginally significantly predicted accuracy when added to the base model: words with a more strongly associated word in the lexicon were more likely to be incorrectly named by SJS at baseline. A summary of the coefficients for each predictor in the base model and each semantic neighbourhood variable is provided in Table 7.

Table 7. Summary of the coefficients for each control predictor in the base model and each critical predictor in logistic regressions on accuracy at baseline for SJS:

Variable	*β*	*SE β*	*z*-value	*p*-value	AIC
Base model					256.19
(Intercept)	-5.296	1.450	-3.653	<0.001	
Log lemma frequency	0.694	0.347	2.001	0.045	
Phonological neighbourhood density	0.038	0.019	1.999	0.046	
Name agreement	2.979	1.313	2.269	0.023	
Models including the semantic neighbourhood variables					
SND	-0.020	0.026	-0.745	0.456	257.63
CloseSim	-0.105	0.593	-0.178	0.859	258.16
MeanSim	-2.718	2.525	-1.076	0.282	206.12
WFSNDh	-0.036	0.081	-0.438	0.661	258.00
LSNDh	-0.073	0.079	-0.919	0.358	257.32
CloseWFF	-0.043	0.266	-0.161	0.872	207.85
CloseLF	-0.151	0.262	-0.575	0.565	207.55
NumAss	0.043	0.046	0.929	0.353	257.33
StrengthAss	*-0.020*	*0.012*	*-1.730*	*0.084*	255.13
WFFAss	-0.268	0.220	-1.219	0.223	257.13
LFAss	-0.264	0.228	-1.159	0.246	257.28

Marginally significant effects of the critical predictors are in italics

Summary: Baseline analyses

DEH and SJS show different patterns of sensitivity to semantic neighbourhood measures on naming, although for both participants there is some evidence of inhibition from semantic neighbours. DEH's latencies were slowed by the presence of many feature-based neighbours and to a greater extent those neighbours that were higher in frequency than the target, while words with semantic neighbours of high mean similarity were more likely to be faster named than words whose semantic neighbours had a low mean similarity than the target. On the other hand, DEH's accuracy at baseline was not affected by any semantic neighbourhood variable. In contrast, SJS's accuracy was (marginally) hindered by the presence of a strongly associated word in the lexicon, while his response latency was unaffected by any semantic neighbourhood predictor.

Analyses of the effects of facilitation

Several measures of semantic neighbourhood density have been shown to affect both participants' performance in some way at baseline (and mostly in an inhibitory fashion). The following analyses examined whether semantic neighbourhood measures influenced the changes in picture naming of the participants following the facilitation task.

DEH: effect of semantic neighbourhood variables on the effects of facilitation: response time analyses

DEH correctly named 71 words at both baseline and post-test. The Boxcox test indicated that an inverse transformation was the most appropriate (power -1.5) for the data to approximate normal distribution. After data transformation, one naming time remained more than 2.5 standard deviations away from the mean and was hence removed, along with the other instance when this word was named (two data points were therefore removed, resulting in 140 observations to be analysed).

Table 8. Summary of the base linear mixed effects model on latencies at pre-and post-test for DEH, and interactions in each model involving a critical predictor: Variance and Standard Deviation (SD) for the random effect, coefficient estimates (*β*), standard errors (*SE*), *t*- and *p*-values for the fixed effects in the base model, and for the interactions in their respective models, and AIC for each model.

Random effect		Variance		*SD*	AIC
Target item (intercept)		<0.001		<0.001	-36.172
Residual		0.038		0.195	
Fixed effects	*β*	*SE β*	*t*-value	*p*-value	
Control predictors					
intercept	*0.230*	*0.079*	2.896	0.004	
Time (Post-test)	*0.207*	*0.033*	6.277	<0.001	
Lemma frequency	0.072	0.028	2.521	0.013	
Interaction between time and …					
SND	0.000	0.004	0.024	0.981	-18.620
CloseSim	-0.078	0.116	-0.676	0.500	-26.860
meanSim	*-1.785*	*0.557*	*-3.204*	*0.002*	-19.391
WFSNDh	0.014	0.014	1.028	0.306	-24.166
LSNDh	0.009	0.014	0.649	0.517	-23.632
CloseWFF	-0.007	0.056	-0.128	0.898	-0.809
CloseLF	0.007	0.054	0.132	0.895	-0.636
NumAss	0.001	0.009	0.081	0.936	-15.506
StrengthAss	-0.001	0.002	-0.570	0.570	-10.740
WFFAss	-0.074	0.045	-1.645	0.102	-25.996
LFAss	-0.068	0.045	-1.499	0.136	-25.912

Significant interaction is in italics

The base model constructed in a stepwise fashion (see Table 8) included target item as a random factor, and the significant predictors time and lemma frequency (latencies were shorter at post-test and, same as at baseline, were shorter for more frequent words). The factor ”time” was highly significant: items that were correct both at baseline and at post-test were significantly faster named after facilitation, showing the overall effectiveness of facilitation for DEH on response time.

Next, the critical predictors were entered one at a time to the base model, together with the interaction of that variable with time. A summary of these results is presented in Table 8. There was a significant interaction between time and MeanSim, indicating that items with more similar neighbours benefited less from facilitation (see Figure 1[3]). No other semantic neighbourhood variable showed an interaction with time.

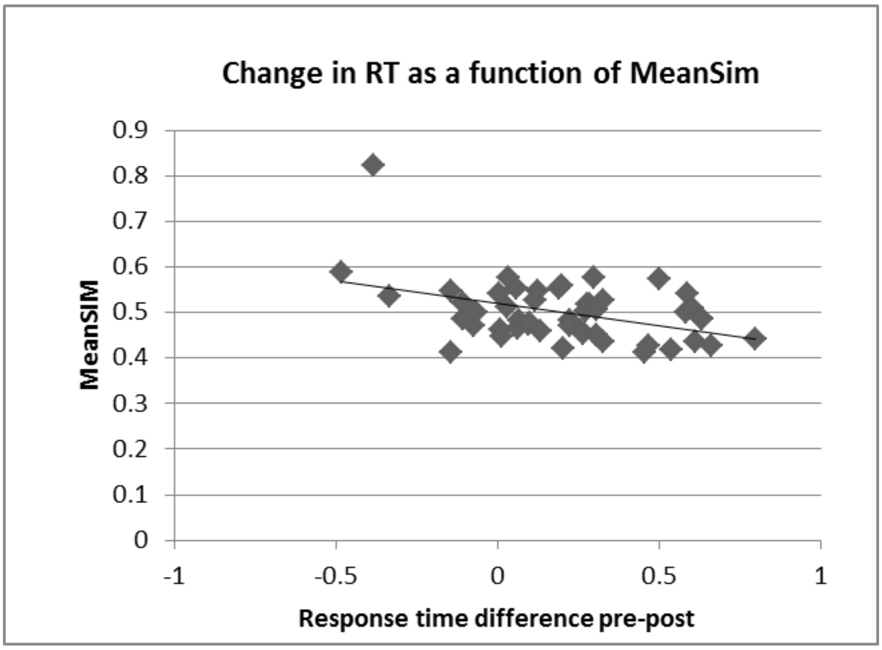


Figure 1. DEH: Effect of mean similarity on change of RT from pre to post facilitation

DEH: effect of semantic neighbourhood variables on the effects of facilitation: accuracy analyses

The base model for DEH's accuracy scores at both baseline and post-test included time and trial as control predictors (see Table 9).

[3] The same pattern was observed when the one outlying item with extremely similar neighbours was removed.

Table 9. Summary of the base generalized linear mixed effects model for binomial outcomes on accuracy at pre-and post-test for DEH, and interactions in each model involving a critical predictor: Variance and Standard Deviation (SD) for the random effect, coefficient estimates (*β*), standard errors (*SE*), *t*- and *p*-values for the fixed effects in the base model, and for the interactions in their respective models, and AIC for each model.

Random effect		Variance		*SD*	AIC
Target item (intercept)		0.537		0.732	525.178
Fixed effects	*β*	*SE β*	*t*-value	*p*-value	
Control predictors					
intercept	0.466	0.269	1.734	0.083	
Time (Post-test)	0.721	0.229	3.154	0.002	
Trial	-0.005	0.002	-2.275	0.023	
Interaction between time and …					
SND	-0.048	0.037	-1.297	0.195	526.900
CloseSim	0.526	0.829	0.634	0.526	526.500
meanSim	*6.266*	*3.695*	*1.696*	*0.090*	419.500
WFSNDh	-0.157	0.096	-1.638	0.102	525.100
LSNDh	-0.148	0.091	-1.631	0.103	524.600
CloseWFF	0.382	0.354	1.080	0.280	421.500
CloseLF	0.340	0.349	0.975	0.330	421.700
NumAss	0.040	0.064	0.630	0.529	528.100
StrengthAss	-0.026	0.017	-1.546	0.122	526.400
WFFAss	0.273	0.318	0.857	0.391	528.200
LFAss	0.293	0.331	0.885	0.376	528.000

(Marginally) significant interaction is in italics.

Time had a significant effect on accuracy: DEH's accuracy was significantly better at post-test than at baseline, showing again the overall beneficial effect of facilitation. In addition, words were more likely to be more accurately named if they appeared earlier in the experiment.

Next, each critical predictor was added to the base model, in an interaction with time. Only one interaction with time reached (marginal) significance in these analyses, and this was the same variable as for latency: the average similarity of the target's neighbours (MeanSim). Although this is the same measure that interacted with time for DEH in response time analyses, the effect was in the opposite direction. Words with more similar semantic neighbours showed greater benefit from facilitation (mean similarity of items incorrect at baseline, and a) correct at post-test = 0.527 (SD= 0.102), b) incorrect at -test = 0.500 (SD=0.053)). None of the other semantic neighbourhood measures interacted significantly with time.

SJS: effect of semantic neighbourhood variables on the effects of facilitation: response time analyses

Table 10. Summary of the base linear mixed effects model on latencies at pre-and post-test for SJS, and interactions in each model involving a critical predictor: Variance and Standard Deviation (SD) for the random effect, coefficient estimates (*β*), standard errors (*SE*), *t*- and *p*-values for the fixed effects in the base model, and for the interactions in their respective models, and AIC for each model.

Random effect		Variance		*SD*	AIC
Target item (intercept)		0.065		0.256	199.036
Residual		0.110		0.331	
Fixed effects	*β*	*SE β*	*t*-value	*p*-value	
Control predictors					
intercept	*0.332*	*0.198*	1.677	0.097	
Time (Post-test)	*0.085*	*0.050*	1.705	0.092	
Lemma frequency	0.180	0.071	2.545	0.013	
Interaction between time and …					
SND	0.001	0.009	0.115	0.909	218.558
CloseSim	-0.137	0.176	-0.779	0.438	205.700
meanSim	-0.618	0.920	-0.671	0.505	131.398
WFSNDh	-0.003	0.028	-0.116	0.908	213.697
LSNDh	0.004	0.025	0.171	0.865	214.477
CloseWFF	0.074	0.071	1.033	0.305	137.748
CloseLF	0.064	0.071	0.904	0.369	138.235
NumAss	-0.001	0.014	-0.093	0.927	216.824
StrengthAss	0.000	0.004	0.075	0.940	222.026
WFFAss	0.059	0.064	0.930	0.355	210.059
LFAss	0.065	0.067	0.963	0.339	209.715

Analyses were performed as for DEH. 88 words were correctly named both at baseline and at post-test for SJS. Following response time transformation (BoxCox test indicated latencies should be taken to the -1.15 power), no data point was more than 2.5 standard deviation away from the mean and hence all 176 observations were used in the analyses. The base model included target item as a random factor, and time and lemma frequency as control predictors (see Table 10). Time was only marginally significant (words tended to be faster named at post-test than at baseline), and more frequent words were faster named. None of the semantic neighbourhood measures significantly interacted with time for response latency (all *p*s>.1) (see Table 10).

SJS: effect of semantic neighbourhood variables on the effects of facilitation: accuracy analyses

Table 11. Summary of the base generalized linear mixed effects model for binomial outcomes on accuracy at pre-and post-test for SJS, and interactions in each model involving a critical predictor: Variance and Standard Deviation (SD) for the random effect, coefficient estimates (*β*), standard errors (*SE*), *t*- and *p*-values for the fixed effects in the base model, and for the interactions in their respective models, and AIC for each model.

Random effect		Variance		*SD*	AIC
Target item (intercept)		1.673		1.293	476.748
Fixed effects	*β*	*SE β*	*t*-value	*p*-value	
Control predictors					
intercept	-2.708	0.897	-3.018	0.002	
Time (Post-test)	1.326	0.291	4.555	<0.001	
Lemma frequency	0.935	0.366	2.556	0.011	
Phonological neighbourhood density	0.036	0.020	1.787	0.074	
Interaction between time and …					
SND	0.015	0.041	0.362	0.718	479.300
CloseSim	-0.885	1.021	-0.867	0.386	476.300
meanSim	-0.671	3.892	-0.172	0.863	390.300
WFSNDh	0.080	0.111	0.720	0.472	479.900
LSNDh	0.099	0.107	0.924	0.356	479.000
CloseWFF	-0.235	0.401	-0.586	0.558	393.600
CloseLF	-0.196	0.395	-0.496	0.620	392.800
NumAss	-0.066	0.075	-0.886	0.375	479.900
StrengthAss	*0.040*	*0.021*	*1.938*	*0.053*	476.700
WFFAss	0.349	0.378	0.924	0.356	479.300
LFAss	0.342	0.391	0.874	0.382	479.200

(Marginally) significant interaction is in italics

The base generalised linear mixed effect model for binomial outcomes that was defined using the same procedure as for DEH, included time, lemma frequency and phonological neighbourhood density, and is summarised in Table 11.

There was a significant effect of time, showing the beneficial effect of facilitation on accuracy for SJS (as opposed to its marginal effect on latencies), and, as at baseline, words that were more frequent and had more phonological neighbours were more accurately named overall.

A summary of the coefficients for interactions involving each semantic neighbourhood variable and time is presented in Table 11. Only the strength of the association between the target and its first associate was involved in a marginally significant interaction with time. At baseline, this variable had a marginally significant detrimental effect on accuracy. However, words with a strongly associated word in the lexicon were more likely to benefit from facilitation (mean association strength of items incorrect at baseline, and a) correct at post-test = 26.509 (SD= 15.137), b) incorrect at post-test = 23.902 (SD=13.376)).

Discussion

In this study, we used a facilitated naming paradigm with two individuals with aphasia with different levels of impairment, in order to investigate the potential influence of several semantic neighbourhood variables on priming. Our main aim was to understand better the role of semantic neighbours in the course of spoken word production.

Different measures relating to the semantic neighbourhood of words were used: measures based on the featural overlap between the target and semantically related words (and within this category of semantic neighbours, measures pertaining to the number of neighbours, to their similarity with the target, or to their frequency), and measures based on associative relationships between the target and words in the lexicon (number of associates, strength of the association and frequency of the closest associate)

Both participants' picture naming behaviour was affected by some measures of semantic neighbourhood. The effect on picture naming was mostly, and for both participants, in the direction of inhibition (in line with Mirman, 2011, and partly in line with Mirman and Graziano, 2013, but inconsistent with Kittredge et al., 2007a, and Hameau et al., present thesis, Chapter 2).

DEH had no semantic impairment, but rather impaired lexical retrieval (lemma and phonological form) and phonological buffer impairments. His naming behaviour was only affected by feature-based measures of semantic neighbourhood overall (and not association-based). Within these measures, a higher number of feature-based semantic neighbours (all semantic neighbours (SND) and particularly those that were higher in frequency than the target, WFSNDh) slowed response time.

Apart from this inhibitory effect of the number of semantic neighbours, there was, surprisingly, at baseline, a facilitatory effect of mean similarity of neighbours on DEH's response time. The more similar the neighbours were, the faster naming occurred. We remind the reader here that MeanSim was not correlated with any of the semantic neighbourhood density measures. How can these opposite effects be explained? An inhibitory effect of neighbours is generally hypothesised to be caused by the activation of the neighbours at the lexical level slowing selection because of increased lateral inhibition or the effects on total level of activation in the Luce choice rule (as in Dell et al., 1997, although the effect is described on errors in that model, not on response time) as more neighbours means more activation. Intuitively, then, it seems as though this inhibition should be stronger when neighbours are more similar, as the activation levels for these neighbours would be more similar to the target's due to their greater proportion of shared features. However, one could speculate that, in a featural account, the more semantic neighbours, the more shared features but also the more non-target, irrelevant features activated (via feedback from the lemma node of the neighbours). In turn these non-target, irrelevant features will cause activation of an even wider range of (non-neighbour) lemmas. Hence it is possible that the balance of shared to irrelevant features is critical, with neighbours with large numbers of shared features having a net facilitatory effect (the additional activation overcoming the increased competition) but when the number of irrelevant features increases, the balance is tipped to result in a net inhibitory effect. Moreover, it is also possible that the extent and location of the impairment in aphasia could change this balance compared to unimpaired participants. However, computational simulation is required in order to fully test these hypotheses.

When we look at the effects of facilitation for DEH, while facilitation improved performance for both latency and accuracy, there was no interaction between the semantic neighbourhood density measures and time. This shows that the outcomes of the facilitation task were not affected by the number of feature-based neighbours for DEH. In other words, that it is unlikely that these semantic neighbours were primed in a way that it created a different pattern of activation of the target and its semantic neighbours compared to baseline, nor that the priming of the target increased activation of the neighbours sufficiently to change the overall effects of neighbours. In contrast, there was an interaction between mean similarity of neighbours and time: the response latencies of words with very similar neighbours benefitted less from facilitation than those with less similar neighbours. A trend to the reverse effect for accuracy with targets with very similar neighbours being more likely to be correct. Hence, unlike the neighbourhood density measures, the featural similarity between the target and its neighbours, appeared to create a different balance in the activation of the target and its neighbours with priming, resulting in these very similar neighbours slowing down the correct production of the target. It might be that, with priming, only the most similar representations also benefit from the priming mechanism. As in Howard et al. (2006b), priming could consist in a strengthening of the mapping from semantics to lexical units. In that case this would also benefit to a great extent those words that share many features with the target and consequently it might be more difficult to select the relevant word form if a semantically related word is highly activated or competes with the target.

For SJS, there were no effects of any semantic neighbourhood variable on response latencies nor any interactions between these variables and time. Even for accuracy, SJS only showed a marginally significant effect of one semantic neighbourhood variable, which was different to that influencing DEH's performance: there was no influence of the number of neighbours, but his accuracy appeared to be hindered at baseline by the strength of the first associate. Moreover, there was an interaction between this variable and time, in such a way that after facilitation, words with a strong first associate were more likely to have benefited and be accurately named. The fact that for SJS, who has a semantic impairment, the only variable that had a significant effect at baseline and then interacted with facilitation was the strength of the association between the first associate and the target word, is in line with Hameau et al. (present thesis, chapter 2) who found an effect of a similar variable (based on associates) specifically in people with a semantic impairment. In that study, we argued that the measure also reflected the presence of a great proportion of feature-based neighbours amongst associates of the target. Here the associate-based measures are based on different norms, and the proportion of feature-based neighbours is smaller (17% of all association neighbours in the present study are also coordinates or superordinates of the target, compared to 26% in the other study). The associate-based measure used in this study most probably *does* reflect how strongly some words are associated rather than being a proxy for semantic similarity.

How could the strength of association between words be implemented in speech production theories? The most influential models (e.g. Levelt et al., 1999; Dell et al., 1997) do not address associative relationships. Some authors have suggested that associative relationships are represented at the lexical level, by lateral connections (e.g., Weigl & Bierwisch, 1970, cited in Nickels, 1997). Activation of a node sends activation to all those nodes with which it is associated. One would have to postulate interactivity between these associated nodes to explain how the strength of associates would affect processing. For SJS words with a strong associate were somewhat more difficult to retrieve at baseline, which could be explained by the presence of competition at the lexical level that was difficult to resolve, but with the priming task these words tended to be more accurately named compared to words with a less strong first associate: it could be the case that, when the target lexical form is accessed in the facilitation task, it sends lateral activation to its associates. The stronger these associates are linked to the target, the more activation they will send back to the target. Hence if priming led to a better activation of the relevant lexical form, those words with strong associates would benefit from further activation from these associates, but unlike at baseline where their influence was inhibitory, if (as in Howard et al., 2006b) priming led to a stronger mapping between the semantic features and the lexical form, the overall effect of associates might not be strong enough compared to the boosted target to interfere with its correct production. What is unclear is whether these strong associates are primed at the moment of the task, or whether they receive more activation after the task because of a better activation of the target lexical representation.

Conclusion

A facilitated naming paradigm was used with two individuals with aphasia in order to investigate how the number, similarity, frequency or strength of association of semantic neighbours in the lexicon interact with priming. Semantic neighbourhood is a complex notion and the influence of having many similar words in the lexicon on the ease of naming is still not well understood. The results of this study once again show this complexity. Both individuals benefitted from the facilitation task, and both showed effects of some semantic neighbourhood variables. However, DEH clearly showed an influence of feature-based neighbours only, while SJS showed some effect of a measure reflecting how strongly words were associated with the target. Moreover, although the overall influence of most semantic neighbourhood measures was in the direction of inhibition for both participants, the interactions with time showed that the effect of the similarity of semantic neighbours (for DEH) and of the strength of association of some neighbours (for SJS) changed direction with the facilitation task. Results are best accommodated by theories that assume interactivity between the semantic and the lexical level, but computational modelling is required in order to fully understand the complex interactions of factors in lexical access, especially with language impairment. With respect to aphasia naming therapy, these results evoke the possibility that individuals with a semantic impairment can improve more on those words that have strong associates, but for individuals without a semantic impairment, the presence of very similar semantic neighbours might interfere with the ease of lexical access after facilitation.

References

Alario, F. X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers*, *36*(1), 140-155.

Alario, F.X., Segui, J., & Ferrand, L. (2000). Semantic and associative priming in picture naming. *The Quarterly Journal of Experimental Psychology: Section A*, *53*(3), 741-764.

Baayen, R. H., & Milin, P. (2015). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12-28.

Baayen, R., Piepenbrock, R., & Gulikers, L. (1995). CELEX2 LDC96L14. *Web Download. Philadelphia: Linguistic Data Consortium*.

Best, W., Howard, D., Bruce, C., & Gatehouse, C. (1997). A treatment for anomia combining semantics, phonology and orthography. *Language disorders in children and adults: Psycholinguistic approaches to therapy*, 102-129.

Biedermann, B., Beyersmann, E., Mason, C., & Nickels, L. (2013). Does plural dominance play a role in spoken picture naming? A comparison of unimpaired and impaired speakers. *Journal of Neurolinguistics*, *26*(6), 712-736.

Blanken, G., Dittmann, J., and Wallesch, C-W. (2002). Parallel or serial activation of word forms in speech production? Neurolinguistic evidence from an aphasic patient. *Neuroscience Letters, 325,* 72-74.

Bormann, T. (2011). The role of lexical-semantic neighbourhood in object naming: implications for models of lexical access. *Frontiers in Psychology, 2(127),* 1-11.

Bormann, T., Kulke, F., Wallesch, C.W., and Blanken, G. (2008). Omissions and semantic errors in aphasic naming: Is there a link? *Brain and Language, 104,* 24-32.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977-990.

Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, *16*(1), 22-29.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 33(4)*, 497-505, DOI: 10.1080/14640748108400805

Costa, A., Santesteban, M., & Caño, A. (2005). On the facilitatory effects of cognate words in bilingual speech production. *Brain and Language*, *94*(1), 94-103.

Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, *37*(1), 65-70.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*(4), 801.

Ferraro, F. R., & Hansen, C. L. (2002). Orthographic neighborhood size, number of word meanings, and number of higher frequency neighbors. *Brain and Language*, *82*(2), 200-205.

Fieder, N., Nickels, L., Biedermann, B., & Best, W. (2015). How 'some garlic'becomes 'a garlic'or 'some onion': Mass and count processing in aphasia. *Neuropsychologia*, *75*, 626-645.

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

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*(3), 474-496.

Grainger, J., O'regan, J. K., Jacobs, A. M., & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Perception & Psychophysics*, *45*(3), 189-195.

Grainger, J., & Segui, J. (1990). Neighborhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies. *Perception & Psychophysics*, *47*(2), 191-198.

Hickin, J., Best, W., Herbert, R., Howard, D., & Osborne, F. (2002). Phonological therapy for word-finding difficulties: A re-evaluation. *Aphasiology*, *16*(10-11), 981-999.

Hillis, A. E., & Caramazza, A. (1994). Theories of lexical processing and rehabilitation of lexical deficits. *Cognitive Neuropsychology and Cognitive Rehabilitation*, 449-484.

Howard D. (2000) Cognitive neuropsychology and aphasia therapy: the case of word retrieval. In: *Acquired neurogenic communication disorders: A clinical perspective* (Ed. I. Papathanasiou). Whurr, London.

Howard, D., Hickin, J., Redmond, T., Clark, P., & Best, W. (2006a). Re-visiting “semantic facilitation” of word retrieval for people with aphasia: Facilitation yes but semantic no. *Cortex*, *42*(6), 946-962.

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006b). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, *100*(3), 464-482.

Howard, D., & Patterson, K. E. (1992). *The Pyramids and Palm Trees Test: A test of semantic access from words and pictures*. Thames Valley Test Company.

Howard, D., Patterson, K., Franklin, S., Orchard-lisle, V., & Morton, J. (1985). The facilitation of picture naming in aphasia. *Cognitive Neuropsychology*, *2*(1), 49-80.

Jescheniak, J. D., & Levelt, W. J. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 824.

Johnston, R. A., Dent, K., Humphreys, G. W., & Barry, C. (2010). British English norms and naming times for a set of 539 pictures: The role of age of acquisition. *Behavior Research Methods, 42*, 461-469.

Kay, J., Lesser, R., & Coltheart, M. (1996). Psycholinguistic assessments of language processing in aphasia (PALPA): An introduction. *Aphasiology*, *10*(2), 159-180.

Kiran, S., & Thompson, C. K. (2003). The Role of Semantic Complexity in Treatment of Naming Deficits. Training Semantic Categories in Fluent Aphasia by Controlling Exemplar Typicality. *Journal of Speech, Language, and Hearing Research, 46*(3), 608-622.

Kiss, G.R., Armstrong, C., Milroy, R., and Piper, J. (1973). An associative thesaurus of English and its computer analysis. In Aitken, A.J., Bailey, R.W. and Hamilton-Smith, N. (Eds.), *The Computer and Literary Studies*. Edinburgh: University Press. http://www.eat.rl.ac.uk/

Kittredge, A.K., Dell, G.S., and Schwartz, M.F. (2007a). Omissions in aphasic picture naming: Late age-of-acquisition is the culprit, not low semantic density. *Brain and Language (103),* 132-133.

Kittredge A. K., Dell G. S., Schwartz M. F. (2007b). "*Omissions in aphasic picture naming: AoA is the culprit, as well as low semantic density*," in Poster Presented at the 45th Meeting of the Academy of Aphasia, Washington, DC.

Kohnert, K. (2004). Cognitive and cognate-based treatments for bilingual aphasia: A case study. *Brain and Language*, *91*(3), 294-302.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(01), 1-38.

Mason, C., Nickels, L., McDonald, B., Moses, M., Makin, K., & Taylor , C. (2011). Treatment of word retrieval impairments in aphasia: Evaluation of a self-administered home programme using personally chosen words. *Aphasiology*, 25(2), 245-268. doi:10.1080/02687038.2010.489258

Mazzoni, D., Brubeck, M., & Haberman, J. (2005). Audacity: Free audio editor and recorder. *URL http://audacity.sourceforge.net*.

McRae, K., Cree, G.S., Seidenberg, M.S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, 37*, 547-559.

Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, *38*(11), 39-41.

Mirman, D. (2011). Effects of Near and Distant Semantic Neighbors on Word Production. *Cognitive, Affective, and Behavioral Neuroscience, 11(1),* 32-43.

Mirman, D., & Graziano, K.M. (2013). The Neural Basis of Inhibitory Effects of Semantic and Phonological Neighbors in Spoken Word Production. *Journal of Cognitive Neuroscience, 25(9),* 1504-1516.

Morton, J. (1980). The logogen model and orthographic structure. *Cognitive Processes in Spelling*, 117-133.

Nickels, L. (2002). Therapy for naming disorders: Revisiting, revising, and reviewing. *Aphasiology*, *16*(10-11), 935-979.

Nickels, L. A. (1997). *Spoken word production and its breakdown in aphasia*, Hove, UK: Psychology Press.

Nickels, L.A. (2000). Semantics and therapy in aphasia. In W. Best, K. Bryan & J.Maxim (Eds.), *Semantic Processing in Theory and Practice* (pp. 108-124). London: Whurr.

Nickels, L., Biedermann, B., Fieder, N., & Schiller, N. O. (2015). The lexical-syntactic representation of number. *Language, Cognition and Neuroscience, 30*(3), 287-304.

Nickels, L., & Cole-Virtue, J. (2004). Reading tasks from PALPA: How do controls perform on visual lexical decision, homophony, rhyme, and synonym judgements? *Aphasiology*, *18*(2), 103-126.

Nickels, L., & Howard, D. (1994). A frequent occurrence? Factors affecting the production of semantic errors in aphasic naming. *Cognitive Neuropsychology, 11*(3), 289-320.

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*(2), 227-252.

Plaut, D. C. (1995). Semantic and associative priming in a distributed attractor network. *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society, 17,* 37-42.

Porter, G., & Howard, D. (2004). *CAT: comprehensive aphasia test*. Psychology Press.

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, URL http://www.R-project.org.

Rabovsky, M., Schad, D. J., & Abdel Rahman, R. (2016). Language production is facilitated by semantic richness but inhibited by semantic density: Evidence from picture naming. *Cognition*, *146*, 240-244.

Romanova, A. (2015). *Word class effects on representation and processing in non-brain damaged speakers and people with aphasia.* Doctoral dissertation, Macquarie University, Sydney.

Székely, A., & Bates, E. (2000). Objective visual complexity as a variable in studies of picture naming. *Center for Research in Language Newsletter, 12(2),* 1-33

Székely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., … Bates E. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language, 51,* 247–250.

Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S.* Fourth edition. Springer.

Weigl, E., & Bierwisch, M. (1970). Neuropsychology and linguistics: Topics of common research. *Foundations of language*, 1-18.

Appendix A

Samples of speech production: picture description from the CAT (Porter and Howard, 2004).

The picture is set in a living room, a man is asleep in an armchair with his feet on a coffee table, and behind him are bookshelves. A cat is sitting on the top shelf trying to get fish from the fishbowl on the shelf below. It has dislodged some books that are about to fall on the man's head. A young child/baby who is playing on the floor is pointing at the cat to warn the man.

DEH:

’The…the cat is…[gesture]…and the books is…[gesture]…boom boom and the…[gesture] sleeping. Uhm…the uh baby is…screaming. Uhm coffee [points at the picture] and the …uhm…he…yeah and the uh...uhm…the uh…no…[gesture] plants and the uh…T…no the uh [gesture]…yeah…no…’

SJS:

’Sleep…tea…and gone…books…table…up there…em…birds and kitty cat, and and … ’eheh!’. Flowers…nice, em…baby, girl, man, car uh…radio, and up there [sound of falling], and books…chair…school, books, up there and books [gesture] up there, work, guy [gesture]…yeah, book, up there…[gesture].’

CHAPTER 4

Paper 3

Effects of phonological neighbourhood density and frequency in picture naming:

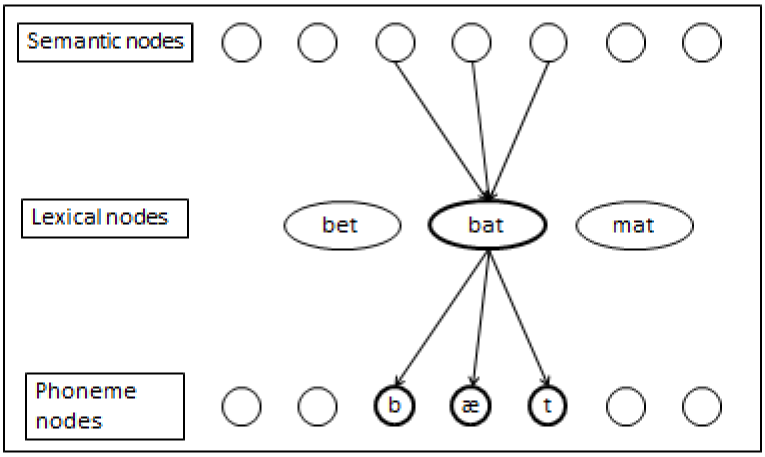
English monolinguals and French-English late bilinguals.

Introduction

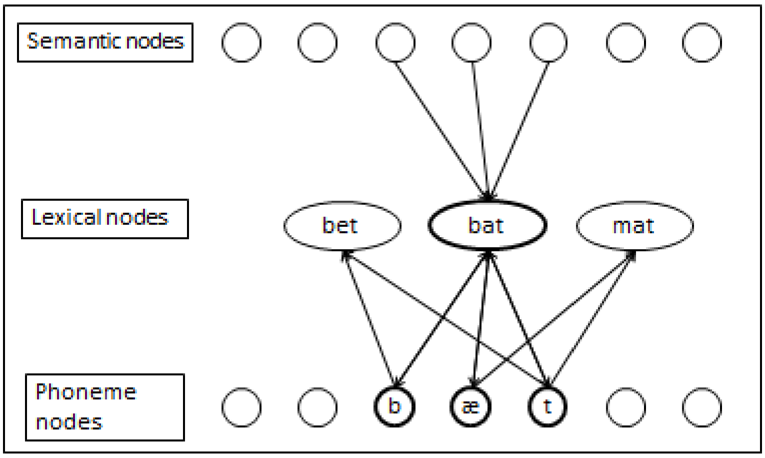
Every act of verbal communication requires access to the mental lexicon. We need to select the word that best matches the meaning we want to convey, and this task is more likely to be facilitated if the mental lexicon is organised along several critical dimensions. For instance, the phonological lexicon is believed to be organised by phonological similarity (Luce & Pisoni, 1998), and one common metric to characterise this phonological similarity is phonological neighbourhood density (PND).

The phonological neighbourhood density of a given word refers to the number of words in the lexicon that only differ from that word by one phoneme, either substituted, added, or deleted (Luce, 1986). For example, phonological neighbours of ’cat‘ include, amongst others, ’fat‘, ’cab‘, ’kit‘, ’scat‘, ’at‘. Some words have many phonological neighbours (high PND, or ’dense’ phonological neighbourhoods), others have few (low PND, or ’sparse‘ neighbourhoods). The frequency of these phonological neighbours is referred to as phonological neighbourhood frequency (PNF): a word’s phonological neighbourhood can be of high or low frequency (high or low PNF).

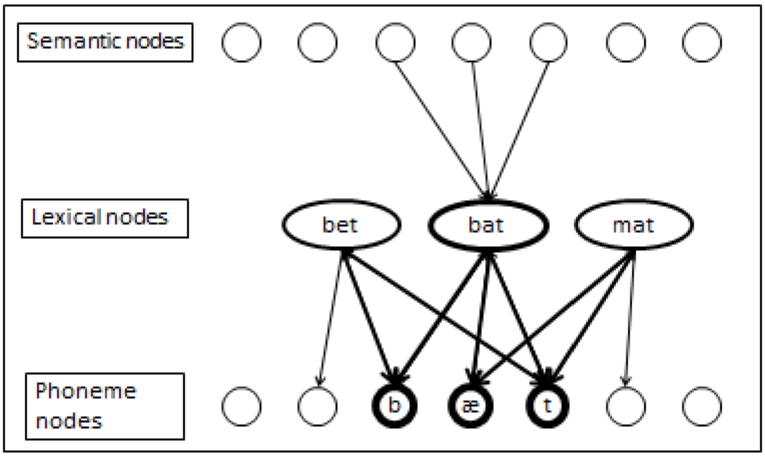
It is generally assumed that when a word is activated, the phonological neighbours of this word are also activated (e.g., Luce, Pisoni & Goldinger, 1990). This is inevitably true for spoken word perception, but some theories of production also hypothesise that these phonological neighbours are activated in the process of word production and even that they might affect the lexical selection process (e.g., Chen & Mirman, 2014; Dell & Gordon, 2003). The activation of phonological neighbours in spoken word production is best accounted for by models that assume interactivity between the word form level (phonological lexicon) and the phoneme level. For example, within the lexical network account proposed by Dell, Schwartz, Martin, Saffran and Gagnon (1997), activation of phonological neighbours occurs by feedback from the phoneme level to the word form level: activation flows back, not only to the target lexical item, but also to its phonological neighbours (see Figure 1). An alternative (but less straightforward) account of the activation of phonological neighbours in the course of speech production, is given by discrete models of lexical access (e.g., Levelt, Roelofs, & Meyer, 1999). In this framework, although phonological activation is restricted to



A. Activated target lexical node activates the corresponding phoneme nodes



B. Target phonemes send activation back to the target lexical node AND its phonological neighbours



C. Target and neighbours send activation back to the phonological nodes which are then selected for articulation?

Figure 1. Graphic representation of the activation of phonological neighbours within an interactive model (e.g., Dell et al., 1997)

the selected lexical node, some influence (an inhibitory effect on accuracy) of phonological neighbours could be explained by monitoring processes (this account of phonological neighbourhood effect within discrete models of lexical access is provided by Sadat, Martin, Costa, and Alario,2014). Before uttering speech, the speaker can attend to his own internal speech and ’monitor‘ the to-be delivered message. Phonological neighbours are similar to the intended word and therefore are more likely to be mistaken for it and to slip through the monitor, but also, words from dense neighbourhoods are more ’word-like‘ than words from sparse neighbourhoods, and hence would be more likely to pass the monitor easily, resulting in a phonological error. While it is less clear how this account could explain effects of phonological neighbours on latencies, if the potential phonological neighbour is successfully detected, then the eventual production of the target would be expected to be slowed.

Although in the area of spoken word perception, consistent inhibitory effects of PND have been attested (e.g., Luce & Pisoni, 1998), and a facilitatory effect of phonological neighbourhood density has been found on reading aloud (e.g., Mulatti, Reynolds, & Besner, 2016), there is not yet any consensus regarding the effect of PND and PNF on the word production (picture naming) of healthy speakers. Findings differ with respect to the presence and the direction of any effect, and in addition, effects seem to vary across languages and even vary depending on the age of the participants. Sadat et al. (2014) review the findings regarding the influence of PND on picture naming latencies, and this will also be discussed below.

In this paper, we are also interested in the influence of phonological neighbours on the speech production of people who speak more than one language (in this case, late French-English bilinguals), and more specifically, whether phonological neighbours of the non-target language are active and play a role when retrieving a given word. This possibility would entail that activation from the conceptual/ semantic system flows to both lexicons of each language, and that word forms from both languages become active. More specifically, non-target language neighbours could only play a role in a framework that assumes that, in bilingual speech production, representations from both languages are active up to the phoneme level. Although this is still under debate, there is evidence of activation of the non-target language up to the word form level (e.g., Colomé, 2001; Colomé & Miozzo, 2010) and even feedback between word form and phoneme levels across both languages of a bilingual (e.g., Costa, Roelstraete, & Hartsuiker, 2006). Building on these observations, it could be the case that cross-language neighbours play a role in bilingual speech production. However, research is very limited on that particular topic, with the exception of a recent study by Sadat, Martin, Magnuson, Alario, and Costa (2015) who found no effect of the number of Catalan neighbours on the Spanish picture naming latencies of Spanish-Catalan bilinguals. Hence, this study aims to gather evidence with respect to the influence of phonological neighbourhood(s) in bilinguals.

Before going further into the potential effects of phonological neighbourhoods, some consideration of the nature of phonological neighbours and phonological neighbourhood measures is warranted.

What counts as a phonological neighbour? As noted above, a phonological neighbour is commonly defined as a word that differs from the target by only one phoneme, with no further restrictions with respect to other characteristics of that word. What follows from this definition is that ’cats‘ counts as a phonological neighbour of ’cat‘, and ’home‘ and ’homes‘ both count as phonological neighbours of ’hose‘. Both N-Watch (Davis, 2005) and Clearpond (Marian, Bartolotti, Chabal, & Shook, 2012) are programs that allow the calculation of phonological neighbourhood density, and both use the one phoneme rule, such that ’cats‘ is returned as a neighbour of ’cat‘. Although this definition of neighbours is widely accepted and used in the speech production literature (e.g., Vitevitch, 2002; Sadat, et al., 2014), Sadat et al. (2015) mention the choice of ’lemmas‘ as neighbours of a target, which we assume means that only lemmas (which can be understood as headwords: uninflected forms, like dictionary entries) were selected among possible neighbours: For example, ’homes‘ would not be considered a neighbour of ’hose‘ then, because the headword ’home‘ is already a neighbour - this is a hypothetical example, as Sadat et al.’s study focused on Catalan neighbours of Spanish words.

The reason for choosing only lemmas over (potentially inflected) word forms as possible neighbours of a given word was not explained in Sadat et al., but yet this raises an interesting question. Should morphological variants of a given word be treated as different words, in which case, potential neighbours of a word can include several word forms that have the same lemma? Or, alternatively, should they be considered as inflected forms of the same word, and only words that do not share the same lemma as the target and as other words in the neighbourhood, could be considered as neighbours? This depends on how we conceptualise the speech production process, and more specifically what kind of representations make up the phonological word form level – the level at which phonological neighbourhood effects are considered to arise. This level could consist of the final holistic word form representation (including the stem[4] and the inflection, such as the plural suffix), as suggested in full-listing morphological theories (e.g., Butterworth, 1983). According

[4] Authors refer to “stem” when talking about morphological decomposition. The stem of “cats” is “cat” and the lemma of “cats” is also “cat”. There are cases where the two differ: for example the lemma of “producing” is “produce” but its stem is produc-. For that reason, we use the word “stem” instead of lemma when referring to morphological decomposition.

to this view, phonological neighbours of a word can share the same lemma (‘cats‘ is a neighbour of ’cat‘). Alternatively, representations at this level could be decomposed, consisting of, for example, for a word like “cats”, the stem “cat” and the plural suffix “-s”, as proposed by decompositional morphological theories (e.g., Taft, 2004). In this case, representations at the phonological level do not include inflected word forms, and, phonological neighbours under this account can only be lemmas. Hence, depending on one’s theory, a word’s neighbourhood could consist of neighbours that are calculated either based on (inflected) word-forms, or based on lemmas.

Neighbourhood frequency is the other type of potential predictor of interest. This refers to the frequency of a word’s neighbours. Unfortunately, however, this is also not as straightforward as it might initially appear. First, following on from the discussion above, the frequency of neighbours could be either related to word form frequency or to lemma frequency, depending on whether we choose “word form neighbours” or “lemma neighbours”. In addition, the question of how to compute a measure of neighbourhood frequency has to be addressed. Most picture naming studies with a focus on neighbourhood frequency use the average of the frequencies of each neighbour as a measure of neighbourhood frequency (e.g., Baus, Costa & Carreiras, 2008; Chan & Vitevitch, 2010; Vitevitch, 2002; Vitevitch & Sommers, 2003), although it is unclear why average frequency of a word’s neighbourhood should be preferred over the sum of the frequencies of each neighbour. Some authors have, however, used the summed frequency of the neighbours as a metric (e.g., Coady & Aslin, 2003; Mirman & Graziano, 2013). In the computational implementation of Levelt et al’s (1999) theory, WEAVER++, the probability of lexical selection is determined by the Luce ratio, which is the activation of the target divided by the sum of the activation of the competitors and the target. Consequently, this view suggests that what matters is the *sum* of phonological neighbours. In this paper, we examine this issue empirically by examining both average and summed neighbourhood frequency as predictors.

The motivation for taking phonological neighbourhood frequency into account is that what is important might not only be the number of similar representations in the lexicon, that is, phonological neighbourhood density, but also the frequency of these items, with higher frequency items more likely to have larger effects on processing. Indeed, in the visual lexical decision literature, the “neighbourhood frequency effect” refers to the fact that when at least one neighbour has a higher frequency relative to the target’s frequency, this target is harder to process (e.g., Grainger, O’Regan, Jacobs, & Segui, 1989). These findings, observed with orthographic neighbours in visual word recognition, have also been replicated for phonological neighbours in spoken word recognition (e.g., Luce & Pisoni, 1998). In the field of word recognition, these findings are supported by both activation-based models that postulate an intra-level inhibitory mechanism operating between lexical representations (as in McClelland & Rumelhart, 1981) and frequency-ordered serial search models (as in Forster, 1976 ; both cited in Grainger, O'regan, Jacobs, & Segui, 1989). It could also be the case that in speech production a phonological neighbour of higher frequency than the target has a stronger influence on its production than a neighbour that is lower in frequency than the target. To our knowledge, this possibility has not been addressed yet in picture naming studies.

Given the number and complexity of the measures involved, and the relative lack of investigation in spoken word production, it seems that the effects of PND /PNF on spoken word production merit further investigation. What are the specific types or characteristics of phonological neighbourhoods that most affect spoken word production? Which theories of spoken word production or of lexical organisation can account best for any such effects?

In this paper, we will examine, and attempt to dissociate, the influence of several types of phonological neighbourhood density and frequency on the picture naming behaviour of English monolingual young adults, and French-English late bilinguals.

Experiment 1: English monolinguals

As noted earlier, there is still no consensus in the literature regarding whether or how phonological neighbourhood density and frequency affect picture naming, with different results found in different languages. For example, English PND studies (see below) show different effects to the inhibitory effects found in Spanish (e.g., Baus, Costa, & Carreiras, 2008; Pérez, 2007; Sadat et al., 2014; Vitevitch and Stamer, 2006) and Dutch (e.g., Jescheniak & Levelt, 1994; Tabak, Schreuder, & Baayen, 2010). It has been postulated that the morphological structure of languages may modulate the influence of PND (e.g., Vitevitch and Stamer, 2006), in such a way that differences in the PND effect are to be expected with different languages. Therefore, we will focus our short review of current findings on picture naming in the English language, acknowledging that PND/PNF effects are not necessarily generalizable across languages.

In English speaking young adults, PND seems to exert either a facilitatory effect on latencies (Vitevitch, 2002: Experiments 3, 4, and 5; see also Gordon & Kurczek, 2013 for a facilitatory trend), or no significant effect (Vitevitch, Armbrüster & Chu, 2004); while the effect on accuracy has also been either facilitatory (Newman & Bernstein Ratner, 2007; Vitevitch, 2002: Experiment 3) or non-significant (Gordon& Kurczek, 2013; Vitevitch, 2002: Experiment 4; Vitevitch et al., 2004: Experiment

3). A different pattern of results has been found in other age groups: in children, Bernstein Ratner, Newman, and Strekas (2009) found no significant effect of PND on latencies despite facilitation on accuracy. Arnold, Conture and Ohde (2005), on the contrary, found inhibitory effects both on latencies and on accuracy, and Newman and German (2002; 2005) observed a detrimental effect of high PND on accuracy. In older adults, Gordon and Kurczek (2013) found inhibitory effects of PND on latencies but accuracy levels were not affected by this variable.

Regarding the effects of PNF (the average frequency of a word's phonological neighbours), facilitation has been found for accuracy in children (e.g., Bernstein Ratner et al., 2009; Newman & German, 2002). In young adults (e.g., Newman and Bernstein Ratner, 2007) and in older adults (e.g., Vitevitch & Sommers, 2003) facilitatory effects of PNF have been found on both accuracy and response latency.

All these studies but one (Gordon & Kurczek, 2013) used a factorial design (i.e., controlled sets of stimuli with a dense / sparse neighbourhood or high frequency /low frequency neighbourhood condition), which, because of the problems in precisely matching the item sets, usually leads to small numbers of items, resulting in a reduced number of trials. The present study investigated the influence of several measures of PND and PNF on picture naming (latency and accuracy) on young English speaking adults, using a large number of different items in a continuous design, and attempting to take into account individual variation induced by different participants and different items. By using this procedure, we hope to have increased power to determine which aspects of PND / PNF are most critical in predicting picture naming behaviour.

Method

Participants

40 monolingual English-speaking participants (29 females) were recruited from Macquarie University, Australia. All of them were students, and, if applicable, received course credit for their participation. All had English as their native language and none was exposed to another language at home. They were aged 18 to 36 (mean 20.7, *SD* 3.64) and had normal or corrected-to-normal vision.

Stimuli

Stimuli consisted of 386 black and white drawings taken from the International Picture Naming Project (IPNP) picture database (Szekely et al., 2004). Stimuli were selected when they had a target name that was a single word, and high name agreement in English monolingual speakers (75% or more). Following the experiment the item set was reduced further. Additional items were excluded from the analyses, due to low (less than 50%) Australian (Experiment 1), or bilingual (Experiment 2) accuracy scores (56 items), or because the Australian name for this target was made of two words (e.g., skipping rope, spinning top, coat hanger; 10 items). Two additional words (backpack and bandaid) were not present in the CELEX database and were also removed. This led to a final set of 318 items.

Procedure

DMDX (Forster & Forster, 2003) was used for presentation of the stimuli. Instructions were presented on the screen and explained further by the examiner: participants were asked to name pictures as quickly and accurately as possible, with a single word. Each trial started with a 200ms fixation cross, followed by a blank screen for 600ms, which was then followed by the target picture presented in the centre of the screen for 2000ms. Recording started upon stimulus presentation and continued for 2000ms after the picture disappeared. A new trial was initiated 1500ms after timeout. There were 10 practice items, followed by the 386 picture stimuli, organised in four blocks of approximately 11 minutes each, separated by a break. The amplifier was calibrated for each participant. The experiment lasted approximately an hour.

Analysis

Vocal response latencies were manually adjusted to get the most precise response time onset, using CheckVocal (Protopapas, 2007). A response was coded as correct when it corresponded exactly to the target word, with no fluency error. Our criterion was strict because the phonological neighbourhood density values do not hold true for synonyms. For example “refrigerator” for fridge or “bathtub” for bath were coded as incorrect. ‘Incorrect’ responses included fluency errors, synonyms, visual errors, other error types that did not fit in these categories, and no responses (or timeout).

Our goal was to assess, a) how phonological neighbourhood predicts picture naming behaviour (response time and accuracy), while controlling for other variables that have proved to be influential in picture naming, b) if applicable, which type of phonological neighbourhood measure is the best predictor and, and c) whether phonological neighbourhood interacts with other properties of words. To do so, we use linear mixed effect modelling to assess the specific respective influence of each PND/PNF value, allowing the effect to vary across participants and items.

We first describe the set of predictors that were used in these analyses, starting with our control predictors, and then define our predictors of interest (PND and PNF measures).

Control Predictors:

Our control predictors included the properties of the words or trials that have commonly been shown to influence the speed of picture naming. These include the response time of the preceding trial, name agreement, visual complexity, familiarity, age of acquisition, word frequency, imageability, word length in phonemes and phonotactic probability[5]. These variables are described below.

Standardized **response time of the preceding trial (preceding RT):** In addition to the fact that preceding RT is a good predictor of latencies, including preceding RT as a predictor can also potentially help attenuate some problems with autocorrelation, resulting in a better fit and promoting a clarified role of the predictors of interest (e.g., Baayen & Milin, 2010).

Name agreement refers to the degree to which participants agree on the name of the picture. Name agreement measures can be obtained offline on a separate group of participants similar to the experimental group (e.g., Alario & Ferrand, 1999; Barry, Morrison & Ellis, 1997; Ellis & Morrison, 1998). Other authors have calculated name agreement values on the basis of their experimental data (e.g., Sadat, Martin, Costa, & Alario, 2014; Severens, Van Lommel, Ratinckx, & Hartsuiker, 2005). The stimuli used here were selected on the basis of high name agreement values from the IPNP, but these were from American English participants. As name agreement can differ across English varieties (as shown for example by the relatively low correlation ($r<.5$) between British and American English name agreement norms in Barry et al., 1997), we also used the mean accuracy

[5] We did not include semantic neighbourhood density as this has not been found to significantly affect standard picture naming in unimpaired adults (see e.g., Hameau et al., this thesis, Chapter 2)

of our participants as a measure of Australian speeded name agreement (mean per item accuracy =88%, *SD*=13).

Objective **visual complexity** values were retrieved from the IPNP: this consists of the size of the digitized stimuli picture files in Kbytes (mean visual complexity =16665 Kbytes, *SD*=8911). This measure has been suggested to be preferable to subjective ratings of visual complexity which have been shown to be often confounded with familiarity (Szekely and Bates, 2000).

Familiarity and **Age of Acquisition** ratings were drawn from a British English norming study (Johnston, Dent, Humphreys, & Barry, 2010). Familiarity values were obtained by asking participants to rate how usual or unusual the concept/object is in their realm of experience, on a scale ranging from very unfamiliar to highly familiar (average familiarity=5.02, *SD*=1.15). The same participants were also asked to estimate the age at which they thought they had first learned the name of the object, choosing between seven age bands (mean age of acquisition = 2.91, *SD*=0.79).

Values of summed spoken and written word form and lemma **frequency** were obtained from the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995). There is still debate regarding where the frequency effect arises in the process of producing words. Some evidence points to the word form level (e.g., Jescheniak & Levelt, 1994), but also to the lemma level (Gahl, 2008). Hence, we included both word form and lemma frequency as they might predict response time and accuracy performance (Brysbaert & New, 2009). Written word form frequencies represent the number of times the word occurs in a 16,600,000 word corpus, and word form spoken frequencies represent the number of times the word occurs in a 1,300,000 word corpus. Lemma frequencies represent the occurrence of the headword for a given word in the same corpus. For example, the word ‘apple’ has a lemma frequency of 546, which is the sum of the word-form frequency of ‘apple’ (315), and that of ‘apples’ (231). Frequency values were log transformed to reduce the influence of extreme values. (mean log word form frequency=2.38, *SD*=0.72; mean log lemma frequency=2.56, *SD*=0.66)

Ratings of **imageability** (the ease with which a word gives rise to a sensory mental image) were obtained from the MRC database (Coltheart, 1981), and were available for 248 of the final 318 experimental items (mean= 593, *SD*=33).

Word length was the number of phonemes in each target word (mean=4.25, *SD*=1.58).

Phonotactic probability was calculated using Vitevich & Luce (2004)’s algorithm: average unigram or bigram positional probabilities across a word (mean=0.44, *SD*=0.33). The measure was computed using an online program (Phonological Corpus Tools: Hall, Allen, Fry, Mackie, & McAuliffe, 2015).

Experimental Predictors

Phonological neighbourhood density (PND) was calculated using the online program CLEARPOND (Marian, Bartolotti, Chabal, & Shook, 2012), which uses the one-phoneme difference rule (words were neighbours if they shared all but one phoneme, either substituted, added or deleted)[6]. We calculated both word form neighbours, and lemma neighbours. Lemma neighbours were selected by considering only one neighbour per headword, and discarding words that had the same headword as the target. For example, "apples" is a word form neighbour of "apple" since the two words only differ by one phoneme, but it is not a lemma neighbour of "apple" because both words share the same lemma. Within the neighbours, we also considered more specifically the number that were higher in frequency than the target word. All measures were log transformed to limit the influence of extreme values.

Four different measures of **PND** were therefore included:

- Word form PND (WF.PND): number of word form neighbours (mean = 0.98, *SD*= 0.53)
- Lemma PND (L.PND): number of lemma neighbours (mean= 0.92, *SD*= 0.59)
- Word form higher frequency PND (WF.PND.h): number of word form neighbours that are higher in frequency than the target (mean=0.46, *SD*=0.40)
- Lemma higher frequency PND (L.PND.h): number of lemma neighbours that are higher in frequency than the target (mean 0.49, *SD*=0.44)

In addition, we considered several measures of **PNF** as discussed in the Introduction (frequencies were taken from the CELEX database, Baayen et al., 1995):

- Summed frequency of word-form neighbours (WF.sPNF) (mean=3.54, *SD*=1.33)
- Average frequency of word-form neighbours (WF.aPNF) (mean=2.56, *SD*=0.87)

[6] The CLEARPOND online program is designed such that possible neighbours are limited to words belonging to an educated monolingual adult's lexicon (a frequency threshold was used, so that the number of words in the corpus – 27,751 for English – is a reasonable vocabulary size estimate). This is to ensure that there are no rare words that are unlikely to be known by most adult monolinguals.

- Summed frequency of lemma neighbours (L.sPNF) (mean=4.10, *SD*=1.12)
- Average frequency of lemma neighbours (L.aPNF) (mean=2.29, *SD*=1.44)
- Summed frequency of higher frequency word-form neighbours (WF.sPNF.h) (mean=3.92, *SD*=1.04)
- Average frequency of higher frequency word-form neighbours (WF.aPNF.h) (mean=3.39, *SD*=0.81)
- Summed frequency of higher frequency lemma neighbours (L.sPNF.h) (mean=4.30, *SD*=0.97)
- Average frequency of higher frequency lemma neighbours (L.aPNF.h) (mean=3.61, *SD*=0.86)

All measures were log transformed to avoid the influence of extreme values, and then predictors were centred so that the interpretation of intercepts is not affected by interaction terms between the variables at play (Schielzeth, 2010) by inducing multicollinearity. Correlations between explanatory variables in multiple regressions can create multicollinearity, which can affect the calculated importance of these explanatory variables by inflating their standard errors resulting in larger p-values. This can be a problem for coefficients of interest as both of two intercorrelated variables can lose statistical significance, leading to the potentially incorrect inference that neither has an effect. The variance inflation factor (VIF) is an indicator of multicollinearity. It has a lower bound of 1, and no upper bound. Depending on the author, a VIF greater than 2.5, which corresponds to an R^2 of .60 (e.g., Allison, 2010), or a VIF greater than 5, which corresponds to an R^2 of .80 (e.g. Hutcheson & Sofroniou, 1999) is a sign of potentially problematic multicollinearity. Multicollinearity is not a problem between control predictors as we are not interested in which of these predictors are significant, however, it may be an issue if the explanatory variables of interest are affected.

Inspection of the pairwise correlations between all predictors revealed that several of the predictors were correlated, especially and unsurprisingly all the PND/PNF variables. More importantly, our PND/PNF predictors were generally strongly negatively correlated with length: longer words had fewer phonological neighbours (see Table 1). Hence, after first glance at the correlation coefficients, a certain degree of multicollinearity between length and (at least some of) our predictors of interest is observed.

Table 1. Pairwise Pearson correlation coefficients between the control predictors and the PND/PNF predictors (monolinguals), and within the PND/PNF predictors.

	WF. PND	L. PND	WF. PND.h	L. PND.h	WF. sPNF	WF. aPNF	L. sPNF	L. aPNF	WF. sPNF.h	WF. aPNF.h	L. sPNF.h	L. aPNF.h
Name agreement	.02	.01	.01	.01	.05	.06	.05	.06	.02	.02	.05	.07
Visual Complexity	-.04	-.06	-.04	.01	-.05	-.06	-.03	-.02	-.05	-.05	-.01	-.01
Familiarity	.29**	.28**	.08	.08	.31**	.30**	.30**	.29**	.21**	.23**	.21**	.21**
Age of acquisition	-.40**	-.37**	-.18**	-.17**	-.44**	-.42**	-.42**	-.42**	-.32**	-.34**	-.31**	-.31**
Worm frequency	.55**	.52**	.13*	.24**	.59**	.57**	.55**	.53**	.33**	.36**	.40**	.40**
Lemma frequency	.13*	.12*	.11*	.11	.15**	.15**	.18**	.18**	.20**	.21**	.20**	.19**
Imageability	-.16*	-.15*	-.08	-.21**	-.15*	-.12	-.17**	-.16**	-.07	-.06	-.17**	-.16*
Length (phonemes)	-.84**	-.82**	-.67**	-.69**	-.82**	-.74**	-.85**	-.82**	-.72**	-.71**	-.76**	-.72**
Phonotactic probability	-.33**	-.31**	-.32**	-.30**	-.35**	-.33**	-.36**	-.36**	-.35**	-.35**	-.33**	-.32**
WF.PNDensity		.96**	.76**	.82**	.93**	.83**	.94**	.89**	.80**	.77**	.86**	.82**
L.PNDensity			.74**	.81**	.88**	.77**	.93**	.88**	.78**	.76**	.85**	.80**
WF.PNDens.higher				.85**	.77**	.72**	.77**	.75**	.89**	.83**	.81**	.76**
L.PNDens.higher					.80**	.72**	.82**	.78**	.79**	.73**	.89**	.82**
WF.sPNFrequency						.97**	.95**	.94**	.86**	.84**	.88**	.85**
WF.aPNFrequency							.88**	.89**	.83**	.82**	.82**	.80**
L.sPNFrequency								.99**	.85**	.84**	.92**	.89**
L.aPNFrequency									.85**	.84**	.91**	.88**
WF.sPNFreq.higher										.99**	.87**	.84**
WF.aPNFreq.higher											.85**	.82**
L.sPNFreq.higher												.97**

*** Correlation is significant at p<0.01 (2-tailed); *Correlation is significant at p<0.05 (2-tailed)*

Abbreviations: WF=word form, L=lemma, PND=phonological neighbourhood density, PNF=Phonological neighbourhood frequency, s=summed, a=average, h=of higher frequency than the target.

A common procedure that is used as an attempt to isolate the variance specific to the predictor of interest over and above the effect of another correlated variable, is to residualize predictors one against the other by running a linear regression, and use the obtained residuals as a new orthogonalized variable. In our case, it would mean for example to residualize length in phonemes against measures of PND/PNF. Although this seems like a useful and appropriate technique, it presents with some dangers for misinterpretation, as pointed out in a recent study by Wurm and Fisicaro (2014). These authors warn that the new, residualized variable should not be understood as "an improved, purified, or corrected version of [the original variable], it is simply the errors of prediction with which one is left when predicting [one variable] from [the other]" (p.40). In addition, the β coefficients for the residualized variable do not change when compared to a situation with both unresidualized variables; it is the standard error for the unresidualized variable that decreases, artificially making statistical significance more likely for that new variable. Wurm and Fisicaro (2014) do not recommend this procedure, suggesting that it does nothing but complicates the interpretation of results, potentially leading to interpretation errors. As regressions already allow us to determine the effect of one predictor while holding the level of another predictor constant, we decided not to residualize our predictors and use the true values instead.

Results

Response time analyses

Incorrect responses were removed from response time (RT) analyses. This resulted in 11,141 remaining observations. Data was analysed using R (R Development Core Team, 2011).

In order to approximate the normal distribution for RT, we investigated the optimal data transformation using the BoxCox function (Box & Cox, 1964) from the R package MASS (Venables & Ripley, 2002). Inspection of the log-likelihood plot resulted in determining that the power-.6 transformation was the most appropriate for our data. Hence, our RTs were raised to the power of -.6. This procedure results in lower values representing longer RTs, which is important to bear in mind when interpreting the direction of the effects of our predictors.

Linear mixed effects models were used to assess the contribution of our variables of interest, using the lme4 software package (Bates, Maechler, Bolker, & Walker, 2015). Parameter specific p-values were computed using normal approximation (Mirman, 2014). We first searched for the best base model excluding PND/PNF variables. Our first model included by-item and by-participant random intercepts in order to account for the random variation induced by specific words or speakers, and all control variables (preceding RT, visual complexity, word form frequency or lemma frequency[7], age of acquisition, familiarity, imageability, length in phonemes, phonotactic probability and Australian speeded name agreement). Next, we removed non-significant fixed factors, which resulted in a model including preceding RT, visual complexity, word form frequency, age of acquisition, familiarity, and Australian speeded name agreement as "control" fixed factors[8]. A summary of the full model including all control predictors, and of this base model is shown in Appendix A.

Consistent with previous literature on the predictors of picture naming latencies, longer preceding RT, higher visual complexity, and later age of acquisition predicted slower reaction times, and words with better name agreement and higher familiarity predicted faster reaction times. However, and oddly, word frequency showed an effect in the opposite direction to that usually observed: more frequent words were named more slowly, that is, the effect of word form frequency, when holding preceding RT, visual complexity, age of acquisition, name agreement and familiarity constant, is in the direction of inhibition. Nonetheless, this only holds when the other control variables are in the model: the zero-order correlation between word frequency and response time shows the expected facilitatory effect on response times (more frequent words were faster named). Hence, this is, most likely, a case of negative suppression that can occur in regression when the magnitude of an effect is greater (although the sign is opposite) in the presence of suppressor variables (Tabachnick & Fidell, 2013). We note here that an effect of frequency in the same unexpected direction was found by Sadat et al. (2015) in their analyses of picture naming RT in bilinguals: words with higher frequency values predicted longer latencies.

Number of phonemes did not significantly predict RT, either with all variables entered simultaneously, or on its own (zero-order correlation with RT). However, since length is strongly correlated with most of the PND/PNF variables, when a PND/PNF variable was a significant predictor (or close to significant $p<.1$), we then ran a subsequent model including length as a predictor.

[7] Both base models with word form and lemma frequency were compared, in order to choose the one that was the best predictor of RT. Word form frequency was found to improve the model better (β=-3.69x10^{-4}, *SE*= 1.11x10^{-4}, *t*= -3.33, *p*=8.601x10^{-4} for Word form frequency, β=-3.38x10^{-4}, *SE*=1.20x10^{-4}, *t*=-2.81, *p*=4.898x10^{-3} for lemma frequency) and was therefore used for all subsequent analyses.

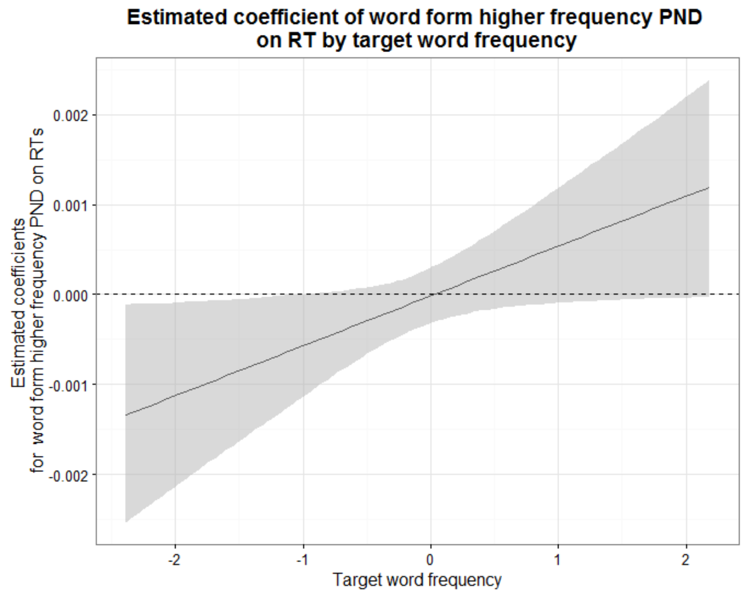
[8] The same base model was run without imageability (as it led to performing analyses on 318 items vs 248 if imageability is in the model), and the same predictors were found to be significant.

Table 2. Summary of the effects of each PND/PNF predictor on Response time

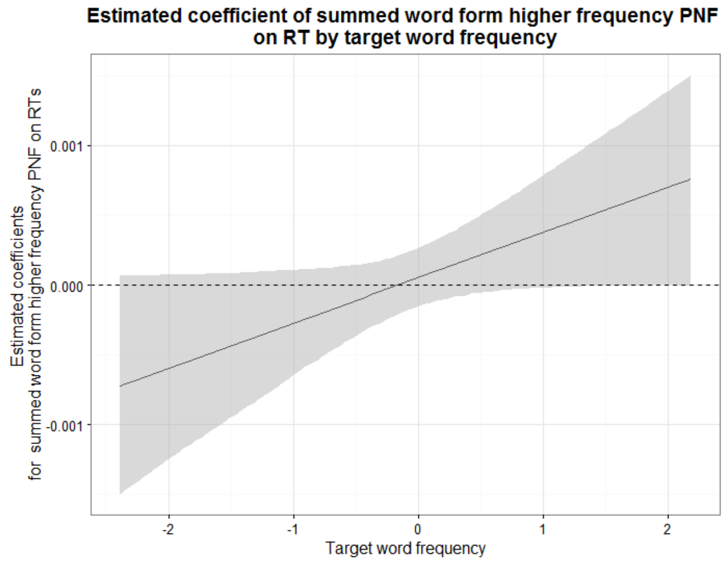
Fixed effects	*β*	*SE β*	*t*-value	*p*-value	VIF
Phonological Neighbourhood Density					
Word form: WF.PND	-1.43E-04	1.46E-04	-0.98	0.329	1.5
Lemma: L.PND	-1.18E-04	1.29E-04	-0.91	0.362	1.5
Word-form higher frequency: WF.PND.h	-6.70E-05	1.60E-04	-0.42	0.676	1.0
Lemma higher frequency: L.PND.h	-1.14E-04	1.49E-04	-0.77	0.444	1.1
Phonological Neighbourhood Frequency					
Word form summed: WF.s.PNF	-3.43E-05	1.27E-04	-0.27	0.787	1.6
Word form average: WF.aPNF	-1.43E-04	1.46E-04	-0.98	0.329	1.5
Lemma summed: L.sPNF	-1.12E-04	1.07E-04	-1.04	0.296	1.5
Lemma average: L.aPNF	*-4.23E-04*	*2.48E-04*	*-1.70*	*0.088*	*1.2*
L.aPNF: model including length	*-3.62E-04*	*2.68E-04*	*-1.35*	*0.177*	*1.4*
WF higher frequency summed: WF.sPNF.h	3.36E-05	1.10E-04	0.31	0.760	1.1
WF higher frequency average: WF.aPNF.h	2.29E-04	2.17E-04	1.05	0.293	1.1
Lemma higher frequency summed: L.sPNF.h	-7.57E-05	1.03E-04	-0.73	0.463	1.1
Lemma higher frequency average: L.aPNF.h	-8.54E-05	2.28E-04	-0.37	0.708	1.2

Each PND/PNF predictor was added to the base model including preceding RT, name agreement, familiarity, word form frequency, age of acquisition and visual complexity, with random intercepts for participant and item. When a predictor was significant or close to significant, an additional model was run also including length (number of phonemes). Table provides coefficient estimates (β), standard errors (SE), t- and p-values, with the variance inflation factor (VIF). WF=word form, L=lemma, PND=phonological neighbourhood density, PNF=phonological neighbourhood frequency, h=of higher frequency than the target, s=summed, a=average. (Marginally) significant effects are in italics

As we were interested to learn about which of the measures of PND/PNF affected performance, we introduced each of these measures independently into the base model. Beta coefficients for each of the PND/PNF variables that were added separately to the base model are shown in Table2. None of the measures were significant as main effects. We only note a marginally significant inhibitory effect of the average frequency of lemma neighbours that disappeared with the introduction of length as a covariate.



A.



B.

Figure 2. Graphic representation of significant interactions (grey ribbon=confidence intervals) showing the effect of PND (panel A) and PNF (panel B) on response time moderated by target word frequency.

Table 3. Summary of the fixed effects of each PND/PNF predictor on accuracy: coefficient estimates (*β*), standard errors (*SE*), *z*- and *p*-values, with the variance inflation factor (VIF).

	Models without length					Models including length				
Main effects	β	SE β	z-value	p-value	VIF	β	SE β	z-value	p-value	VIF
Phonological neighbourhood density										
Word form: WF.PND	*-3.05E-01*	*1.73E-01*	*-1.76*	*0.078*	*1.2*	-5.28E-02	2.99E-01	-0.18	0.860	3.6
Lemma: L.PND	*-2.82E-01*	*1.54E-01*	*-1.83*	*0.067*	*1.2*	-8.45E-02	2.59E-01	-0.33	0.745	3.3
Word-form higher frequency: WF.PND.h	5.77E-03	2.13E-01	0.03	0.978	1.0	*5.26E-01*	*2.83E-01*	*1.86*	*0.063*	*1.8*
Lemma higher frequency: L.PND.h	-1.28E-01	1.95E-01	-0.66	0.510	1.0	3.02E-01	2.73E-01	1.11	0.268	2.1
Phonological neighbourhood frequency										
Word-form summed: WF.sPNF	*-2.49E-01*	*1.48E-01*	*-1.68*	*0.094*	*1.2*	-2.15E-02	2.46E-01	-0.09	0.930	3.4
Word-form average: WF.aPNF	*-3.05E-01*	*1.73E-01*	*-1.76*	*0.078*	*1.2*	1.03E-01	3.12E-01	0.33	0.740	3.6
Lemma summed: L.sPNF	*-2.33E-01*	*1.28E-01*	*-1.83*	*0.068*	*1.2*	-6.30E-02	2.21E-01	-0.29	0.776	3.6
Lemma average: L.aPNF	*-3.05E-01*	*1.73E-01*	*-1.76*	*0.078*	*1.1*	1.03E-01	3.12E-01	0.33	0.740	1.3
WF higher frequency summed: WF.sPNF.h	3.70E-02	1.44E-01	0.26	0.798	1.1	2.37E-01	1.97E-01	1.20	0.229	1.9
WF higher frequency average: WF.aPNF.h	-8.81E-04	1.34E-01	-0.01	0.995	1.1	1.83E-01	1.92E-01	0.96	0.339	1.6
Lemma higher frequency summed: L.sPNF.h	-1.01E-01	2.81E-01	-0.36	0.718	1.1	8.23E-02	3.38E-01	0.24	0.808	2.2
Lemma higher frequency average: L.aPNF.h	-4.14E-02	2.96E-01	-0.14	0.889	1.1	2.35E-01	3.94E-01	0.60	0.550	2.1

Abbreviations: WF=word form, L=lemma, PND=phonological neighbourhood density, PNF=phonological neighbourhood frequency, h=of higher frequency than the target, s=summed, a=average. (Marginally) significant effects are in italics.

When we included interactions in the models, we found a significant interaction between target word frequency and the number of word form neighbours of higher frequency (WF.PND.h): *β*=5.22x10^{-4}, *SE*=2.61x10^{-4}, *t*[11141]=2.003, *p*=0.046, VIF=1.22) , such that, on targets of low frequency, the effect was inhibitory, but evolved towards a facilitatory effect on high frequency targets. The same type of interaction, although only marginally significant (*β*=2.86x10^{-4}, *SE*=1.64x10^{-4}, *t*[11141]=1.74, *p*=0.083, VIF=1.17) was found between target word frequency and the summed frequency of the higher frequency word-form neighbours (h.WF.sPNF). A table summarizing coefficients for all interactions involving PND/PNF variables and frequency in monolingual response time analyses is available in Appendix B). For a better understanding of interactions, we visualized them with plots using the R package Interplot (Solt & Hu, 2015): these are presented in Figure 2.

Accuracy analyses

Accuracy analyses were also performed, by means of generalised linear mixed effect models, using the lme4 software package (Bates, Maechler, Bolker, & Walker, 2015). Accuracy was coded as error = 0, correct = 1, and overall, accuracy was 88%. For accuracy analyses, we considered the same set of predictors as for RT analyses, but excluding preceding RT, and name agreement (as our name agreement is a measure of mean accuracy per-item across subjects). Among the remaining control variables, only age of acquisition was found to be significant (see Appendix C), and in the expected direction: words acquired later were more likely to be named incorrectly.

Table 4. Summary of the interactions between target frequency and each PND/PNF variable: coefficient estimates (*β*), standard errors (*SE*), *t*- and *p*-values, and the AIC for the models with significant and marginally significant interactions.

Variable in the interaction with target frequency	β	SE β	t-value	p-value	AIC
Phonological neighbourhood density					
Word-form higher frequency: WF.PND.h	0.81	0.34	2.38	0.017	7865.5
Lemma higher frequency: L.PND.h	0.47	0.27	1.74	0.081	7869.2
Phonological neighbourhood frequency					
Lemma summed: L.sPNF	0.37	0.17	2.11	0.035	7868.7
WF higher frequency summed: WF.sPNF.h	0.52	0.22	2.40	0.017	7865.9
WF higher frequency average: WF.aPNF.h	0.87	0.35	2.47	0.013	7867.1
Lemma higher frequency summed: L.sPNF.h	0.39	0.19	2.06	0.040	7867.6
Lemma higher frequency average: L.aPNF.h	0.96	0.40	2.40	0.016	7866.8

We added each PND/PNF predictor individually to the base model that included by-item and by-participant random intercepts and age of acquisition as a fixed factor. Several measures (word form and lemma PND, word form and lemma summed PNF, word form average PNF) showed a marginally significant inhibitory main effect on accuracy, but generally only before length was introduced as a covariate (see Table 3). Only one predictor showed any suggestion of a facilitatory effect: higher frequency word form PND was marginally significant (p=0.06) in the direction of facilitation, but only when length was in the model and length becomes significant in this model (*β*=0.22, *SE*=0.08, *t*=2.76, p=0.006). The effect of length was always in the unexpected direction with longer words being more accurately named, although this effect was rarely significant. We note that when length is entered as a single predictor of accuracy, its effect (although not significant: *β*=-0.07, SE= 0.06, *t*= -1.236, *p*=0.216) is in the other direction: longer words are less accurately named. It seems likely therefore, that the introduction of length, either resulted in suppression (in the case of higher frequency word form PND) or created some multicollinearity (VIF range from 1.3 to 3.6 for the PND/PNF predictors when length is in the model, which can be problematic or not according to different authors, inflating p-values for these predictors, It is clear that if there are main effects of PND/PNF measures on accuracy, they are not reliable.

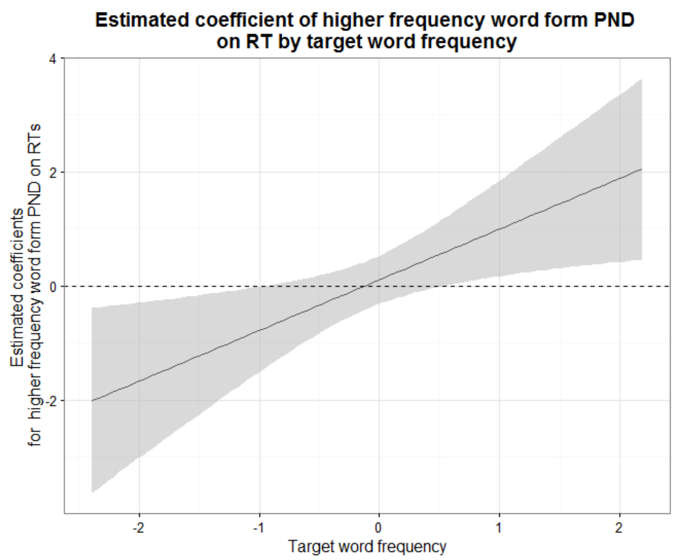


Figure 3. Graphic representation of the most significant interaction (grey ribbon=confidence interval) showing the effect of higher frequency WF PND on response time moderated by target word frequency.

When we examine the interactions, all higher frequency measures of PND/PNF, as well as summed lemma PNF, showed a significant interaction with target word frequency following the same pattern (see Table 4 for significant and marginally significant effects, Appendix D for the full set of interactions, and Figure 2 for depiction of the most significant interaction): target words of low frequency were less accurately named if they had many higher frequency neighbours or / if the higher frequency neighbours were of higher frequency, and words of high frequency were more accurately named if they had many higher frequency neighbours or if the frequency of these neighbours was particularly high. Models that included the interaction between PND/PNF measures and target word frequency were compared by examining the Akaike Information Criterion (AIC): "the AIC is related to the evidence ratio, which expresses the relative probability that the model with the lowest AIC is more likely to provide a more precise model of the data" (Lahmann, Steinkrauss, & Schmid, 2015). Although AICs were similar, the interactions that best predicted accuracy in monolinguals (those with the lowest AIC) were those that included higher frequency word form PND (see Figure 3), and summed higher frequency word form PNF.

Discussion

This experiment examined the effect of several measures of phonological neighbourhood on picture naming in monolingual English speakers. In neither response time nor accuracy analyses was there a clear or robust main effect of any measure of phonological neighbourhood density or phonological neighbourhood frequency. The null effect of PND measures on latencies is consistent with Vitevitch et al. (2004), but inconsistent with Vitevitch's often cited earlier study (Vitevitch, 2002: Experiments 3, 4, and 5). The absence of an effect on accuracy is consistent with several studies (Gordon& Kurczek, 2013, Vitevitch, 2002: Experiment 4; Vitevitch et al., 2004: Experiment 3) but inconsistent with others that found facilitation from high PND on accuracy (Newman & Bernstein Ratner, 2007; Vitevitch, 2002: Experiment 3). The absence of an effect of PNF for this type of task and participant sample is inconsistent with the facilitatory effect found by Newman and Bernstein Ratner (2007) for both accuracy and response time. However, all these studies but one (Newman & Bernstein Ratner, 2007) used factorial designs with controlled sets of stimuli. Here, we used a continuous design, whereby effects are studied allowing individual variability of items and participants. Indeed, as pointed by Rabovsky, Schah and Abdel Rahman (2015, p.241), "dichotomizing continuous variables can result in a substantial loss of statistical power due to reducing the amount of experimental variance. Furthermore, the excessive matching of other variables required by this dichotomization strategy can result in the selection of unusual materials". Hence, we are confident that our procedure led to a more powerful and fine-grained analysis.

However, despite this absence of a clear main effect, critically, we found interactions between predictors of interest and the frequency of the target word, both with RT and accuracy as a dependent variable. In response time analyses, having many neighbours of higher frequency led to slower RTs for low frequency targets, but speeded naming on targets of high frequency, while there was no effect of these neighbours on targets of medium frequency. Similar effects were found in accuracy analyses, where several measures of higher frequency neighbours and neighbourhood frequency interacted with target word frequency in the same fashion. For low frequency targets, having many higher frequency neighbours or a higher (summed or average) frequency of the higher frequency neighbours made the word more likely to be incorrectly named. In contrast, if the target was of high frequency it was more likely to be correctly named if it had many higher frequency neighbours. This is consistent with the spoken word recognition literature where neighbours of higher frequency than the target have an inhibitory effect on targets of low frequency (e.g., Grainger et al., 1989).

Hence, our data shows that although there is no effect of PND or PNF measures overall, phonological neighbourhood does have an effect when modulated by target word frequency: neighbours of higher frequency than the target (when the target is low in frequency), and to a lesser extent, the frequency of these neighbours, would appear to generate competition in such a way that these target words are slower and less accurately named.

Our results provide additional evidence that phonological neighbours generate "opposing facilitatory and inhibitory forces" as demonstrated by Dell and Gordon (2003), and Chen and Mirman (2012). However, here these opposite forces do not depend on the task (inhibitory effect in reception vs facilitatory effect in production) but, within speech production, on the relative frequency of a word compared to its neighbours. Therefore, it is possible that, depending on the overall frequency of a given experimental set of items, a net null, inhibitory or facilitatory main effect of (higher frequency) PND/PNF may emerge: if overall, items are low in frequency, the effect of PND / PNF on that whole set may be inhibitory, but alternatively, if the data set is of high overall frequency, PND/PNF might facilitate picture naming latency and/or accuracy. Here, we made sure we kept only items with high name agreement, a measure that is usually correlated with frequency. Since we wanted to have the same set of items across both populations (monolinguals and late bilinguals), only items that were correctly named by at least 50% of the bilinguals were included, and consequently it is possible that our item set is overall of relatively high frequency compared to other studies. Moreover, in the majority of other studies, a different approach is used to obtain better accuracy levels: naming accuracy is increased by a previous "rehearsal" of the target names of the experimental items by the participant before the experimental trials in a familiarisation phase, allowing the inclusion of lower frequency / more "difficult" items. Hence, depending on the procedure used in the study to select items, the set of experimental items might be more likely to generate inhibitory, null, or facilitatory effects of PND/PNF.

When we compared statistical models that used lemma frequency and those that used word form frequency measures, those with word form neighbours and word form frequency were usually a better fit of our data. It seems then that what matters more is the number of inflected word forms that are similar to the target, and less so the headwords that are similar to the target: the neighbourhood of "cat" is then likely to include the word "cats". These results tend therefore to favour spoken word production theories that include a representation of the full, morphologically complex form (e.g., Butterworth, 1983) over decompositional morphological theories (e.g., Taft, 2004).

Finally, with regards to the question of whether "summed" or "average" frequency of the neighbours should be chosen, the pattern that emerges from our data seems to be more in favour of the "summed" option. In response time analyses, only the summed frequency of higher frequency word form neighbours showed a marginal effect on latencies. Moreover, in accuracy analyses, it seems that, although the difference in the fit of the respective models is small, summed higher frequency word form PNF was a better predictor than average higher frequency word form PNF, and that summed higher frequency lemma PNF was better than average higher frequency lemma PNF. Therefore, there is some evidence that summed neighbourhood frequency is a better predictor than average neighbourhood frequency, consistent with the predictions of a theory that includes the Luce ratio as a competitive mechanism (e.g. Levelt et al., 1999).

The emerging picture here is that in several different ways, frequency is an important factor when investigating the effect of phonological neighbours on monolingual speech production.

In bilinguals, frequency of use is divided between two languages, consequently frequency of a lexical item is most probably lower than for monolingual speakers of the same language. In addition, phonological neighbourhood is affected by the fact that two different phonological systems are at play. In Experiment 2, we therefore investigate how bilingualism affects the influence of different types of phonological neighbourhood density on word retrieval in a population of French-English late bilinguals.

Experiment 2: French-English late bilinguals

Introduction

In this paper, for the sake of simplicity, we will refer to the population under investigation as late bilinguals or just bilinguals, being aware of the debate around what constitutes a bilingual (some other ways to refer to our population would be second language -or L2- learners), and with no intention to generalize our findings to bilingualism in general. Here, we investigate how knowing (even with different proficiency levels) two languages affects speaking in one of these languages, looking more specifically into the effects of PND and PNF.

The influence of cross-language neighbourhood density (the number of similar words relative to a given target, that belong to the other language of a bilingual: for example French neighbours of English words) in bilinguals has been investigated in the field of visual word recognition, and has revealed significant effects of cross-language orthographic neighbourhood density in visual recognition tasks (e.g., Coltheart, Davelaar, Jonasson, & Besner, 1977; van Heuven, Dijkstra & Grainger, 1998), but in the area of speech production, research is more scarce. De Groot, Borwald, Bos, and van der Eijnden (2002) looked at the influence of within- and cross-language orthographic neighbourhood density on reading aloud in English-Dutch bilinguals: both within- and cross-language orthographic neighbourhood density correlated negatively with latencies, and significant effects of both within- and cross-language orthographic neighbours were observed in regressions.

In bilingual picture naming, few studies have examined the influence of PND (and to our knowledge, none has targeted the influence of PNF). As far as we know, only two picture naming studies have documented the influence of PND in bilinguals. First, Sadat, Martin, Magnuson, Alario, and Costa (2015) examined picture naming performance in highly proficient early bilinguals from Catalunia, a bilingual area of Spain. Although not the primary focus of their study, their predictors of picture naming included both within and cross language phonological neighbourhood density, calculated with the one phoneme difference rule (Luce, 1986). They found an inhibitory effect of within language (Spanish) phonological neighbourhood density on latencies, and no effect on accuracy. There was no effect of cross-language phonological neighbourhood density (number of Catalan neighbours) on response time or on accuracy. Second, Marian and Blumenfeld (2006) investigated the role of language status on the phonological neighbourhood density effect in picture naming. Two groups of late bilingual English-German participants were asked to name pictures in

German. One group had English as a native language, the other, German. Pictures represented words that had either high or low German neighbourhood density. Overall, accuracy was better in both groups on high neighbourhood density targets, whereas latencies were shorter in the high neighbourhood density condition, but only for non-native speakers. The authors concluded that the group differences in latencies on high vs low neighbourhood density items were possibly due to a difference in proficiency that made L2 words overall lower in frequency and thus more sensitive to phonological neighbourhood effects. We note here that their findings on speakers naming pictures in their first language are consistent with our own findings on (monolingual) first language speakers of English (no main effect of PND on latencies).

Marian and Blumenfeld (2006) did not look at the influence of cross-language PND. However, they did mention that, although it would be informative to take it into account, it was not possible to do so, "because differences in phonetic features between German and English precluded meaningful computations of cross-linguistic phonological neighbourhoods in English". With respect to Sadat et al.'s (2015) study the calculation of Catalan-Spanish cross-language neighbours did not trigger this type of concern as there are probably no dramatic differences between the Spanish and Catalan phonological inventories (although there is not complete phonological overlap). However, it is true that there are probably important differences between English and German's respective phonology, and the concern would grow the more distant the phonological systems of language pairs.

In addition to this difficulty in calculating cross-language neighbours between two languages with different phoneme repertoires, the same problem applies with the calculation of within language neighbours in a second language, as we will discuss later. Vitevitch (2012) addressed this question of how to take into account the differences between two phoneme repertoires when calculating cross-language PND. Vitevitch attempted to calculate the number of cross-language phonological neighbours between English and Spanish in order to understand the relative importance of cross-language neighbours compared to within language neighbours. He first calculated cross-language neighbours based on the International Phonetic Alphabet (IPA) transcription of words of both languages (and using the one phoneme rule; Luce, 1986), and found very few cross-language neighbours. He subsequently tried to take "perceptual assimilation" into account, which he defined as the fact that "phonemic contrasts that exist in a second language are difficult to perceive if they do not exist in the native language"(p.169). Indeed, there is not much overlap between the English and the Spanish sets of vowels (respectively 20 and five different vowels, all five Spanish vowels are included within the 20 English), hence, an adaptation was made to take into account how Spanish speakers are likely to perceive English vowels: based on a study by García Lecumberri & Cenoz Iragui (1997), Vitevitch replaced all English vowels with the Spanish vowel that was perceived as most similar in order to get a more plausible estimate , for Spanish speakers, of the English words that are similar to Spanish words and vice versa, He re-calculated the proportion of Spanish neighbours of English neighbours with this new coding of vowels, and estimated that this proportion was still very small (the proportion increased from about 5% to about 13%). His conclusion was that, since they represent such a small proportion of neighbours, cross-language neighbours should not influence processing to a great extent.

This brings up the question of the best way to calculate not only cross-language PND in bilinguals but also within-language PND. Typically, in monolinguals, the calculation of a word's neighbourhood follows the one phoneme difference rule, based on the IPA transcription of this word. However, phoneme categories are language-specific, that is, some contrasts are distinctive (allowing discrimination between two words) in one language but not in another (this is particularly true for vowels). This is an issue when studying phonological neighbourhood density in a non-native language or between two languages, as the phonological categories for a non-native speaker may not be equal to those of a monolingual. For example, in French (as in many Romance languages), there is no distinction between /i/ and /ɪ/, in such a way that French natives who speak English might have difficulties distinguishing between words like "meat" and "mitt". Here, some clarification of how phonemes are possibly represented for a non-native speaker seems warranted.

In the course of first language acquisition, infants start from being able to discriminate every possible low-level speech property in any language, and then adjust their perceptual system so that they are only able to perceive the phonological properties of their native language. As a consequence, it is well established that monolingual adults have trouble categorizing, discriminating and producing phonemes from other languages when these phonemes are not used to differentiate words of their native language.

The perception of non-native phonemes is shaped by native phonological contrasts, that is, for example, the perception of English phonemes by monolingual French speakers is shaped by French language phonological contrasts. This is not only the case for naïve listeners of the non-native language (as demonstrated by the Perceptual Assimilation Model: Best, 1995), but also for speakers who are learning a second language, be it by formal instruction or by immersion in a second language (L2) environment (as shown by Best and Tyler, 2007; or Fledge, 1995). Categorization and discrimination performance levels in an L2 vary across L2 contrasts and across native languages (L1s). Some learning of L2 contrasts does occur when an individual is exposed to an L2 environment, but it rarely equals native performance and there are different ways an L2 learner can incorporate L2 phonemes in his phonological repertoire. Most of this learning seems to occur within 6-12 months of immersion in an L2 environment (Best & Tyler, 2007). However, even very proficient bilinguals can fail to differentiate between non-native phonological contrasts (Pallier, Colomé & Sebastián-Gallés, 2001). In an auditory lexical decision experiment, with highly proficient speakers of both Spanish and Catalan (living in a bilingual environment – Barcelona, and all having acquired their second language before age 7), these authors found a repetition priming effect between targets that differed by a non-native phonological contrast, just as if these two targets were identical. This finding shows that the representation of phonological categories in a non-native language, even in very highly proficient speakers of that second language, is certainly different from that of a native speaker of that language. Although the evidence reviewed here covers phoneme perception, perceptual skill level has been shown to be positively correlated with accuracy in producing the L2 phonemes (at least for vowels; Best & Tyler, 2007). In other words, representations in the input lexicon are likely to be similar to the representations in the output lexicon.

What are the implications for the calculation of phonological neighbours in non-native speakers, here, in the case of an L2 speaker whose native language is French? While the English language has twenty different vowels, French only has thirteen (plus four nasal vowels). There is clearly no full overlap between vowel categories in these two languages. Take, for example, the English target word “bit” (/bɪt/). As we pointed out earlier, there is only one French /i/ vowel (at least in the European French varieties), therefore French speakers might not have two clear and distinct representations for the English phonemes /ɪ/ (as in “bit”) and /i/ (as in “beat”), and both might be perceived as equivalent (i.e., "bit" and "beat" will be homophones). If that is the case, then for a French native speaker, the English words “bit” and “meat” only differ by one phoneme and, hence, unlike in English monolinguals, count as phonological neighbours. In the same vein, the French word “vite” (/vit/) would count as a cross language neighbour of “bit”, although strictly following the one phoneme difference rule based on IPA, they differ by two phonemes. Identifying the phonemes that are likely not to be fully acquired by French natives who speak English as a second language seems a good starting point in order to provide a more accurate calculation of the phonological neighbourhood of English words for this population.

Iverson and Evans (2007) investigated how native speakers of different languages perceived English vowels’ acoustic, phonetic and phonological properties. In particular, the authors asked French native speakers to rate some aurally presented English vowels. All the participants grew up in France but had started learning English in childhood and were tested in London where they had been living for up to 17 years. They heard the English vowel embedded in a /b/-Vowel-/t/ word, and had to identify which French vowel (also embedded in a /b/-V-/t/ word) was most similar to it. Then, the participants had to rate how similar the French vowel they chose was to the English target vowel. Table 5 displays the English target vowels, the French vowel that was most often chosen as similar, and an average percentage of similarity as shown by the ratings.

Table 5. English vowels, the most similar corresponding French vowel, and the average similarity rating between the English and the French vowels (adapted from Iverson & Evans, 2007).

English vowel	Closest French vowel	Average similarity rating
I	I	0.86
ɪ	I	0.87
eɪ	E	0.70
E	E	0.86
A	A	0.91
aɪ	a	0.57
aʊ	a	0.47
ɑ	ɑ	0.80
ɒ	ɔ	0.83
ɔ	o	0.84
əʊ	o	0.74
ɜ	ø	0.80
ʌ	ø	0.85
U	u	0.82

Iverson and Evans' (2007) findings allow us to identify how French native speakers map the English vowels to their phonological repertoire, namely which English vowels are often mapped on to the same French vowel by French native speakers. Using these results, and building on Vitevitch (2012)'s idea, for our investigation of phonological neighbourhood effects in French-English bilinguals, we developed a new additional "French-speaker" transcription of English targets that took these mappings into account. English vowels that were paired to a different French vowel with a similarity rating of 70% or more, were replaced by that vowel. For instance, IPA transcription for the word "boat" (/bəʊt/) was recoded as /bot/ as French speakers rated the English vowel /əʊ/ 74% similar to the French vowel /o/. This new transcription was used for the calculation of both English and French phonological neighbours of the English targets, in order to take into account the specificities in the French native speakers' phonological repertoire (see below for further details).

In summary, in this study, we investigated the effects of cross-language neighbourhood density and frequency on bilingual spoken word production. To our knowledge this is the first study that attempts to examine neighbourhood in bilingual speakers while taking into account the specificities of late bilinguals' phonological system.

Method

Participants

50 individuals (32 females) with French as a native language and who had lived in in an English-speaking country for more than 2 years were recruited through several networks of French speakers in Sydney (e.g., Alliance Française, French Lycée, and Facebook groups) and were given 15AUD for their participation. All participants grew up in a French monolingual environment, either in France or in the French speaking part of Belgium, and moved to an English speaking country after age 16. All had learned English at school for at least four years before living in an English speaking country, and were living in Sydney at the time of the experiment. They were aged between 18 and 65 (mean age 37 years, *SD* 10.92). The number of years living in Australia or in other English speaking countries ranged between 2 and 42 years (mean 9.92 years, *SD* 8.87). All of the participants had normal or corrected to normal vision and no reported history of learning difficulties or language impairment.

Although all participants acquired English in similar contexts (formal instruction in a non-English-speaking context first, then informal exposure to English by living in an English-speaking country), this group was heterogeneous with regards to English proficiency, daily exposure to English relative to French, language dominance patterns, age and occupation. We believe they represent a good sample of the French native speakers living in the Sydney area. In addition, this sample is similar to the one used by Iverson and Evans (2007) in terms of the type and amount of exposure to English.

In this study, (late) bilingualism is seen as a continuum rather than a category, and we differentiate between bilinguals using the number of years of exposure to English, rather than with self-ratings or formal tests. Our study targets speech production at the single word level, while self-ratings and formal tests take into account many other dimensions of language proficiency that are not relevant here. Instead, the number of years of exposure to English is likely to affect an individual's vocabulary range and how the coexistence of two lexicons affects processing.

Materials

The same materials were used as in Experiment 1.

Predictors

Control predictors

All the non-phonological neighbourhood predictors in Experiment 1 were also used here with the exception of phonotactic probability. Specifically we used: standardized preceding RT, Australian speeded name agreement, visual complexity, familiarity, age of acquisition, target word form and lemma frequencies, word length and imageability. We did not include phonotactic probability as this seems an inappropriate measure for non-native speakers. Indeed, it is linguistic experience that allows a speaker to exploit phonotactic rules that are specific to the vocabulary of a particular language, hence these rules are not likely to be fully internalised by a late bilingual, especially by those speakers with less exposure to the second language.

We also added *s*ome additional measures of specific relevance to bilingual speakers:

Language exposure in years was the number of years the participant had been exposed to English by living in an English speaking country (for the majority of participants, Australia, but also New Zealand, the UK and the USA).

It is common in bilingual studies of picture naming to make sure there are no cognates in the experimental items (e.g., Costa, Miozzo, & Caramazza, 1999; Muñoz & Marquardt, 2003) unless cognates are the focus of the study. Cognates are translation equivalents that are similar in (written and/or phonological) form, like 'kangaroo' (English) and 'kangourou' (French), and they have been shown to influence bilingual language processing (e.g., Costa, Caramazza,& Sebastián-Gallés, 2000; Dijkstra, Grainger, & van Heuven, 1999; Kohnert, 2004; Roberts & Deslauriers, 1999). However, it is unclear how similar the two members of a translation pair should be in order to be considered cognates. When the procedure to define cognates is specified, authors have used subjective ratings (e.g., de Groot, Dannenburg, & van Hell, 1994), or a phonological overlap cut-off (a minimum 70% of features in common for example in Kohnert (2004) and Roberts & Deslauriers, 1999). Other authors have attempted to characterize the form overlap between the two members of the translation pair, using different methods of calculation (see Sadat, Martin, Magnuson, Alario, & Costa, 2015 for a review). Here, following the idea of a continuum between translation pairs, we did not document cognate status nor did we try to exclude these items (there are many cognates between English and French, especially between the written forms of translation equivalents), but instead we used two different measures to characterize the formal overlap between translation pairs: Levensthein edit distance and phonological edit distance, obtained from the Phonological corpus Tools (PCT) online program (Hall et al., 2015).

Levensthein edit distance measures similarity between the written form of (English) target items and their written French translations (mean=0.38, *SD*=0.33). Levensthein edit distance is defined as the minimum number of letter deletions, additions, and substitutions necessary to turn the English target word into its French translation. Higher Levensthein edit distance scores mean more formal overlap between the two members of the translation pair. For instance, the written words ’piano‘ (English) and ’piano‘ (French) are identical and therefore have a Levensthein edit distance score of 1, whereas ’fountain‘ and ’fontaine‘ have a score of 0.75: 2 letter changes are needed to turn ’fountain‘ into ’fontaine‘: delete ’u‘ and add ’e‘, which makes 6 letters identical out of the 8 letters of the English word (6/8=0.75).

Phonological edit distance compares the French translation’s phonology and the “French-speaker” (cf. above) IPA transcription of the English target word (mean= 18.64, *SD*=10.41). This measure is similar to Levensthein Edit distance in that it also calculates the number of one symbol (letter/phoneme) changes, but here, to turn the (English) *phonological* form to the French *phonological* form. Changes are here weighted based on featural similarity. For this measure, *higher* scores indicate *less* overlap between each member of the translation pair. For instance, ’saw‘ (/sɔ/) and ’scie’(/si/) have a phonological edit distance score of 3, which is obtained as follows: the consonant is identical, but as for the vowel, there are three different featural contrasts: Open-mid vs close, back vs front, and rounded vs unrounded, hence the phonological edit distance value of 3. The more featural contrasts between each phoneme of a word and the corresponding phoneme of its translation, the higher is the phonological edit distance.

Predictors of interest

The measures of **phonological neighbourhood density** used in this experiment were different to those in Experiment 1 and specific to our bilinguals. There are three reasons for this: first, as we said earlier, a target word can have neighbours in both the target language (English) and the non-target language (French); second, as noted earlier, even within English, some English phonemes are likely to be assimilated to other phonemes and hence some English words might count as neighbours of an English target for French native speakers while they would not for an English native speaker (e.g. ’bit‘ and ’meat‘ are not ‘English-speaker’ neighbours but are ‘French-speaker' neighbours, by virtue of the assimilation of /ɪ/ and /i/); third, in late bilinguals the limited size of the bilingual's vocabulary in English reduces the number of possible English neighbours for bilinguals compared to monolingual native speakers of English.

Following this reasoning, we first added to the PND of the English word, the neighbours that were formed using the 'French-speaker' transcription of these words. Then, of these potential neighbours, we only kept words that were likely to be present in the majority of our French native speakers' English lexicon. To achieve this, we estimated a frequency cut off as follows: 27000 words is the average number of words known by a native speaker of English aged 30 (levels of vocabulary can go up to 40000). For second language learners after 3-4 years spent in an English speaking country, the number of words is estimated to 13000 words (so 32 to 48% of a native speaker's vocabulary)[9]. Words in the CELEX database most likely reflect the highest level of vocabulary a native speaker of English can achieve. Hence, we reckon that the 32% of words with the highest CELEX frequency are those that are most likely present in most of our bilingual participants' English lexicon. These words correspond to words with a log CELEX combined (sum of spoken and written) word-form frequency greater than 3. French neighbours of the English targets (using both the English target's IPA transcription and the 'French-speaker' transcription) were obtained using Clearpond (Marian et al., 2012)[10]. We defined separate sets of word form neighbours and lemma neighbours, as we did for monolinguals. French word form and lemma frequencies were drawn from the Lexique 2 database (New, Pallier, Brysbaert, & Ferrand, 2004), with only one neighbour per headword selected for the lemma neighbours.

Several measures of **Phonological Neighbourhood Density** were used for bilinguals:

Some PND measures were identical to those used in the monolingual experiment (but with different values, as described above):

- English word form PND (WF.PND)(mean=0.85, *SD*=0.52)
- English lemma PND (L.PND) (mean=0.74, *SD*=0.54)
- English higher frequency word form PND (WF.PND.h) (mean=0.51, *SD*=0.43
- English higher frequency lemma PND (L.PND.h) (mean=0.50, *SD*=0.44)

Other PND measures were specific to bilinguals:

[9] Estimates of vocabulary sizes for both monolinguals and bilinguals were taken from the "Test your vocab" website: http://testyourvocab.com.

[10] In CLEARPOND, French neighbours are drawn from the Lexique 2 database (New, Pallier, Brysbaert, & Ferrand, 2004), using a similar frequency cut-off as for English neighbours. This is to ensure that the French neighbours are likely to be words from an educated French monolingual adult's vocabulary (i.e. excluding rare words that potentially very few participants would know).

- French word form PND (F.WF.PND): number of word form neighbours of the English target in French (mean=0.63, *SD*=0.68)
- French Lemma PND (F.L.PND) number of lemma neighbours of the English target in French (mean=0.53, *SD*=0.59)
- Combined English and French word form PND (C.WF.PND): Total number of word form neighbours of the English target in both English and French (sum of WF.PND and F.WF.PND) (mean=1.02, *SD*=0.65)
- Combined English and French lemma PND (C.L.PND): Total number of lemma neighbours of the English target in both English and French (sum of L.PND and F.L.PND) (mean=0.88, *SD*=0.64)

Similarly, several **Phonological Neighbourhood Frequency** measures were defined:

The same set of PNF measures as used in the monolingual experiment were calculated (with different values that take into account the bilingual coding and filters):

- English word form summed PNF (WF.sPNF) (mean=3.77,*SD*=1.17)
- English Lemma summed PNF (L.sPNF) (mean=4.22, *SD*=1.02)
- English word form average PNF (WF.aPNF) (mean=2.85, *SD*=0.74)
- English Lemma average PNF (L.aPNF) (mean=3.31, *SD*=0.66)
- English higher frequency word form neighbours summed PNF (WF.sPNF.h)(mean=4.03, *SD*=1.02)
- English higher frequency lemma neighbours summed PNF (L.sPNF.h) (mean=3.70, *SD*=0.71)
- English higher frequency word form neighbours average PNF (WF.aPNF.h) (mean=3.43, *SD*=0.76)
- English higher frequency lemma neighbours average PNF (L.aPNF.h) (mean=4.32, *SD*=0.96)

And four different measures of French PNF:

- French word form summed PNF (F.WF.sPNF)(mean=2.74, *SD*=1.18)
- French lemma summed PNF (F.L.sPNF) (mean=2.97, *SD*=1.20)
- French word form average PNF (F.WF.aPNF) (mean=1.66, *SD*=0.76)
- French lemma average PNF (F.L.aPNF) (mean=2.10, *SD*=0.79)

Table 6. Pairwise Pearson correlation coefficients (2-tailed) between the control predictors and the PND/PNF predictors (Bilinguals).

		Name agreement	Levenshtein edit distance	Phonological edit distance	Lemma frequency	Word form frequency	Age of acquisition	Visual complexity	Imageability	Familiarity	Length in phonemes	Phonotactic probability
English PND	WF.PND	.092	-.408**	-.097	.606**	.572**	-.409**	-.039	-.151*	.282**	-.828**	-.340**
	L.PND	.084	-.415**	-.093	.575**	.550**	-.393**	-.038	-.157*	.280**	-.835**	-.346**
	WF.PND.higher	.065	-.334**	-.077	.255**	.188**	-.228**	-.061	-.102	.101	-.715**	-.357**
	L.PND.higher	0.058	-.316**	-.095	.250**	.224**	-.193**	.002	-.102	.091	-.708**	-.350**
English PNF	WF.sPNF	.096	-.412**	-.095	.627**	.589**	-.429**	-.046	-.155*	.300**	-.837**	-.348**
	WF.aPNF	0.103	-.347**	-.052	.632**	.570**	-.440**	-.074	-.081	.302**	-.700**	-.301**
	L.sPNF	.089	-.409**	-.111*	.583**	.556**	-.403**	-.025	-.163*	.283**	-.843**	-.358**
	L.aPNF	.096	-.369**	-.104	.557**	.524**	-.413**	.006	-.144*	.266**	-.787**	-.346**
	WF.sPNF.higher	.079	-.349**	-.095	.330**	.271**	-.280**	-.070	-.083	.143*	-.747**	-.371**
	WF.aPNF.higher	.023	-.087	-.096	.168**	.143*	-.123*	-.056	.054	.041	-.288**	-.156**
	L.sPNF.higher	.080	-.344**	-.096	.333**	.306**	-.252**	-.006	-.096	.141*	-.748**	-.362**
	L.aPNF.higher	.128*	-.356**	-.094	.501**	.476**	-.378**	-.024	-.054	.256**	-.740**	-.335**
Combined PND	Combined.WF.PND	.089	-.399**	-.135*	.588**	.559**	-.389**	-.053	-.177**	.281**	-.828**	-.352**
	Combined.L.PND	.080	-.405**	-.127*	.568**	.545**	-.378**	-.049	-.185**	.275**	-.837**	-.354**
French PND	French.WF.PND	.061	-.335**	-.148**	.440**	.425**	-.269**	-.059	-.194**	.214**	-.673**	-.295**
	French.L.PND	.063	-.331**	-.150**	.437**	.423**	-.264**	-.057	-.196**	.214**	-.670**	-.285**
French PNF	French.WF.sPNF	.078	-.346**	-.140*	.438**	.424**	-.281**	-.061	-.196**	.224**	-.685**	-.299**
	French.WF.aPNF	.101	-.291**	-.123*	.393**	.373**	-.269**	.010	-.176**	.158**	-.631**	-.278**
	French.L.sPNF	.071	-.344**	-.139*	.441**	.429**	-.277**	-.053	-.200**	.226**	-.688**	-.297**
	French.L.aPNF	.062	-.300**	-.110	.402**	.391**	-.262**	.007	-.180**	.172**	-.637**	-.299**

***Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).*

Abbreviations: WF=word form, L=lemma, PND=phonological neighbourhood density, PNF=phonological neighbourhood frequency, higher=of higher frequency than the target, s=summed, a=average, Combined=English and French neighbours summed.

Table 7. Pairwise Pearson correlation coefficients (2-tailed) between the PND/PNF predictors (Bilinguals).

	L.PND	WF.PND higher	L.PND higher	WF sPNF	WF aPNF	L.sPNF	L.aPNF	WF.sPNF. higher	WF.aPNF. higher	L.sPNF higher	L.aPNF higher	Combined WF.PND	Combined .L.PND	French.W F.PND	French.L. PND	F.renchW F.sPNF	French. WF.aPNF	French.L. sPNF	French.L. aPNF
English PND																			
WF.PND	.987**	.841**	.848**	.992**	.812**	.978**	.857**	.868**	.332**	.882**	.832**	.960**	.961**	.750**	.747**	.755**	.674**	.759**	.681**
L.PND		.846**	.860**	.976**	.746**	.990**	.861**	.873**	.331**	.894**	.842**	.952**	.972**	.757**	.755**	.763**	.684**	.766**	.689**
WF.PND.higher			.947**	.843**	.676**	.848**	.738**	.980**	.345**	.939**	.759**	.813**	.824**	.660**	.657**	.675**	.608**	.674**	.601**
L.PND.higher				.846**	.650**	.865**	.759**	.928**	.290**	.986**	.786**	.821**	.838**	.671**	.670**	.683**	.601**	.686**	.606**
English PNF																			
WF.sPNF					.858**	.981**	.887**	.875**	.339**	.886**	.852**	.952**	.950**	.737**	.734**	.743**	.668**	.746**	.675**
WF.aPNF						.774**	.826**	.718**	.307**	.698**	.727**	.771**	.730**	.531**	.525**	.533**	.502**	.534**	.514**
L.sPNF							.909**	.880**	.342**	.906**	.873**	.945**	.964**	.746**	.742**	.752**	.682**	.755**	.687**
L.aPNF								.788**	.350**	.816**	.860**	.829**	.842**	.606**	.597**	.611**	.593**	.612**	.600**
WF.sPNF.higher									.486**	.943**	.827**	.837**	.849**	.664**	.662**	.679**	.618**	.679**	.612**
WF.aPNF.higher										.324**	.387**	.306**	.312**	.197**	.197**	.204**	.189**	.205**	.187**
L.sPNF.higher											.874**	.853**	.871**	.685**	.684**	.698**	.625**	.701**	.629**
L.aPNF.higher												.801**	.820**	.609**	.607**	.620**	.585**	.624**	.587**
Combined PND																			
Combined.WF.PND													.990**	.888**	.883**	.887**	.749**	.887**	.765**
Combined.L.PND														.870**	.869**	.868**	.746**	.871**	.755**
French PND																			
French.WF.PND															.995**	.991**	.761**	.990**	.791**
French.L.PND																.986**	.745**	.990**	.764**
French PNF																			
French.WF.sPNF																	.803**	.996**	.820**
French.WF.aPNF																		.793**	.964**
French.L.sPNF																			.818**

Abbreviations: WF=word form, L=lemma, PND=phonological neighbourhood density, PNF=phonological neighbourhood frequency, higher=of higher frequency than the target, s=summed, a=average, Combined= English and French neighbours summed

As for Experiment 1, all measures were log transformed to avoid the influence of extreme values, and predictors were then centred. Unsurprisingly, once again, most PND/PNF measures were strongly intercorrelated, and correlated with length, as shown in Tables 6 and 7.

Procedure

The same procedure as in Experiment 1 was used for picture naming with our bilingual participants.

Results

Responses were analysed using a similar method to Experiment 1.

Response time analyses

Incorrect responses (see Experiment 1 for definition) were removed from reaction time (RT) analyses, resulting in 12,198 observations for analysis. Data was analysed using R (R Development Core Team, 2011). For this analysis, the optimal data transformation resulting from the BoxCox test (Box & Cox, 1964) was to raise RTs to the power of -.36. Hence, once again, in the transformed data lower values represent longer RTs. Linear mixed effects models were run to assess the contribution of our variables of interest, using the lme4 software package (Bates, Maechler, Bolker, & Walker, 2015). As for Experiment 1, we first defined the best base model, excluding PND/PNF variables, and including by-item and by-participant random intercept, and all control variables (preceding RT, visual complexity, word form or lemma frequency[11], age of acquisition, familiarity, imageability, length in phonemes, Australian speeded name agreement, number of years of exposure to English, Levensthein edit distance, and phonological edit distance).

Next, we removed non-significant fixed factors, which resulted in a base model including preceding RT, lemma frequency, age of acquisition, familiarity, Australian name agreement, English experience, Levensthein edit distance and phonological edit distance as “control” fixed factors. A summary of the full model with all control predictors, as well as the base model with only significant control predictors is shown in Appendix E.

Consistent with previous studies of predictors of picture naming behaviour (e.g. Alario at

[11] Models with lemma and word form frequency were run and compared. Unlike in our monolingual analyses, l emma frequency was chosen here over word form frequency as it showed a significant effect on response time (β=8.60x10^{-4}, *SE*=3.67x10^{-4}, *t*=2.34, *p*= 0.019), whereas word form frequency did not (β=2.53x10^{-4}, *SE*=3.42x10^{-4} , *t*=0.74, *p*=0.46).

al. 2004), words with shorter preceding RTs, higher (lemma) frequency, higher name agreement, higher familiarity, were named faster, and words with later age of acquisition ratings had longer latencies. In line with Sadat et al. (2015) (even though their participants were highly proficient early bilinguals), words with higher written and phonological overlap with their translation equivalent were named faster, and finally, participants with less exposure to English were slower (although the effect was marginal, *p*=0.053).

Table 8. Summary of the fixed effects of each PND/PNF predictor on response time: coefficient estimates (*β*), standard errors (*SE*), *t*- and *p*-values, with the variance inflation factor (VIF).

Fixed effects	*β*	*SE β*	*t*-value	*p*-value	VIF
Phonological neighbourhood density					
English word form: WF.PND	-7.12E-04	5.08E-04	-1.40	0.161	2.0
English lemma: L.PND	-6.16E-04	4.64E-04	-1.33	0.184	1.9
English higher frequency word form: WF.PND.h	-3.77E-04	4.92E-04	-0.77	0.444	1.3
English higher frequency lemma: L.PND.h	-5.80E-04	4.82E-04	-1.20	0.229	1.3
French word form: F.WF.PND	-1.85E-04	3.42E-04	-0.54	0.589	1.5
French lemma: F.L.PND	-2.42E-04	3.85E-04	-0.63	0.530	1.5
Combined word form: C.WF.PND	-4.02E-04	4.09E-04	-0.98	0.326	2.0
Combined lemma: C.L.PND	-4.28E-04	4.01E-04	-1.07	0.286	2.0
Phonological neighbourhood frequency					
English word form, summed: WF.sPNF	-4.85E-04	4.14E-04	-1.17	0.241	2.1
English lemma, summed: L.sPNF	-3.59E-04	3.60E-04	-1.00	0.319	2.0
English higher frequency WF, summed: WF.sPNF	-1.44E-04	3.49E-04	-0.41	0.679	1.4
English higher frequency lemma, summed: L.sPNF	-2.75E-04	3.37E-04	-0.82	0.414	1.4
English word form, average: WF.aPNF	2.54E-05	2.12E-04	0.12	0.904	1.8
English lemma, average: L.aPNF	2.70E-05	2.87E-04	0.09	0.925	1.9
English higher frequency WF, average: WF.aPNF	6.08E-06	5.25E-05	0.12	0.908	1.1
English higher frequency lemma,average: L.aPNF	8.24E-05	1.37E-04	0.60	0.549	1.6
French word form, summed: F.WF.sPNF	-2.05E-04	3.25E-04	-0.63	0.529	1.5
French word form,average F.WF.aPNF	-2.57E-04	3.39E-04	-0.76	0.449	1.4
French lemma, summed: F.L.sPNF	-2.20E-04	3.39E-04	-0.65	0.516	1.5
French lemma,average F.L.aPNF	-8.42E-05	2.66E-04	-0.32	0.751	1.4

Abbreviations: WF=word form, L=lemma, PND=phonological neighbourhood density, PNF=phonological neighbourhood frequency, h=of higher frequency than the target, s=summed, a=average, C=combined (English and French), F=French.

Moreover, some interactions were observed within control variables[12]: Levensthein edit distance and phonological edit distance interacted with language experience. Both measures of formal overlap between the target word and its French translation equivalent showed an overall facilitatory influence on response time[13] but, interestingly, this influence decreased with language exposure (interaction between English exposure and Levensthein edit distance: β=-4.86x10^{-5}, *SE*=1.89x10^{-5}, *t*=-2.57, *p*=0.01, and between English exposure and Phonological edit distance: β= 1.45x10^{-6}, *SE*=6.35x10^{-7}, *t*= 2.29, *p*=0.022). There was a significant interaction between Levensthein edit distance and target word frequency (β=-1.80x10^{-3}, *SE*= 8.52x10^{-4}, *t*= -2.10, *p*= 0.036): the facilitatory influence of Levensthein edit distance on RT was greatest on low frequency targets, and decreased while target frequency increased. Put differently, the facilitative influence of target word frequency on RT was most important for targets with little written overlap with their French translation equivalent and diminished as the overlap increased. The relationship between target frequency and overlap was further specified by a significant three way interaction between target word frequency, English exposure and Levensthein edit distance (β=7.43x10^{-5}, *SE*=2.88x10^{-5}, *t*= 2.58, *p*<0.01). In order to understand this interaction between three continuous variables better, we split the items into three groups of 106 items each, of respectively low, medium and high word form frequency, and examined the interaction between Levensthein edit distance and English exposure for each frequency subgroup: for low frequency targets, the written overlap yielded a facilitatory effect that decreased over the length of English exposure. On medium frequency targets, the facilitatory effect of overlap also decreased with English exposure, but the effect was weaker (lower intercept), whereas on high frequency targets, the interaction between overlap and English exposure was not significant, and the effect was close to zero. The interaction between phonological edit distance and target word frequency was not significant (β=1.43x10^{-5}, *SE*=2.08x10^{-5}, *t*=0.69, *p*= 0.491), nor was the interaction between target word frequency and the number of years of exposure to English significant (β= 3.44x10-6, SE= 9.58x10-6, t=0.36, p=0.720).

Here again, length in phonemes was not significant when included in the model with the other control variables, but when entered as a single predictor of RTs, with no other independent variable, it did show a significant effect, in the expected direction of inhibition: longer words had longer latencies (β= -3.82x10^{-4}, *SE*=1.56x10^{-4}, *t*=-2.44, *p*=0.015). Since length is mostly strongly

---

[12] Interactions were examined by means of plots using the R package Interplot (Solt & Hu, 2015). Some examples of these plots are provided in Figure 1 (earlier).

[13] Recall that, although both variables measure the overlap between the English target word and its French translation equivalent, they work in opposite directions: the higher the Levensthein edit distance value, the larger the overlap, and the lower the phonological edit distance value, the larger the overlap. This explains why the coefficient estimates are of opposite signs for these two measures, although the true effects work in the same direction.

correlated with our predictors of interest, once again, we decided to include it to the model when a significant effect of our predictors of interest was found, aware that, besides possibly acting as a suppressor variable, it might also yield some multicollinearity between predictors.

Table 9. Summary of the significant (or marginally significant) interactions involving PND predictors on response time in bilinguals: Coefficient estimates (*β*), standard errors (*SE*), *t*- and *p*-values, and Akaike Information Criterion (AIC) of the model.

Interactions: PND variables	*β*	*SE β*	*t*-value	*p*-value	AIC
English exposure and...					
English word form: WF.PND	2.80E-05	1.22E-05	2.30	0.021	-88369
English lemma: L.PND	2.63E-05	1.13E-05	2.33	0.020	-88369
English word form, higher frequency: WF.PND.h	3.00E-05	1.48E-05	2.02	0.043	-88368
English lemma: L.PND	4.07E-05	1.46E-05	2.79	0.005	-88372
Combined word form: C.WF.PND	2.29E-05	9.76E-06	2.35	0.019	-88369
Combined lemma: C.L.PND	2.37E-05	9.66E-06	2.45	0.014	-88370
French word form: F.WF.PND	1.77E-05	9.29E-06	1.91	0.057	-88367
French lemma: F.L.PND	2.01E-05	1.05E-05	1.92	0.055	-88367
English exposure, Levensthein distance and...					
English word form: WF.PND	7.01E-05	3.71E-05	1.89	0.058	-88370
English lemma: L.PND	7.03E-05	3.46E-05	2.03	0.042	-88371
English word form, higher frequency: WF.PND.h	7.83E-05	4.65E-05	1.69	0.092	-88371
English lemma, higher frequency: L.PND.h	8.13E-05	4.54E-05	1.79	0.073	-88373
Combined word form: C.WF.PND	7.29E-05	3.14E-05	2.32	0.020	-88372
Combined lemma: C.L.PND	6.89E-05	3.05E-05	2.26	0.024	-88372
French word form: F.WF.PND	9.86E-05	3.37E-05	2.93	0.003	-88375
French lemma: F.L.PND	1.03E-04	3.83E-05	2.69	0.007	-88374
English exposure, target word frequency and...					
French word form: F.WF.PND	-3.30E-05	1.50E-05	-2.19	0.028	-88367
French lemma: F.L.PND	-3.60E-05	1.69E-05	-2.13	0.033	-88367

WF=word form, L=lemma, PND=phonological neighbourhood density, PNF=phonological neighbourhood frequency, h=of higher frequency than the target, s=summed, a=average, C=combined (English and French), F=French.

Each measure of PND/PNF was introduced one at a time into the base model. None of the PND/PNF measures showed a main effect on bilingual latencies. Coefficient estimates for each PND and PNF variable are listed in Table 8. In contrast to this absence of a main effect, there were many significant interactions involving our predictors of interest (presented in Table 9 for PND measures, and Table 10 for PNF measures).

All measures of PND showed an interaction with English exposure: the effect of PND on RT was inhibitory in participants with less exposure to English, but this effect diminished over time or was even slightly facilitatory in the most experienced speakers. This interaction was only marginally significant for French PND (and models including French PND had the highest AIC, which is a sign that these models were not as good a fit to our data).

In addition, all measures of PND showed a significant three way interaction with English exposure and Levensthein edit distance. In order to understand this three way interaction, we split the items in three groups of low (0-0.167, n=111), medium (0.182-0.44, n=98) and high (0.5-1.00) written overlap (Levensthein edit distance). The "high" written overlap group contains several cognates (36 have identical spelling to their French counterpart, even though the phonological form never overlaps completely). We examined the interaction between PND and English exposure on each of these three "overlap" groups. The interaction operates as follows: on words with little overlap, PND did not significantly interact with language exposure, irrespective of language (whether the neighbours were English neighbours, French neighbours or a combination of both), and the effect of PND tended to be inhibitory; on words with medium overlap, the interaction was also not significant and here there was no effect on latencies; finally, and in contrast, on words with high overlap (i.e. cognates), the influence of PND on RT changed from null to facilitatory with increasing exposure to English. Models with this three-way interaction that included French PND were better fits to the data than those with English or combined PND, as shown by their lower Akaike information criterion (AIC).

Moreover, French PND (both word form and lemma) interacted with target word frequency and language experience: there was an *inhibitory* effect of French PND on low frequency targets that was present in participants who had little exposure to English but was no longer significant in participants with extensive exposure to English.

PNF measures relating to English neighbours of higher frequency than the target (WF.sPNF.h, L.sPNF.h, WF.aPNF.h, L.aPNF.h) all showed an interaction with target word frequency: the effect of the frequency of neighbours that were higher in frequency than the target was different depending on the frequency of the target: *inhibitory* for low frequency targets, *facilitatory* effect for those targets that had a particularly high frequency.

Some PNF measures (English lemma and word form summed PNF, English lemma higher frequency summed PNF, French word form and lemma summed PNF, French lemma average PNF) also interacted with English exposure: their influence tended to be *inhibitory* with low English exposure, and evolved with language exposure in the direction of *facilitation*. Most French PNF measures were only marginally significant and, within the interactions, they were not as good predictors as the English measures, as shown by their higher AIC.

Table 10. Summary of the significant (or marginally significant) interactions involving PNF predictors on response time in bilinguals: Coefficient estimates (*β*), standard errors (*SE*), *t*- and *p*-values, and AIC of the model.

Interactions: PNF variables	*β*	*SE β*	*t*-value	*p*-value	AIC
Target word frequency and...					
English WF, higher frequency, summed: WF.sPNF.h	1.18E-03	4.97E-04	2.37	0.018	-88370
English lemma, higher frequency, summed: L.sPNF.h	9.28E-04	4.90E-04	1.89	0.058	-88368
English WF, higher frequency, average: WF.aPNF.h	3.31E-04	1.59E-04	2.08	0.037	-88374
English lemma, higher frequency, average: L.aPNF.h	3.06E-04	1.56E-04	1.96	0.050	-88190
English exposure and...					
English word form, summed: WF.sPNF	2.20E-05	9.73E-06	2.27	0.024	-88369
English lemma, summed: L.sPNF	1.95E-05	8.69E-06	2.24	0.025	-88369
English lemma, higher frequency, summed: L.sPNF.h	2.49E-05	9.82E-06	2.53	0.011	-88370
French word form, summed: F.WF.sPNF	1.58E-05	8.80E-06	1.79	0.073	-88185
French lemma, summed: F.L.sPNF	1.76E-05	9.20E-06	1.91	0.056	-88186
French lemma, average: F.L.aPNF	1.54E-05	7.65E-06	2.01	0.044	-88185
English exposure, target word frequency and...					
French word form, summed: F.WF.sPNF	-3.07E-05	1.42E-05	-2.16	0.030	-88367
French lemma, summed: F.L.sPNF	-3.23E-05	1.47E-05	-2.19	0.029	-88367
English exposure, Levensthein distance and...					
English word form, summed:WF.sPNF	5.24E-05	2.93E-05	1.79	0.074	-88138
English lemma, summed: L.sPNF	4.95E-05	2.63E-05	1.88	0.060	-88137
English lemma, higher frequency, summed: L.sPNF.h	5.48E-05	3.05E-05	1.80	0.073	-88124
French word form, summed: F.WF.sPNF	9.08E-05	3.22E-05	2.82	0.005	-88374
French lemma, summed: F.L.sPNF	8.87E-05	3.33E-05	2.67	0.008	-88374
French word form, average: F.WF.aPNF	9.16E-05	3.32E-05	2.76	0.006	-88374
French lemma, average: F.L.aPNF	7.15E-05	2.48E-05	2.88	0.004	-88375

WF=word form, L=lemma, PND=phonological neighbourhood density, PNF=phonological neighbourhood frequency, h=of higher frequency than the target, s=summed, a=average, C=combined (English and French), F=French.

As for French PND, a significant 3-way interaction was found between the summed frequency of French (lemma or word form) neighbours, target word frequency and language exposure: French summed neighbourhood frequency had an *inhibitory* effect on RTs for low frequency targets, but this effect decreased with English exposure, so that there was no inhibitory influence of French neighbourhood frequency for individuals with longer exposure to English.

Finally, there was a significant 3-way interaction between English exposure, Levensthein distance and all measures of French PNF: once again, the *inhibitory* effect of French PNF that was found on targets with little similarity to their French counterparts was constant over the extent of exposure to English, there was no effect of French PNF on RTs on targets with medium overlap, and French PNF had a progressively more *facilitatory* effect with more exposure to English when the targets were very similar to their French translation. Some English PNF measures showed similar patterns of interaction with English exposure and Levensthein distance, but these interactions were only marginally significant and the interaction models including these English measures fit the data less well than those that included French PNF measures.

Accuracy analyses

Generalised linear mixed effect models were also run for accuracy analyses, using the lme4 software package (Bates et al., 2015). Accuracy was coded as follows: error = 0, correct = 1. Overall accuracy was 77%. For accuracy analyses, we considered the same set of predictors as for RT analyses, minus preceding RT, and name agreement (as name agreement here is simply a measure of mean accuracy per-item). Entering all the predictors caused the model to fail to converge. Hence, we removed: those predictors that did not show significant effects; age of acquisition, which was correlated with familiarity, and was not as meaningful for bilinguals whose age of acquisition of English words is different to the rated English sample; and imageability, to increase power and allow analysis of exactly same items as for monolinguals and the bilingual RT analyses. The remaining predictors in the base model were familiarity, log lemma frequency[14], English exposure, and phonological edit distance. A summary of the model with all control predictors in and of the base model only including significant control predictors is provided in Appendix F. Targets that had higher familiarity ratings, were more frequent and had a more similar phonological form to their French translation equivalent were more accurately named and participants with more extensive exposure to English named pictures more accurately. We note here that, although not significant, the effect of length in phonemes when entered as a single fixed factor, trended in the direction of longer words being less accurately named (β=-0.05, *SE*=0.05, *t*=-0.99, *p*=0.323).

[14] As in the bilingual RT analyses, log lemma frequency was chosen over log word form frequency as the fit of the model was significantly better: β=0.7542, *SE*=0.1154, *t*=6.533, *p*<0.0001 for log lemma frequency, vs *B*=0.5939, *SE*=0.1136, *t*=5.227, *p*<0.0001 for log word form frequency.

Table 11. Summary of the main effects of PND/PNF predictor on accuracy in bilinguals: Coefficient estimates (*β*), standard errors (*SE*), *t*- and *p*-values.

	Models excluding length					Models including length				
	β	*SE β*	*t*-value	*p*-value	VIF	*β*	*SE β*	*t*-value	*p*-value	VIF
Phonological neighbourhood density										
English Word form: WF.PND	**-0.68**	**0.17**	**-3.96**	**<0.001**	**1.6**	**-0.56**	**0.26**	**-2.14**	**0.032**	**3.7**
English Lemma: L.PND	**-0.65**	**0.16**	**-4.14**	**<0.001**	**1.5**	**-0.58**	**0.24**	**-2.4**	**0.017**	**3.6**
English WF, higher frequency: WF.PND.h	**-0.49**	**0.17**	**-2.87**	**0.004**	**1.1**	-0.17	0.24	-0.72	0.474	2.1
English Lemma, higher frequency: L.PND.h	**-0.58**	**0.17**	**-3.42**	**0.001**	**1.1**	-0.35	0.23	-1.49	0.136	2.1
Combined word form: C.WF.PND	**-0.5**	**0.14**	**-3.68**	**<0.001**	**1.5**	*-0.36*	*0.21*	*-1.72*	*0.085*	*3.6*
Combined lemma: C.L.PND	**-0.53**	**0.13**	**-3.99**	**<0.001**	**1.5**	**-0.45**	**0.21**	**-2.17**	**0.030**	**3.6**
French word form: F.WF.PND	**-0.34**	**0.12**	**-2.79**	**0.005**	**1.3**	-0.16	0.15	-1.07	0.283	1.9
French Lemma: F.L.PND	**-0.39**	**0.14**	**-2.85**	**0.004**	**1.3**	-0.19	0.16	-1.16	0.246	1.9
Phonological neighbourhood frequency										
English word form, summed: WF.sPNF	**-0.53**	**0.14**	**-3.78**	**<0.001**	**1.7**	*-0.41*	*0.22*	*-1.85*	*0.064*	*4.0*
English lemma, summed: L.sPNF	**-0.47**	**0.12**	**-3.84**	**<0.001**	**1.5**	*-0.37*	*0.19*	*-1.94*	*0.053*	*3.8*
English WF, higher frequency, summed: WF.sPNF.h	**-0.31**	**0.12**	**-2.54**	**0.011**	**1.1**	-0.04	0.17	-0.24	0.814	2.3
English lemma, higher frequency, summed: L.sPNF.h	**-0.36**	**0.12**	**-3.08**	**0.002**	**1.1**	-0.16	0.17	-0.98	0.329	2.3
English WF, higher frequency, average: WF.aPNF.h	-0.01	0.02	-0.46	0.644	1.0	0.01	0.02	0.3	0.764	1.1
English lemma, higher frequency, average: L.aPNF.h	-0.06	0.05	-1.3	0.195	1.3	0.07	0.06	1.14	0.252	2.3
French word form, summed: F.WF.sPNF	**-0.31**	**0.11**	**-2.69**	**0.007**	**1.3**	-0.13	0.14	-0.9	0.370	1.9
French lemma, summed: F.L.sPNF	**-0.33**	**0.12**	**-2.72**	**0.007**	**1.3**	-0.14	0.15	-0.93	0.352	1.9
French word form, average: F.WF.aPNF	-0.23	0.12	-1.82	0.068	1.2	-0.01	0.15	-0.05	0.959	1.7
French lemma, average: F.L.aPNF	-0.17	0.1	-1.7	0.090	1.2	0.01	0.11	0.12	0.905	1.7

Abbreviations: WF=word form, L=lemma, PND=phonological neighbourhood density, PND=phonological neighbourhood frequency, h=of higher frequency than the target, s=summed, a=average, C= combined French and English, F=French. Significant effects are in bold, marginally significant effects are in italics.

All PND measures showed a significant inhibitory effect on accuracy, but when length was added in the model, the French PND measures, and the higher frequency PND measures were no longer significant (see Table 11). English and combined word form and lemma PND still showed the main inhibitory effect on accuracy when length was included compared to the models excluding length. For PNF measures, all but two (the average frequency of higher frequency neighbours, lemma and word form) showed an inhibitory effect on accuracy when length was not part of the model, although only word form and lemma PNF showed a (marginal) effect after the introduction of length. In all these models, length had an effect in the opposite direction to its zero order correlation with RT, which suggests suppression effects between length and other variables.

Several significant (or close to significant) interactions were also found in the bilingual accuracy analyses[15] (Table 12). First, six different measures of PND interacted with target word frequency, in the same fashion as for RT analyses: words with many phonological neighbours were more likely to be incorrectly named, but it was driven by low frequency targets, with the effect moving towards less inhibition/more facilitation on high frequency targets. The English PND measures were those that showed the effect most strongly (as shown by their respective AICs, and compared to the combined and French PND measures).

Furthermore, the higher frequency PND measures as well as the French PND measures interacted with English exposure: their inhibitory effect on accuracy decreased with more exposure. Comparing the models that investigated the same interaction with different types of PND measure revealed that the measures involving English neighbours had the best predictive power for both the interaction with frequency and the interaction with English exposure.

Finally, a 3 way interaction was observed between English exposure, target word frequency and almost all of the PND measures. This means that for low frequency targets, words with many neighbours had an inhibitory effect on accuracy that decreased with more English exposure. The interaction was no longer significant for targets of medium / high frequency. Once again, English PND measures improved the fit of the model more than French measures.

[15] In some models, we obtained convergence warnings from R, which is a sign that the model may be too specified to fit our data. In these instances, we changed optimizer, and/or dropped one or more predictors (that were not involved in the interaction) until we achieved convergence. This always resulted in very similar coefficient estimates and significance levels for the interaction of interest, while having a model that converged.

Table 12. Summary of the significant (or marginally significant) interactions involving PND/PNF predictors on accuracy in bilinguals: Coefficient estimates (*β*), standard errors (*SE*), *t*- and *p*-values, and model AIC.

	Interaction terms	*β*	*SE β*	*t*-value	*p*-value	AIC
Phonological neighbourhood density	*Target word frequency and…*					
	English Lemma: L.PND	0.42	0.20	2.05	0.041	14172
	English word form, higher frequency : WF.PND.h	0.86	0.30	2.92	0.004	14173
	English lemma, higher frequency : L.PND.h	0.70	0.29	2.42	0.015	14174
	Combined lemma: C.L.PND	0.34	0.17	2.00	0.046	14174
	French word form: F.WF.PND	0.29	0.17	1.67	0.095	14178
	French lemma: F.L.PND	0.33	0.19	1.69	0.090	14178
	English exposure and…					
	English word form, higher frequency : WF.PND.h	0.01	0.01	2.16	0.031	14177
	English lemma, higher frequency : L.PND.h	0.01	0.01	1.70	0.089	14177
	French word form: F.WF.PND	0.01	0.00	1.89	0.059	14178
	French lemma: F.L.PND	0.01	0.00	1.76	0.078	14178
	English exposure and target word frequency and…					
	English word form: WF.PND	-0.03	0.01	-4.49	0.000	14155
	English lemma: L.PND	-0.03	0.01	-4.54	0.000	14153
	Combined word form: C.WF.PND	-0.02	0.01	-4.43	0.000	14154
	Combined lemma: C.L.PND	-0.03	0.01	-4.43	0.000	14153
	French word form: F.WF.PND	-0.03	0.01	-4.34	0.000	14156
	French lemma: F.L.PND	0.01	0.00	1.77	0.076	14158
Phonological neighbourhood frequency	*Target word frequency and…*					
	English lemma, summed: L.sPNF	0.30	0.16	1.89	0.059	14175
	English word form, higher frequency, summed: WF.sPNF.h	0.62	0.19	3.26	0.001	14172
	English lemma, higher frequency, summed: L.sPNF.h	0.52	0.19	2.78	0.005	14174
	English word form, higher frequency, average :WF.aPNF.h	0.19	0.05	3.68	0.000	14169
	English lemma, higher frequency, average: L.aPNF.h	0.16	0.06	2.62	0.009	14174
	French word form, summed: F.WF.sPNF	0.28	0.17	1.71	0.088	14178
	French lemma, summed: F.L.sPNF	0.31	0.17	1.78	0.075	14178
	French lemma, average : F.L.aPNF	0.26	0.15	1.73	0.083	14179
	English exposure and…					
	English WF, higher frequency, summed: WF.sPNF.h	0.01	0.00	2.25	0.024	14177
	French word form, summed: F.WF.sPNF	0.01	0.00	1.86	0.063	14178
	French lemma, summed: F.L.sPNF	0.01	0.00	1.84	0.065	14178
	French word form, average: F.WF.aPNF	0.01	0.00	1.65	0.100	14180
	French lemma, average: F.L.aPNF	0.01	0.00	2.17	0.030	14178
	Target word frequency, English exposure and…					
	English word form, summed: WF.sPNF	-0.02	0.01	-4.35	0.000	14157
	English lemma, summed: L .sPNF	-0.02	0.01	-4.56	0.000	14156
	English lemma, higher frequency, average: L .aPNF.h	-0.01	0.00	-2.64	0.008	14168
	French word form, summed: F.WF.sPNF	-0.02	0.01	-4.27	0.000	14157
	French lemma, summed: F.L.sPNF	-0.02	0.01	-4.25	0.000	14157
	French word form, average: F.WF.aPNF	-0.03	0.01	-4.26	0.000	14161
	French lemma, average: F.L.aPNF	-0.02	0.01	-4.49	0.000	14154

Similar patterns of interaction were observed with PNF. Eight measures of PNF interacted with target word frequency in predicting accuracy, in the same manner: the inhibitory effect of PNF on accuracy was found on low frequency targets, but it diminished and even in some cases moved towards a facilitatory effect with increasing frequency. Among these eight measures, those that produced the best fitting models were related to the frequency of English higher frequency neighbours (WF.sPNF.h, L.sPNF.h, WF.aPNF.h, L.aPNF.h).

In addition, the summed frequency of word-form higher frequency neighbours, as well as all French PNF measures (although the effect was marginal for most French measures) interacted with English exposure, such that there was decreasing inhibition/ increasing facilitation on accuracy from these PNF measures over the time course of exposure to English.

Finally, a 3-way interaction was found between target frequency, English exposure and eight measures of PNF (including the four French measures). In every case, the effect of PNF on accuracy changed from inhibitory to increasingly facilitatory over the course of exposure, but this interaction was only significant with low frequency targets. A comparison of the respective AIC of the models revealed that the "best" PNF measure in this interaction was the average frequency of French lemma neighbours.

Discussion

This experiment examined picture naming in English of French native speakers who had acquired English later in life but had been immersed in an English speaking environment for at least the previous two years. We used a novel measure of phonological neighbourhood in an attempt to capture the specific phonological representations of non-native words in a bilingual's lexicon, and taking into account the range of vocabulary of our sample of late bilinguals.

Unsurprisingly, bilinguals with more extensive exposure to English were more accurate and somewhat faster than less experienced English learners. We also replicated the usual cognate facilitation effect on response time and on accuracy but using more nuanced measures of orthographic and phonological overlap, in the same vein as Sadat et al. (2015): the more similar an English word was to its French counterpart, the faster and the more accurately it was named. Our analyses also make it possible to further specify this effect: the influence of the overlap between the members of the translation pair was mostly present for low frequency targets, and the heterogeneity of our participant sample in terms of second language experience made it possible, too, to show that this effect decreased with English exposure.

Regarding the influence of our neighbourhood predictors, we believe our results can be explained both in terms of processing and of representation of phonological neighbours, and more broadly, of lexical and phonological representations in late bilinguals.

First, we found no main effect of any PND/PNF measure on response time. However, we did find a main inhibitory effect of several English PND and PNF variables on accuracy (with a stronger effect of English compared to French neighbours): words with more English neighbours/neighbours of higher frequency were less accurately named. This effect could be explained by the fact that an English word with many English neighbours is potentially more difficult to acquire because, as we noted earlier, phonological representations in a second language may not be as well defined as in the native language, consequently, it might be more difficult to differentiate a word from its neighbours under these conditions. The direction of the effect is, however, inconsistent with Marian and Blumenfeld (2006) who found a *facilitatory* effect of within language PND on accuracy when the task was in the bilinguals' second language. Their sample consisted, at least for the English native speakers, of proficient late bilinguals, with longer exposure time to the other language compared to our study – these participants had started "learning" their second language 17 years on average prior to the experiment, while our sample had been exposed to the second language for ten years on average. These findings are also inconsistent with Sadat et al. (2015), who found no effect on accuracy of within or cross-language PND in Spanish, in highly proficient Spanish-Catalan bilinguals that had been exposed to their second language very early in their childhood (both when the target language was the participants' first language or when it was their second language). One possibility is that the discrepancy between our results and both Marian and Blumenfeld (2006), and Sadat et al. (2015) might be explained by the fact that our participants included non-native speakers with very little exposure to their second language compared to these two studies. The fact that a further interaction was found between the effect of L2 phonological neighbourhood density on L2 picture naming accuracy and language exposure fits with this interpretation: The inhibitory effect of L2 neighbours was only present for participants with less exposure to English, and evolved towards a null (similar to Sadat et al., 2015), then facilitatory (similar to Marian & Blumenfeld, 2006) effect for those participants whose exposure time was more similar to that of participants of these two studies. With more exposure and better proficiency, it is possible that the lexical / phonological representations in the second language resemble more those of a native speaker of that language, in a way that there is less confusability between similar words of the second language.

A similar inhibitory effect of the number and the frequency of neighbours was found on response time for participants with less exposure to English, that changed to a facilitatory effect with longer exposure. This was mostly the case for English neighbours whose number and frequency slowed down picture naming latency when participants had less exposure to English. In an interactive activation framework whereby activation from lexical forms spreads to the phonological level and back to the target lexical item and its phonological neighbours (see Figure 1, earlier), it might be the case that there are initially weak mappings both from semantics to the lexical nodes and to these non-native phonemes. Therefore more time steps are required for selection. This relatively slow selection results in greater impact of feedback and, hence, more susceptibility to inhibition from neighbours.

A similar effect was also observed with French neighbours of the target (even if it was not as strong as that of English neighbours): the number and frequency of French neighbours interacted with proficiency such that if participants had little exposure to English there were slower latencies the more/higher frequency the French neighbours, but the effect disappeared with more experience in English. This effect could be explained by the same mechanism as the effect of English neighbours, but could also be due to relative differences in activation levels of representations in the two languages. In the dominant language, connections between the different levels of processing are supposed to be stronger compared to in the non-dominant language. Costa and colleagues (2000) argued that the amount of activation of the phoneme nodes level is proportional to the lexical node's activation level, and that this activation level is likely to be higher for representations from a dominant language: they observed a greater cognate facilitation effect in picture naming when the non-response language was the dominant language. It is therefore likely that the links between the lexical level and the phoneme nodes (to use Costa et al.'s terminology) are stronger in the dominant language than in the non-dominant language, and it is then possible that in our participants with less exposure to English (hence most probably dominant in French), French representations are more likely to be strongly activated when naming English words, compared to participants with more extensive experience in English.

There is strong evidence that, in bilingual speech production, activation is language-independent, that is, the non-target language is activated down to the level of phonological representations (e.g., Colomé, 2001; Rodriguez-Fornells et al., 2005). These studies demonstrate that the phonemes from the non-target language are activated via the translation equivalent of the target word: for instance, when naming a picture in one language, activation from semantics flows to the lexical nodes of both the target language and the non-target language, then each of these lexical nodes send activation to their respective sublexical / phoneme nodes. However, it is one step further to suppose activation of phonological neighbours of the other language, because it implies that phoneme nodes from the non-target language can also be activated via a lexical representation from the target language, and not only via activation of its translation equivalent. This would entail that, in bilinguals, activation from a lexical representation that is specific to one language is sent to phoneme nodes in a language non-specific way (which means that the phonological repertoire is shared between languages). Furthermore, these phoneme nodes would send activation back to lexical representations of *both* languages, meaning there would be interactivity between levels of processing, both within and across languages. If connections are particularly strong between the phonemes and the lexical representation of the dominant language, then in the case of our experiment, French neighbours could be serious competitors of the English target, resulting in an inhibitory effect of these French neighbours. But when language dominance is shifted to the target language, less effect of cross-language neighbours might be expected, as we saw in this experiment.

We also found an interaction between phonological neighbourhood frequency and target word frequency on response time, in the same way as for monolinguals: there was more influence of the frequency of neighbours of the same language when the target was low in frequency, that is, the higher the difference in frequency in favour of the neighbours, the slower the response time. The effect was similar with French neighbours, although this effect disappeared with higher time of exposure, i.e., when French was no longer as dominant.

Finally, the effect of phonological neighbourhood density was also modulated by the similarity between the target word and its French translation equivalent: the more distant the English word and its French equivalent, the slower the response time for words with many phonological neighbours, and conversely, the more similar (as in the case of cognates), the faster the response time if these words had many phonological neighbours. This interaction might be explained in similar terms as the one with frequency, as for some authors, “the cognate effect may reflect a word frequency effect in disguise, with cognate words behaving as high-frequency words and non-cognate words as low-frequency words” (Strijkers, Costa, & Thierry, 2010, p.925). These authors used behavioural and electrophysiological evidence to show that, in bilinguals, the effect of word frequency and of cognate status (words with a very high overlap with their translation equivalent) were confounded. In that view, the interaction between translation similarity and phonological neighbourhood density, and between target frequency and phonological neighbourhood density, could be explained by similar mechanisms.

General discussion

We have reported two picture naming experiments in English, one with a population of English monolinguals, the other with a group of late bilinguals with French as a native language and English as a second language, examining the effect on response time and accuracy of several phonological neighbourhood measures. The monolingual group showed no main effect of any measure, but we observed an interaction between the frequency of the target and the number and frequency of those phonological neighbours that were higher in frequency than the target: low frequency words were slower named if they had many or very frequent neighbours, but there was facilitation from these neighbours if the target was higher in frequency. We suggested that the discrepancy between our results and some previous findings in the literature could have been due to our set of items being relatively high in frequency. In the bilingual group, effects of phonological neighbours were also present in the form of interactions. For these bilinguals, as for the monolinguals, if the target was low in frequency, words were slower and less accurately named when they had dense or frequent neighbourhoods. In addition, more neighbourhood effects in this direction were found in participants with little exposure to English, and also on targets that had very little form overlap with their French translation equivalent. For English neighbourhood measures, we argued that those English words that had many English neighbours were harder to retrieve because it is difficult to develop distinctive phonological representations for words that are similar in a non-native phonological system. For French neighbours, we speculated that if the non-target language was dominant (as is the case for the participants with little exposure), then the French representations might be relatively highly activated compared to the English representations, increasing the effects of neighbours from the non-target language.

Our results suggest that, for both monolinguals and our sample of late bilinguals, the influence of phonological neighbours on latencies and accuracy was conditional on how active these neighbours are in the lexicon in comparison with the target word: for example, neighbours have an inhibitory effect on word production if their frequency is substantially higher than that of the target. On the other hand, the number and frequency of neighbours may exert a facilitatory effect if the target is high in frequency.

How could these effects be interpreted within theories of speech production? In the case of monolinguals, it has been argued (e.g., Gordon & Dell, 2001; Vitevitch, 2002) that an influence of phonological neighbourhood density can be best explained by models that postulate cascading and interactivity between the lexical level and the phonological level (e.g., Dell et al., 2007): a facilitatory effect of neighbours could appear in cases when activation of phonemes is stronger because activation originates from both the lexical node of the target and that of its phonological neighbours via feedback mechanisms. Hence, the more neighbours, the more extra activation is sent to target phonemes. Conversely, in a theory where there is competitive lexical selection (e.g., Howard, Nickels, Coltheart & Cole-Virtue, 2006) because of this feedback from phonemes, the high level of activation of the target's neighbours, could hinder selection of the target lexical form, resulting in more inhibition the more neighbours there are. Neighbourhood frequency effects in speech production have not received as much attention as neighbourhood density. In Dell et al.'s model, the frequency of a given word could be implemented in the strength, as a product of learning, of the connections between levels of representations. If high frequency neighbours are therefore strongly connected to their phonemes, their activation at the lexical level will be high as a result of feedback mechanisms. In this view, the frequency of neighbours could hinder word selection because the activation levels of these neighbours would potentially be closer to that of the target, and particularly if the target is much lower in frequency than its neighbours. If the target is particularly high in frequency, it is possible that activation of neighbours would not be sufficiently high compared to the target to create inhibition, but instead, may facilitate the production of the target because of the supplementary activation of common phonemes.

In bilinguals, interaction of phonological neighbourhood variables with frequency within the target language could be explained similarly, and the influence of two factors could be thought of as similar to frequency effects. First, low proficiency (as operationalized in our study by recent exposure to the second language) could be seen as a frequency effect, because items in a non-proficient language have been used much less often. In that case, words from the native / dominant / more proficient language are overall higher in frequency than items from the second language. Second, if cognate status is confounded with high frequency status, then items in which the form is closer to their translation equivalent in the native language may also 'inherit' some of their frequency benefits (cf, accounts of frequency inheritance in homophones, e.g., Dell, 1990). As hypothesised above, inhibitory effects of phonological neighbours would be more likely to occur if the frequency of the target is significantly lower than that of its phonological neighbours. This frequency difference can be due to actual frequency differences within the language, to a low level of overlap between the target word and its translation equivalent, or to the overall status of the target language in comparison to the non-target language, or a combination of all of these. In contrast, facilitation effects can be caused from activation of phonological neighbours when these neighbours are not substantially higher in frequency than the target and contribute to a stronger activation of the target's phonemes but do not provide substantial lexical competition.

Finally, to explain the inhibitory effects of neighbours from the target language for bilinguals, we raised the possibility that mappings of phonological representations in a less proficient language may not be fully specified, leading in confusability between a word and its phonological neighbours. Within an interactive model, this could be implemented by adding some noise between the lexical level and the phoneme nodes level in the case of less proficient speakers. In that case it would require more time steps to select the relevant lexical representation, and these words are more likely to result in an error compared to those with few phonological neighbours.

Conclusion

In this study we investigated the influence of phonological neighbourhood density measures in English picture naming for monolingual English speakers and late bilinguals. In order to capture better the set of words that are similar to a given word for this population of bilinguals, we defined a novel measure of phonological neighbourhood density and frequency that aimed to take into account phonological representations and vocabulary range. In both groups, we identified a critical interaction between phonological neighbourhood measures and the frequency of the target word, with an effect of these neighbours in the direction of inhibition when the target was significantly lower in frequency than its neighbours, while the effect was facilitatory when both the target word and its neighbours were high in frequency (and the difference in frequency was then not large enough to create competition). In our late bilinguals, phonological neighbourhood measures (including those of the non-target language) also interacted with the length of exposure to English and the overlap between the target and its translation equivalent in the same fashion. We argued that both language exposure and translation overlap could be assimilated to frequency effects. We suggest that our results can be best accommodated within interactive activation models with interactivity between the lexical level and the phoneme level, and that in the bilingual case activation and interactivity exist within and across languages. While we have attempted to provide plausible accounts of the range of neighbourhood effects found in this study, clearly these accounts are speculative. Only full computational implementation would reveal whether it is possible to achieve a set of parameters that successfully enables simulation of all of these effects in the ways we have claimed. Finally, our results underline the dynamic nature of the bilingual lexicon, with neighbourhood effects that evolve with increasing proficiency.

References

Alario, F-X., Ferrand, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers, 31*(3), 531-552.

Alario, F. X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers*, *36*(1), 140-155.

Allison, P. (2010, September 10). When Can You Safely Ignore Multicollinearity? [Web log post]. Retrieved from http://statisticalhorizons.com/multicollinearity

Arnold, H. S., Conture, E. G., & Ohde, R. N. (2005). Phonological neighborhood density in the picture naming of young children who stutter: Preliminary study. *Journal of Fluency Disorders*, *30*(2), 125-148.

Baayen, R. H. & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3*(2), 12-28.

Baayen, R, R Piepenbrock, and L Gulikers. CELEX2 LDC96L14. Web Download. Philadelphia: Linguistic Data Consortium, 1995.

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *Quarterly Journal of Experimental Psychology,50A*, 560–585.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01.

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

Bernstein Ratner, N., Newman, R., & Strekas, A. (2009). Effects of word frequency and phonological neighborhood characteristics on confrontation naming in children who stutter and normally fluent peers. *Journal of Fluency Disorders*, *34*(4), 225-241.

Box, G.E.P., & Cox, D.R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological). 26*(2), 211-252.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), *Development, writing and other language processes, Vol. 2*, London: Academic Press.

Chan, K. Y., & Vitevitch, M. S. (2010). Network structure influences speech production. *Cognitive Science*, *34*(4), 685-697.

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

Coady, J. A., & Aslin, R. N. (2003). Phonological neighbourhoods in the developing lexicon. *Journal of Child Language*, *30*(02), 441-469.

Colomé, Á. (2001). Lexical activation in bilinguals' speech production: Language-specific or language-independent? *Journal of Memory and Language, 45,* 721–36.

Colomé, Á., & Miozzo, M. (2010). Which words are activated during bilingual word production? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 96–109.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33A*, 497-505.

Costa, A., Caramazza, A. & Sebastián-Gallés, N. (2000). The cognate facilitation effect: implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26*(5), 1283-1296.

Costa, A., Miozzo, M, & Caramazza, A. (1999). Lexical Selection in Bilinguals: Do Words in the Bilingual's Two Lexicons Compete for Selection? *Journal of Memory and Language, 41,* 365–397.

Costa, A., Roelstraete, B., & Hartsuiker, R. J. (2006). The lexical bias effect in bilingual speech production: Evidence for feedback between lexical and sublexical levels across languages. *Psychonomic Bulletin & Review*, *13*(6), 972-977.

de Groot, A. M., Borgwaldt, S., Bos, M., & Van den Eijnden, E. (2002). Lexical decision and word naming in bilinguals: Language effects and task effects. *Journal of Memory and Language*, *47*(1), 91-124.

de Groot, A.M.B., Dannenburg, L., & van Hell, J.G. (1994). Forward and Backward Word Translation by Bilinguals. *Journal of Memory and Language, 33*, 600-629.

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, *27*(2), 124-142.

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, *5*(4), 313-349.

Dell, G.S., & Gordon, J.K. (2003). Neighbors in the lexicon: Friends or foes? In N.O. Schiller and A.S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities*. New York: Mouton de Gruyter.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104*(4), 801.

Dijkstra, T., Grainger, J., & van Heuven, W.J.B. (1999). Recognition of cognates and interlingual homographs: the neglected role of phonology. *Journal of Memory and Language, 41*, 496-518.

Ellis, A. W., &Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24*, 515–523.

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

Gordon, J. K., & Dell, G. S. (2001). Phonological neighborhood effects: Evidence from aphasia and connectionist modeling. *Brain and Language, 79*(1), 21-23.

Gordon, J. K., & Kurczek, J. C. (2014). The ageing neighbourhood: phonological density in naming. *Language, Cognition and Neuroscience*, *29*(3), 326-344.

Grainger, J., O'regan, J. K., Jacobs, A. M., & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Perception & Psychophysics, 45*(3), 189-195.

Hall, K.C., Allen, B., Fry, M., Mackie, S., & McAuliffe, M. (2015). Phonological Corpus Tools, Version 1.1.0 [Computer Program]. Available from https://github.com/PhonologicalCorpusTools/CorpusTools/releases

Hauk, O., Davis, M. H., Ford, M., Pulvermüller, F., & Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage*, *30*(4), 1383-1400.

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, *100*(3), 464-482.

Hutcheson, G., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models*. London: Sage Publications.

Jescheniak, J. D., & Levelt, W. J. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 824.

Johnston, R. A., Dent, K., Humphreys, G. W., & Barry, C. (2010). British English norms and naming times for a set of 539 pictures: The role of age of acquisition. *Behavior Research Methods, 42*, 461-469.

Kohnert, K. (2004). Cognitive and cognate-based treatments for bilingual aphasia: A case study. *Brain and Language, 91*, 294–302.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1-38.

Luce, P. A. (1986). Neighborhoods of Words in the Mental Lexicon. Research on Speech Perception. Technical Report No. 6.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*(1), 1.

Luce, P.D., Pisoni, D.B., & Goldinger, S.D. (1990). Similarity neighborhoods of spoken words. In G. Altmann (ed.), *Cognitive models of speech perception: Psycholinguistic and computational perspectives.* (pp. 122-147).Cambridge, MA: MIT Press.

Marian, V., Bartolotti, J., Chabal, S., Shook, A. (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE 7*(8): e43230. doi:10.1371/journal.pone.0043230.

Marian, V., & Blumenfeld, H. (2006). Phonological neighborhood density guides lexical access in native and non-native language production. *Journal of Social and Ecological Boundaries, 2,* 3–35.

Mirman, D. (2014, February 11). Three ways to get parameter-specific p-values from lmer. [Web log post]. Retrieved from http://mindingthebrain.blogspot.com.au.

Mirman, D., & Graziano, K. M. (2013). The neural basis of inhibitory effects of semantic and phonological neighbors in spoken word production. *Journal of Cognitive Neuroscience*, *25*(9), 1504-1516.

Mirman, D., Kittredge, A. K., & Dell, G. S. (2010). Effects of near and distant phonological neighbors on picture naming. In *Proceedings of the 32nd annual conference of the Cognitive Science Society* (pp. 1447-1452). Austin, TX: Cognitive Science Society.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*(5), 375.

Mulatti, C., Reynolds, M. G., & Besner, D. (2006). Neighborhood effects in reading aloud: New findings and new challenges for computational models. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(4), 799.

Muñoz, M., & Marquardt, T. (2003). Picture naming and identification in bilingual speakers of Spanish and English with and without aphasia. *Aphasiology, 17*(12), 1115-1132, doi: 10.1080/02687030344000427

New, B., Pallier, C., Brysbaert, M., Ferrand, L. (2004) Lexique 2: A New French Lexical Database. *Behavior Research Methods, Instruments, & Computers*, *36* (3), 516-524.

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. *Journal of Speech, Language, and Hearing Research*, *50*(1), 196-213.

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. *Language and Speech*, *45*(3), 285-317.

Newman, R. S., & German, D. J. (2005). Life span effects of lexical factors on oral naming. *Language and Speech*, *48*(2), 123-156.

Pérez, M. A. (2007). Age of acquisition persists as the main factor in picture naming when cumulative word frequency and frequency trajectory are controlled. *The Quarterly Journal of Experimental Psychology*, *60*(1), 32-42.

Protopapas, A. (2007). CheckVocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods, 39,* 859‑862. doi:10.3758/BF03192979

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, URL http://www.R-project.org.

Rabovsky, M., Schad, D. J., & Rahman, R. A. (2016). Language production is facilitated by semantic richness but inhibited by semantic density: Evidence from picture naming. *Cognition*, *146*, 240-244.

Roberts, P. M., & Deslauriers, L. (1999). Picture naming of cognate and non-cognate nouns in bilingual aphasia. *Journal of Communication Disorders, 32,* 1–23.

Rodriguez-Fornells, A., Van Der Lugt, A., Rotte, M., Britti, B., Heinze, H. J., & Münte, T. F. (2005). Second language interferes with word production in fluent bilinguals: brain potential and functional imaging evidence. *Journal of Cognitive Neuroscience*, *17*(3), 422-433.

Sadat, J., Martin, C.D., Costa, A., Alario, F-X. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology, 68,* 33-58.

Sadat, J., Martin, C. D., Magnuson, J. S., Alario, F. X., & Costa, A. (2015). Breaking down the bilingual cost in speech production. *Cognitive Science*. doi: 10.1111/cogs.12315

Schielzeth, H. (2010), Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution, 1*, 103–113. doi:10.1111/j.2041-210X.2010.00012.x

Severens, E., Van Lommel, S., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica (Amst)*, *119*(2), 159–187. doi:10.1016/j.actpsy.2005.01.002

Solt , F., & Hu, Y. (2015). Interplot: Plot the Effects of Variables in Interaction Terms. available at The Comprehensive R Archive Network (CRAN). URL http://CRAN.R-project.org/package=interplot.

Strijkers, K., Costa, A., & Thierry, G. (2009). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex, 20,* 912–928.

Székely, A., & Bates, E. (2000). Objective visual complexity as a variable in studies of picture naming. *Center for Research in Language Newsletter, 12*(2), 1-33

Székely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., … Bates E. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language, 51,* 247–250.

Tabachnick, B. G., and Fidell, L. S. (2013). *Using multivariate statistics, 6th ed.* Boston: Pearson.

Tabak, W., Schreuder, R., & Baayen, R. H. (2010). Producing inflected verbs: A picture naming study. *The Mental Lexicon*, *5*(1), 22-46.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology Section A*, *57*(4), 745-765.

Test your vocab (2013). Retrieved July 7, 2016, from http://www.testyourvocab.com.

Venables, W.N. & Ripley, B.D.(2002). Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 735.

Vitevitch, M. S., Armbrüster, J., & Chu, S. (2004). Sublexical and lexical representations in speech production: effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 514.

Vitevitch, M.S. & Luce, P.A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, and Computers, 36,* 481-487.

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

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

Wurm, L. H., & Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does *not* do). *Journal of Memory and Language, 72,* 37-48.

APPENDIX A

Experiment 1: Base models to which predictors of interest were added one at a time: Monolingual response time base model

Summary of the base linear mixed effect model on monolingual response time: variance and standard deviation (*SD*) for the random effects, and coefficient estimates (*β*), standard errors (*SE*), *t*- and *p*-values, with the variance inflation factor (VIF) for each fixed effect.

- For the model including all control predictors:

Random effects	Variance	SD
item (intercept)	8480	92.09
participant (intercept)	18177	134.82
Residual	45229	212.67

Fixed effects	*β*	*SE β*	*t*-value	*p*-value	VIF
Intercept	1555.782	52.591	29.582	<0.001	
preceding RT	9.171	2.335	3.927	<0.001	1.0
Name agreement	-7.092	0.541	-13.103	<0.001	1.2
Familiarity	-19.471	7.655	-2.544	0.012	1.8
Word Form frequency	39.333	21.074	1.866	0.063	4.3
Age of acquisition	37.933	12.585	3.014	0.003	2.3
Visual complexity	65.517	30.127	2.175	0.031	1.1
Imageability	389.931	257.332	1.515	0.131	1.1
Length in phonemes	-3.412	5.158	-0.662	0.509	1.4
Lemma frequency	-15.231	23.502	-0.648	0.518	4.7
Phonotactic probability	-7.860	48.822	-0.161	0.872	1.0

- For the base model including only significant control predictors:

Random effects	Variance	SD
item (intercept)	$1.13x10^{-6}$	<0.01
participant (intercept)	$2.40x10^{-6}$	<0.01
residual	$4.22x10^{-6}$	<0.01

Fixed effects	*β*	*SE β*	*t*-value	*p*-value	VIF
intercept	$1.11x10^{-2}$	$5.22x10^{-4}$	0.21	<0.001	
preceding RT	$-1.37x10^{-4}$	$1.99x10^{-5}$	-6.88	<0.001	1.0
Name agreement	$7.10x10^{-5}$	$5.19x10^{-6}$	13.68	<0.001	1.2
Familiarity	$3.14x10^{-4}$	$7.40x10^{-5}$	4.24	<0.001	1.8
Word Form frequency	$-3.69x10^{-4}$	$1.11x10^{-4}$	-3.33	<0.001	1.5
Age of acquisition	$-3.38x10^{-4}$	$1.17x10^{-4}$	-2.89	0.004	2.2
Visual complexity	$-6.25x10^{-4}$	$2.98x10^{-4}$	-2.10	0.036	1.0

APPENDIX B

Experiment 1: interactions involving PND/PNF variables and frequency in monolingual response time

Coefficients for each interaction involving PND/PNF measures and frequency in the monolingual response time analyses: coefficient estimates (*β*), standard errors (*SE*), *t*- and *p*-values, with the variance inflation factor (VIF).

Interactions with word form frequency and:	*β*	*SE β*	*t*-value	*p*-value	VIF
Phonological Neighbourhood Density					
Word form: WF.PND	1.34×10^{-4}	1.71×10^{-4}	0.787	0.432	1.1
Lemma: L.PND	1.37×10^{-4}	1.59×10^{-4}	0.863	0.389	1.1
Word-form higher frequency: WF.PND.h	$\mathbf{5.22 \times 10^{-4}}$	$\mathbf{2.61 \times 10^{-4}}$	**2.003**	**0.046**	**1.2**
Lemma higher frequency: L.PND.h	3.09×10^{-4}	2.11×10^{-4}	1.460	0.145	1.1
Phonological Neighbourhood Frequency					
Word form summed: WF.s.PNF	1.50×10^{-4}	1.47×10^{-4}	1.026	0.306	1.1
Word form average: WF.aPNF	1.34×10^{-4}	1.71×10^{-4}	0.787	0.432	1.8
Lemma summed: L.sPNF	1.01×10^{-4}	1.35×10^{-4}	0.744	0.457	1.1
Lemma average: L.aPNF	2.20×10^{-5}	2.99×10^{-4}	0.073	0.942	1.1
WF higher frequency summed: WF.sPNF.h	2.86×10^{-4}	1.65×10^{-4}	*1.737*	*0.083*	*1.2*
WF higher frequency average: WF.aPNF.h	3.65×10^{-4}	2.73×10^{-4}	1.333	0.183	1.1
Lemma higher frequency summed: L.sPNF.h	2.12×10^{-4}	1.47×10^{-4}	1.442	0.150	1.1
Lemma higher frequency average: L.aPNF.h	4.83×10^{-4}	3.05×10^{-4}	1.583	0.114	1.1

Significant interaction is in bold, marginally significant interaction is in italics.

APPENDIX C

Experiment 1: Base models to which predictors of interest were added one at a time: Monolingual accuracy base model

Summary of the base linear mixed effect model on monolingual accuracy: variance and standard deviation (*SD*) for the random effects, and coefficient estimates (*β*), standard errors (*SE*), *z*- and *p*-values, for each fixed effect:

- For the model including all control predictors:

Random effects		Variance		*SD*
Item (intercept)		1.65		1.29
Participant (intercept)		0.36		0.60
Fixed effects	β	SE β	z-value	p-value
Intercept	2.71	0.14	19.44	<0.001
Visual complexity	0.37	0.44	0.85	0.398
Word form frequency	-0.10	0.29	-0.34	0.734
Lemma frequency	-0.07	0.33	-0.22	0.827
Familiarity	-0.05	0.11	-0.43	0.668
Imageability	5.88	3.61	1.63	0.103
Length in phonemes	-0.01	0.07	-0.20	0.844
Phonotactic probability	0.71	0.75	0.96	0.339
Age of acquisition	-0.83	0.17	-4.80	<0.001

- For the base model including only significant control predictors:

Random effects		Variance		*SD*
Item (intercept)		1.7591		1.3263
Participant (intercept)		0.3783		0.6151
Fixed effects	*β*	*SE β*	*t*-value	*p*-value
Intercept	2.7527	0.1316	20.9230	<0.001
Age of acquisition	0.7471	0.1052	-7.1010	<0.001

APPENDIX D

Experiment 1: interactions involving PND/PNF variables and frequency in monolingual accuracy

Coefficients for each interaction involving PND/PNF measures and frequency in the monolingual accuracy analyses: coefficient estimates (*β*), standard errors (*SE*), *t*- and *p*-values, with the Akaike Information Criterion (AIC) for each model.

Interaction with target frequency and:	*β*	*SE β*	*z*-value	*p*-value	AIC
Phonological neighbourhood density					
Word form: WF.PND	0.35	0.22	1.56	0.119	7870.61
Lemma: L.PND	0.31	0.21	1.51	0.132	7869.67
Word-form higher frequency: WF.PND.h	**0.81**	**0.34**	**2.38**	**0.017**	7865.50
Lemma higher frequency: L.PND.h	*0.47*	*0.27*	*1.74*	*0.081*	7869.20
Phonological neighbourhood frequency					
Word form summed: WF.sPNF	0.31	0.19	1.64	0.101	7869.81
Word form average: WF.aPNF	0.35	0.22	1.57	0.116	7869.72
Lemma summed: L.sPNF	**0.37**	**0.17**	**2.11**	**0.035**	7868.70
Lemma average: L.aPNF	0.40	0.39	1.01	0.310	7870.84
WF higher frequency summed: WF.sPNF.h	**0.52**	**0.22**	**2.40**	**0.017**	7865.90
WF higher frequency average: WF.aPNF.h	**0.87**	**0.35**	**2.47**	**0.013**	7867.10
Lemma higher frequency summed: L.sPNF.h	**0.39**	**0.19**	**2.06**	**0.040**	7867.60
Lemma higher frequency average: L.aPNF.h	**0.96**	**0.40**	**2.40**	**0.016**	7866.80

Significant interactions are in bold, marginally significant interactions are in italics

APPENDIX E

Experiment 2: Bilingual response time base model

Summary of the base linear mixed effect model on bilinguals' response time: Variance and Standard Deviation (*SD*) for the random effects, and coefficient estimates (*β*), standard errors (*SE*), *t*- and *p*-values, with the variance inflation factor (VIF) for each fixed effect.

- Model including all control predictors

Random effects		Variance			*SD*
item (intercept)		9.94×10^{-6}			3.15×10^{-3}
participant (intercept)		9.48×10^{-6}			3.08×10^{-3}
Residual		3.89×10^{-5}			6.24×10^{-3}
Fixed effects	*β*	*SE β*	*t*-value	*p*-value	vif
(Intercept)	6.91×10^{-2}	1.67×10^{-3}	41.45	<0.001	
Preceding RT	-3.78×10^{-4}	6.71×10^{-5}	-5.63	<0.001	1.0
Lemma frequency	2.20×10^{-3}	7.93×10^{-4}	2.78	0.006	4.8
Word form frequency	-1.41×10^{-3}	7.12×10^{-4}	-1.99	0.048	4.4
Exposure to English	9.51×10^{-5}	4.96×10^{-5}	1.92	0.061	1.0
Levensthein edit distance	2.14×10^{-3}	9.63×10^{-4}	2.23	0.027	2.0
Phonological edit distance	-5.98×10^{-5}	2.78×10^{-5}	-2.15	0.032	1.8
Visual Complexity	-1.93×10^{-3}	1.01×10^{-3}	-1.90	0.058	1.1
Age of acquisition	-1.32×10^{-3}	4.24×10^{-4}	-3.12	0.002	2.3
Imageability	9.47×10^{-3}	8.71×10^{-3}	1.09	0.278	1.1
Familiarity	6.68×10^{-4}	2.61×10^{-4}	2.56	0.011	1.8
Length in phonemes	1.08×10^{-4}	1.96×10^{-4}	0.55	0.583	1.8
Name agreement	1.38×10^{-4}	1.82×10^{-5}	7.59	<0.001	1.2

- Base model including only significant control predictors

Random effects		Variance			*SD*
Item (intercept)		9.91×10^{-6}			0.003148
Participant (intercept)		9.62×10^{-6}			0.003101
Residual		3.86×10^{-5}			0.006215
Fixed effects	*β*	*SE β*	*t*-value	*p*-value	VIF
(Intercept)	7.02×10^{-2}	1.43×10^{-3}	49.19	<.001	
Preceding RT	-3.63×10^{-4}	5.97×10^{-5}	-6.08	<.001	1.0
Lemma frequency	8.60×10^{-4}	3.67×10^{-4}	2.35	.019	1.8
Exposure to English	9.65×10^{-5}	4.98×10^{-5}	1.94	.053	1.0
Levensthein edit distance	2.38×10^{-3}	7.67×10^{-4}	3.10	.002	1.8
Phonological edit distance	-5.44×10^{-5}	2.34×10^{-5}	-2.32	.020	1.7
Age of acquisition	-1.05×10^{-3}	3.59×10^{-4}	-2.94	.003	2.3
Familiarity	8.05×10^{-4}	2.20×10^{-4}	3.66	<.001	1.8
Name agreement	1.27×10^{-4}	1.53×10^{-5}	8.28	<.001	1.2

APPENDIX F

Experiment 2: Bilingual accuracy base model

Summary of the base generalized linear mixed effect model on accuracy in bilinguals: Variance and Standard Deviation (SD) for the random effects, and coefficient estimated (β), standard errors (SE), z- and p-values, for the fixed effects.

- Base model with all control predictors

Random effects		Variance		*SD*
Item (intercept)		1.1465		1.0707
Participant (intercept)		0.2857		0.5345

Fixed effects	β	SE *β*	*z*-value	*p*-value
Intercept	1.64	0.11	15.07	<0.001
Visual Complexity	0.03	0.36	0.09	0.927
Familiarity	0.17	0.09	1.85	0.064
Word form frequency	-0.20	0.24	-0.81	0.417
Lemma frequency	0.63	0.27	2.36	0.018
Age of acquisition	-0.74	0.14	-5.25	<0.001
Imageability	9.84	3.10	3.17	0.002
Length in phonemes	0.15	0.08	1.94	0.053
Phonological edit distance	-0.02	0.01	-2.04	0.042
Levensthein edit distance	0.20	0.34	0.60	0.548
Phonotactic probability	0.29	0.29	1.00	0.317
English exposure	0.04	0.01	3.93	<0.001

- Base model including only significant control predictors

Random effects		Variance		*SD*
Item (intercept)		1.4932		1.2219
Participant (intercept)		0.2995		0.5473

Fixed effects	*β*	*SE β*	*z*-value	*p*-value
Intercept	1.70	0.11	15.81	< 0.001
Familiarity	0.35	0.08	4.60	<0.001
Lemma frequency	0.41	0.13	3.16	0.001
Phonological edit distance	-0.03	0.01	-4.17	<0.001
English exposure	0.03	0.01	3.82	<0.001

CHAPTER 5

General Discussion

The research presented in this thesis aimed to investigate what constitutes a word's meaning and form neighbourhood in the lexicon, and whether and how this neighbourhood might affect the production of the intended target word.

Most theories of speech production postulate that in the process of producing words, representations that are similar to the target word with respect to its meaning and, for some theories, in terms of form, also become active. However, in previous speech production research, findings are still inconclusive with regards to the influence of the number of these neighbours, with data supporting the full range of possibilities - no effect, inhibitory and facilitatory effects. In this thesis, I investigated the influence of two broad types of neighbourhoods, semantic and phonological, on speech production. I compared several measures of semantic neighbourhood density, frequency, and similarity, and observed their influence on the picture naming performance of English monolinguals and individuals with aphasia (Chapter 2 – Paper 1), and on a facilitated naming paradigm in a case study with two participants with aphasia (Chapter 3 – Paper 2). In Chapter 4 (Paper 3), I examined the effects of different types of phonological neighbourhood density and frequency measures on the English picture naming performance of English monolinguals, and of French-English late bilinguals (using a novel phonological neighbourhood density metric adapted to these bilingual speakers).

Paper 1 is the first report in the literature to compare the different measures used to account for the number of semantically related words in the lexicon within the same study, and to assess the respective influence of these conceptually different measures on picture naming performance. Moreover, compared to other studies investigating the influence of semantic neighbourhood density in individuals with aphasia, this study involved a larger number of participants, allowing comparison of the effects across participants with different impairment levels.

In Paper 2, I once again distinguished between several measures of semantic neighbourhood density, but on top of those used in Paper 1, added measures that aim to take into account the similarity, frequency and strength of association of semantic neighbours relative to the target. This time, I examined how these variables interacted with the outcomes of a task commonly used in aphasia treatment (repetition in the presence of a picture). This study is one of very few studies that have looked at whether item-related properties could potentially influence the outcomes of aphasia therapy, and the first to look at semantic neighbourhood factors.

Finally, Paper 3 features an investigation of the effects of the number and frequency of phonological neighbours on the picture naming performance of English monolinguals and of French-English late bilinguals. This study uses a regression-based methodology involving naming of a large number of items, allowing better control of confounding factors compared to factorial designs that have been used in the vast majority of phonological neighbourhood studies. It used statistical analysis that took into account specific item- and participant- related variation. This is the first study to use such methods to assess the influence of phonological neighbourhood *frequency* in picture naming. In addition, the bilingual portion of the paper is the first to investigate the influence of cross-language phonological neighbourhood factors using a novel measure adapted to the phonology of this particular bilingual population. It is also the first study to reveal a critical interaction between phonological neighbourhood factors and the target word's frequency.

In the remainder of this chapter, I first summarise the key findings of each chapter and then discuss some of the methodological choices I made, namely comparing between several measures in each chapter, using picture naming tasks, and choosing continuous designs as opposed to factorial designs. Then, some limitations and future directions will be proposed.

Chapter 2: Paper 1: Effects of semantic neighbourhood density on unimpaired and aphasic spoken word production

In this chapter, three experiments were conducted in order to investigate the influence of semantic neighbourhood density on the picture naming performance of unimpaired and aphasic monolingual speakers of English.

In Experiment 1, six different ways were presented that had been previously used in the literature to measure the size of the set of semantically related words in the lexicon: two measures representing the number of words in the lexicon that shared semantic features with the target according to McRae, Cree, Seidenberg and McNorgan's (2005) feature norms (Near- and distant-feature neighbourhood density: for the word *butterfly*, an example of near- feature neighbour is *wasp*, and of distant-feature neighbour is *canary*); two measures representing the set of words that are commonly found in the same semantic contexts, and hence were considered as very similar, based on Latent Semantic Analysis (Landauer, Foltz, & Laham, 1998), with a raw version (including words like *shedding* for butterfly), and a category-trimmed version that included only those words that belonged to the same semantic category (*moth* for butterfly) (raw and categorical contextual neighbourhood density); a subjective measure reflecting the number of coordinates of a given word, obtained by ratings (rated competitors: e.g., Blanken, Dittmann, & Wallesch, 2002); and a measure using free association norms (Nelson, McEvoy, & Schreiber, 1998) to represent the set of words that are commonly associated to the target word (association neighbourhood density, for butterfly, the word *flower*). Experiment 1 investigated how these six measures were related to each other and to other psycholinguistic predictors of picture naming latencies and accuracy. Correlations revealed surprisingly little intercorrelation between these critical measures and between them and control psycholinguistic predictors, and a principal component analysis allowed the identification of four broad types of semantic neighbourhood measures based on how the initial measures loaded on a common component: *feature-based* (the component on which rated competitors and near-feature neighbours loaded the highest), *contextual* (raw and categorical contextual measures), and an individual component for distant feature neighbours (*distant*) and for association neighbours (*associates*). None of the control psycholinguistic measures loaded on the same component as semantic neighbourhood measures, showing the importance and the specificity of these semantic neighbourhood density measures.

In Experiment 2, multiple regression analyses were performed to assess the influence of the four semantic neighbourhood components on average response time and response type of 50 English monolinguals (data was drawn from the International Picture Naming Project: Szekely et al., 2003). None of the semantic neighbourhood factors had any effect, neither on latencies, nor on accuracy nor rate of semantic errors.

In Experiment 3, the influence of the four critical measures on the accuracy and error type of a group of 246 speakers with aphasia was evaluated, with further analyses of four subgroups of participants based on their degree of semantic and phonological impairment. Results of the analyses (Generalised linear mixed effect models) showed that a high feature-based semantic neighbourhood density predicted better accuracy overall, a finding that was inconsistent with the previous literature (e.g., Mirman, 2011). It also increased the probability of a semantic error over an omission, in line with the findings of several authors (Blanken et al., 2002; Bormann, 2011; Bormann, Kulke, Wallesch, & Blanken, 2008; Kittredge Dell, & Schwartz 2007b). An effect of association-based neighbourhood density was found on errors: words with fewer association neighbours (that likely equated to words with more strongly associated neighbours) were more likely to result in a semantic error than in an omission, but this was particularly the case for participants with a semantic impairment.

I suggested that a facilitation effect of semantic neighbours could best be explained within interactive activation theories (e.g., Dell, Martin, Schwartz, & Gagnon, 1997), with the addition of within level links related to the strength of associations between words. Results suggested that while effects of semantic neighbourhood density may be too small to be easily detected in unimpaired picture naming, investigation of aphasic picture naming performance allows detection of effects that are modulated by the type of impairment, and that the type of metric used to represent the set of semantically related words does matter, with better predictive power from feature-based and association measures.

Chapter 3: Paper 2: Investigation of the effects of semantic neighbours in aphasia: a facilitated naming study

In Paper 2, the influence of semantic neighbourhood predictors was observed on the outcomes of a facilitated naming task in two individuals with chronic aphasia. We used a new set of semantic neighbourhood measures, that were chosen in order to take into account, not only the number of semantic neighbours, but also the similarity of these neighbours, their frequency and how strongly they are related to the target. This was motivated by recent evidence showing that what might matter is not only the number of semantic neighbours but also measures related to potential strength of activation of some of these neighbours (e.g., Britt, Ferrara, & Mirman, 2016; Rabovsky, Schad, & Abdel Rahman, 2016). Seven feature-based measures were defined (based on feature norms), that represented either the number of words sharing many features with the target, and among these, the number of neighbours that were higher in frequency than the target, or the proportion of features shared between the target and its most similar neighbour or all its neighbours on average, or alternatively the frequency of the most similar neighbour. Four measures of association-based neighbourhood were defined (based on free association norms), based on the number of different associates (as in Paper 1), the strength of association of the first associate, and the frequency of this first associate.

Two individuals with chronic aphasia, one with a semantic impairment (SJS) and one with a post-semantic impairment (DEH) participated in a facilitated naming experiment. A facilitated naming paradigm consists of investigation of the effects of a single application of a therapy task (here, repetition of the target name in the presence of a corresponding picture) on naming a few minutes later. This is thought to assess priming mechanisms in the case of aphasia and has the potential to predict future therapy outcomes. The influence of the semantic neighbourhood variables was assessed on picture naming latency and accuracy before facilitation, and the effects of these variables on the benefit from naming after facilitation.

Results showed that DEH's performance was slowed at baseline if words had many feature-based neighbours, and SJS was less accurate on words that had a strong first associate, showing sensitivity to different types of semantic neighbourhood measures for each participant, but both in the direction of neighbours hindering word retrieval. Following the facilitation task, for DEH words with least similar neighbours showed the greatest improvement in response time after the task. For SJS, words with a strong first associate, despite being less accurately named at baseline compared to words with a weak first associate, benefitted from the facilitation task in such a way that they were more likely to be better named at post-test compared to words with a weak first associate.

It was argued that these results were best explained by theories that assume interactivity between the semantic and the lexical level (e.g., Dell et al., 1997), and by theories postulating that priming consists of a stronger mapping between the semantic and the lexical level (Howard, Nickels, Coltheart, & Cole-Virtue, 2006). Results are also a confirmation of the complexity of the influence of semantic neighbourhood factors on speech production, with different effects depending on the measure (feature-based versus association-based), the type of task (picture naming versus a task analogous to priming) and the level of impairment in aphasia (presence or not of semantic processing deficits).

Chapter 4: Paper 3: Effects of phonological neighbourhood density and frequency in picture naming: English monolinguals and French-English late bilinguals.

Paper 3 reports an investigation of the effect of phonological neighbourhood factors in picture naming, in a population of English monolinguals and French-English bilinguals. There were inconsistencies in the speech production literature as to whether dense phonological neighbourhoods facilitated or hindered speech production in monolinguals. Indeed, although it is commonly assumed that phonological neighbourhood has a facilitatory effect (e.g., Vitevitch, 2002), it seems that in fact there is not much agreement among studies, with effects in different directions found in different populations and on different languages. Evidence relating to the influence of phonological neighbourhood frequency is not more consistent. In bilinguals, the influence of phonological neighbourhood density in picture naming has received very little attention so far (as far as we know, only two studies address the topic: Marian & Blumenfeld, 2006, and Sadat, Martin, Magnuson, Alario, & Costa, 2015). The literature is even more restricted for phonological neighbours of the non-target language (only one recent study: Sadat et al., 2015), and, to date, there are no studies investigating the influence of phonological neighbourhood frequency (neither summed nor averaged) in bilingual picture naming.

I attempted to address these matters, using regression methodology, and extending the focus, for monolinguals, on additional phonological neighbourhood density measures, namely the number of phonological neighbours that are higher in frequency than the target (building on previous evidence from the visual recognition field that neighbours of higher frequency than the target might be critical, e.g., Grainger, O'Regan, Jacobs, & Segui,1989), and several measures of phonological neighbourhood frequency. The question of what counts as a phonological neighbour for a late bilingual motivated the implementation of a novel measure, taking into account how non-native phonemes might be represented in late bilinguals, and an estimate of their vocabulary range.

In Experiment 1, monolingual speakers of English performed a picture naming task. Analysis with linear mixed models showed that, while no main effect of any phonological neighbourhood density or frequency measure was found, there was a significant interaction between the number of phonological neighbours that were higher in frequency than the target, and the target's frequency: words were slower/less accurately named if they were low in frequency while their phonological neighbours were higher in frequency. Similar interactions involving the frequency of these phonological neighbours were also observed.

In Experiment 2, individuals who were native French speakers and were later exposed to English in an English-speaking environment for at least two years, performed the same English picture naming task. Similar analyses as in Experiment 1 were run, with the addition of predictors specific to the bilingual group, namely the number of years of exposure to English (used as a proxy for proficiency), the amount of formal overlap between the target word and its French translation equivalent, and phonological neighbourhood measures in both languages. As in monolinguals, there was no main effect of any measure on response time, but in contrast, an inhibitory effect of most phonological neighbourhood density measures, including cross-language neighbourhood measures, was found on accuracy. In addition, many critical measures were involved in similar interactions with frequency as well as with the length of exposure and the amount of overlap between the target and its translation equivalent. I suggested that, in late bilinguals, the frequency of a word is affected by several variables: words in a non-proficient, non-dominant language have an overall low frequency status compared to words of the dominant language, and the difference between overall frequency levels in both languages decreases as proficiency / dominance increase. Moreover, I postulated that words of the non-dominant language that have a very high overlap with their translation equivalent (i.e., cognates) inherit the higher frequency status of the dominant language, while words with little overlap with their translation equivalent have a lower frequency status. Hence, it was inferred that in bilinguals, interactions between phonological neighbourhood measures and frequency, English exposure, and translation overlap, followed the same mechanism as in the monolingual case. Words with many phonological neighbours, or phonological neighbours of high frequency relative to the target word, exerted an inhibitory effect on response time and accuracy, while the effect was facilitatory when the target and its neighbours were more similar in *frequency*. This *frequency* of the target word could however be an actual effect of word frequency, or a consequence of the proficiency level or the formal overlap with the translation equivalent.

These results were again best explained by models that postulate interactivity between the lexical level and the phonological level, and in which frequency is represented in the connection strengths (e.g., Dell et al., 1997), as well as such theories where there is competitive lexical selection (e.g., Howard et al., 2006). I speculated that, to explain further the results of our bilingual participants, one would have to hypothesise the existence of interactivity *within* as well as *across* languages between the lexical and the phoneme levels, and of insufficiently specified mappings between the lexical level and the phoneme level in case of less proficiency in the second language of a bilingual.

In the remainder of this chapter, I will address some of the main choices that were taken in this thesis, the reasons for these choices and difficulties encountered. These include comparing between different measures, adopting a picture naming paradigm, and using continuous instead of factorial designs. I will then propose some potential future directions for this research.

The comparability of different measures

In each paper of this thesis, we systematically contrasted measures that have been used to capture neighbourhoods. These include, in Paper 1, several measures of the number of semantic neighbours; in Paper 2, measures of both the number, and the similarity and potential level of activation (association strength and frequency of closest neighbours) of semantic neighbours; and measures of the number and frequency of within and cross-language phonological neighbours in Paper 3.

These investigations were motivated by the fact that, when reviewing the literature, it became apparent that studies often investigate the influence of a word’s neighbourhood using different metrics. Nevertheless, an effort is sometimes made to allow comparison with studies that have used alternative measures. For instance, when studying the influence of semantic neighbourhood density on the performance of individuals with aphasia, Kittredge, Dell, and Schwartz (2007) used a novel measure in latent semantic analysis (Landauer et al., 1998). In order to show how this measure approached that used by Bormann, Kulke, Wallesch, and Blanken (2008; number of semantic competitors), they demonstrated that items in their "low semantic density" condition when compared to items in their "high semantic density" condition also significantly differed on the Bormann et al. (2008) measure. We believed that this comparison was important but could be extended. Consequently, we took it a step further by performing more fine-grained comparisons between different measures of semantic neighbourhood density. Indeed, it turns out that (at least for our unselected set of items) these two different measures are not so similar: semantic neighbourhood density values defined with Bormann et al.'s measure were only weakly correlated to those defined by Kittredge et al.'s (2007) latent semantic analysis measure ($r(84)=.25$, $p<0.05$), and when submitted to a principal component analysis, these two measures did not load on the same component, which is additional evidence that they are likely to measure different things. Hence, this procedure underscores the importance of awareness of the similarity of different metrics used in the literature. Researchers need to be cautious when interpreting results against previous findings if the measures used are different.

In some cases the concern is different: in order to use the measure that has the best predictive power, several measures are combined to define a "better" measure. For instance, Mirman and Graziano (2013) investigated the influence of phonological neighbourhood density on the picture naming performance of individuals with aphasia, and chose to use a particular measure: "the summed log frequency of the target word and all words that share the word onset (i.e., same initial two phonemes; called "cohort density"(…))" (p.1510) as a measure of phonological neighbourhood density. While this choice was based on previous research showing that it had the biggest impact on a similar population as their experimental group, it is unclear what the separate role of the number of phonological neighbours is and that of their frequency. Hence, once again, in order to be able to better specify the relative independent contribution of these aspects, we used separate measures of the number and the frequency of phonological neighbours in analyses of the same data.

A "simple" picture naming task

The act of referring to concepts with names is central to human communication. The task of picture naming is thought to reflect closely this natural act. In all the papers included in this thesis, a picture naming task was the method chosen to examine spoken word representation and processing across all papers in this thesis. Despite its apparent simplicity, picture naming relies on complex cognitive processes. Three broad stages are required: *identification* of the concept represented in the picture, *activation* of the target name from among the thousands of possible words in our lexicon, and finally *selection* of the phonemes and preparation of the articulatory commands for the specific response. These operations must occur rapidly and for many words in quick succession in fluent speech (Johnson, Paivio, & Clark, 1996). While picture naming is a commonly used task for the investigation of the influence of phonological neighbourhood density (e.g., Bernstein Ratner, Newman, & Strekas, 2009; Sadat, Martin, Costa, & Alario, 2014; Vitevitch, 2002), "semantic factors in language production research are often investigated by manipulating the contexts in which identical messages are produced, rather than contrasting item-inherent attributes of different utterances" (Rabovsky, Schad, & Abdel Rahman, 2016: p.240). Indeed, substantial evidence for the co-activation of semantically related words in speech production comes from studies using the picture-word interference paradigm (e.g., Schriefers, Meyer, & Levelt, 1990), the semantic blocking paradigm (e.g., Kroll & Stewart, 1994), or priming studies (e.g. Howard et al., 2006). These studies have often concluded that there was competition between semantically related representations in the course of speech production. The common factor between these paradigms is that the activation of semantically related words may be "induced" by the paradigm, and hence it is unclear whether the competition is revealed by the study design, or instead induced by it. Using "simple" picture naming (without manipulating the presence of semantically related representations) and investigating item-related semantic properties allows one an alternative method to investigate the influence of semantic factors without inducing them. Moreover, this method avoids the difficulty of some of these paradigms (except for the continuous naming priming paradigm (e.g., Howard et al, 2006) in that the possibility of task-specific strategies or effects that may be present (see for example the response exclusion account of semantic effects in picture-word interference, e.g., Finkbeiner & Caramazza, 2006).

Continuous versus factorial designs

In this thesis, a regression methodology was chosen for all papers, whereby, using a continuous design, I evaluated whether the dependent variable (response time or type of error) could be predicted by the predictors of interest (neighbourhood measures) while controlling for the potential influence of control variables.

There are some advantages in adopting this design: first, it does not require carefully balanced sets of stimuli, but only that every measure of interest is available for all stimuli. This allows for the inclusion of a high number of items and therefore increases power. In contrast, when using a factorial design matching stimuli can result in relatively small numbers of items, and indeed may be close to impossible when closely related factors are manipulated (such as neighbourhood density and frequency were here).

Second, it makes it possible to treat neighbourhood density measures as continuous variables. Indeed, it is not always recommended to dichotomise continuous variables, as, although it simplifies analyses, this procedure might once again result in reduced statistical power (e.g., Cohen, 1983). Moreover, the matching of other variables that is required by this procedure can result in selecting unusual materials (e.g., Hauk, Davis,Ford, Pulvermuller, & Marslen-Wilson, 2006).

However, the use of regression techniques also comes with some problems, one of them being the common presence of multicollinearity among predictors. Multicollinearity arises when there are substantial correlations between independent variables, and it is still not clear how to best deal with any multicollinearity in psycholinguistic experiments. Although the presence of multicollinearity does not change the overall fit of the model, it can increase the variance of the coefficient estimates and make them unstable and difficult to interpret. This is not a major issue for control predictors if one is not interested in their effects, but merely wish to ensure they are not confounds for the variables of interest. However, it is a concern when the critical predictors are involved in correlations with other predictors. This was the case in several instances in this thesis, and we used different procedures to reduce problems related to multicollinearity. In Paper 1, a principal component analysis was used to understand better the structure of correlations between predictors. The resulting factor scores for each component were used as new predictor variables that had the advantage of being uncorrelated, hence, multicollinearity was not a problem in regression analyses in Paper 1. In Paper 2, little multicollinearity was present between our critical predictors and the control predictors because they showed relatively low correlations, and while the measures of semantic neighbourhood density were often strongly intercorrelated, they were always included in separate models. In Paper 3, multicollinearity was more of an issue because phonological neighbourhood density measures are typically strongly (negatively) correlated with length.

Some measures can help reducing multicollinearity: namely, the effects of multicollinearity are weaker with increased power. Consequently, we attempted to ensure we had a relatively high number of items (386) and participants (40 for Experiment 1, 50 for Experiment 2). A common procedure to circumvent the issue of multicollinearity is to residualize a measure against the one it is strongly correlated with, and use the obtained residuals as a new uncorrelated measure. However, there is recent evidence showing that this procedure is not an ideal solution as it can pose some interpretation problems (Wurm & Fisicaro, 2014). Hence, when assessing the influence of a phonological neighbourhood measure, the chosen approach to multicollinearity in Paper 3 was to run two different models, one excluding length, the other including it, and see whether an effect of the critical measure "survived the inclusion of length" in the model, while checking for the degree of multicollinearity by looking at variance inflation factors of the critical variables. If an effect of a critical variable was observed only with or without length in the model, we considered this effect as not reliable.

Perhaps, an ideal procedure would be to combine both a continuous and a factorial design within the same experiment, as recommended by Ellis, Lum, and Lambon Ralph (1996). One could use regression techniques as a first step, using a maximal number of items, and then within these items, two different sets of items matched for the control variables but differing on the critical predictor could be compared in a factorial design (although this would be a particularly difficult task to do when matching for 10 or more critical variables). An additional possibility would be to gain more insight into the stability and reliability of the effects by running the same stimuli more than once with the same participants, once again as recommended by Ellis et al (1996) and implemented by Sadat et al. (2014).

Limitations: Other neighbourhoods?

Although many different measures representing the possible neighbourhoods of a word were studied in this thesis, other types of neighbourhoods could (or have been shown to) play an important role in the process of producing words (e.g., Goldrick, Folk & Rapp, 2010).

First, I have not addressed the influence of the position of overlap in phonological neighbours, although this has been shown to be an influential characteristic of phonological neighbourhoods in speech perception (e.g., Vitevitch, 2007) and production (e.g., Vitevitch, Armbrüster, & Chu, 2004). Indeed, Vitevitch et al. (2004) observed an inhibitory effect of onset density (the number of phonological neighbours that share their first phoneme with the target).

Sadat et al. (2014) replicated this effect, although, as they noted, it could be the case that the effect of onset density was confounded with that of (a classic measure of) phonological neighbourhood density. Another measure is the “cohort density”, or number of words in the lexicon that share the same initial two phonemes with the target, a measure that has been shown to influence speech recognition (e.g., Magnuson, Dixon, Tanenhaus, & Aslin, 2007), but that has also used been to predict speech production outcomes (e.g., Mirman & Graziano, 2013).

An interesting approach was proposed by Chen and Mirman (2012), related to the clustering of neighbours: these authors propose that the flow of activation from the target to its neighbours could also extend from the neighbours of a word to these words’ neighbours, in such a way that, if neighbours of a word are also neighbours of each other, there is stronger activation of the neighbourhood overall. Although the evidence they provided came from simulations, there is in fact evidence that clustered neighbours are particularly inhibitory in spoken word recognition (Chan & Vitevitch, 2009). The influence of the clustering of neighbours is yet to be tested in production. It could be the case that this type of measure has a greater impact on speech production compared to other phonological neighbourhood density measures.

Another possibility is that the set of neighbours that are activated during speech production in the context of phrases and sentences is constrained by the grammatical category of the target. This was demonstrated by Heller and Goldrick (2014): words that had many phonological neighbours from the same word class were slower named than words that had few of these neighbours in the context of sentences but there was no difference in a simple picture naming task. This finding underscores the limits of the picture naming task in representing what happens when we produce words within in the context of sentences.

Goldrick et al. (2010) raised the interesting possibility that, in line with theories that assume graded activation, related word representations might not be simply “on or off” but could be partially activated. In this view, neighbourhood status would not be considered as a binary distinction but more as a dimension along which non-target words vary. Although it seems very challenging to test this hypothesis, Goldrick et al. (2010) tried to determine along which dimensions the neighbourhood of a word (words that are likely to be co-activated in processing) vary. To do so, they looked at the formal (phonologically related word) errors that were produced by a group of participants with aphasia and compared each error with its target. They found that these errors had a high position-specific form overlap, similar lexical frequency, belonged to the same category and had similar length compared to the target. Based on these findings, they inferred that these dimensions were critical in defining the actual neighbourhood of a word, and they created a novel measure (the LexForm composite measure) that incorporated these dimensions.

Both semantically and phonologically similar representations could be represented in the form of morphological neighbours. Morphological families consist of clusters of words that all differ from one morpheme but share another (e.g., Bertram, Schreuder, & Baayen, 2000, cited by Marian & Blumenfeld, 2006). For example, the morpheme *work* is present in the words *homework*, *housework*, and *workable* as well as *working*, *worked* and *works* (examples provided by Marian & Blumenfeld, 2006, p.6). Interestingly, the morphological family size has been shown to influence the ease of word recognition in the non-native language of a bilingual (Dijkstra, Moscoso del Prado Martín, Schulpen, Schreuder and Baayen, 2005). Once again, an extension to production would be interesting.

Finally, while I have made a distinction between semantic and phonological neighbours, I did not explicitly address their influence within the same study (although phonological neighbourhood density was a control variable in both semantic neighbourhood papers). However, there is evidence that words that are similar in meaning are more likely to share phonological structure than words that have very different meanings (e.g., O'Toole, Oberlander, & Shillcock, 2001). Moreover, there is other evidence suggesting a strong association between some particular sound sequences and meanings (e.g., Vigliocco & Kita, 2006, cited by Goldrick et al., 2010). Hence, further work looking at the intersection of these two forms of neighbourhood, and others, while challenging, may prove important.

Concluding remarks

In this thesis, the influence of both semantic and phonological neighbourhood density and frequency on the picture naming performance of English monolinguals, individuals with aphasia, and late French-English bilinguals was investigated. Many different neighbourhood measures were defined and compared, allowing a better understanding of the relative contribution of aspects of semantic and phonological relatedness in the effect of neighbours on spoken word production. The findings showed that not only the number of neighbours, but also their similarity and strength of association to the target, as well as their frequency, in comparison to that of the target, have an influence on spoken word production. We also observed that, while some effects of neighbours were absent or small in unimpaired monolingual performance, they appeared or were amplified in the case of people with aphasia and late bilingual participants, showing the importance of studying different populations to reveal or magnify subtle processes in speech production.

Overall, this research enabled an in depth investigation of what constitutes a word's neighbourhood, allowing a better specification of relevant metrics, and of what types of dynamics these neighbourhoods of a given word are involved in, using converging evidence from different populations and methodologies. It has underlined the complexity of the role of neighbourhood factors and their interactions with other variables. The next step towards building a full theory that can account for these complex dynamics would be to attempt simulation using computational modelling.

References

Allison, P. (2010, September 10). When Can You Safely Ignore Multicollinearity? [Web log post]. Retrieved from http://statisticalhorizons.com/multicollinearity

Bernstein Ratner, N., Newman, R., & Strekas, A. (2009). Effects of word frequency and phonological neighborhood characteristics on confrontation naming in children who stutter and normally fluent peers. *Journal of fluency disorders*, *34*(4), 225-241.

Bertram, R., Schreuder, R., & Baayen, R. H. (2000). The balance of storage and computation in morphological processing: The role of word formation type, affixal homonymy, and productivity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(2), 489.

Blanken, G., Dittmann, J., and Wallesch, C-W. (2002). Parallel or serial activation of word forms in speech production? Neurolinguistic evidence from an aphasic patient. *Neuroscience Letters, 325,* 72-74.

Bormann, T., Kulke, F., Wallesch, C.W., and Blanken, G. (2008). Omissions and semantic errors in aphasic naming: Is there a link? *Brain and Language, 104,* 24-32.

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

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

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7*, 249-253.

Dell, G.S., Schwartz, M.F., Martin, N., Saffran, E.M., & Gagnon, D.A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104(4),* 801-838.

Dijkstra, T., Moscoso del Prado Martín, F., Schulpen, B., Schreuder, R., & Harald Baayen, R. (2005). A roommate in cream: Morphological family size effects on interlingual homograph recognition. *Language and Cognitive Processes*, *20*(1-2), 7-41.

Ellis, A. W., Lum, C., & Lambon Ralph, M. A. (1996). On the use of regression techniques for the analysis of single case aphasic data. *Journal of Neurolinguistics*, *9*(3), 165-174.

Finkbeiner, M., & Caramazza, A. (2006). Now you see it, now you don't: On turning semantic interference into facilitation in a Stroop-like task. *Cortex*, *42*(6), 790-796.

Goldrick, M., Folk, J. R., & Rapp, B. (2010). Mrs. Malaprop’s neighborhood: Using word errors to reveal neighborhood structure. *Journal of Memory and Language*, *62*(2), 113-134.

Grainger, J., O’regan, J. K., Jacobs, A. M., & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Perception & Psychophysics, 45*(3), 189-195.

Hauk, O., Davis, M. H., Ford, M., Pulvermüller, F., & Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage, 30*(4), 1383-1400.

Heller, J. R., & Goldrick, M. (2014). Grammatical constraints on phonological encoding in speech production. *Psychonomic Bulletin & Review*, *21*(6), 1576-1582.

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, *100*(3), 464-482.

Johnson, C. J., Paivio, A., & Clark, J. M. (1996). Cognitive components of picture naming. *Psychological Bulletin*, *120*(1), 113.

Kittredge, A.K., Dell, G.S., and Schwartz, M.F. (2007a). Omissions in aphasic picture naming: Late age-of-acquisition is the culprit, not low semantic density. *Brain and Language (103),* 132-133.

Kittredge A. K., Dell G. S., Schwartz M. F. (2007b). “*Omissions in aphasic picture naming: AoA is the culprit, as well as low semantic density*,” Poster Presented at the 45th Meeting of the Academy of Aphasia, Washington, DC.

Landauer, T.K., Foltz, P.W., and Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes, 25,* 259-284.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, *31*(1), 133-156.

Marian, V., & Blumenfeld, H. (2006). Phonological neighborhood density guides lexical access in native and non-native language production. *Journal of Social and Ecological Boundaries, 2,* 3–35.

McRae, K., Cree, G.S., Seidenberg, M.S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, 37*, 547-559.

Mirman, D. (2011). Effects of Near and Distant Semantic Neighbors on Word Production. *Cognitive, Affective, and Behavioral Neuroscience, 11(1),* 32-43.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. http://w3.usf.edu/FreeAssociation/

O'Toole, M., Oberlander, J., & Shillcock, R. (2001). The age-complicity hypothesis: A cognitive account of some historical linguistic data. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 716-719).

Rabovsky, M., Schad, D. J., & Abdel Rahman, R. (2016). Language production is facilitated by semantic richness but inhibited by semantic density: Evidence from picture naming. *Cognition*, *146*, 240-244.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*(3), 460.

Sadat, J., Martin, C.D., Costa, A., & Alario, F-X. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology, 68,* 33-58.

Sadat, J., Martin, C. D., Magnuson, J. S., Alario, F. X., & Costa, A. (2015). Breaking down the bilingual cost in speech production. *Cognitive Science*.

Székely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., Jacobsen, T., & Bates, E. (2003). Timed picture naming: Extended norms and validation against previous studies. *Behavior Research Methods, Instruments, & Computers, 35(4),* 621-633.

Vigliocco, G., & Kita, S. (2006). Language-specific properties of the lexicon: Implications for learning and processing. *Language and Cognitive Processes*, *21*(7-8), 790-816.

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 735.

Vitevitch, M. S. (2007). The spread of the phonological neighborhood influences spoken word recognition. *Memory & Cognition*, *35*(1), 166-175.

Vitevitch, M. S., Armbrüster, J., & Chu, S. (2004). Sublexical and lexical representations in speech production: effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 514.

Wurm, L. H., & Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does *not* do). *Journal of Memory and Language, 72,* 37-48.

Appendix

Final ethics approval letter

MACQUARIE UNIVERSITY

Lyndsey Nickels <lyndsey.nickels@mq.edu.au>

Approved- Ethics application- Nickels (Ref No: 5201200905)

Ethics Secretariat <ethics.secretariat@mq.edu.au> 12 December 2012 11:17
To: Prof Lyndsey Nickels <lyndsey.nickels@mq.edu.au>

Dear Prof Nickels

Re: "Understanding language processing, its breakdown and treatment"
(Ethics Ref: 5201200905)

Thank you for your recent correspondence. Your response has addressed the issues raised by the Human Research Ethics Committee and you may now commence your research.

This research meets the requirements of the National Statement on Ethical Conduct in Human Research (2007). The National Statement is available at the following web site:

http://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/e72.pdf.

The following personnel are authorised to conduct this research:

Dr Anna Elisabeth Beyersmann
Dr Antje Lorenz
Dr Britta Biedermann
Dr Karen Croot
Dr Kati Renvall
Dr Samantha Siyambalapitiya
Dr Saskia Kohnen
Miss Anastasiia Romanova
Miss Catherine Mason
Miss Trudy Geertruida Krajenbrink
Ms Belinda McDonald
Ms Emily Church
Ms Fransizka Bachmann
Ms Jennifer Cole-Virtue
Ms Nora Fieder
Polly Barr
Prof Lyndsey Nickels
Prof Niels Schiller
Professor David Howard
Rimke Groenewold
Shiree Heath
Solene Hameau
Tina Marusch

NB. STUDENTS: IT IS YOUR RESPONSIBILITY TO KEEP A COPY OF THIS APPROVAL EMAIL TO SUBMIT WITH YOUR THESIS.

Please note the following standard requirements of approval:

1. The approval of this project is conditional upon your continuing compliance with the National Statement on Ethical Conduct in Human Research (2007).

2. Approval will be for a period of five (5) years subject to the provision of annual reports.

Progress Report 1 Due: 12 December 2013
Progress Report 2 Due: 12 December 2014
Progress Report 3 Due: 12 December 2015
Progress Report 4 Due: 12 December 2016
Final Report Due: 12 December 2017

NB. If you complete the work earlier than you had planned you must submit a Final Report as soon as the work is completed. If the project has been discontinued or not commenced for any reason, you are also required to submit a Final Report for the project.

Progress reports and Final Reports are available at the following website:

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/forms

3. If the project has run for more than five (5) years you cannot renew approval for the project. You will need to complete and submit a Final Report and submit a new application for the project. (The five year limit on renewal of approvals allows the Committee to fully re-review research in an environment where legislation, guidelines and requirements are continually changing, for example, new child protection and privacy laws).

4. All amendments to the project must be reviewed and approved by the Committee before implementation. Please complete and submit a Request for Amendment Form available at the following website:

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/forms

5. Please notify the Committee immediately in the event of any adverse effects on participants or of any unforeseen events that affect the continued ethical acceptability of the project.

6. At all times you are responsible for the ethical conduct of your research in accordance with the guidelines established by the University. This information is available at the following websites:

http://www.mq.edu.au/policy/

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/policy

If you will be applying for or have applied for internal or external funding for the above project it is your responsibility to provide the Macquarie University's Research Grants Management Assistant with a copy of this email as soon as possible. Internal and External funding agencies will not be informed that you have final approval for your project and funds will not be released until the Research Grants Management Assistant has received a copy of this email.

Please retain a copy of this email as this is your official notification of final ethics approval.

Yours sincerely
Dr Karolyn White
Director of Research Ethics
Chair, Human Research Ethics Committee