Effects of modality of input on word recognition

Diogo Almeida

Cognitive Neuroscience of Language

Department of Linguistics

University of Maryland, College Park

Final 895 Draft: December 8, 2006

Revised: February 20, 2007

Introduction

Any theory of linguistic knowledge and linguistic behavior must posit a lexicon, a repository of information that is used in the different kinds of linguistic computations our minds must perform in the course of language comprehension and production. Understanding the workings of this mental structure and how it satisfies the different requirements imposed by comprehension and production mechanisms is therefore understandably a central research topic in psycholinguistics.

In order for language comprehension and production to occur at all, one must assume that the information stored in the lexicon is properly retrieved. Different units of representation have been proposed to compose the mental lexicon, such as morphemes and idioms, but most of the research done in the field has assumed the *word*[1] as the privileged unit of analysis (Balota, 1994). Therefore, recognizing words and retrieving the information they store is considered to be a central process in language comprehension, and it has received

[1] While the concept of *word* is intuitively accessible and easy to manipulate, there is considerable disagreement in linguistic theories about what a word is, and whether it plays a central role in the organization of the lexicon. However, for the present purposes, I will ignore this controversy.

a commensurate level of attention in the psycholinguistic community.

The main problem researchers in the area of word recognition identify is that the way the lexicon is organized is very conducive to large scale perceptual confusability: an average adult is estimated to know between 30.000 to 50.000 different words (Seidenberg & McClelland, 1990; Monsell, Doyle, & Haggard, 1989), but the whole lexicon is based on a small amount of building blocks (around 40 phonetic segments for spoken language) that are concatenated in different unique combinations to serve as representational units (Goldinger, Luce, & Pisoni, 1989; cf. Balota, 1994; Huntsman & Lima, 2002; Dehaene, Cohen, Sigman, & Vinckier, 2005 for similar arguments for the representation of written words in cultures that use alphabetical writing systems). This entails that overlap of building blocks amongst different words is rampant in the lexicon, which means that for each word actually present in the input, there will be several others partially compatible with it (and arguably partially activated by it) to different degrees of fit. Yet, despite this potential perceptual confusability issue, human word recognition is extremely fast (as fast as 250ms for spoken and written words) and robust to noise.

However, despite the extreme speed of the process, and despite how early lexical access can occur in the processing stream, much of the research done in the field has relied on reaction time latencies in lexical decision tasks as their primary source of empirical evidence. These latencies normally figure in the range of 500-800ms for written words and up to 1000ms for spoken words. Even though reaction time latencies can and have been used to constrain and test different models of lexical access, they have the drawback of only providing a snapshot of the process at the output of the decision making routines. This is considered by some researchers to be a major problem in using reaction time latencies to tease apart different hypotheses about early stages of lexical access (eg. Balota & Chumbley, 1984), and constitutes the main rationale underlying lexical decision research turning increasingly more often to high–temporal resolution electrophysiological techniques (such as EEG and MEG), which can record brain activity from stimulus presentation up to the behavioral response with milisecond accuracy.

In this work, I will report one MEG experiment designed to investigate stages of lexical access in which mapping between sensory input and lexical representations occur, but also stages where selection mechanisms ensure convergence to the (hopefully correct) response. In order to tap into these processes, I will attempt to replicate and extend the findings of two previous studies (Pylkkänen, Stringfellow, & Marantz, 2002; Stockall, Stringfellow, & Marantz, 2004), in which two different kinds of lexical variables were manipulated in simple lexical decision tasks. The first kind of variable, the frequency of subunits that compose words (their phonotactic and orthotactic probability), is thought to facilitate lexical activation of possible candidates. The second kind, lexical neighborhood size and frequency, is thought to influence selection within the candidate set.

The ultimate goal of this work is to gain better understanding of early neural correlates of lexical access across modalities such that they can be subsequentially used to disentangle different hypotheses about lexical representations (cf. Pisoni & Levi, in press; Poeppel, Idsardi, & van Wassenhove, in press for discussion, and Almeida, 2005 for a proposal).

Neural Correlates of Word Recognition

The use of electrophysiological techniques such as EEG has been successfully and fruitfully used in lexical access research for over twenty–five years, since the pioneering work of Kutas and Hillyard (1980, 1984), and although potential neural correlates of word recognition have been identified, I will argue that the standard lexically–induced response reported in the EEG literature (the N400) is in fact not entirely adequate for research focused on the precise time course of lexical access.

Kutas and Hillyard (1980, 1984) reported an Event–Related Potential (ERP) to sentence–final visual words whose amplitude was modulated by its semantic fit in the context in which it was presented. This ERP was identified as a large negative polarity deflection ocurring between 200ms and 600ms post stimulus onset. Since its peak activity normally occurs at around 400ms, it was named N400 (cf. Kutas & Federmeier, 2000 for a review).

Subsequent research has found that this ERP displays many of the characteristics we would expect a lexically–induced ERP to reflect. First, it is reliably elicited by words or wordlike stimuli, either in sentential context (Van Petten & Kutas, 1990) and in isolation or pairs (Bentin, McCarthy, & Wood, 1985), but *not* for non word–like written stimuli, like strings of consonants (Rugg & Nagy, 1987). Second, it is found across a variety of tasks (passive reading Kutas and Hillyard (1980, 1984), lexical decision (Rugg, 1983), semantic classification (Young & Rugg, 1992), and repetition judgment (Rugg, Brovedani, & Doyle, 1992; Rugg, Cox, Doyle, & Wells, 1995; Wilding, 2000)), suggesting a high degree of task–independence. Third, the N400 seems to reflect meaning retrieval, given that (i) its amplitude inversely correlates with how appropriate a word is in particular sentential contexts (Kutas & Hillyard, 1980, 1984), and (ii) its amplitude is modulated in the same way that lexical–level variables such as frequency (Smith & Halgren, 1987; Van Petten & Kutas, 1990), and semantic relatedness or identity of prime in word pair lists influence recognition times at the behavioral level (Bentin et al., 1985; Holcomb, 1988, 1993; Rugg, 1987, 1990): higher N400 amplitudes are found for lower frequency words, which elicit longer RTs, and smaller N400 amplitudes are observed for semantic relatedness or identity between prime and target words in pair lists, which in turn elicit shorter RTs in behavioral tasks. Fourth, the ERP is found in both the visual and auditory modalities of oral language (Holcomb & Neville, 1990, 1991; Bentin, Kutas, & Hillyard, 1995) as well as for signed languages (Kutas, Neville, & Holcomb, 1987), suggesting a high–degree of modality independence.

However, there are two main reasons to think that the N400 is actually not entirely adequate as a dependent measure to specifically tap into the precise time course of lexical activation. First, the response is broad (it can span up to 400ms), but only its amplitude is reliably manipulated, not its peak–latency (eg. Kounios & Holcomb, 1992, p. 469). While this still allows the N400 to probe degrees of relatedness of lexical items in memory and perhaps aspects of semantic processing, it diminishes its usefulness as a tool for probing the fine–grained temporal structure of lexical access. Second, and perhaps most importantly, is that despite an apparent consensus in the field (eg. Kutas & Federmeier, 2000), there is to my knowledge no coherent functional interpretation of the N400 that is able to account for all the facts. I will address this second concern in more detail in the following section.

The functional interpretation of the N400

The standard interpretation of the N400 is that it is a neural correlate of post-lexical meaning processing (Kutas & Federmeier, 2000). This interpretation is backed up by the fact that the N400 is elicited by words that don't provide a good fit to the local semantic context of the sentence, as well as by semantic priming in word pairs. There is also reason to believe that the N400 is not necessarily a lexically–induced response, since it is not uniquely elicited by lexical information. Modulation of the N400 has been found for matching pictures (Barrett & Rugg, 1990; McPherson & Holcomb, 1999), line–drawings (Holcomb & McPherson, 1994), faces (Barrett, Rugg, & Perrett, 1988; Barrett & Rugg, 1989; Bobes, Valdés-Sosa, & Olivares, 1994) and other non–linguistic material (see Kutas & Federmeier, 2000 for review). This has led some researchers to posit that the N400 reflects meaning integration and semantic expectancy across different levels of processing.[2]

However, this view cannot readily accomodate two families of results in the N400 literature: (i) *lack* of semantic fit effects in sentences with simple linguistic operators, such as negation and different quantifiers, that can modify the truth value of propositions by being inserted in what would otherwise be the exact same structure, and (ii) N400 priming effects for unconsciously perceived words.

Lack of N400 effects for semantic anomalies. Soon after the first descriptions of the N400 and its association with semantic processing, Fischler, Bloom, Childers, Roucos, and Perry (1983) conducted a study that found that for simple sentences like "*a robin is a bird*" and "*a robin is **not** a bird*", the N400 elicited by the final word in the two sentences was equivalent in its temporal profile and morphology. The truth value of the sentences, however, is exactly the opposite and therefore the semantic fit of the final word

[2] *Cloze probability*, a variable that had been found to correlate with N400 amplitude modulation in sentential contexts (Kutas, Lindamood, & Hillyard, 1984), served as a surrogate for these elements in the literature regarding sentence comprehension.

in the second sentence should be very low. If the interpretation of the N400 as meaning processing is correct, the false sentences should elicit an N400 effect, contrary to fact.

Moreover, the N400 amplitudes for both sentences were smaller than the ones obtained by pairs like "*a robin is a vehicle*" and "*a robin is* ***not*** *a vehicle*", which were in turn identical to each other. This strongly suggests that the N400 in those cases reflects the degree of association between the subject and the object of the predicates, but *not* the actual integrated linguistic meaning of the sentence.

This effect was later replicated using different quantifiers by Kounios and Holcomb (1992), who reported lexical association, but not sentence meaning effects using materials like"*[all/some/no] [dogs/apples] are [animals/fruit]*" and "*[all/some/no] [animals/fruit] are [dogs/apples]*". These results have recently been replicated and extended in French (Noveck & Posada, 2003), German (Drenhaus, Graben, & Frisch, 2006) and Spanish (Beltrán, Carreiras, Alvarez, & Santamaría, 2006), and are hard to accomodate if one takes the N400 to reflect semantic or expectancy–based integration.

N400 priming for unconsciously perceived words. One of the results that lent support to a semantic processing account of the N400 response was the lack of semantic priming in masked priming environments (Brown & Hagoort, 1993). However, several subsequent studies revealed that masked words that are unconsciously perceived can indeed elicit N400 semantic priming effects (Deacon, Hewitt, Yang, & Nagata, 2000; Kiefer, 2002; Kiefer & Brendel, 2006; Kreher, Holcomb, & Kuperberg, 2006), as can words presented under the attentional blink (Luck, Vogel, & Shapiro, 1996; Vogel, Luck, & Shapiro, 1998; Rolke, Heil, Streb, & Hennighausen, 2001).

All the results reviewed above are striking under a strictly sentential meaning processing or contextual integration account of the N400. This is especially true in light of recent work reporting discourse–level N400 effects (St. George, Mannes, & Hoffman, 1994; van Berkum, Hagoort, & Brown, 1999; van Berkum, Zwitserlood, Hagoort, & Brown, 2003; Nieuwland & van Berkum, 2006; Swaab, Camblin, & Gordon, 2004; Ledoux, Camblin, Swaab, & Gordon, 2006; Coulson & Wu, 2005; Coulson, Federmeier, Van Petten, & Kutas, 2005). The obvious question then is how come we have high-level contextual information and very low level subliminal information eliciting N400 effects, but not simple linguistic manipulations, such as negation and quantification, that are readily available and easy to interpret? If anything, those should be the bread and butter of semantic integration, and not the exception.

Subliminal priming and the lack of semantic fit N400 effects in simple sentences argue instead for a association/spreading activation account of the N400, a view that has recently received new empirical support (Rhodes & Donaldson, 2006). This interpretation was first proposed by Fischler and Raney (1989), who argued that

> "... the reduction of N400 in linguistic contexts had little to do with sentence structure or meaning, and was closely tied to the lexical association of prime and target words" (p. 38, cited in Kounios and Holcomb (1992), p. 461)

Although this would in principle explain most, if not all, the semantically–anomalous N400 effects, it still fails to explain several of the aforementioned discourse–related N400 effects, in which lexical association was, intentionally or not, controlled for (St. George et al., 1994; van Berkum et al., 2003; Nieuwland & van Berkum, 2006; Coulson & Wu, 2005, to name a few particularly hard cases).

Therefore, the functional interpretation of the N400 still eludes a unified coherent account, and all we are left with is the generalization that it has "something to do with meaning". Therefore, even in the absence of the temporal coarseness issue, the fact that a coherent interpretation of the N400 is currently lacking would definitely obscure any potential results one might find using this ERP as a dependent measure.

Using MEG to investigate early aspects of lexical access

Recent investigation using magnetoencephalography (MEG) has identified a candidate response that shows very good prospects for the study of the precise time course of lexical access (Pylkkänen & Marantz, 2003 for review). There are three main reasons for this. The first is that this response, which peaks at around 350ms for visually presented words (therefore within the same temporal window of the N400), has a much more narrower temporal window than its EEG counterpart, and does not span the whole 250-300ms that the N400 does; in fact, the response normally assumes the form of a very narrow peak of activity. The second reason is that this component has been reported to be sensitive to much the same kind of manipulations that the N400 is sensitive to, such as semantically anomalous sentential contexts (Helenius, Salmelin, Service, & Connolly, 1998; Halgren et al., 2002), lexical frequency (Embick, Hackl, Schaeffer, Kelepir, & Marantz, 2001; Halgren et al., 2002), stimulus repetition (Sekiguchi, Koyama, & Kakigi, 2000; Pylkkänen, Stringfellow, Flagg, & Marantz, 2001) and multiple meaning ambiguities (Beretta, Fiorentino, & Poeppel, 2005; Pylkkänen, Llinás, & Murphy, 2006). The third reason is that it is the peak latency, and not its amplitude, that has been shown to systematically vary according to all these manipulations. Therefore, there is hope that this component will allow research on lexical access to incorporate a more accurate and precise temporal dimension.

Furthermore, MEG has the added benefit of being able to tell something about the underlying sources of activity, and could be useful in establishing spatio-temporal maps and profiles of different linguistic processes (Dhond, Buckner, Dale, Marinkovic, & Halgren, 2001; Dhond, Marinkovic, Dale, Witzel, & Halgren, 2003; Marinkovic et al., 2003; Dhond, Witzel, Dale, & Halgren, 2005, 2006; Halgren et al., 2006).

Due to the fact that the M350 could prove to be a very useful window into the time course of lexical access, we decided to focus our attention in this response. Moreover, three recent studies underscored the potential of the M350 to serve as an index of activity in early lexical activation stages. Pylkkänen et al. (2002), building on work of Vitevitch and Luce (1999), was able to show that the M350 latency tracks solely ease of lexical activation, which is thought to be dependent on the frequency of sublexical constituents, but not lexical selection, which is thought to be affected by lexical neighborhood frequency, and was found to be tracked by reaction time latencies. Stockall et al. (2004) replicated and extended these results, with a very similar design, and Fiorentino and Poeppel (in press) found M350 latency manipulation for compound words of exact same lexical frequency based on the frequency of their constituents being high or low, *regardless* of the behavioral outcome. These results strongly suggest that the M350 is indeed tracking lexical activation routines, but not decisional processes.

Neighborhood Density and Phonotactic Probability influences in Word Recognition

As noted in the introduction, lexical access is often conceptualized as a perceptual confusability problem. Given the small number of building blocks used to store lexical representations and the large number of words in the lexicon, this entails that words will normally have a high degree of partial overalp with other words in terms of the sub-units that compose them. The questions that arises then is what kind of effect (if any) does *similarity* have in the processing of words? Does similarity inhibit processing due to the possible competition amongst candidates? Or on the contrary, does it have a facilitatory effect (for instance, familiarity), due to possible synergistic effects of highly overlapping sub–units?

In this sense, lexical neighborhoods are of high theoretical interest because they provide a way of quantifying the similarity space for words. With such a metric, it is possible to investigate how much of an effect (either inhibitory or facilitatory) similarity might have in lexical processing. On this point, different models make different predictions about what the effects of similarity might be. Models of lexical access have historically espoused the idea that the process of lexical retrieval rests primarily on either parallel search or multiple simultaneous activation of independent units (cf. Morton, 1969; Coltheart, Davelaar, Jonasson, & Besner, 1977; Rumelhart & McClelland, 1981, 1982, but see K. I. Forster, 1976 for a serial search proposal).

These models normally require a selection stage where, amongst all the possible alternatives, only one (and hopefully the right) candidate is selected (see Marslen-Wilson, 1987, 1989; Luce & Pisoni, 1998; Luce, Goldinger, Auer, & Vitevitch, 2000 for auditory word recognition and Seidenberg & McClelland, 1989; Grainger & Jacobs, 1996 for visual word recognition). Further, most of these models assume that the mechanism that ensures convergence to the final decision is competition between the different candidates (sometimes called *lateral inhibition*). Therefore, all else being equal, these models uniformly predict that words with few neighbors should be recognized *faster* than words with lots of neighbors. Likewise, all else being equal, neighbors of lower frequency should impair recognition *less* than neighbors of higher frequency. In summary, it is a straightforward prediction of most *multiple activation*–based models that effects of lexical competition should occur, and that more competition (either in terms of number or frequency of competitors) should *inhibit* recognition. To the extent that inhibition effects from both *neighborhood size* and *neighborhood frequency* have been reliably found in spoken word recognition (Goldinger et al., 1989; Goldinger, Luce, Pisoni, & Marcario, 1992; Luce & Pisoni, 1998), these models have some empirical support.

However, Vitevitch, Luce, Charles-Luce, and Kemmerer (1997) have recently found that nonwords with highly frequent sub–units (co-occurrences of phonemes that compose the word) were named more quickly and accurately than nonwords with less frequent sub–units. This result was surprising given the high degree of correlation between phonotactic probability and neighborhood size that is reported for different languages. Therefore, this could be an example where similarity between words might not impair recognition, but actually *facilitate* it. However, these researchers posited that this facilitatory effect was instead due to sublexical processing, and didn't originate from lexical processing per se. Given that a parametric independent manipulation of phonotactic probability and neighborhood density was deemed impossible by the authors (due to the aforementioned high correlation between phonotactic probability and neighborhood density), they decided to follow up on the issue in a series of studies (Vitevitch & Luce, 1998, 1999) adopting a different strategy. They used materials in which these both phonotactic probability and neighorhood frequency were highly correlated, but in which the use of different tasks (naming, same–different judgment and lexical decision) would bias subjects' mode of processing to focus either at a sublexical or lexical level.

Vitevitch and Luce (1998, 1999) were able to show then that when tasks were biased towards sublexical processing, facilitatory effects in latencies and accuracy in both the naming and same–different judgment tasks were found for nonwords with high phonotactic probability, *even though* these very same words also had more frequent neighbors. However, when the experimental task was biased towards lexical processing (lexical decision task), then the standard inhibition effect was found again for both words and nonwords with more frequent neighbors. These authors then concluded that the locus of the two different effects were at two distinct processing levels, and one can still maintain that neighborhood size and frequency has inhibitory effects due to lexical competition.

Assuming a version of the *activation–selection* paradigm, Pylkkänen et al. (2002) interpreted Vitevitch and Luce (1998, 1999) results as an indication that phonotactic probability might facilitate lexical activation of multiple possible candidates, and in situations where lexical processing or competition is not so strong (as in the processing of nonwords), these facilitatory effects might be visible at the behavioral level. These researchers hypothesized then that since effects of facilitation of phonotactic probability and inhibition due to lexical competition seem to be in principle dissociable and act at different levels of representation, one might be able to observe them in earlier periods of the processing stream if one uses the right tools.

By using MEG and converting a subset of Vitevitch and Luce (1998, 1999)'s materials to written form, Pylkkänen et al. (2002) were able to show that although reaction time latencies displayed the standard neighborhood frequency inhibition effect, the latency of the electrophysiological response elicited by written words (the M350) was modulated in exactly the *opposite* way, displaying a *facilitatory effect* for words and nonwords with highly probable sub–units, even though they also had more neighbors than words with lesser probable sub–units. It is worth stressing that the latter also had smaller neighborhoods, and were responded to faster than their counterparts at the behavioral level. Pylkkänen et al. (2002) interpreted these results as a demonstration that the M350 selectively tracks lexical activation without the effects of lexical competition. The electrophysiological results were later replicated by Stockall et al. (2004) in a similar, albeit extended manipulation of lexical frequency, phonotactic probability and neighborhood density.

These results, and if correct, imply that we now have a powerful tool that allows us to look at the fine temporal structure of lexical access. However, before these results can be taken at face value, some discussion of the potential problems of the studies is warranted.

First of all, it is worth keeping in mind that the original Vitevitch and Luce (1998, 1999)'s results were found for auditory word recognition, while Pylkkänen et al. (2002)'s and Stockall et al. (2004)'s were found for visually presented stimuli. This unfortunately introduces problems for the interpretation of the results.

Although compelling given the nature of their predictions and findings, Pylkkänen et al. (2002)'s behavioral results are somewhat surprising in the light of an extensive line of research that investigates the effect of neighborhoods for written words. For instance, unlike *phonological* neighborhood size, *orthographic* neighborhood size has either been found to systematically fail to cause inhibition (Coltheart et al., 1977) or has been found to elicit the exact opposite effect, i.e., *facilitation* for items with more neighbors (Andrews, 1989, 1992; Snodgrass & Minzer, 1993; Sears, Hino, & Lupker, 1995; K. I. Forster & Shen, 1996; Sears, Lupker, & Hino, 1999; Siakaluk, Sears, & Lupker, 2002; Huntsman & Lima, 2002; Sears, Campbell, & Lupker, 2006).

Inhibitory effects of orthographic neighborhood *frequency*, on the other hand, is only occasionaly found in English (Huntsman & Lima, 1996; Perea & Pollatsek, 1998), and has also been found to elicit null or facilitatory effects (Sears et al., 1995; Sears, Lupker, & Hino, 1999; Huntsman & Lima, 2002; Sears et al., 2006). In fact, most of the evidence supporting the existence of inhibitory orthographic neighborhood frequency comes from other languages, such as French, Spanish and Dutch (see Andrews, 1997; Sears, Hino, & Lupker, 1999; Perea & Rosa, 2000 for reviews). Finally, Balota, Cortese, Sergent-Marshall, Spieler, and Yap (2004), using an innovative approach to word recognition, performed a large scale hierarchical regression analysis on reaction time and naming latencies on 2,428 words, directly comparing a large number of variables that have been shown to influence reading and naming times. They reported *facilitatory* effects of orthographic neighborhood size for low frequency words, but no reliable effect of facilitation nor inhibition for high frequency words[3].

One possible and potentially interesting interpretation for this pattern of results is that, by virtue of simply adapting Vitevitch and Luce (1998, 1999) materials, Pylkkänen et al. (2002) actually performed the first experiment in which written words had their *phonological* neighborhood frequency controlled. Research of neighborhood effects in reading has practically entirely ignored the issue of whether or not phonological neighborhoods might actually play a role in visual word recognition. This might stem from the fact that, given the extremely complicated way English orthography maps into phonetics, many researchers do not believe that reading (at least in English) involves a mandatory phonological recoding stage (see summary of arguments in Balota, 1994), and can therefore be autonomously studied without reference to phonology.

A unifying argument could be made then that the reason why one consistently fails to find reliable inhibitory orthographic neighborhood frequency effects in English is due to the fact that phonology is indeed accessed in reading. Given that English orthography does not map straightfowardly to phonology, there could be considerable mismatch between orthographic neighborhoods of English words and their phonological neighborhoods. In other languages where phonology is more transparently represented in the orthography, phonological and orthographic neighborhoods of words will mostly overlap. Incidentally, it is precisely in these languages that effects of neighborhood inhibition are found for written words. If one claims that the source of all those effects is in the phonological neighborhoods, one is then able to unify all the discrepant inhibitory effects found across languages and modalities: one finds reliable inhibitory neighborhood effects in spoken word recognition because phonological neighborhoods are always at play in the auditory modality. One finds reliable

[3]Westbury and Buchanan (2002, p. 1) note, this unpervasivenss of high frequency words in displaying different kinds of effects while low frequency words do is entirely compatible with the rest of the literature for written word recognition in English.

inhibitory neighborhood frequency effects in languages with transparent orthographies because orthographic neighborhoods in these languages mirror phonological neighborhoods, and since phonological forms are accessed in reading, phonological neighborhoods exert their inhibitory effect. Finally, the mixed pattern of orthographic neighborhood effects in English is found because orthographic neighborhoods of English words do not necessarily match their phonological neighborhoods. Pylkkänen et al. (2002)'s behavioral results then are entirely compatible with a direct prediction of this hypothesis: once *phonological* neighborhood frequency is controlled, English should start behaving like other languages.

Furthermore, this proposal would also be compatible with the tacit assumption upon which Pylkkänen et al. (2002)'s phonotactic facilitation results rests: that reading activates phonological units. Without positing this, there is no clear way in which to understand their phonotactic probability facilitation effect. There is to my knowledge no evidence for facilitatory *orthotactic* effects in reading available in the literature (but see Westbury & Buchanan, 2002).

Unfortunately, the little evidence on the relationship between orthographic and phonological neighborhoods in reading that is available in the literature goes entirely *against* this hypothesis. Yates (2005; Yates, Locker, & Simpson, 2004) has found *facilitatory* phonological neighborhood effects in English for written words. Moreover, Ziegler and Perry (1998), Peereman and Content (1997) and Grainger, Muneaux, Farioli, and Ziegler (2005) found that when orthographic neighborhoods included phonological neighbors in French, *facilitatory* neighborhood effects were found for visually presented words. On that note, it is interesting to notice that Stockall et al. (2004) themselves failed to find phonological neighborhood size *inhibition* effects on reaction time latencies for their visually presented words. Instead, they reported a null neighborhood size effect for high frequency words and a *facilitatory* effect for low frequency words.

Therefore, the fact that Pylkkänen et al. (2002) did find *inhibitory* effects of phonological neighborhood frequency in reading is surprising and goes against most of the reported results in the field. Given the importance of their findings, this issue merits further investigation.

Another potential problem that was introduced by Pylkkänen et al. (2002)'s translation of (Vitevitch & Luce, 1998, 1999) materials from auditory form to visual representation is that stimuli length ceased to be controlled: Low probability/density items were significantly longer in their visual form than high probability/density items. Although Pylkkänen et al. (2002) do acknowledge this fact, they assume that this is actually biasing the materials *against* their hypothesis of longer reaction times for high probability/density words. This logic assumes that length has inhibitory effects in visual word recognition and works as follows: all else being equal, low probability/density items should be responded to faster than high probability/density items. However, if the low probability/density items are longer than high probability/density items, and length is inhibitory, then the predicted difference in reaction times due to phonotactics might disappear purely due to the effect of length.

This reasoning is problematic for three reasons. First, it assumes that length is inhibitory. Although several inhibitory results have indeed been found, there is great discrepancy in results reported in the literature (see New, Ferrand, Pallier, & Brysbaert, 2006 for a review), and null effects and even facilitatory effects have also been found. For instance, New et al. (2006) report facilitatory effects of length for short words (3 to 5 letters), null effects for middle sized words (5 to 8 letters), and inhibitory effects only for long words (8 to 11 letters). In a similar fashion, Balota et al. (2004) report facilitatory effects for high frequency words in university students (as opposed to elders) performing lexical decision tasks. They also report null effects for middle range frequency words. Balota et al. (2004)'s results are particularly important because only monosyllabic words were used in their experiment, and therefore the length effect is entirely independent from the effect of syllabic length, much like (Vitevitch & Luce, 1998, 1999) materials in their written form.

The second reason to think that length being biased in Pylkkänen et al. (2002)'s materials is a problem is that, even though the authors do not make the direct connection, the main effect of length is entirely confounded with their reported main effect of probability. In other words, it is impossible to rule out based on their data alone the interpretation that the effect they find on the M350 is due to phonotactic probability rather than orthographic length. This is a serious concern for the interpretation of their data, since recent electrophysiological studies have shown how word length has surprisingly early effects (60–100ms) that can persist for quite some time (from 200 to 600ms), and interact with other lexical level variables, such as frequency (Assadollahi & Pulvermüller, 2003; Wydell, Vuorinen, Helenius, & Salmelin, 2003; Hauk & Pulvermüller, 2004). A possible counter argument to that concern, however, is that Stockall et al. (2004) did report a phonotactic probability main effect on the M350, and orthographic length was controlled in their materials.

Finally, the last issue with Pylkkänen et al. (2002)'s study has to do with their conclusion and whether or not it is consistent with their results. They concluded that the M350 tracks lexical activation but not lexical competition. However, they did report that lexicality interacted with probability, but the effect was marginal. However, they only tested 10 people, and the effect was marginal at $p = 0.05$. It is very likely that, had more subjects been tested, the interaction would turn out to have been a main effect. If this is true, then it seems that Pylkkänen et al. (2002) have to account for why lexicality would be interacting with probability in pre-decisional, lexical activation stages, where only potential candidates are being generated.

In summary, both studies that reported M350 modulation due to phonotactic probability (Pylkkänen et al., 2002; Stockall et al., 2004) present very interesting results, but also challenging problems that raises questions about the interpretation of their results. On the one hand, the fact that they used a different modality for their stimuli rather than the one used by the original studies they based their work on (Vitevitch & Luce, 1998, 1999) complicates the interpretation of the results in two ways. First, competition effects in the visual modality are not routinely found, especially in English. Second, it is not yet clear that (a) accessing phonology from reading is a mandatory step in visual word recognition, (b) that the effects of phonology on reading, which are mostly *facilitatory*, are similar to the effects of phonology in spoken word recognition, which seem to be mostly inhibitory. In the absence of a clear phonological activation, one could claim that the effect found by Pylkkänen et al. (2002) was due not necessarily to phonotactic probability, but to *orthotactic* probability. However, such effects have not been reported in the literature about visual word recognition. Moreover, one of the studies (Pylkkänen et al., 2002) has a potentially damaging visual word length confound.

Given the theoretical interest that a pre-decisional measure of lexical activation has for models of written and spoken word recognition, and in the light of all these considerations, we decided to replicate the design of Vitevitch and Luce (1998, 1999) and Pylkkänen et al. (2002) and extend the results to the auditory modality, while addressing the issues raised for the confounds for the visual modality.

Experiment 1: Modality of presentation and the effects of Neighborhood Density and Phonotactic Probability

The first problem identified in the studies by Pylkkänen et al. (2002) and Stockall et al. (2004) is that when it comes to neighborhood effects, phonological units unexpectedly seem to have largely opposite influences according to the modality of stimulus presentation. For spoken words, inhibitory effects for larger and more frequent neighborhoods are found, whereas for written words, facilitatory effects are found for larger number of neighbors, and conflicting results are found for neighborhood frequency. By choosing to change the modality of the presentation of their stimuli, the interpretation of their results became muddled by these conflicting effects that modality of stimulus presentation seem to have on lexical competition. The only way to address this issue is to try to replicate Pylkkänen et al. (2002)'s findings in both modalities at the same time and see how convergent or divergent they are for the same set of materials. This problem is further complicated by recent studies (eg. Ziegler, Muneaux, & Grainger, 2003) showing that orthography actually plays a modulatory role in the recognition of *auditorily* presented words. Therefore, if one wishes to attempt a direct comparison between auditory and visual word recognition, great attention should be given in the construction of the materials. For instance, one should choose visual words with very transparent pronunciations and straightforward mapping to phonology.

The second issue we wanted to address relates to the confounding effect of length in Pylkkänen et al. (2002). Although one could in principle argue this issue was settled since Stockall et al. (2004) controlled their materials for orthographic length and still found the same M350 modulation that Pylkkänen et al. (2002) did, there are good reasons to think otherwise. Orthographic length in Stockall et al. (2004) was only controlled in their conditionwise average. In other words, while Stockall et al. (2004)'s materials were all strictly matched for phonetic length (they were all CVC words and nonwords), they were allowed to vary in orthographic length (3 to 5 characters), and only the averaged length was controlled across lists. However, given the nature of electrophysiological experiments, which require artifact rejection on top of error rejections, up to 30% of trials for any given condition could be thrown out for each subject. The data that is averaged across participants then could end up being, in the worst case scenario, 60% different. This fact, together with the observation that *means* are very non–robust *measures of location* of any distribution of values (see for instance Wilcox, 1997, chapter 1 for a particularly dramatic illustration) makes it very possible that real differences might simply not be detected if one relies on simple null hypothesis testing in order to assess whether the differences between two populations of values is statistically significant[4]. Therefore, variables such as orthographic length, that were only matched on their average over *complete* lists, could very well turn out to be uncontrolled and biased in unkonwn ways in the data that *actually* ends up being averaged. In order to sidestep this issue, one should control orthographic length in the same way that phonetic length was.

Finally, in order to increase our chances of replicating the Pylkkänen et al. (2002) and Stockall et al. (2004) results in both visual and auditory modality simultaneously, we need to make sure that both orthotactic and phonotactic probability are indeed very distinct in our materials. Due to all these constraints, we were unable to find enough CVC words per condition such as to obtain an adequate signal–to–noise ratio required by an

[4]In fact, it is a statistical fallacy to assume that a lack of a statistically significant result between two populations implies equality of the populations. See (Tryon, 2001) for details.

MEG experiment. Therefore, we chose to use slightly longer words (CV.CVC).

With such materials, we expected to be able to replicate Pylkkänen et al. (2002) and Stockall et al. (2004)'s results, and extend their findings to the auditory modality, which still has not been extensivley done in MEG (see Bowles & Poeppel, 2005 for the only other attempt to investigate spoken word recognition using MEG).

Methods

Participants. 20 native English speakers (8 women) participated in the experiment (Mean age: 21, age range: 18-28). All were right-handed (Oldfield, 1971), had normal or corrected to normal vision, and reported no history of hearing problems, language disorders or mental illness. All participants gave their written informed consent to take part in the study, which was approved by the University of Maryland Institutional Review Board. Subjects were paid for their participation.

Design. A 2x2 design with Lexicality (levels: word vs non–word) and Phonotactic/Orthotactic Probability / Neighborhood Density (Probability/Density) (levels: High/Dense vs Low/Sparse) as factors was used, yielding the following four experimental conditions: (1) High Probability / Dense Neighborhood words, (2) Low Probability / Sparse Neighborhood words, (3) High Probability/ Dense Neighborhood non–words, (4) Low Probability / Sparse Neighborhood non–words.

Materials. Forty-two items in each condition were used. All were five letters long in their visual form and five segments long in their phonetic form (as transcribed by the Carnegie Mellon Pronouncing Dictionary v0.6, 1998). All were of CV.CVC syllabic structure, with first syllable stress. Word items were all primarily nouns and adjectives, according to their frequency count in the CELEX database. Nonword items, due to their syllabic structure, could have other pronunciations in their visual form rather than the one intended by the experimenters and therefore three native speakers were asked to read them aloud; all used the pronunciation that matched the string of segments intended by the experimenters.

Auditory stimuli were recorded in a sound attenuated room by a female native English speaker into a digital recorder at a sampling rate of 44.1 kHz, and saved as 16-bit mono WAV format. The stimuli were low-pass filtered at 10 kHz, subsequently edited into individual sound files, whose duration and RMS amplitude were equated across conditions, and then gated on and off using 10 ms cosine-squared ramps.

Lexical frequency Lexical frequency (in *log*) was controlled for words across the Probability/Density bin. Equality of the frequency means of the High/Dense condition ($\bar{x} = 7.8$, $SE = 0.2$) with those of the Low/Sparse condition ($\bar{x} = 7.8$, $SE = 0.2$) was assessed by generating a 95% Confidence Interval for each condition via the bias corrected and accelerated bootstrap (BCa) method (Efron & Tibshirani, 1993) based on 10000 bootstrap samples. There was almost perfect overlap in their confidence intervals (see figure 1), and therefore we considered the two means statistically equal (Tryon, 2001).

Phonotactic and Orthotactic Probability One of our goals was to directly compare the recognition process of the same set of stimuli varying only on their modality of presentation. Therefore, we had to manipulate both the orthotactic and the phonotactic probabilities of our materials. Orthotactic probability refers to the co–occurence frequency of subsets of characters that comprise the word in its orthographic form. Phonotactic probability refers to the co-occurence frequency of subset of segments (or phones) that comprise the word in its phonetic form. We used three subset sizes (also know as *n–grams*): unigrams, bigrams and trigrams, referring respectively to a subset of size one (one character or segment), two (a sequence of two characters or segments) and three (a sequence of three characters or segments).

Moreover, there are a number of dimensions along which one could compute phonotactic and orthotactic probabilities. We chose to use both type and token counts, both in position–independent and position–dependent contexts. *Type counting* consists in determining how many other words in the corpus share a specific set of characters / segments. Token counting consists in adding the log–frequency of the other words in the corpus that

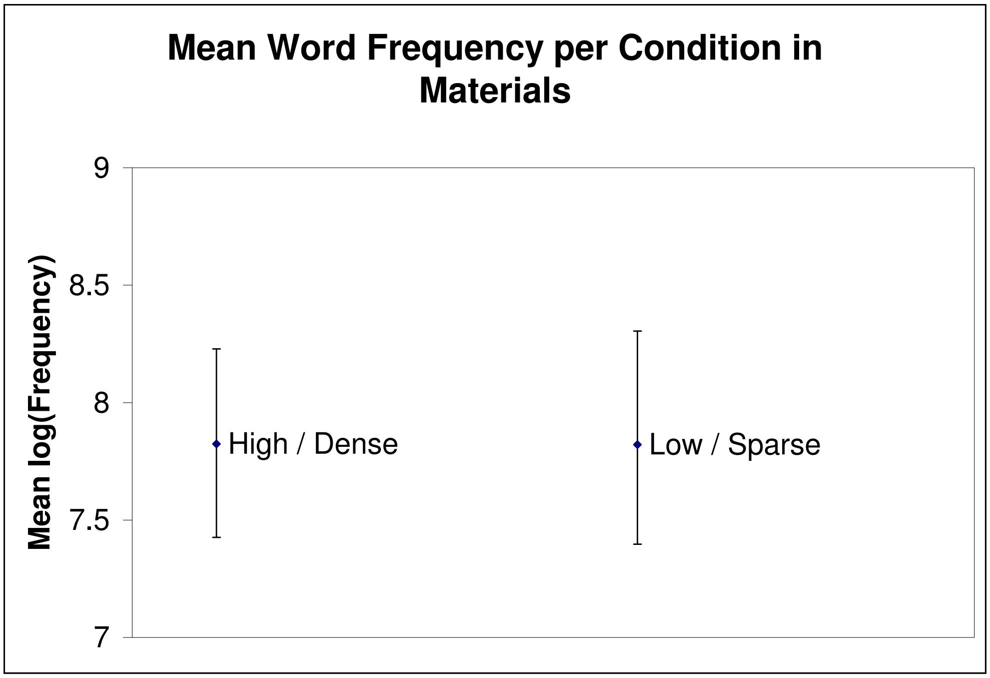

Figure 1. Materials: Mean lexical frequency of High/Dense and Low/Sparse words. The bars denote 95% Confidence Intervals derived by BCa bootstrap (10000 samples). Since their overlap is almost perfect, we can consider the two mean lexical frequency of the two conditions statistically equivalent, and therefore claim that lexical frequency was controlled in our materials.

share a specific set of characters / segments[5]. Position–dependent counts take into consideration not only the set of characters / segments, but also their position in the word. Position–independent counts take into consideration only the set of characters / segments, independent on the position they occur in the word.

For instance, consider the word *cat*. Suppose we are calculating the overall phonotactic probability of this word, and we are analyzing its bigrams. This word has two bigrams *ca* and *at*. Consider the first bigram, *ca*. If we are establishing how frequent this bigram is by

[5]This serves as a surrogate to counting every instance of the word in a corpus

type counting, we would look for other words that share the same bigram. For each word we find, like ***ca**p* and ***ca**n*, we would add 1 to our count. By the end of the count, we will have the number of words that share the bigram *ca* with the word *cat*. If we are establishing how frequent this bigram is by *token counting*, we do the same thing, but instead of adding 1 to our count, we add the corpus frequency of the words that share the bigram under consideration (for instance, we would add 6.4 for *cap* and 8.5 for *can*). Furthermore, if we are limiting ourselves to Position–Dependent counting, we would only consider matches for the bigram *ca* of *cat* other words that have the same bigram in the same position. In our example the words ***ca**p* and ***ca**n* would count as matches, but not the words *Jamai**ca*** or *lo**ca**l*. If we are counting in a Position–Independent way, however, all of the aforementioned words would count as matches for the bigram *ca* in *cat*.

Therefore, for each *n–gram* size we used, we had four different ways of computing orthotactic and phonotactic probability, and all were explicitely manipulated across the Probability conditions. Thus, items in the High Probability bin had both significantly higher phonotactic and orthotactic probabilities than their counterparts in the low probability bin according to type and token unigram, bigram and trigram counts, both in their position–dependent and position–independent variety, totaling 12 different dimensions of phonotactic probability and Neighborhood Density (see tables 1 and 2 for more details). Statistical significance was established by multiple one–tailed t–tests for each relevant contrast (High vs Low words and High vs Low nonwords). *P*–values were adjusted by Holm's method for multiple comparisons (Holm, 1979).

Moreover, effort was put into matching these same parameters for items in the same phonotactic and orthotactic probability bin. Therefore, words and non–words in each probability bin were highly similar in their phonotactic and orthotactic composition.

Neighborhood Density Similarity neighborhoods were also manipulated in both their orthographic and phonetic dimensions. Two measures of neighborhood density were estimated for each modality: the number of words in the corpus that could be found by either

Table 1: Materials: Phonotactic Probability for WORDS. Each score represents the mean difference between the Probability levels, and is obtained by subtracting the Low Probability mean from its corresponding High Probability mean, for each particular *n–gram*. Thus, each score in the table quantify the difference between Probability conditions. Values larger than zero indicate that the High Probability mean is larger than the Low Probability mean. The larger the value, the larger the difference. Statistical significance for each difference score was verified by multiple one–tailed t–tests for the appropriate contrasts (High vs Low words). The p–values for each comparison are displayed in parentheses. All p–values are adjusted for multiple comparisons by Holm's method (Holm, 1979). All p–values were significant at the $\alpha = 0.05$ level, except those marked by an asterisk (*). Legend: Position–Independent Type counts refers to the summed amount of other words that share n-grams with a particular word. Position–Dependent Type counts refers to the summed amount of other words that share n-grams in the same position with a particular word. Position–Independent Token counts refers the summed frequency of all tokens that share n-grams with a particular word. Position–Dependent Token counts refer to the summed frequency of all tokens that share n-grams in the same position with a particular word.

	Unigrams			
	PI-Type counts	PD-Type counts	PI-Token counts	PD-Token counts
Orthographic	24320 ($p < 0.001$)	7061 ($p < 0.001$)	119813 ($p < 0.001$)	36825 ($p < 0.001$)
Phonetic	18105 ($p < 0.001$)	6206 ($p < 0.001$)	111768 ($p < 0.001$)	38549 ($p < 0.001$)
	Bigrams			
	PI-Type counts	PD-Type counts	PI-Token counts	PD-Token counts
Orthographic	4956 ($p < 0.001$)	838 ($p = 0.002$)	25333 ($p < 0.001$)	4509 ($p = 0.003$)
Phonetic	3294 ($p < 0.001$)	1109 ($p < 0.001$)	20446 ($p < 0.001$)	6926 ($p < 0.001$)
	Trigrams			
	PI-Type counts	PD-Type counts	PI-Token counts	PD-Token counts
Orthographic	364 ($p = 0.031$)	69 ($p = 0.031$)	1852 ($p = 0.031$)	354 ($p = 0.031$)
Phonetic	221 ($p = 0.001$)	88 ($p = 0.002$)	1350 ($p = 0.001$)	531 ($p = 0.002$)

adding, deleting or substituting one character or segment from a given word (Coltheart et al., 1977; Vitevitch & Luce, 1999), and the summed log–frequency of these same words. We refer to the first measure as the *neighborhood size*, and to the second measure as the *neighborhood frequency* of a word. Thus, items in the dense neighborhood (DN) bin had a significantly larger number of orthographic and phonetic neighbors which were also significantly more frequent than their counterparts in the sparse neighborhood (SN) bin (see table 3 for more details). Statistical significance was established by multiple one–tailed t–tests for each contrast (Dense vs Sparse words and Dense vs Sparse nonwords). *P*–values were adjusted by

Table 2: Materials: Phonotactic Probability for NONWORDS. Each score represents the mean difference between the Probability levels, and is obtained by subtracting the Low Probability mean from its corresponding High Probability mean, for each particular *n–gram*. Thus, each score in the table quantify the difference between Probability conditions. Values larger than zero indicate that the High Probability mean is larger than the Low Probability mean. The larger the value, the larger the difference. Statistical significance for each difference score was verified by multiple one–tailed t–tests for the appropriate contrasts (High vs Low nonwords). The p–values for each comparison are displayed in parentheses. All p–values are adjusted for multiple comparisons by Holm's method (Holm, 1979). All p–values were significant at the $\alpha = 0.05$ level, except those marked by an asterisk (*).Legend: Position–Independent Type counts refers to the summed amount of other words that share n-grams with a particular word. Position–Dependent Type counts refers to the summed amount of other words that share n-grams in the same position with a particular word. Position–Independent Token counts refers the summed frequency of all tokens that share n-grams with a particular word. Position–Dependent Token counts refer to the summed frequency of all tokens that share n-grams in the same position with a particular word.

	Unigrams			
	PI-Type counts	PD-Type counts	PI-Token counts	PD-Token counts
Orthographic	27379 ($p < 0.001$)	7315 ($p < 0.001$)	135336 ($p < 0.001$)	35921 ($p < 0.001$)
Phonetic	17733 ($p < 0.001$)	5292 ($p < 0.001$)	109128 ($p < 0.001$)	32758 ($p < 0.001$)
	Bigrams			
	PI-Type counts	PD-Type counts	PI-Token counts	PD-Token counts
Orthographic	5188 ($p < 0.001$)	1124 (p= 0.001)	26580 ($p < 0.001$)	6001 (p= 0.001)
Phonetic	3217 ($p < 0.001$)	1247 (p< 0.001)	19868 ($p < 0.001$)	7803 (p< 0.001)
	Trigrams			
	PI-Type counts	PD-Type counts	PI-Token counts	PD-Token counts
Orthographic	244 ($p = 0.065$*)	114 ($p = 0.001$)	1179 ($p = 0.065$*)	583 ($p = 0.001$)
Phonetic	190 ($p = 0.001$)	69 ($p = 0.001$)	1160 ($p = 0.001$)	415 ($p = 0.001$)

Holm's method for multiple comparisons (Holm, 1979)[6]. Moreover, the size and frequency of the neighborhoods was kept constant across words and non–words within each neighborhood bin.

Uniqueness Point Uniqueness point refers to the point where a given string of characters or phonemes becomes unambiguous, i.e., compatible with only one entry in the mental lexicon. We tried to make the Uniquenss Point of our materials always be at the end of the word. The mean Uniqueness Point was always controlled across conditions in both

[6]The neighborhoods and phonotactic and orthotactic comparisons were all carried out simultaneously, and all the p–values were adjusted accordingly

Table 3: Materials: Neighborhood Densities. Each score represents the mean difference between the Density levels. In other words, each score is obtained by subtracting the Sparse Neighborhood mean from its corresponding Dense Neighborhood mean, for each particular estimate of neighborhood size. Therefore, each score in the table counts as a difference score between Density conditions. Values larger than zero indicate that the Dense mean is larger than the Sparse mean. The larger the value, the larger the difference. Statistical significance for each difference score was verified by multiple one–tailed t–tests for the appropriate contrasts (Dense vs Sparse words and Dense vs Sparse nonwords). The p–values for each comparison are displayed in parentheses. All p–values are adjusted for multiple comparisons by Holm's method (Holm, 1979). All p–values were significant at the $\alpha = 0.05$ level, except those marked by an asterisk (*).

WORDS – Neighborhoods: Mean Differences ($Dense - Sparse$)		
	Differences in Lexical Neighborhoods	
	Number of Neighbors	Summed Frequency of Neighbors
Orthographic	3 ($p = 0.001$)	18 ($p = 0.001$)
Phonetic	3 ($p < 0.001$)	21 ($p < 0.001$)

NONWORDS – Neighborhoods: Mean Differences ($Dense - Sparse$)		
	Differences in Lexical Neighborhoods	
	Number of Neighbors	Summed Frequency of Neighbors
Orthographic	3 ($p = 0.001$)	15 ($p = 0.001$)
Phonetic	4 ($p < 0.001$)	20 ($p < 0.001$)

their orthographic and their phonetic form (see table 4 for details). Statistical equivalence of the means was assessed by deriving 95% Confidence Intervals for each condition via the bias corrected and accelerated bootstrap (BCa) method (Efron & Tibshirani, 1993), based on 10000 bootstrap samples. There was almost perfect overlap in their confidence intervals (see figures 2 and 3), and therefore we considered the two means statistically equal (Tryon, 2001). However, a word of caution is warranted about the relevance and interpretation of the Uniqueness Point. The Uniqueness Point is not the same thing as the Recognition Point. It has been known for a long time that speakers can reliably recognize spoken words with only partial information (Marslen-Wilson, 1973; Marslen-Wilson & Tyler, 1975, 1980), so the Uniqueness Point should be taken as an index of how much *possible* ambiguity there is in the signal, not as an index of how much ambiguity there actually is. Moreover, the way we computed Uniqueness Points was very coarse. For instance, according to our

Table 4: Materials: Mean Uniqueness Points.

	Mean Orthographic UP		Mean Phonetic UP	
	Word	Nonword	Word	Nonword
High/Dense	5	5	4.9	4.9
Low/Sparse	4.9	4.9	4.9	4.9

metric, virtually all singular nouns would have their Uniqueness Point at the end of the word due to the simple fact that most nouns have plural forms. Whether or not this is a problem or a feature of our way of quantifying Uniqueness Points depends on how plurals and singular nouns are represented. If plurals and singulars are independently represented in the lexicon, then one should expect competition between the two entries, and therefore considering plurals in our way of quantifying Uniqueness Points is warranted. If plurals and singulars are represented under one entry alone, then we might expect that competition between the forms would not arise, and therefore our way of computing Uniqueness Points might be biased. On this issue, there is a lot of cross–linguistic discrepant results reported in the literature (Baayen, Dijkstra, & Schreuder, 1997; Dominguez, Cuetos, & Segui, 1999; New, Brysbaert, Segui, Ferrand, & Rastle, 2004), but there is evidence that singulars and plurals are indeed independently represented in the lexicon at least for English (Sereno & Jongman, 1997, but see Alegre & Gordon, 1999 for some qualifications about this claim).

Procedure. Subjects were placed horizontally in a dimly lit magnetically shielded room (Yokogawa Corporation, Tokyo, Japan) and were screened for MEG artifacts due to dental work, and excessive eye-blinking. A scout scan was performed with the purposes of verifying the presence of identifiable MEG responses to 1 kHz and 250 Hz pure tones (M100) and determining adequate head positioning inside the machine.

Stimulus presentation and experiment control was carried out by the DMDX program (K. L. Forster & Forster, 2003). In both sections of the experiment (visual and auditory), each presentation of a word or non-word was preceded by the display of a fixation point projected onto the center of a rear-projection screen for 1s. In the auditory section,

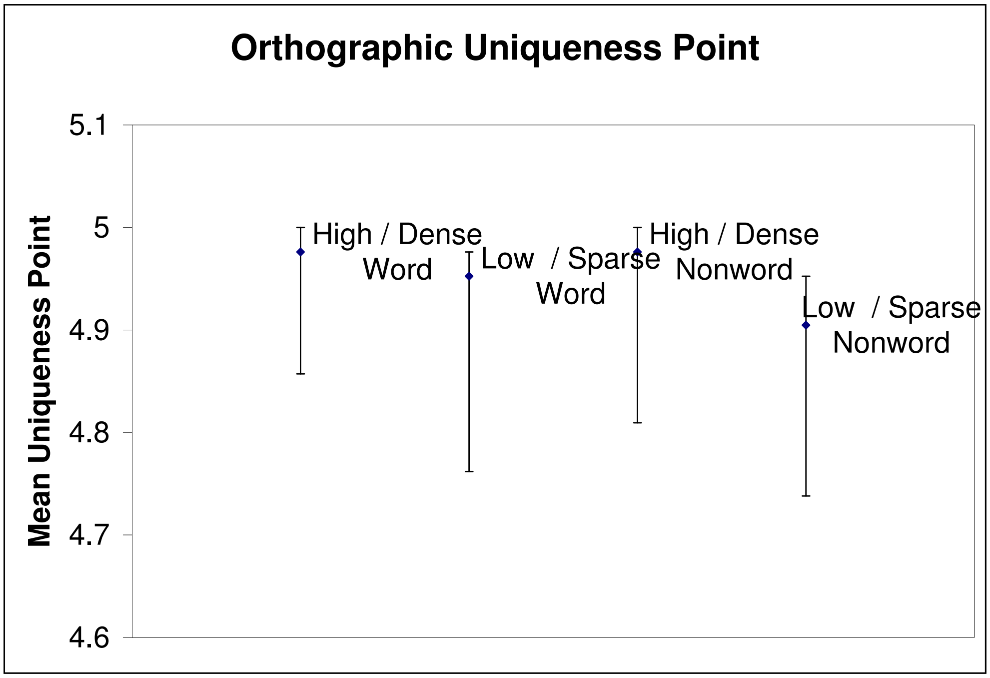

Figure 2. Materials: Orthographic Mean Uniqueness Point for each experimental condition. The bars denote 95% Confidence Intervals derived by BCa bootstrap (10000 samples). Since their overlap is almost perfect, we can consider all condition means to be statistically equivalent, and therefore claim that Uniqueness Point was controlled in our materials.

subjects were asked to keep their eyes closed to reduce the likelihood of eye-blinking artifacts. The interstimulus interval pseudorandomly varied between 400 and 1500ms in both sections.

Subjects were instructed to decide whether each stimulus item was a real word or not (lexical decision), and to respond as quickly and accurately as possible using a button box placed in their right hand. Visual stimuli remained on the screen for 3000ms or until subjects responded. Auditory stimuli were always played in their entirety, and subjects were given 3500ms from the onset of presentation to respond. Accuracy and reaction times

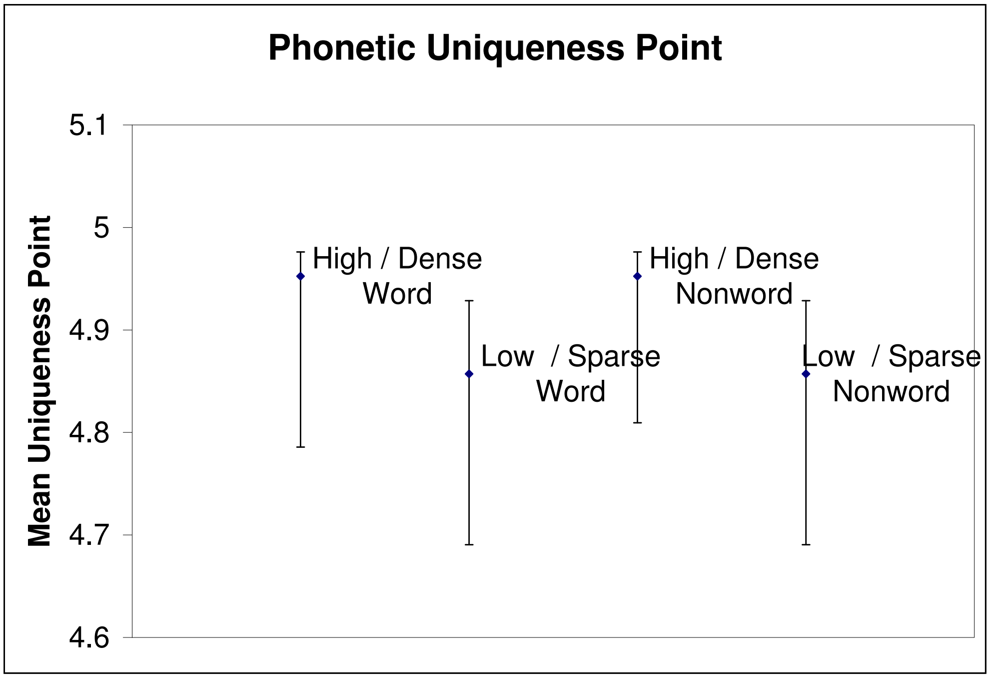

Figure 3. Materials: Phonetic Mean Uniqueness Point for each experimental condition. The bars denote 95% Confidence Intervals derived by BCa bootstrap (10000 samples). Since their overlap is almost perfect, we can consider all condition means to be statistically equivalent, and therefore claim that Uniqueness Point was controlled in our materials.

from the onset of stimulus presentation were recorded.

Auditory stimuli were presented binaurally at 60–70dB SPL over E-A-RTONE® 3A (Aearo Company Auditory Systems, Indianapolis, IN) earphones attached to E-A-RLINK® foam plugs inserted into the ear canal. Visual stimuli were presented using the Courier New font (size 12), in yellow over a black background.

Subjects were first given a practice session of 8 items to help familiarize them with the task. Each section of the experiment was administered in two blocks, in between which participants could take a break. The order of presentation of sections was counter-balanced across subjects. A distracter face recognition task was inserted between sections, and lasted approximately 25 minutes.

MEG data acquisition and analysis. MEG recordings were conducted using a 160–channel axial gradiometer whole-head system (Kanazawa Institute of Technology, Kanazawa, Japan). Data were sampled at 500 Hz and acquired continuously with a bandwidth between DC to 200 Hz. In order to remove external sources of noise artifacts a multi–shift PCA filter (de Cheveigné & Simon, submitted) was used. Epochs with artifacts exceeding ±2 pT in amplitude were removed before averaging. Incorrect behavioral responses and trials where subjects failed to respond were also excluded from both behavioral and MEG data analysis. Subjects with behavioral error rates larger than 10% were excluded from any further analysis. Data from one subject was eliminated based on this criterion. Subjects whose data had less than 30 trials (70%) in any condition surviving error and artifact rejection were also excluded from further analysis. Data from three other subjects were eliminated based on this criterion. Data from 16 subjects remained, with 9.5% of the data from the visual modality section and 10% from the auditory modality section being discarded. Following averaging, data were baseline corrected using a 100ms prestimulus interval and were lowpass filtered at 20 Hz.

For each subject, five sensors in the source (outgoing magnetic field) and five sensors in the sink (ingoing magnetic field) were chosen to represent the M350 component based on which sensors best captured the dipolar fields on the left hemisphere when all trials of each condition for the subject were averaged together. For the visual section, we followed the literature, and only the first root mean square (RMS) peak across the chosen 10 channels for each stimulus condition was analyzed. The latency, amplitude, rise time (time from valley to peak) and overall peak duration (time from the valley preceding the first peak to the valley following the peak) were recorded and used in the data analysis.

For the auditory section, a similar procedure was attempted. However, the auditory evoked response displayed several peaks starting from around 250ms and persisting until

Table 5: Mean reaction time latencies (in ms) for both experimental sections. Standard Errors are presented betweeen parentheses.

Behavioral Results ($N = 16$)				
	Visual		Auditory	
	Word	Nonword	Word	Nonword
High / Dense	647 (22)	687 (19)	1007 (17)	1082 (23)
Low / Sparse	670 (21)	671 (16)	1012 (20)	1039 (20)

around 800ms, with mostly the same dipole distribution. Moreover, the number of peaks was not necessarily consistent across conditions for individual subjects. Therefore, selecting only the first peak in the 300-420ms window with the right dipolar field, as was done for the visual section, would not yield consistent results. Instead, we visually inspected the RMS wave of all the four conditions for each subject, identified the M100 response for each condition, and then proceeded from there, inspecting for each condition each subsequent peak and its distribution, looking for analogues in the other conditions. Peaks that (i) were very salient in all four conditions, (ii) were optimally spatially selected in terms of the dipolar distributions of the 10 selected channels, and (iii) were considered to be analogues of each other based on their comparative temporal morphology since the M100 response were selected for analysis. Like in the visual section, their latency, amplitude, rise time and overall duration were recorded and used in the data analysis.

Results

Behavioral.

Separate Lexicality X Probability repeated–measures ANOVAs were performed on subjects' mean reaction times for each modality section (visual and auditory). Table 5 shows the mean results for both sections.

Visual section. The main effects of Lexicality (F(1,15)= 2.3414; *ns*) and Probability/Density ($F(1,15) = 0.3098$; *ns*) did not reach statistical significance, but the interaction between them did ($F(1,15) = 12.662$; $p = 0.003$). Two post-hoc comparisons (two–tailed

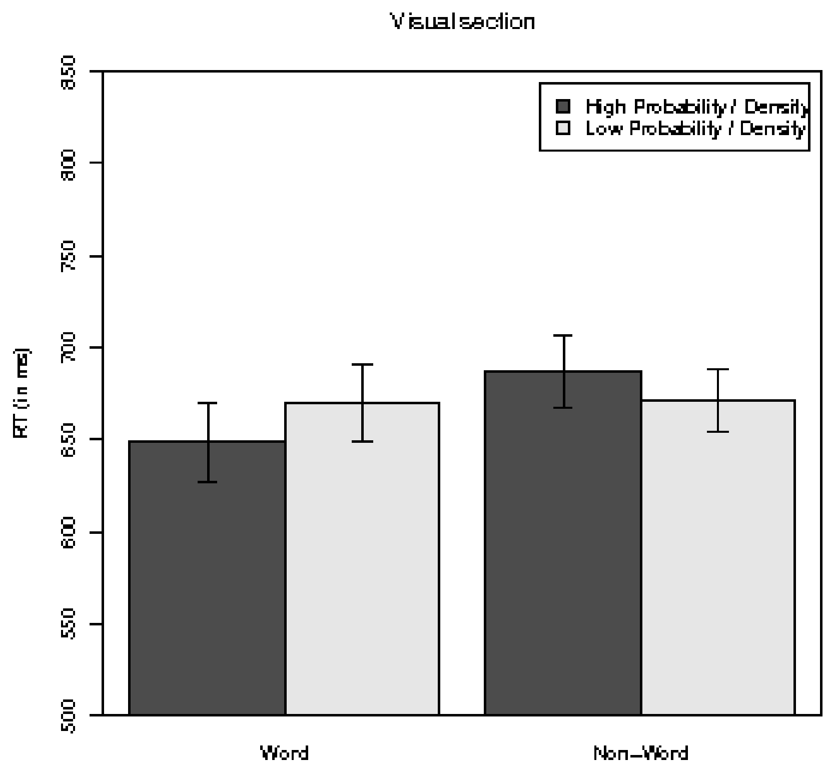

Figure 4. Experiment 1: Visual section mean reaction times (in ms). Error bars show Standard Errors.

paired t–tests, Bonferroni corrected p–values) revealed that reaction times for words in the High/Dense bin ($\bar{x}$ = 647ms) was statistically different from the mean reaction time for words in the Low/Sparse bin ($\bar{x}$ = 670ms; mean difference= 16ms; $t(15) = 2.5442$, $p = 0.044$), the same being true for non–words as well, but in the opposite direction, with High/Dense nonwords ($\bar{x}$ = 687ms) being responded to slower than nonwords in the Low/Sparse condition ($\bar{x}$ = 671ms, mean difference= 22ms; $t(15) = 2.5575$, $p = 0.045$). A summary of the results is presented in figure 4.

Auditory section. The effects of Lexicality ($F(1, 15) = 11.671$; $p = 0.004$) and Probability/Density ($F(1, 15) = 15.511$; $p = 0.001$) were statistically significant, as well as their

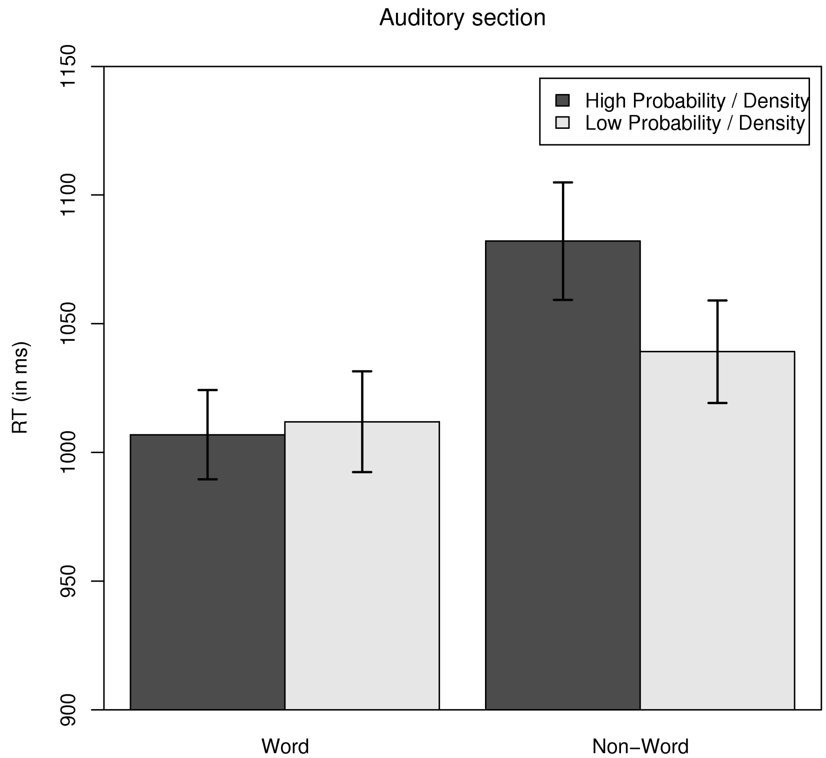

Figure 5. Experiment 1: Auditory section mean reaction times (in ms). Error bars show Standard Errors.

interaction ($F(1,15) = 10.873$; $p = 0.004$). Two post-hoc comparisons (two–tailed paired t–tests, Bonferroni corrected p–values) revealed that reaction times for non–words in the High/Dense condition ($\bar{x} = 1082$ms) were statistically different from reaction times for non–words in the Low/Sparse condition ($\bar{x} = 1039$ms; mean difference= 43ms; $t(15) = 3.8378$, $p = 0.0032$), which was not the case in for words in the two Probability conditions (mean difference= 5ms; $t(15) = 0.9728$, *ns*), as can be seen in figure 5.

M350.

Separate Lexicality X Probability repeated–measures ANOVAs were performed on subjects' M350 peak latencies, amplitudes, rise times and overall durations, for each modality section (visual and auditory).

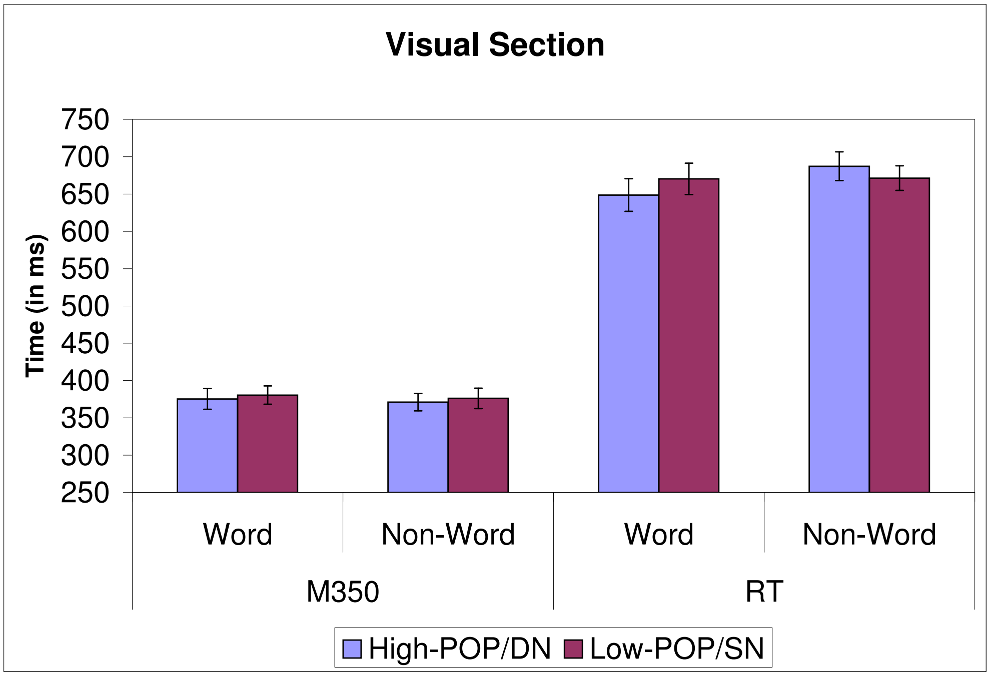

Figure 6. Experiment 1: Visual section mean M350 latencies and mean reaction times (in ms). Error bars show Standard Errors.

Visual section. The data for the M350 peak analysis of the visual section is summarized in table 6. None of the main effects nor interactions turned out significant, as can be seen below:

Peak Latency Main effects of Lexicality ($F(1,15) = 2.7611$, ns) and Probability ($F(1,15) = 2.1429$, ns) were not significant, and neither was their interaction (Lexicality x Probability, $F(1,15) < 0.0001$, ns). See figure 6 for a comparison between reaction time latencies and the M350 latencies.

Table 6: Experiment 1. Visual section M350 peak analysis. Standard Errors are presented in parentheses. Legend: *H/D*=High Probability / Dense Neighborhood, *L/S*=Low Probability / Sparse Neighborhood, *Lat.*=Peak Latency, *Ampl.*=Peak Amplitude, *Rise*=Peak rise time, *Dur.*=Overall Peak Duration.

Visual M350 Peak analysis – Different parameters								
	Lat. (in ms)		Ampl. (in pT)		Rise (in ms)		Dur. (in ms)	
	Word	N-Word	Word	N-Word	Word	N-Word	Word	N-Word
H/D	375 (14)	371 (12)	136 (11)	145 (13)	68 (6)	71 (7)	122 (8)	128 (13)
L/S	381 (12)	376 (14)	139 (10)	137 (14)	66 (7)	67 (8)	125 (16)	117 (10)

Peak Amplitude Main effects of Lexicality ($F(1,15) = 0.3999$, *ns*) and Probability ($F(1,15) = 0.1277$, *ns*) were not significant, and neither was their interaction (Lexicality x Probability, $F(1,15) = 1.294$, *ns*).

Peak Rise Time Main effects of Lexicality ($F(1,15) = 0.1271$, *ns*) and Probability ($F(1,15) = 0.2667$, *ns*) were not significant, and neither was their interaction (Lexicality x Probability, $F(1,15) = 0.0776$, *ns*).

Overall Peak Duration Main effects of Lexicality ($F(1,15) = 0.0089$, *ns*) and Probability ($F(1,15) = 0.17$, *ns*) were not significant, and neither was their interaction (Lexicality x Probability, $F(1,15) = 0.5995$, *ns*).

Auditory section. The data for the M350 peak analysis of the auditory section is summarized in table 7. None of the main effects nor interactions turned out significant, as can be seen below:

Peak Latency Main effects of Lexicality ($F(1,15) = 2.1799$, *ns*) and Probability ($F(1,15) = 1.0299$, *ns*) were not significant, and neither was their interaction (Lexicality x Probability, $F(1,15) = 0.2297$, *ns*). See figure 7 for a comparison between reaction time latencies and the M350 latencies.

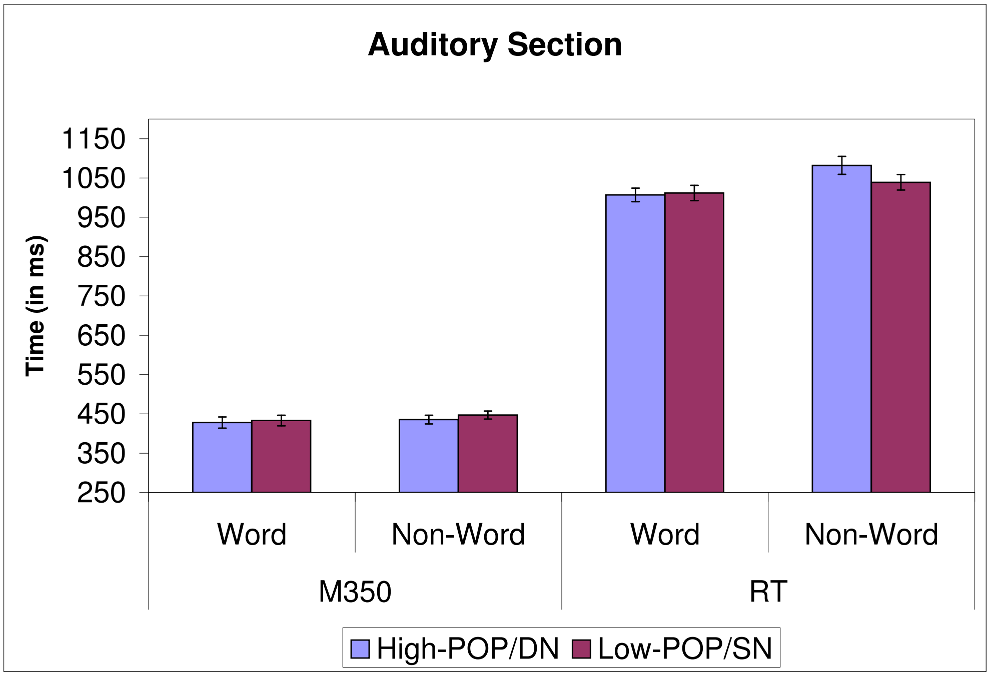

Figure 7. Experiment 1: Auditory section mean M350 latencies and mean reaction times (in ms). Error bars show Standard Errors.

Peak Amplitude Main effects of Lexicality ($F(1,15) = 0.382$, ns) and Probability ($F(1,15) = 0.006$, ns) were not significant, and neither was their interaction (Lexicality x Probability, $F(1,15) = 0.032$, ns).

Peak Rise Time Main effects of Lexicality ($F(1,15) = 1.7572$, ns) and Probability ($F(1,15) = 0.0589$, ns) were not significant, and neither was their interaction (Lexicality x Probability, $F(1,15) = 1.7447$, ns).

Overall Peak Duration Main effects of Lexicality ($F(1,15) = 0.3953$, ns) and Probability ($F(1,15) = 0.5088$, ns) were not significant, and neither was their interaction

Table 7: Experiment 1. Auditory section M350 peak analysis. Standard Errors are presented in parentheses. Legend: *H/D*=High Probability / Dense Neighborhood, *L/S*=Low Probability / Sparse Neighborhood, *Lat.*=Peak Latency, *Ampl.*=Peak Amplitude, *Rise*=Peak rise time, *Dur.*=Overall Peak Duration.

Auditory M350 Peak analysis – Different parameters								
	Lat. (in ms)		Ampl. (in pT)		Rise (in ms)		Dur. (in ms)	
	Word	N-Word	Word	N-Word	Word	N-Word	Word	N-Word
H/D	428 (14)	436 (11)	132 (15)	136 (12)	60 (6)	64 (9)	99 (9)	101 (8)
L/S	433 (14)	447 (10)	133 (15)	136 (12)	54 (5)	74 (12)	102 (8)	113 (12)

(Lexicality x Probability, $F(1, 15) = 0.4149$, *ns*).

Discussion

Behavioral results – Visual section. In the visual modality, our results replicated the standard neighborhood size facilitatory effect that is commonly found in the literature on English visual word recognition (see Andrews, 1997 and Perea & Rosa, 2000 for reviews). Moreover, since in our materials nighborhood size and neighborhood frequency were correlated, and neighborhood frequency is reported to have inhibitory effect on recognition (at least in other languages), we would not have been surprise to find a null or inhibitory effect of neighborhood frequency, since the inhibition would offset the facilitation boost given by neighborhood size. That is however not what we find for words in our data. Therefore this experiment adds to the long list of studies that fail to find inhibitory effects of neighborhood frequency for written words in English(Sears et al., 1995; Sears, Lupker, & Hino, 1999; Sears, Hino, & Lupker, 1999; Huntsman & Lima, 2002; Sears et al., 2006).

Moreover, since phonological and orthographical neighborhoods were correlated as well in our materials, we can compare them to the results reported by Yates; (2005, Yates et al., 2004) for English. Yates (2005 Yates et al., 2004) controlled orthographic neighborhood and manipulated only the phonological neighborhood size of his written words, and did report a *facilitatory* effect of neighborhood size. Our results are compatible with these, since we also found facilitatory neighborhood size results. However, orthographic neighborhood size was confounded with phonological neighborhood size in our materials, so any claims that our effect is a direct result of phonological neighborhood size is definitely not warranted.

On the other hand, Peereman and Content (1997), Ziegler and Perry (1998), and Grainger et al. (2005) found that when orthographic neighborhoods included phonological neighbors in French, *facilitatory* neighborhood effects were found for visually presented words. It is unclear whether this effect is due to the overlap consistency between the two kinds of neighborhoods (are both neighborhoods inhabited by the same words, or different words?), or whether the phonological neighborhood size alone is driving the effect in these languages, like Yates, (2005 Yates et al., 2004) reported for English. Our data cannot answer this question without a thorough evaluation of the phonological–orthographic consistency of our neighborhoods. However, we found the same pattern of results they report.

Furthermore, our results are also partially compatible with Stockall et al. (2004)'s, who reported a facilitatory effect of phonological neighborhood in the recognition of visually presented words in the low frequency range.

The picture that emerges from the result of all these studies is that, at least in English, phonological neighborhood size and frequency do not seem to compete with the effects of orthographic neighborhood size and frequency. Pylkkänen et al. (2002)'s behavioral result remains to my knowledge the only study to have found *inhibitory* phonological neighborhood frequency effects in visually presented words.

We did however find the exact *opposite* effect for our nonwords. This *does* replicate Pylkkänen et al. (2002)'s behavioral findings for nonwords, and indicate that perhaps there is a yet unknown asymmetry in the way phonological neighborhoods influence recognition of visual words and nonwords.

Behavioral results – Auditory section. Although our nonword data does replicate the inhibitory effect of neighborhood size and frequency in auditory word recognition (eg. Goldinger et al., 1989, 1992; Luce & Pisoni, 1998), we failed, much to our surprise, to replicate it in our word data, finding instead a null effect.

After careful analysis of some of the results in the literature, however, we came under the impression that *even in the auditory domain*, the *inhibitory* neighborhood size and frequency effects are not so well established, especially when it comes to reaction time latencies in auditory lexical decision task. Many of the studies that report inhibitory effects of phonological neighborhood side and frequency use disparate dependent measures, and there are often discrepancy and inconsistencies between them. Some studies find effects of inhibition only in error rates, but not in reaction time latencies (Cluff & Luce, 1990). Vitevitch and Luce (1999), for example, finds inhibition effects for words in naming latencies, but not in speeded same–different judgment tasks, which Luce and Large (2001) do find. Even more puzzling, Vitevitch and Luce (1999, experiment 3) report a main inhibitory effect of neighborhood frequency for reaction time latencies in the recognition of words and nonwords, much like we did. However, they do not report the planned comparisons within each lexicality bin, as they do for all their other experiments. Given their graphs, there is a possibility that this main inhibition effect might be like ours, driven exclusively (or mainly) by the nonword results. Finally, in a recent study Vitevitch and Rodríguez (2005) found *facilitatory* effects of neighborhood size and family in spoken word recognition in Spanish. This suggests that to the extent that the neighborhood inhibition effect for spoken words is real, it is weaker and less reliable than what has been previously suggested.

M350 data. We were expecting to replicate Pylkkänen et al. (2002)'s and Stockall et al. (2004)'s studies and find reliable phonotactic and orthotactic probability effects in our electrophysiological data in both modalities. Regardless of how the behavioral data turned out, we antecipated large probability effects, due to the magnitude of the manipulation of this variable in our materials, and simultaneous control of possible interefering factors. However, instead of finding a robust facilitatory effect on the M350 latency, we failed to find in either modality of presentation *any* effect of phonotactic and orthotactic probability in *any* of the four different M350 parameters we analyzed. In fact, the M350 latency looks exactly the same across conditions in our experiment. This strongly suggests that contrary to Pylkkänen et al. (2002)'s interpretation, the M350 is indeed *not* tracking phonotactic probability.

It is somewhat surprising that no effect of lexicality has been observed in the M350 latency, since effects of lexical frequency have been reported to modulate the peak of the response (Embick et al., 2001). However, Stockall et al. (2004) has also failed to observe a lexical frequency–induced modulation of the M350 latency.

Although our results do not replicate Pylkkänen et al. (2002) and Stockall et al. (2004), we have at least two good reasons to believe our results over theirs. First, their results could be due to either uncontrolled variables (such as Orthographic Length, that is confounded with Phonotactic Probability in Pylkkänen et al. (2002)), or to a condition–wise averaged controlled variable ended up uncontrolled and biased due to the fact that artifact and error rejection procedures can discard large portions of the materials. The second (and perhaps most important) reason is that it is very hard to accomodate the central assumption that phonotactic probability will play an effect in reading in the exact same way that it does in auditory word recognition with the findings reported in the literature that phonology affects reading in the opposite way it affects spoken word recognition. We believe our study had less possible sources of interference and more power to detect a phonotactic probability effect (if there was one), due to both our attempt to use auditorily presented words, and our careful manipulation of sublexical probability in our materials.

General Discussion and Future Directions

This work has focused on a very specific topic, a putative neural correlate of word recognition. Given the potentially central role that such a dependent measure could have in current research in lexical access, and the possibilities that it would offer to researchers interested in the nature of lexical representations and how they are retrieved and put to use, we felt justified in trying to replicate and extend to the auditory modality the exciting results reported by Pylkkänen et al. (2002) and Stockall et al. (2004), and while doing so, we tried to solve some problems and inconsistencies we observed in their results and other current findings in the behavioral literature.

However, the final result was mixed at best. First, our word data was entirely compatible and in line with previous findings in the visual word recognition literature, whereas our auditory data added more ambiguity to a series of conflicting findings previously reported about the effect of neighborhood density and frequency in the recognition time of spoken words. Our nonword data in both modalities seemed to pattern with the standard claims that neighborhood size and frequency inhibit recognition. At present we do not have any explanation for this, but we do note that several studies have identified variables that interact with lexical frequency (eg. Westbury & Buchanan, 2002; Andrews, 1997) in both linear and nonlinear ways (New et al., 2006), and this could be a source of some of the discrepancy between our word and nonword data (and perhaps in fact of much of the discrepancy in the literature).

However, if that is the case, then it becomes extremely difficult to perform any kind of factorial parametric investigation of lexical properties, especially in electrophysiological experiments where artifact and error rejection can exclude large portions of materials, thereby greatly increasing the risk of eliminating conditionwise averaging controls over materials that are scant to begin with. This concern is not new (Cutler, 1981), and perhaps it is time that approaches such as the ones proposed by Baayen (2004) and Balota et al. (2004) of using more sophisticated modeling in multiple regression designs become more widespread (see Max & Onghena, 1999; Baayen, 2004 for reviews, and Hauk, Davis, Ford, Pulvermüller, and Marslen-Wilson (2006) for an example of this kind of analysis on electrophysiological data)[7].

[7] Indeed, most of the experiments about lexical properties such as neighborhood frequency rely on different lists of materials for each condition. Each word has inherently different and unique characteristics, and for any arbitrary grouping of words, differences are to be expected. Therefore, it is not surprising at all that differences between conditions will be found in this kind of experiment. The crux of the matter is that any such difference is only interpretable to the extent that all other possibly confounding variables have been controlled. However, in the case of word recognition, the list of variables that has been shown to influence performance is huge, but only a handful are ever controlled in any single experiment. Perhaps that is why results are so discrepant across languages and manipulations. If we can sidestep this kind of design, perhaps we will get data that is more easily interpretable.

Moreover, our MEG data found absolutely no effect of phonotactic and orthotactic probability in the M350. We propose that this response does not track phonotactic probability at all. On a more positive note, we were able to identify and study an auditory correlate of the M350 response, which strongly suggests that this response is independent of the modality of input. Moreover, the auditory M350 is very reliably elicited, which is encouraging for future studies.

References

Alegre, M., & Gordon, P. (1999). Frequency Effects and the Representational Status of Regular Inflections. *Journal of Memory and Language*, *40*(1), 41-61.

Almeida, D. (2005). *LING646 Final paper Experiment Proposal.* (Unpublished manuscript)

Andrews, S. (1989). Frequency or neighborhood effects on lexical access: activation or search? *Journal of Experimental Psychology: Learning, Memory & Cognition*, *15*, 802-814.

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory & Cognition*, *18*, 234-254.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, *4*(4), 439-461.

Assadollahi, R., & Pulvermüller, F. (2003). Early influences of word length and frequency: a group study using MEG. *Neuroreport*, *14*(8), 1183–1187.

Baayen, R. H. (2004). Statistics in Psycholinguistics: A critique of some current gold standards. In *Mental lexicon working papers* (Vol. 1, p. 1-45). Edmonton.

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and Plurals in Dutch: Evidence for a Parallel Dual-Route Model. *Journal of Memory and Language*, *37*(1), 94-117.

Balota, D. A. (1994). Visual word recognition: The journey from features to meaning. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (p. 303-358). New York: Academic Press.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *J Exp Psychol Hum Percept Perform*, *10*(3), 340–357.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. (2004). Visual word recognition of single-syllable words. *J Exp Psychol Gen*, *133*(2), 283–316.

Barrett, S. E., & Rugg, M. D. (1989). Event-related potentials and the semantic matching of faces. *Neuropsychologia*, *27*(7), 913–922.

Barrett, S. E., & Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain Cogn*, *14*(2), 201–212.

Barrett, S. E., Rugg, M. D., & Perrett, D. I. (1988). Event-related potentials and the matching of familiar and unfamiliar faces. *Neuropsychologia*, *26*(1), 105-117.

Beltrán, D., Carreiras, M., Alvarez, M., & Santamaría, C. (2006). Verification of quantified sentences: An ERP study. In *Poster presented at the cognitive neuroscience society annual meeting* (p. 233). San Francisco. (Poster G105 – Presented at the Cognitive Neuroscience Society annual meeting)

Bentin, S., Kutas, M., & Hillyard, S. A. (1995). Semantic processing and memory for attended and unattended words in dichotic listening: behavioral and electrophysiological evidence. *J Exp Psychol Hum Percept Perform*, *21*(1), 54–67.

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalogr Clin Neurophysiol*, *60*(4), 343–355.

Beretta, A., Fiorentino, R., & Poeppel, D. (2005). The effects of homonymy and polysemy on lexical access: an MEG study. *Brain Res Cogn Brain Res*, *24*(1), 57–65.

Bobes, M. A., Valdés-Sosa, M., & Olivares, E. (1994). An ERP Study of Expectancy Violation in Face Perception. *Brain and Cognition*, *26*, 1-22.

Bowles, A., & Poeppel, D. (2005). Modality-independent and modality-specific aspects of the MEG response during lexical decision. In *Poster presented at human brain mapping.*

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, *5*, 34-44.

Cluff, M. S., & Luce, P. A. (1990). Similarity neighborhoods of spoken two-syllable words: retroactive effects on multiple activation. *J Exp Psychol Hum Percept Perform*, *16*(3), 551–563.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (p. 535-555). Hillsdale, NJ: Erlbaum.

Coulson, S., Federmeier, K. D., Van Petten, C., & Kutas, M. (2005). Right hemisphere sensitivity to word- and sentence-level context: evidence from event-related brain potentials. *J Exp Psychol Learn Mem Cogn*, *31*(1), 129–147.

Coulson, S., & Wu, Y. C. (2005). Right hemisphere activation of joke-related information: an event-related brain potential study. *J Cogn Neurosci*, *17*(3), 494–506.

Cutler, A. (1981). Making up materials is a confounded nuisance: or Will we be able to run any psycholinguistic experiments at all in 1990? *Cognition*, *10*, 6570.

de Cheveigné, A., & Simon, J. Z. (submitted). MEG denoising based on Multi-Shift PCA.

Deacon, D., Hewitt, S., Yang, C., & Nagata, M. (2000). Event-related potential indices of semantic priming using masked and unmasked words: evidence that the N400 does not reflect a post-lexical process. *Brain Res Cogn Brain Res*, *9*(2), 137–146.

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: a proposal. *Trends Cogn Sci*, *9*(7), 335–341.

Dhond, R. P., Buckner, R. L., Dale, A. M., Marinkovic, K., & Halgren, E. (2001). Spatiotemporal maps of brain activity underlying word generation and their modification during repetition priming. *J Neurosci*, *21*(10), 3564–3571.

Dhond, R. P., Marinkovic, K., Dale, A. M., Witzel, T., & Halgren, E. (2003). Spatiotemporal maps of past-tense verb inflection. *Neuroimage*, *19*(1), 91–100.

Dhond, R. P., Witzel, T., Dale, A. M., & Halgren, E. (2005). Spatiotemporal brain maps of delayed word repetition and recognition. *Neuroimage*, *28*(2), 293–304.

Dhond, R. P., Witzel, T., Dale, A. M., & Halgren, E. (2006). Spatiotemporal cortical dynamics underlying abstract and concrete word reading. *Hum Brain Mapp*.

Dominguez, A., Cuetos, F., & Segui, J. (1999). The Processing of Grammatical Gender and Number in Spanish. *Journal of Psycholinguistic Research*, *28*(5), 485-498.

Drenhaus, H., Graben, P. beim, & Frisch, S. (2006). Not all but some ERP results on the scalar expressions 'some' and 'all'. In *Poster presented at the cognitive neuroscience society annual meeting* (p. 219). San Francisco. (Poster G37 – Presented at the Cognitive Neuroscience Society annual meeting)

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap.* New York: Chapman & Hall.

Embick, D., Hackl, M., Schaeffer, J., Kelepir, M., & Marantz, A. (2001). A magnetoencephalographic component whose latency reflects lexical frequency. *Cognitive Brain Research*, *10*, 345-348.

Fiorentino, R., & Poeppel, D. (in press). Compound words and structure in the lexicon. *Language and Cognitive Processes.*

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, *20*(4), 400–409.

Fischler, I., & Raney, G. E. (1989). Language by eye: Behavioral, autonomic and cortical approaches to reading. In J. R. Jennings & M. G. H. Coles (Eds.), *Handbook of cognitive psychology:*

Central and autonomic nervous system (p. 511-574). New York: Wiley.

Forster, K. I. (1976). Accessing the mental lexicon. In R. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (p. 257-287). North–Holland.

Forster, K. I., & Shen, D. (1996). No enemies in the neighborhood: Absence of inhibitory neighborhood effects in lexical decision and semantic categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *22*(3), 696-713.

Forster, K. L., & Forster, J. (2003). DMDX: a Windows display program with millisecond accuracy. *Behavior research methods, instruments, & computers*, *35*(1), 116-124.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, *28*, 501-518.

Goldinger, S. D., Luce, P. A., Pisoni, D. B., & Marcario, J. K. (1992). Form-based priming in spoken word recognition: the roles of competition and bias. *J Exp Psychol Learn Mem Cogn*, *18*(6), 1211–1238.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: a multiple read-out model. *Psychol Rev*, *103*(3), 518–565.

Grainger, J., Muneaux, M., Farioli, F., & Ziegler, J. C. (2005). Effects of phonological and orthographic neighbourhood density interact in visual word recognition. *Q J Exp Psychol A*, *58*(6), 981–998.

Halgren, E., Dhond, R. P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., et al. (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *Neuroimage*, *17*(3), 1101–1116.

Halgren, E., Wang, C., Schomer, D. L., Knake, S., Marinkovic, K., Wu, J., et al. (2006). Processing stages underlying word recognition in the anteroventral temporal lobe. *Neuroimage*, *30*(4), 1401–1413.

Hauk, O., Davis, M., 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.

Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clin Neurophysiol*, *115*(5), 1090–1103.

Helenius, P., Salmelin, R., Service, E., & Connolly, J. F. (1998). Distinct time courses of word and context comprehension in the left temporal cortex. *Brain*, *121 (Pt 6)*, 1133–1142.

Holcomb, P. J. (1988). Automatic and attentional processing: an event-related brain potential analysis of semantic priming. *Brain Lang*, *35*(1), 66–85.

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: implications for the role of the N400 in language processing. *Psychophysiology*, *30*(1), 47–61.

Holcomb, P. J., & McPherson, W. B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain Cogn*, *24*(2), 259–276.

Holcomb, P. J., & Neville, H. J. (1990). Auditory and Visual Semantic Priming in Lexical Decision: A Comparison Using Event-related Brain Potentials. *Language and Cognitive Processes*, *5*(4), 281-312.

Holcomb, P. J., & Neville, H. J. (1991). Natural speech processing: An analysis using event–related brain potentials. *Psychobiology*, *19*(4), 286-300.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 6570.

Huntsman, L. A., & Lima, S. D. (1996). Orthographic neighborhood structure and lexical access. *Journal of Psycholinguistic Research*, *25*, 417-429.

Huntsman, L. A., & Lima, S. D. (2002). Orthographic neighbors and visual word recognition. *J Psycholinguist Res*, *31*(3), 289–306.

Kiefer, M. (2002). The N400 is modulated by unconsciously perceived masked words: further evidence for an automatic spreading activation account of N400 priming effects. *Cognitive Brain Research*, *13*(1), 27–39.

Kiefer, M., & Brendel, D. (2006). Attentional modulation of unconscious "automatic" processes: evidence from event-related potentials in a masked priming paradigm. *J Cogn Neurosci*, *18*(2), 184–198.

Kounios, J., & Holcomb, P. J. (1992). Structure and process in semantic memory: evidence from event-related brain potentials and reaction times. *J Exp Psychol Gen*, *121*(4), 459–479.

Kreher, D. A., Holcomb, P. J., & Kuperberg, G. R. (2006). An electrophysiological investigation of indirect semantic priming. *Psychophysiology*, *43*(6), 550–563.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn Sci*, *4*(12), 463–470.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163.

Kutas, M., Lindamood, T. E., & Hillyard, S. A. (1984). Word expectancy and event-related brain potentials during sentence processing. In S. Kornblum & J. Requin (Eds.), *Preparatory states and processes* (p. 217-237). Hillsdale, NJ: Erlbaum.

Kutas, M., Neville, H. J., & Holcomb, P. J. (1987). A preliminary comparison of the N400 response to semantic anomalies during reading, listening and signing. *Electroencephalogr Clin Neurophysiol Suppl*, *39*, 325–330.

Ledoux, K., Camblin, C. C., Swaab, T. Y., & Gordon, P. C. (2006). Reading words in discourse: the modulation of lexical priming effects by message-level context. *Behav Cogn Neurosci Rev*, *5*(3), 107–127.

Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Percept Psychophys*, *62*(3), 615–625.

Luce, P. A., & Large, N. (2001). Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes*, *16*, 565-581.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear Hear*, *19*(1), 1–36.

Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature*, *383*(6601), 616–618.

Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., & Halgren, E. (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron*, *38*(3), 487–497.

Marslen-Wilson, W. D. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, *244*(5417), 522–523.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1-2), 71–102.

Marslen-Wilson, W. D. (1989). Lexical representation and process. In W. D. Marslen-Wilson (Ed.), *Access and integration: Projecting sounds onto meaning* (p. 3-24). Cambridge, MA: MIT press.

Marslen-Wilson, W. D., & Tyler, L. K. (1975). Processing structure of sentence perception. *Nature*, *257*(5529), 784–786.

Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*(1), 1–71.

Max, L., & Onghena, P. (1999). Some issues in the statistical analysis of completely randomized and repeated measures designs for speech, language, and hearing research. *Journal of Speech, Language, and Hearing Research*, *42*, 261-270.

McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology*, *36*(1), 53–65.

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: where are they? *J Exp Psychol Gen*, *118*(1), 43–71.

Morton, J. (1969). The interaction of information in word recognition. *Psychological Review*, *76*, 165-178.

New, B., Brysbaert, M., Segui, J., Ferrand, L., & Rastle, K. (2004). The processing of singular and plural nouns in French and English. *Journal of Memory and Language*, *51*(4), 568-585.

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: new evidence from the English Lexicon Project. *Psychon Bull Rev*, *13*(1), 45–52.

Nieuwland, M. S., & van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *J Cogn Neurosci*, *18*(7), 1098–1111.

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: an evoked potentials study. *Brain Lang*, *85*(2), 203–210.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh Inventory. *Neuropsychologia*, *1*(9), 97-113.

Peereman, R., & Content, A. (1997). Orthographic and phonological neighborhoods in naming: Not all neighbors are equally influential in orthographic space. *Journal of Memory and Language*, *37*, 382-410.

Perea, M., & Pollatsek, A. (1998). The effects of neighborhood frequency in reading and lexical decision. *J Exp Psychol Hum Percept Perform*, *24*(3), 767–779.

Perea, M., & Rosa, E. (2000). The effects of orthographic neighborhood in reading and laboratory word identification tasks: A review. *Psicológica*, *21*, 327-340.

Pisoni, D. B., & Levi, S. V. (in press). Some observations on representations and representational specificity in speech perception and spoken word recognition. In M. Gaskell (Ed.), *The oxford handbook of psycholinguistics.* Oxford University Press.

Poeppel, D., Idsardi, W., & van Wassenhove, V. (in press). Speech perception at the interface of neurobiology and linguistics. *Proc. Royal Soc. London.*

Pylkkänen, L., Llinás, R., & Murphy, G. L. (2006). The representation of polysemy: MEG evidence.

J Cogn Neurosci, *18*(1), 97–109.

Pylkkänen, L., & Marantz, A. (2003). Tracking the time course of word recognition with MEG. *Trends in Cognitive Sciences*, *7*(5), 187–189.

Pylkkänen, L., Stringfellow, A., Flagg, E., & Marantz, A. (2001). A Neural Response Sensitive to Repetition and Phonotactic Probability: MEG Investigations of Lexical Access. In *Proceedings of biomag 2000* (p. 363-367).

Pylkkänen, L., Stringfellow, A., & Marantz, A. (2002). Neuromagnetic evidence for the timing of lexical activation: an MEG component sensitive to phonotactic probability but not to neighborhood density. *Brain Lang*, *81*(1-3), 666–678.

Rhodes, S., & Donaldson, D. (2006). Meaning is first accessed through association and not shared semantic features: Electrophysiological evidence from the N400. In *Poster presented at the cognitive neuroscience society annual meeting* (p. 219). San Francisco. (Poster G38 – Presented at the Cognitive Neuroscience Society annual meeting)

Rolke, B., Heil, M., Streb, J., & Hennighausen, E. (2001). Missed prime words within the attentional blink evoke an N400 semantic priming effect. *Psychophysiology*, *38*(2), 165–174.

Rugg, M. D. (1983). Further study of the electrophysiological correlates of lexical decision. *Brain and Language*, *19*, 142152.

Rugg, M. D. (1987). Dissociation of semantic priming, word and nonword repetition effects by Event-Related Potentials. *The Quarterly Journal of Experimental Psychology*, *39A*, 123148.

Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high- and low-frequency words. *Mem Cognit*, *18*(4), 367–379.

Rugg, M. D., Brovedani, P., & Doyle, M. C. (1992). Modulation of event-related potentials (ERPs) by word repetition in a task with inconsistent mapping between repetition and response. *Electroencephalogr Clin Neurophysiol*, *84*(6), 521–531.

Rugg, M. D., Cox, C. J. C., Doyle, M. C., & Wells, T. (1995). Event-related potentials and the recollection of low and high frequency words. *Neuropsychologia*, *33*(4), 471-484.

Rugg, M. D., & Nagy, M. E. (1987). Lexical contribution to nonword repetition effects: Evidence from event-related potentials. *Memory and Cognition*, *15*, 473481.

Rumelhart, D. E., & McClelland, J. L. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychol Rev*, *88*, 375-407.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychol Rev*, *89*(1), 60–94.

Sears, C. R., Campbell, C. R., & Lupker, S. J. (2006). Is there a neighborhood frequency effect in English? Evidence from reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(4), 1040–1062.

Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception & Performance*, *21*(4).

Sears, C. R., Hino, Y., & Lupker, S. J. (1999). Orthographic neighborhood effects in parallel distributed processing models. *Canadian Journal of Experimental Psychology*, *53*(3), 220-229.

Sears, C. R., Lupker, S. J., & Hino, Y. (1999). Orthographic neighborhood effects in perceptual identification and semantic categorization tasks: a test of the multiple read-out model. *Percept Psychophys*, *61*(8), 1537–1554.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychol Rev*, *96*(4), 523–568.

Seidenberg, M. S., & McClelland, J. L. (1990). More words but still no lexicon: Reply to Besner et al 1990. *Psychological Review*, *97*, 447-452.

Sekiguchi, T., Koyama, S., & Kakigi, R. (2000). The effect of word repetition on evoked magentic responses in the human brain. *Japanese Psychological Research*, *42*(1), 3-14.

Sereno, J. A., & Jongman, A. (1997). Processing of English inflectional morphology. *Mem Cognit*, *25*(4), 425–437.

Siakaluk, P. D., Sears, C. R., & Lupker, S. J. (2002). Orthographic neighborhood effects in lexical decision: the effects of nonword orthographic neighborhood size. *J Exp Psychol Hum Percept Perform*, *28*(3), 661–681.

Smith, M. E., & Halgren, E. (1987). Event-related potentials during lexical decision: effects of repetition, word frequency, pronounceability, and concreteness. *Electroencephalogr Clin Neurophysiol Suppl*, *40*, 417–421.

Snodgrass, J. G., & Minzer, M. (1993). Neighborhood effects in visual word recognition: Facilitatory or inhibitory? *Memory and Cognition*, *21*, 247-266.

St. George, M., Mannes, S., & Hoffman, J. E. (1994). Global semantic expectancy and language comprehension. *Journal of cognitive neuroscience*, *6*(1), 70-83.

Stockall, L., Stringfellow, A., & Marantz, A. (2004). The precise time course of lexical activation:

MEG measurements of the effects of frequency, probability, and density in lexical decision. *Brain Lang*, *90*(1-3), 88–94.

Swaab, T. Y., Camblin, C. C., & Gordon, P. C. (2004). Electrophysiological evidence for reversed lexical repetition effects in language processing. *J Cogn Neurosci*, *16*(5), 715–726.

Tryon, W. W. (2001). Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: an integrated alternative method of conducting null hypothesis statistical tests. *Psychol Methods*, *6*(4), 371–386.

van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: evidence from the N400. *J Cogn Neurosci*, *11*(6), 657–671.

van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Brain Res Cogn Brain Res*, *17*(3), 701–718.

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Mem Cognit*, *18*(4), 380–393.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, *9*, 325-329.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic Phontactics and Neighborhood Activation in Spoken Word Recognition. *Journal of Memory and Language*, *40*, 374-408.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: implications for the processing of spoken nonsense words. *Lang Speech*, *40 (Pt 1)*, 47–62.

Vitevitch, M. S., & Rodríguez, E. (2005). Neighborhood density effects in spoken word recognition in Spanish. *Journal of Multilingual Communication Disorders*, *3*(1), 64 - 73.

Vogel, E. K., Luck, S. J., & Shapiro, K. L. (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *J Exp Psychol Hum Percept Perform*, *24*(6), 1656–1674.

Westbury, C., & Buchanan, L. (2002). The probability of the least likely non-length-controlled bigram affects lexical decision reaction times. *Brain Lang*, *81*(1-3), 66–78.

Wilcox, R. R. (1997). *Introduction to robust estimation and hypothesis testing*. San Diego: Academic Press.

Wilding, E. L. (2000). In what way does the parietal ERP old/new effect index recollection? *Int J Psychophysiol*, *35*(1), 81–87.

Wydell, T. N., Vuorinen, T., Helenius, P., & Salmelin, R. (2003). Neural correlates of letter-string length and lexicality during reading in a regular orthography. *J Cogn Neurosci*, *15*(7), 1052–1062.

Yates, M. (2005). Phonological neighbors speed visual word processing: Evidence from multiple tasks. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *31*(6), 1385-1397.

Yates, M., Locker, L., & Simpson, G. B. (2004). The influence of phonological neighborhood on visual word perception. *Psychon Bull Rev*, *11*(3), 452–457.

Young, M. P., & Rugg, M. D. (1992). Word frequency and multiple repetition as determinants of the modulation of event-related potentials in a semantic classification task. *Psychophysiology*, *29*(6), 664–676.

Ziegler, J. C., Muneaux, M., & Grainger, J. (2003). Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language*, *2003*, 779-793.

Ziegler, J. C., & Perry, C. (1998). No more problems in Coltheart's neighborhood: resolving neighborhood conflicts in the lexical decision task. *Cognition*, *68*(2), B53–B62.

Appendix

Appendix: List of Items in Experiment 1

Materials

Words		Nonwords	
High/Dense	Low/Sparse	High/Dense	Low/Sparse
bacon	bagel	banel	boreb
baron	civic	banic	boril
basin	comic	belet	davel
bison	fatal	berin	devim
bonus	fetal	beris	folip
cabin	focal	caris	galef
camel	forum	casef	galem
civil	haven	cazel	gapel
colon	havoc	cobin	givel
comet	hazel	delin	havel
coral	humor	fenid	helic
demon	legal	galed	herib
denim	level	galel	janep
facet	libel	gorel	japel
feral	logic	halet	jokeb
latin	madam	janem	jokel
linen	magic	kevil	jural
mason	manic	labin	kelic
melon	mimic	laken	lakeb
merit	modem	lamin	likel
metal	mogul	levim	lipit
minus	nasal	limid	lomic
moral	natal	maled	lorib
panel	naval	malim	mafil
pedal	novel	maron	majil
penal	panic	menon	molec
peril	relic	minel	nomel
petal	rival	monit	nosen
pilot	sonar	navin	polec
rabid	sonic	norel	ronic
rapid	tidal	novin	tamic
rigid	tonic	palil	tanic
robin	topaz	pelen	tavin
roman	topic	ranis	terib
salad	totem	revit	tokez
satin	tulip	sanis	toric
serum	venom	savin	vagul
seven	vigil	solip	valuk
siren	viral	tacin	vanic
solid	vital	telen	verem
tacit	vocal	tened	vivit
tenet	woman	torel	zarin