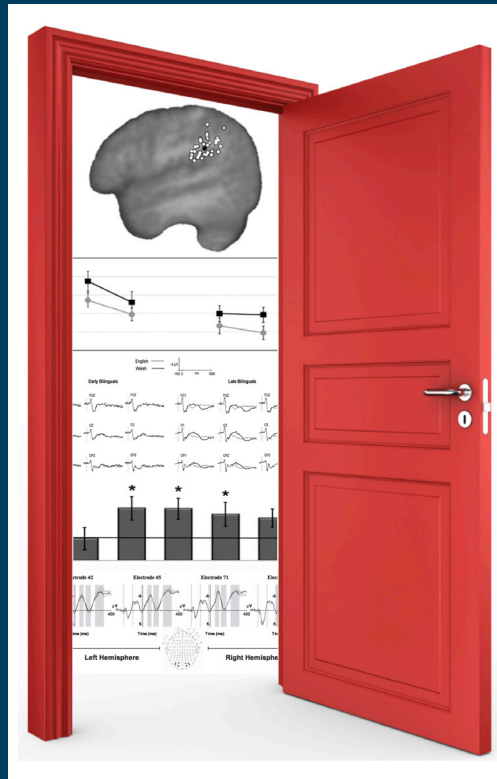# frontiers
RESEARCH TOPICS

AT THE DOORS OF LEXICAL ACCESS: THE IMPORTANCE OF THE FIRST 250 MILLISECONDS IN READING

Topic Editors
Jon Andoni Dunabeitia and Nicola Molinaro

frontiers in
PSYCHOLOGY

frontiers

ABOUT FRONTIERS

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

FRONTIERS JOURNAL SERIES

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing.

All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

DEDICATION TO QUALITY

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

WHAT ARE FRONTIERS RESEARCH TOPICS?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area!

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

AT THE DOORS OF LEXICAL ACCESS: THE IMPORTANCE OF THE FIRST 250 MILLISECONDS IN READING

Topic Editors:
Jon Andoni Dunabeitia, Basque Center on Cognition, Brain and Language, Spain
Nicola Molinaro, Basque center on Cognition, Brain and Language, Spain

The image is a selection of several figures from the articles included in this Research Topic.

Correct word identification and processing is a prerequisite for accurate reading, and decades of psycholinguistic and neuroscientific research have shown that the magical moments of visual word recognition are short-lived and markedly fast. The time window in which a given letter string passes from being a mere sequence of printed curves and strokes to acquiring the word status takes around one third of a second. In a few hundred milliseconds, a skilled reader recognizes an isolated word and carries out a number of underlying processes, such as the encoding of letter position and letter identity, and lexico-semantic information retrieval. However, the precise manner (and order) in which these processes occur (or co-occur) is a matter of contention subject to empirical research. There's no agreement regarding the precise timing of some of the essential processes that guide visual word processing, such as precise letter identification, letter position assignment or sub-word unit processing (bigrams, trigrams, syllables, morphemes), among others. Which is the sequence of processes that lead to lexical access? How do these and other processes interact with each other during the early moments of word processing? Do these processes occur in a serial fashion or do they take place in parallel? Are these processes subject to mutual interaction principles? Is feedback allowed for within the earliest stages of word identification? And ultimately, when does the reader's brain effectively identify a given word? A vast number of questions remain open, and this Research Topic will cover some of

them, giving the readership the opportunity to understand how the scientific community faces the problem of modeling the early stages of word identification according to the latest neuroscientific findings.

The present Research Topic aimed to combine recent experimental evidence on early word processing from different techniques together with comprehensive reviews of the current work directions, in order to create a landmark forum in which experts in the field defined the state of the art and future directions. We were willing to receive submissions of empirical as well as theoretical and review articles based on different computational and neuroscience-oriented methodologies. We especially encouraged researchers primarily using electrophysiological or magnetoencephalographic techniques as well as eye-tracking to participate, given that these techniques provide us with the opportunity to uncover the mysteries of lexical access allowing for a fine-grained time-course analysis. The main focus of interest concerned the processes that are held within the initial 250-300 milliseconds after word presentation, covering areas that link basic visuo-attentional systems with linguistic mechanisms.

Table of Contents

The wide-open doors to lexical access

Jon A. Duñabeitia* and Nicola Molinaro*

Basque Center on Cognition, Brain and Language, Donostia, Spain
*Correspondence: j.dunabeitia@bcbl.eu; n.molinaro@bcbl.eu

Edited by:
Manuel Carreiras, Basque Center on Cognition, Brain and Language, Spain

Reading is an ability that appears simple and automatic to the experienced reader, in the same way that driving a car holds no mysteries for the practiced driver. However, most drivers would recall that the number of operations which needed to be learned to move the car smoothly seemed insurmountable during the first days of driving instruction. Nonetheless, as time passed by, thanks to repetition and practicing, and to the operations progressively becoming automatized, driving was no longer a challenge. Considering that in modern societies reading is typically acquired during early childhood, it is relatively implausible that we remember the hard moments we went through on the road to becoming fluent readers. Still, as is the case with driving, reading requires a substantial number of perceptual, attentional and mnemonic abilities, and a vast array of operations that can appear overwhelming to the neophyte until they become automatized.

Reading requires complex abstraction of the highly variable alphabetic visual input, which ultimately allows the access to the abstract orthographic categories that are in turn the door to the retrieval of phonological, morphological, lexical, and semantic representations. This stimulus-to-meaning mapping has to be robust enough to face font variability, handwriting styles, orthographic errors, contractions, and many other potential alterations in the input. This mapping poses the first paradoxical conundrum for the reader, who on the one hand has to be relatively "blind" to the obvious perceptual differences between multiple fonts, cases or handwriting of the same word (e.g., door, dOoR), and on the other hand needs to be "sighted" enough to detect basic perceptual differences between a given word and other similar items (e.g., door, deer, odor, dear).

The time window in which a given letter string passes from being a mere sequence of printed curves and strokes to acquiring the word status takes around one third of a second. In that fraction of a second the expert reader manages to access the meaning represented by the written symbolic and arbitrary graphic patterns. This phenomenon represents a model of human abstract symbolic thinking, since there is no direct relation between the meaning of a word and its written form. If we consider the concepts of a *door* and a *window*, it seems relatively straightforward to define the semantic relation between them. However, from a linguistic perspective there is no physical or functional relation between the two written codes *door* and *window*. How is it then possible that readers are able to compute the semantic relation between these two written codes through a simple eye fixation of 250 ms? What does reading imply for the human brain? And where and when in the brain does reading take place?

The answers to these questions are still controversial. Nonetheless, in recent years the neurocognitive literature has provided the grounds for constructing the perfect test scenario to help solve this issue. What, where and when? Neuroimaging and behavioral methods have demonstrated that reading implies a complex pattern of feed-forward and feedback interactive activations flowing along the visual recognition system, mainly in ventral regions of the left temporal lobe. Still, the precise way in which all the intermediate representations between a physically concrete printed stimulus and the mentally stored abstract lexico-semantic representation are activated is still debated and needs to be further explored.

The present Research Topic aimed to create a landmark forum in which experts in the field define the state of the art and future directions. A total of 10 excellent articles have been compiled (six Original Research articles, three Review articles and one General Commentary). Su et al. (2012) open the section of Original Research articles with an experiment using ERPs to test the interactions between graphemic similarity, position of the radicals of Chinese characters and lexical access. Next, Sliwinska et al. (2012) present the readership with a study using chronometric TMS devoted to better characterizing the role of the supra-marginal gyrus in phonological processing, and ultimately, in visual-word identification. In the third article, Grossi et al. (2012) present an ERP study exploring the interactions between bilinguals' linguistic experience and orthographic and lexico-semantic effects associated with cross-language orthographic neighborhood effects in two groups of English-Welsh bilinguals. Hand et al. (2012) present an article exploring the early interactions between the orthographic constraint imposed by word-initial letters and context-based predictability effects using eye movement tracking techniques. A similar rationale is followed in the article by Lee et al. (2012), offering electrophysiological data regarding interactions between contextual information and early orthographic processing. Kinoshita and Norris (2012) provide the last Original Research article summarizing recent findings from the visual-word recognition domain and proposing an interpretation of masked priming based on the Bayesian Reader account that explains some controversial task-dependent effects. The Research Topic then continues with three Review articles and one General Commentary. Van Assche et al. (2012) offer an outline of recent data demonstrating that lexical access is language-non-selective in bilinguals, both at the level of recognizing words in isolation and at the level of recognizing words in sentence context. Hyönä (2012) presents an overview of the findings on compound word identification, and provides a physiologically valid for the way in which polymorphemic words are processed in alphabetic languages, based on visual acuity principles. Amenta and Crepaldi (2012) offer the last Review article, which is also related to the processing of polymorphemic words. They summarize benchmark morphological processing effects

and set the scenario for future experimental and theoretical work by highlighting the most consistent and inconsistent findings. The General Commentary by Koester (2012) extends some of the issues raised by Amenta and Crepaldi (2012), and raises other concerns regarding the future of neurocognitive scientific activity on morphological processing (see also the General Commentary by Crepaldi and Amenta; doi: 10.3389/fpsyg.2013.00056).

As the (proud) Editors of this Research Topic, we honestly believe that the initial aims have been fulfilled. The excellence of the Original Research articles is doubtless, and they nicely cover different experimental approaches (i.e., behavioral or eye-tracking techniques, ERPs, TMS) to current questions regarding monolingual and bilingual lexical access. Similarly, the worth of

the Review articles is undeniable. These Review articles represent a compelling updated overview of critical topics for the community investigating lexical access, and they will certainly serve for inspiration for other researchers in the field. Now it is time for the audience to assess the value of all these articles, and we sincerely hope that the reception will be at least as good as it has been during these last months, in which the amount of views and downloads of the articles has been heartening.

REFERENCES

Amenta, S., and Crepaldi, D. (2012). Morphological processing as we know it: an analytical review of morphological effects in visual word identification. *Front. Psychol.* 3:232. doi: 10.3389/fpsyg.2012.00232

Grossi, G., Savill, N., Thomas, E., and Thierry, G. (2012). Electrophysiological cross-language neighborhood density effects in late and early English-Welsh bilinguals. *Front. Psychol.* 3:408. doi: 10.3389/fpsyg.2012.00408

Hand, C. J., O'Donnell, P. J., and Sereno, S. C. (2012). Word-initial letters influence fixation durations during fluent reading. *Front. Psychol.* 3:85. doi: 10.3389/fpsyg.2012.00085

Hyönä, J. (2012). The role of visual acuity and segmentation cues in compound word identification. *Front. Psychol.* 3:188. doi: 10.3389/fpsyg.2012.00188

Kinoshita, S., and Norris, D. (2012). Task-dependent masked priming effects in visual word recognition. *Front. Psychol.* 3:178. doi: 10.3389/fpsyg.2012.00178

Koester, D. (2012). Future morphology. Summary of visual word identification effects draws attention to necessary efforts in understanding morphological processing. *Front. Psychol.* 3:395. doi: 10.3389/fpsyg.2012.00395

Lee, C.-Y., Liu, Y.-N., and Tsai, J.-L. (2012). The time course of contextual effects on visual word

recognition. *Front. Psychol.* 3:285. doi: 10.3389/fpsyg.2012.00285

Sliwinska, M. W., Khadilkar, M., Campbell-Ratcliffe, J., Quevenco, F., and Devlin, J. T. (2012). Early and sustained supramarginal gyrus contributions to phonological processing. *Front. Psychol.* 3:161. doi: 10.3389/fpsyg.2012.00161

Su, I.-F., Mak, S.-C. C., Cheung, L.-Y. M., and Law, S.-P. (2012). Taking a radical position: evidence for position-specific radical representations in Chinese character recognition using masked priming ERP. *Front. Psychol.* 3:333. doi: 10.3389/fpsyg.2012.00333

Van Assche, E., Duyck, W., and Hartsuiker, R. J. (2012). Bilingual word recognition in a sentence

context. *Front. Psychol.* 3:174. doi: 10.3389/fpsyg.2012.00174

Taking a radical position: evidence for position-specific radical representations in Chinese character recognition using masked priming ERP

I.-Fan Su*, Sin-Ching Cassie Mak, Lai-Ying Milly Cheung and Sam-Po Law

Laboratory for Communication Sciences, Division of Speech and Hearing Sciences, The University of Hong Kong, Hong Kong SAR, China

In the investigation of orthographic representation of Chinese characters, one question that has stimulated much research is whether radicals (character components) are specified for spatial position in a character (e.g., Ding et al., 2004; Tsang and Chen, 2009). Differing from previous work, component or radical position information in this study is conceived in terms of relative frequency across different positions of characters containing it. A lexical decision task in a masked priming paradigm focusing on radicals with preferred position of occurrence was conducted. A radical position that encompasses more characters than other positions was identified to be the preferred position of a particular radical. The prime that was exposed for 96 ms might share a radical with the target in the same or different positions. Moreover, the shared radical appeared either in its preferred or non-preferred position in the target. While response latencies only revealed the effect of graphical similarity, both effects of graphical similarity and radical position preference were found in the event-related potential (ERP) results. The former effect was reflected in greater positivity in occipital P1 and greater negativity in N400 for radicals in different positions in prime and target characters. The latter effect manifested as greater negativity in occipital N170 and greater positivity in frontal P200 in the same time window elicited by radicals in their non-preferred position. Equally interesting was the reversal of the effect of radical position preference in N400 with greater negativity associated with radicals in preferred position. These findings identify the early ERP components associated with activation of position-specific radical representations in the orthographic lexicon, and reveal the change in the nature of competition from processing at the radical level to the lexical level.

Keywords: word recognition, Chinese radicals, sub-lexical processing, orthography, spatial specification, N170, P200, N400

INTRODUCTION

There has been a long-standing and intense interest in alphabetic scripts regarding how positional information of letters in a word is coded (see review of different models in Grainger and Van Heuven, 2003). This is understandable as about 34% of all four-letter words in English and French can form other words by rearranging their letters (Shillcock et al., 2001). Therefore for alphabetic scripts, spatial specification of letter position is vital to correct word recognition and production (Grainger et al., 2006). While the letters are arranged linearly, radicals (also referred to as components or constituents) in a Chinese character are arranged in a two-dimensional square shape. The same question can be raised whether spatial information of components is similarly necessary. That is, are radicals specified for position of occurrence in the orthographic lexicon? The present study investigated the relevance of positional information of orthographic units during the early stages of visual character recognition using event-related potential (ERP) with a primed-lexical decision task.

The Chinese writing system is a non-alphabetic script with words composed of characters that represent morphemes. It has often been characterized as morphosyllabic. Each character occupies a constant square-shaped space that is constructed by combination of stroke patterns. Particular grouping of strokes form radicals that may exist as characters. These radicals may combine to form complex characters (Hoosain, 1991). Over 80% of complex characters contain radicals carrying probabilistic phonetic cues (e.g.耶, /je4/[1], is the phonetic radical of 椰,/je4/) or semantic cues (e.g., 木 "*wood/plant*" is the semantic radical of 樹 "*tree*"), and their reliability can vary across characters (Chen et al., 1996). It has been reported that up to 10 different spatial arrangement of radicals, or configurations, can be found in complex characters (Fu, 1993), such as horizontal (AB 說, ABC 辦), vertical (呆), and semi-enclosed configuration (送).

Studies investigating the properties of radicals have shown that they constitute an important level of representation in the orthographic processing system (e.g., Li and Chen, 1997; Taft et al., 2000), and influence response latencies in lexical decision (e.g., Feldman and Siok, 1997, 1999; Taft and Zhu, 1997), naming (e.g.,

---

[1]Phonetic transcriptions of Chinese characters provided in this paper are given in jyutping, a Romanization system developed by the Linguistic Society of Hong Kong. The number in the transcription represents the tone.

Taft et al., 1999; Zhou and Marslen-Wilson, 1999; Lee et al., 2005), semantic judgment (e.g., Chen and Weekes, 2004; Chen et al., 2006), and priming tasks (e.g., Ding et al., 2004). They show that character activation is achieved via radical activation in complex characters. For example, Taft and Zhu (1997) found an effect of radical type frequency (number of characters containing that radical irrespective of function) and radical status (real or invented radical) in a lexical decision task.

Growing evidence from electrophysiological studies using ERP has also shown that radicals modulate the N1/N170, frontal P200, and N400 components. Focusing on the phonetic radical, Hsu et al. (2009) found interactive effects of phonetic combinability (neighborhood size of the phonetic radical) and phonological consistency (degree of agreement in pronunciation among orthographic neighbors having the same phonetic radical) at N170 in the occipital region, and main effects of phonetic combinability and phonological consistency at the frontal P200 component. Similar findings were also reported by Lee et al. (2006a,b, 2007) showing that phonetic radical consistency and regularity of the (phonetic) radical was reflected in the frontal P200 component, which Liu et al. (2003) suggested was related to phonological processing. Consistency effects were also found at the late N400 indicating that radicals are involved in later lexical processing via what the authors proposed to be competition of other phonetically similar neighbors (Lee et al., 2006a,b, 2007).

While a number of studies show that the radical acts as a sub-lexical representational unit in the Chinese script, the theoretical question we put forward is to test whether it is necessary to have separate position-specific representations for radicals as argued by Taft and his colleagues in their Multilevel Interactive-Activation framework (Taft and Zhu, 1997; Taft et al., 1999; Taft, 2006). For example, a radical (田) can be located at different positions in a character, such as on the left (e.g., 略), right (e.g., 細), top (e.g., 胃), or bottom (e.g., 雷; Hoosain, 1991). According to Taft and colleagues, the 田 radical in each of the characters above have their own position-specific representation activated (position-specific view), so that the character 略 would activate a left-田 radical, and the character 胃 would activate a top-田 radical. Although all models of Chinese character recognition allow for the representation of radicals, they differ in opinion about how spatial information of these radicals is characterized (e.g., Perfetti et al., 2005; Perfetti and Liu, 2006; Yang et al., 2009).

The claim for separate position-specific representations was based on a study by Taft and Zhu (1997) showing that character decision latency was influenced by radical position frequency (number of characters containing that radical in that position). Specifically, for two characters of equal frequency, the one with high radical position frequency (e.g., the character 鵡 containing 鳥 which occurs on the right-hand side of many characters) was easier to be recognized than the one with low radical position frequency (e.g., 捫 containing 門 which occurs on the right-hand side of few characters). Taft et al. (1999) also argued that activation of the appropriate position-specific radical unit within a character would lead to lateral inhibition of other inappropriate position-specific radical units as no interference effects were found when recognizing transposable characters using a lexical decision or naming tasks. Thus, characters containing transposable

radicals (e.g., 呆 and 杏) had response latencies and error rates that were comparable to those containing non-transposable radicals (e.g., 孕). This suggested to them that the same radical occurring at different positions were represented independently. To further support their claim, Ding et al. (2004) show significant priming effects when prime-target characters shared a radical in the same position (e.g., 軀-樞) but not when they shared a radical in different positions (e.g., 歐-樞) using a primed-lexical decision task. It was argued that pre-activation of the radical from a similar radical position prime led to faster target lexical decision latencies. However, the priming effect that Ding et al. found is problematic as facilitation in trials with radicals in similar position also shared more visual overlap than when in different positions.

Furthermore, contradictory findings have also been reported, leading some to argue that radicals are not coded for position (position-general view). For instance, the Lexical Constituency Model (Perfetti et al., 2005; Perfetti and Liu, 2006) claims that radicals do not require position information, and postulates instead configuration "slots" to allocate the position-free radicals. This implies that configural information, represented as input units in the model (e.g., left-right or top-bottom), is activated independently from the radicals themselves. Note that Taft's model also includes position-general radicals that consequently activate position-specific radicals. However, the activation of a complex character is achieved via the activation of its position-specific radicals. This line of reasoning has mainly risen from studies using illusory conjunction and visual search tasks. Tsang and Chen (2009) found using a illusory conjunction task that participants would mistakenly perceive the target character as being one of the two preceding "source" characters when target and source characters have shared radical(s). Importantly, no difference was observed when the shared radical occurred in the same (e.g., 拮 preceded by 結 and 栓) or different (頡 preceded by 結 and 栓) positions across source-target. This finding was taken to argue that radicals were position-general and would activate all characters containing it irrespective of their position. Yet, it is unclear in that study why source characters with radicals in the same position as the target were less error prone than ones in different positions, and a latency difference of approximately 40 ms did not reach statistical significance. Using a visual search paradigm, Yeh and Li (2002) found that target characters took longer to identify when embedded in an array of characters sharing a radical and the same configuration but with the common radical in same or different positions than when they appeared in an array of control (unrelated) characters. Nonetheless, the generalizability of its findings may be challenged with only two target characters manipulated to form eight distractor-target pairs, which were not balanced for character frequency (i.e., 5.18–422.53 per million).

Given the limitations and inconsistencies of previous findings, this study separated the effects of visual overlap resulting from position similarity across prime and target in a primed-lexical decision task, and more significantly took a different conceptual approach to specification of radical position. Spatial position of radicals was explicitly conceived in terms of relative frequency across positions of a radical. As mentioned before, a radical can occupy different positions in characters, with some position encompassing more characters than other positions. For example,

the radical 分 can be found in at least four different positions in characters but it occurs more frequently on the right (76%, e.g., 粉) than on the left (9%, e.g., 頒), top (2%, e.g., 盆), or bottom (4%, e.g., 芬), taking into account type-token frequency. Therefore a radical can have a preferred or dominant position (high type-token frequency), while other positions that the radical can also occur in may be considered less preferred or subordinate positions (low type-token frequency). We argue that this conceptualization provides a more controlled test for position-specific radical representations as it relies on the relative frequency of distribution within the radical neighborhood. The contrast between high and low radical position frequency while controlling for character frequency and overall radical frequency in Experiment 2 of Taft and Zhu (1997) can be considered similar to our concept of dominant vs. subordinate positions, although our current design is superior in some important aspects. First, response latencies to characters containing the same radicals were compared as a function of dominant vs. subordinate positions. Unlike in Taft and Zhu where different radicals for high and low radical position frequency were used, without explicitly controlling for factors that may well influence lexical decision latency, including neighborhood size, phonological consistency, and orthographic complexity. Second, our manipulation took into consideration not only type frequency (i.e., number of characters in which a radical appears in a particular position) as in Taft and Zhu, but also token frequency (i.e., the frequency of each of the characters containing the radical in that position), which has been shown to affect the speed of lexical decision (Lee et al., 2005). Finally, the contrast of dominant vs. subordinate positions equally involved the left and right positions or the top and bottom positions of horizontally or vertically structured characters in this study (see details in Materials and Methods), dissimilar to the exclusive focus on the right or bottom position in Taft and Zhu, which may have inadvertently confounded radical position with function. For example, radicals that occur in the right and bottom position are more likely to be phonetic radicals and thus loosely linked to the phonological function. If spatial position information is inherently specified at the radical level, such information should be sensitive to its relative position distribution. Thus, characters with a radical in its dominant position would be recognized faster than those containing a radical in subordinate position, as characters with radicals in preferred (high type-token frequency) locations may require less effort to be activated relative to less preferred (low type-token frequency) locations.

Event-related potential in addition to response latency was collected, as ERP provides excellent temporal resolution and can reveal the unfolding of graphic, phonological, and semantic processes in visual word recognition online (e.g., Perfetti and Tan, 1998; Liu et al., 2003; Hauk et al., 2006; Holcomb and Grainger, 2006). Although previous studies in Chinese character recognition had not investigated ERP components that functionally reflected the processing of position-specific radical representation, ERP components known to reflect semantic and phonetic radical analysis were selected as components of interest, specifically the N1/N170, P200, and N400 components. Using silent naming, Lee and colleagues (Lee et al., 2006a,b, 2007; Hsu et al., 2009) showed that radical processing is associated with the N1/N170, and

frontal P200 component, which they suggested reflects activation of radical processing at the visual word form area during the mapping of orthography-to-phonology. Greater N1/N170 negativity was found at electrodes P5/P6, P7/P8, PO5/PO6, and PO7/PO8 for characters that encompassed radicals with high combinability/neighborhoods size (Hsu et al., 2009). They suggest that the N170 component is an index of orthographic detection during the early perceptual categorization stage (see also Bentin et al., 1999), and greater visual experience from highly combinable characters lead to more efficient and specialized processing, thus, showed greater activation at the N170 than low combinability characters.

Lee and colleagues suggested that this early stage of visual word recognition shapes later orthographic-to-phonological conversion of the character's radical that was reflected in the N170 and more robust at the frontal P200 component (see also Sereno et al., 1998). Sensitivity to the consistency of the (phonetic) radical's pronunciation within a character showed smaller N170 effects and smaller P200 effects were found at electrodes F3/F4, FC3/FC4, C3/C4, CP3/CP4 Fz, FCZ, FC3, Cz, and CPz for characters that were highly consistent, with the greatest significant difference at the left frontal site F3 (Lee et al., 2007; see also Lee et al., 2006b; Hsu et al., 2009). In a separate task, Liu et al. (2003) showed with a primed pronunciation task that characters sharing similar radicals (graphical similarity) between the prime and target elicited smaller P200 at frontal and central electrodes sites. The N170 and (frontal) P200 components were further found to be sensitive to differences between two types of characters having opposite arrangement of semantic and phonetic radicals (one with semantic radical on the left and phonetic radical on the right vs. one with the opposite alignment; Hsiao et al., 2007). Note though that, this finding could also suggest that character recognition may be sensitive to the positions in which radicals are more likely to occupy as 89.9% of phonetic radicals occur on the right side of characters with semantic radicals on the left of horizontally structured complex characters (Hsiao and Shillcock, 2006).

Radical processing has also been shown to affect the N400 component. Lee et al. (2007) suggests that the N400 reflects a later stage of lexical processing after the P200. Greater N400 component was found for high phonologically consistent radicals at the central region electrodes Fz, FCz, Cz, CPz, and Pz. They argued that due to more homophone characters found in the high consistency condition, greater lexical competition would occur leading to an enhanced N400. In addition, Hsu et al. (2009) found using similar electrodes of interest that highly combinable radicals also elicit a greater N400 component, suggesting that highly combinable radicals increase semantic competition at the N400 (see also Holcomb et al., 2002). Liu et al. (2003) also identified that characters preceded by visually similar primes showed smaller N400 amplitudes at the central region during a semantic relatedness judgment task.

In light of previous ERP findings, it is assumed that position information is encoded early (Taft, 2006), and will modulate the early stages of character recognition in which visual-perceptual analysis proceeds to the orthographic stage as reflected in the N1/N170 and P1 visual components, but may also influence the frontal P200 and N400 (Liu et al., 2003; Lee et al., 2007; Hsu et al., 2009). It is predicted that characters with similar radical position to their primes will show a more negative P1, N1/N170, and reduced

N400 due to prior exposure from the prime, while the position effect of dominance may be reflected in the later components, including the N170 and N400 components.

MATERIALS AND METHODS

PARTICIPANTS

Twenty-five native Cantonese speakers aged 18–23 ($M = 20.8$, $SD = 1.7$; female: 14) participated in the study. Three participants were excluded in the ERP analysis due to excessive movements or loss of over 40% of useable trials ($M = 20.95$, $SD = 1.64$; female = 11). All were assessed to be right-handed (Oldfield, 1971), had normal or corrected-to-normal vision, and no prior history of learning difficulties, reading difficulties or head injury. All had completed their secondary education in local mainstream schools and not lived outside of Hong Kong for more than 2 years.

MATERIALS

Seventeen radicals of interest were selected that could appear in dominant and subordinate positions, respectively, in at least four relatively low frequency characters (token frequency less than 100 in a million) using the Hong Kong Corpus of Chinese NewsPaper (HKCCNP; Leung and Lau, 2010) database. The degree of dominance was calculated by dividing the sum frequency of characters that shared the same radical in each possible position by the total frequency of all characters sharing the radical irrespective of position (i.e., dominance = position token frequency/radical token frequency × 100). Radicals that can occur in more than one possible position in characters and appear in one particular position over 60% of the time were classified as radicals having a dominant position of occurrence. The positions in which these radicals appear less than 35% of the time were classified as subordinate. For each of the 17 radicals, four characters containing it were selected, two serving as target and two as prime. The prime and target characters would be paired in such a way that the target radical occurred in either its dominant or subordinate position in the target character, preceded by the prime character containing the radical in either the same or different position. For example, the radical 見 appears on the right-hand side of a character, such as 蜆, 86% of the time (hence its dominant position), and at the bottom of a character, such as 覓, 11% (hence its subordinate position). When 蜆 and 覓 were selected as target characters for the dominant and subordinate position conditions, respectively, they would each be matched with two prime characters with 見 in either the right side 蜆 or the bottom 覽. This is illustrated in **Table 1**.

Of the 17 target radicals, the dominant and subordinate positions appear in characters of the same configuration in 12 cases. The dominant position of eight radicals occurs on the left, seven on the right, one each in the top and the bottom. Six of the radicals serve as a semantic radical in a character, seven as a phonetic radical, and four as either.

All the target characters and prime characters were matched in character frequency [$t(33) = 1.35$, $p = 0.19$] and visual complexity in terms of stroke number [$t(33) = 1.14$, $p = 0.26$]. To represent the variety of character configurations found in Chinese, we included both characters in horizontal and vertical configurations. Phonological and semantic similarities between primes and targets were avoided as much as possible.

Table 1 | Examples of pairs of prime (left) and target (right) characters with mean dominance, frequency (per million), and stroke number in each experimental condition.

	Radical dominance	
	Dominant	**Subordinate**
Position Distribution (SE)	75.88% (2.91)	9.25% (2.25)
	[60–92%]	[0.1–31%]
RADICAL POSITION		
Same	蜆蜆	覽覓
Frequency	14.58–7.92	25.04–7.97
Stroke	12.06–12.12	12.35–11.53
Different	覽蜆	蜆覓
Frequency	25.04–7.92	14.58–7.97
Stroke	12.35–12.12	12.06–11.53

The same number of pseudo characters, created by combining the target radicals with other radicals in their legal positions was used as fillers. These were stroke matched to the target characters. All the prime-target pairs were pseudo-randomized for each participant to avoid successive exposure to the same prime or target. There were 34 primes and 68 targets.

All stimuli were presented as digitized images measuring 3 mm × 33 mm (125×125 pixels) and presented in yellow MingLiu font on a black background. The forward and backward mask consisted of a 125×125 pixels matrix with half of the pixels randomly colored in yellow and the other half in black.

PROCEDURE

The participants took part in a primed-lexical decision task where they were asked to judge whether the target character was a real or pseudo character as quickly and accurately as possible. Each trial began with a fixation cross (500 ms), sequentially followed by a blank page that randomly varied in duration for 500–700 ms ($M = 601$ ms), a forward mask (100 ms), the prime character (96 ms), a backward mask (16 ms), and finally the target character, which remained on the screen until the participant made a response. Once a response was made, a blank screen (500 ms) appeared, followed by an "eye blink" cue (500 ms) and another blank screen for a random duration between 800 and 1000 ms ($M = 897$ ms). The eye blink cue was used to reduce blinking artifacts occurring in the critical time windows of interest. Fifteen practice trials were given to each participant, and a total of 204 experimental trials were divided evenly into four blocks randomized across participants.

Participants were seated in front of a LCD monitor (60 frames/s) at a distance of approximately 100 cm, in an electrically and acoustically shielded room. Before the experiment began, the participants were instructed to minimize their movements and eye blinks to reduce artifactual electroencephalography (EEG) signals. The E-prime 2.0 (Psychological Software Inc.) program was used to present the stimuli, collect behavioral reaction time, and accuracy data. Across all participants, the response hand for lexical decision was counter-balanced.

EEG recording and pre-processing

Electroencephalography/ERP data recorded from 128 Ag/AgCl electrodes (NSL QuikCap, Neuromedical Supplies, Sterling, USA) was digitized online at 1 kHz and amplified with a band pass of 0.05–200 Hz using SynAmps2® (Neuroscan, Inc., El Paso, TX, USA) amplifiers. All electrodes were referenced to a common vertex electrode between electrodes 63 (equivalent to Cz) and 64 (CPz), and ground (GND) was positioned anterior to electrode 60 (Fz). Horizontal eye movement was measured using a pair of bipolar electrodes placed approximately 1 cm lateral to the left and right external canthi (HEOG). Eye blinks and vertical eye movements were monitored using two bipolar electrodes placed on the supra- and infraorbital ridges of the left eye (VEOG). Electrode impedance was maintained below 5 KΩ as much as possible.

In the off-line analysis, channels with bad recording were first removed, ranging from none to three electrodes across participants. The remaining data was subsequently filtered using a zero phase low-pass filter of 30 Hz (12 dB/octave slopes). Channels affected by eye blink were corrected mathematically using individually modeled eye blinks computed from at least 100 eye blink artifacts for each participant based on the ocular artifact reduction procedure implemented in Scan 4.5 (Neuroscan, Inc). Epochs of real character trials (−400 pre-stimulus onset to 1000 ms post-stimulus onset intervals) were then selected and baseline corrected using a 100 ms pre-stimulus interval before the presentation of the forward mask (pre-target stimulus interval of −312–212 ms). Incorrect trials and trials with voltage exceeding ±60 μV or affected by muscle movements were automatically rejected, with equivalent amount of trials being excluded across the experimental conditions (Dominant-same = 22.51%; Dominant-different = 22.76%; Subordinate-same = 24.55%; Subordinate-different = 23.79%). The remaining data was then re-referenced to the average activity across all channels and used to compute grand average waveforms for each condition per participant.

DATA ANALYSIS

A within-participant and between-item two-way analysis of variance (ANOVA) was used to analyze the behavioral reaction time and percentage of error data. Target radical dominance (dominant vs. subordinate) and radical position similarity (same vs. different) between the prime and target served as the independent measures. In both analyses, *post hoc* multiple comparisons were adjusted using the Bonferroni correction.

For the analysis of ERP data, within-participant three-way ANOVA was conducted at the N400 component, with the inclusion of electrode location as the third independent variable. At the occipital (N1, P1, and N170) and frontal component (P200) analyses, within-participants four-way ANOVA was implemented with hemisphere (left vs. right) included. The mean amplitudes at *a priori* selected electrodes served as the dependent variable. Bonferroni adjustment was used to correct the significance threshold for *post hoc* comparisons, and the Greenhouse–Geisser (ε) correction was applied when the assumption of sphericity of variance was violated. The electrodes and components were selected *a priori* based on electrode locations found in previous radical
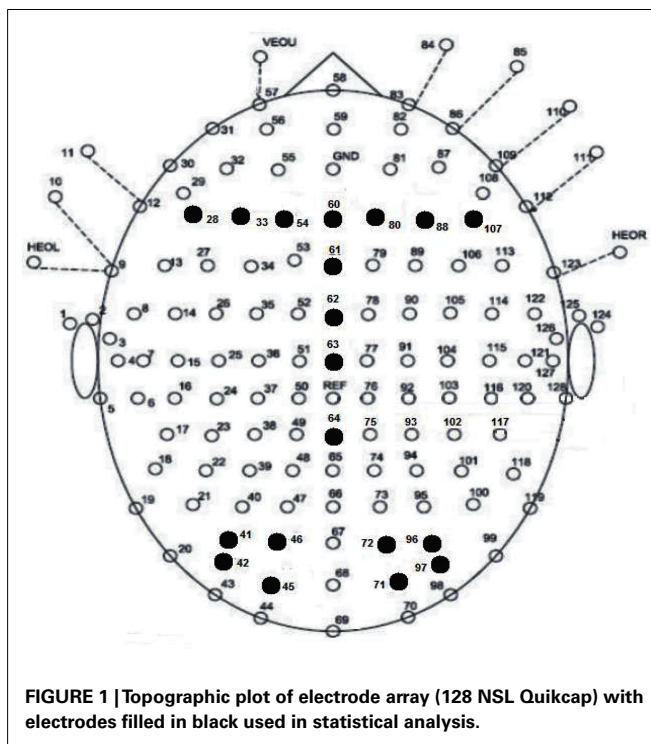
FIGURE 1 | Topographic plot of electrode array (128 NSL Quikcap) with electrodes filled in black used in statistical analysis.

analysis studies and were expected to display maximal amplitudes correlated with visual word form processing at the occipital N1, P1, and N170 component analyses (Holcomb and Grainger, 2006, 2009; Lee et al., 2007; Hsu et al., 2009), phonological processing at the frontal electrodes for the P200 component analysis (Rugg, 1984; Sereno et al., 1998; Lee et al., 2007; Hsu et al., 2009), and semantic processing reflected along the midline electrodes for the N400 component analysis (Liu et al., 2003; Lee et al., 2006a,b, 2007), respectively. The time windows for the components of interest were selected based on the Mean Global Field Power (MGFP) of all trials. Thus, the mean amplitudes at the N1, P1, and N170 components, were computed between 50–100, 120–180, and 225–325 ms, respectively, at the left parietal-occipital electrodes 41, 42 (PO5), 45 and 46 (PO3), and right parietal-occipital electrodes 96, 97 (PO6), 71 and 72 (PO4; see **Figure 1** for electrode array of selected channels). At the frontal P200 component, the mean amplitude was computed between 225 and 325 ms and the electrodes of interest were at the left frontal electrodes 28 (F5), 33 (F3), and 54 (F1), and right frontal electrodes 107 (F6), 88 (F4), and 80 (F2). As the N400 component is maximal at the central region, a time window of 300–450 ms was chose for five electrodes, 60 (Fz), 61 (FCz), 62 (Cz), 63 (Cz), and 64 (CPz) along the midline.

RESULTS

BEHAVIORAL RESULTS

The pseudo character filler trials and five trials with response latencies exceeding 3000 ms (<0.01) were discarded. Trials exceeding ±2.5 SD from the mean of each participant were also excluded from the analysis (1.77%). The mean response latencies for each condition and accuracy are shown in **Table 2**.

With the remaining data, a two-way ANOVA with target radical dominance (dominant or subordinate) and radical position (same or different) was conducted. Only a main effect of radical position was found, $F(1, 24) = 5.37$, $p < 0.05$, $\eta_p^2 = 0.18$, where participants were significantly faster to recognize the target character when the prime and target shared a radical in the same position ($M = 680.62$ ms, SE $= 26.07$) than in a different position ($M = 696.03$ ms, SE $= 23.03$). No main effect of radical dominance, $F(1, 24) = 0.01$, $p = 0.928$, $\eta_p^2 = 0$, or interaction, $F(1, 24) = 1.09$, $p = 0.306$, $\eta_p^2 = 0.04$, was observed, suggesting that characters with radicals in their dominant or subordinate position did not affect the speed of character recognition.

Table 2 | Mean RT and accuracy of each experimental condition.

Dominance	Prime position	Reaction time		Accuracy	
		M (ms)	SE	*M* (%)	SE
Dominant	Same	676	23.28	88.1	0.02
Dominant	Different	702	22.28	90.4	0.02
Subordinate	Same	685	30.63	89.4	0.02
Subordinate	Different	690	25.91	89.7	0.02

Error analysis showed no significant effects indicating that accuracy was not affected by the dominance of the radical's position in the target character, the similarity in radical position between prime and target, or their interaction (all p's > 0.05).

ERP RESULTS

Based on the MGFP, the ERP morphology started its first negative deflection with a maximal peak in the occipital regions at 83 ms from stimulus onset followed by a positive deflection at 151 ms and negative deflection at 279 ms. The frontal electrodes showed a similar pattern to the N170 in the occipital region but with its polarity reversed; hence, occipital N170-frontal P200. A later central negativity peaking at 383 ms and positivity at 585 ms were also observed. **Figures 2–4** show the grand average waveforms for the effects of radical dominance and radical position at various components in the frontal, centro-parietal, and occipito-parietal electrodes. Topographic plots showing scalp distribution and difference amplitude for radical dominance and position similarity effects are shown in **Figure 5**.

N1 (50–100 ms)

No significant effects were found in the four-way repeated measures ANOVA at the first occipital component, all p's > 0.05.

FIGURE 2 | Grand averaged ERP waveforms showing effects of radical dominance and radical position at N1, P1, and N170 components located at the parietal-occipital electrodes and shaded in gray. Dom-Same, dominant radical character with prime radical in same position; Dom-Diff, dominant radical character with prime radical in different position; Sub-Same, subordinate radical character with prime radical in same position; Sub-Diff, subordinate radical character with prime radical in different position. Dotted vertical lines indicate the onset of the forward mask (FM), prime (P), and backward mask (BM).

FIGURE 3 | Grand averaged ERP waveforms showing effects of radical dominance and radical position at the frontal P200 component shaded in gray. Dom-Same, dominant radical character with prime radical in same position; Dom-Diff, dominant radical character with prime radical in different position; Sub-Same, subordinate radical character with prime radical in same position; Sub-Diff, subordinate radical character with prime radical in different position. Dotted vertical lines indicate the onset of the forward mask (FM), prime (P), and backward mask (BM).

Occipital P1 (120–180 ms)

At the next occipital component, the four-way repeated measures ANOVA showed a main effect of radical position, $F(1, 21) = 4.62$, $p < 0.05$, $\eta_p^2 = 0.18$, indicating that target characters preceded by primes with radical in a different position elicited a more positive amplitude ($M = 0.83 \mu V$, SE $= 0.63$) than in the same position ($M = 0.31 \mu V$, SE $= 0.63$), see **Figure 2**. No main effect of radical dominance was observed, $F(1, 21) = 1.28$, $p = 0.27$, $\eta_p^2 = 0.06$, nor interactions with hemisphere or electrode, all F's < 2.37, p's > 0.10.

Occipital N170 (225–325 ms)

At the later occipital N170 component in **Figure 2**, the effect of radical dominance was found $F(1, 21) = 6.21$, $p < 0.05$, $\eta_p^2 = 0.24$, such that characters with radicals in the subordinate position ($M = -6.07 \mu V$, SE $= 0.97$) elicited a more negative going potential than targets with radicals in their dominant position ($M = -5.48 \mu V$, SE $= 0.90$). However, the effect of radical position was not significant, $F(1, 21) = 0.23$, $p = 0.63$, $\eta_p^2 = 0.01$, or any of the interactions, all F's < 3.57, p's > 0.07.

Frontal P200 (225–325 ms)

The frontal P200 component illustrated in **Figure 3** also shows effects of radical dominance similar to the occipital N170, $F(1, 21) = 9.19$, $p < 0.005$, $\eta_p^2 = 0.30$, whereby characters with subordinate radicals ($M = 3.16 \mu V$, SE $= 0.60$) elicited greater positivity than characters with dominant radicals ($M = 2.49 \mu V$, SE $= 0.55$). The effect of radical position was again, not significant, $F(1, 21) = 0.04$, $p = 0.80$, $\eta_p^2 < 0.01$. Moreover, no interactions were observed, all F's < 1.95, p's > 0.17.

N400 (300–450 ms)

The three-way ANOVA with electrodes along the midline revealed main effects of radical dominance, $F(1, 21) = 21.00$, $p < 0.005$, $\eta_p^2 < 0.36$, showing that characters with a dominant radical ($M = 1.27 \mu V$, SE $= 0.54$) elicited a more negative going wave than characters with a subordinate radical ($M = 2.00 \mu V$, SE $= 0.53$). Radical position similarity was also significant, $F(1, 21) = 5.37$, $p < 0.05$, $\eta_p^2 = 0.20$, and characters with radicals in different positions to their primes ($M = 1.36 \mu V$, SE $= 0.48$) showed greater negativity than characters with radicals in the same position ($M = 1.89 \mu V$, SE $= 0.58$).

No interactions were shown to be significant, including the radical dominance-by-position interaction, $F(1, 21) = 0.12$, $p = 0.74$, $\eta_p^2 = 0.01$, as well as the by electrode interactions $F_{\text{Elect-by-position}}$ $(1.72, 36.20) = 1.20$, $p = 0.31$, $\eta_p^2 = 0.05$, $\varepsilon = 0.43$; $F_{\text{Elect-by-dominance}}$ $(1.94, 40.79) = 2.57$, $p = 0.09$, $\eta_p^2 = 0.10$, $\varepsilon = 0.49$.

To summarize the main ERP findings, the occipital P1 component showed a main effect of radical position similarity whereby primes and targets sharing a radical in the same position exhibited a less positive P1 than cases with a common radical in different positions. At the later occipital N170/frontal P200, greater negativity and positivity, respectively, were found for characters with target radicals in their subordinate position compared to characters with target radicals in the dominant position. However, while the pattern of the effect of radical position similarity remained principally the same for the N400 component where radicals appearing in different positions in the prime and target revealed greater negativity than those with radicals in the same

FIGURE 4 | Grand averaged ERP waveforms showing effects of radical dominance and radical position at the N400 component located at the centro-parietal midline electrodes and shaded in gray. Dom-Same, dominant radical character with prime radical in same position; Dom-Diff, dominant radical character with prime radical in different position; Sub-Same, subordinate radical character with prime radical in same position; Sub-Diff, subordinate radical character with prime radical in different position. Dotted vertical lines indicate the onset of the forward mask (FM), prime (P), and backward mask (BM).
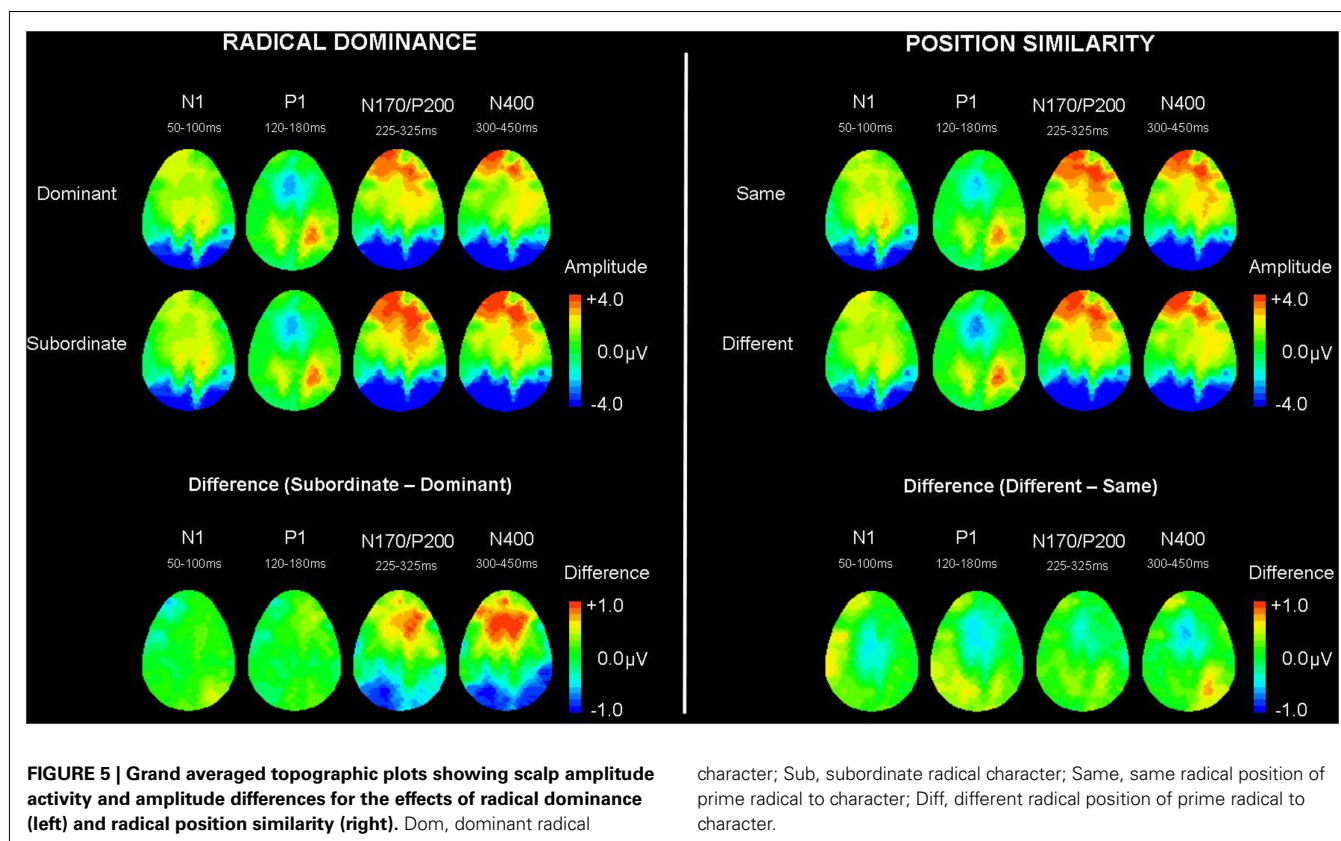
position, the pattern of radical dominance effects changed. More specifically, characters containing radicals in dominant positions elicited a larger N400 than characters with the same radicals in subordinate positions, particularly at the central-frontal sites.

DISCUSSION

The aims of this study were to assess Taft and colleagues' claim for independent representation of position-sensitive radicals and to identify ERP components that may reflect position-specific radical processing and the stage(s) at which the associated effects take place. Unlike previous studies, this investigation separated the effect of visual overlap from that of position specification in radicals by manipulating whether prime-target pairs shared similar radical positions or whether the target's radical differed in relative position frequency factorially. While the behavioral results only showed a position similarity effect, the ERP findings revealed a more complex contribution/relationship of radical processing during the stages of lexical processing. Radical dominance effects were found at the occipital N170-frontal P200 and N400, in addition to the observations of visual similarity effects at the early occipital P1 and later N400 components. Neural sensitivity to radical dominance supports the view that position-specific radicals are activated early, and that the spatial relationship among orthographic units within a character can impact on character recognition. The following discussion considers the position similarity effect reflected in response latency, and examines each of the significant ERP components and attempts to integrate the temporal dynamics of radical processing into Taft's model of character processing (Taft et al., 1999; Taft, 2006).

Based on the behavioral effects, when a target character shared a radical in similar position to its prime, participants were faster to recognize the target due to pre-activation of the relevant position-specific constituent radical in the prime facilitating the recognition of the target character. Previously, this has been taken as evidence for position-specific information of radical representation (Taft and Zhu, 1997; Ding et al., 2004), but such an interpretation may be problematic. First, the ERP findings confirmed that the facilitation was primarily driven by visual overlap between the prime and target radical rather than from independent representations of position-sensitive radicals *per se*. This was because radical position effects were found in early components known to reflect visual analysis at the P150 component in the occipital regions (Grainger and Holcomb, 2009). Less activation was needed when primes and targets shared a radical in the same position compared to different positions, particularly in the left hemisphere. Moreover, although the radical position effects found at the N400 could be indicative of position-specific radical representational processing as radicals in different positions required more effort for integration as they elicited greater N400, one may nevertheless argue that such an observation would suggest that the independent radical representations only influence lexical processing.

Contrasting the effects of radical position and radical dominance in relation to the time course of Chinese character recognition, the findings suggest that visual overlap/similarity is initially processed before access to radical representation as radical position effect precedes the radical dominance effect. The effects of radical dominance provide stronger support for independent

FIGURE 5 | Grand averaged topographic plots showing scalp amplitude activity and amplitude differences for the effects of radical dominance (left) and radical position similarity (right). Dom, dominant radical character; Sub, subordinate radical character; Same, same radical position of prime radical to character; Diff, different radical position of prime radical to character.

representation of position-specific radicals, considering that it is unaffected by the degree of visual overlap between prime and target. Greater negativity at occipital N170 and positivity at frontal P200 for characters with radicals in subordinate position may be taken to reflect increased processing effort to activate the less frequently encountered subordinate radical representation. On the other hand, dominant radical representations are frequently activated because they are connected to more characters containing the radicals in the same positions, and thereby have a lower activation threshold. As such, characters with radicals in dominant positions would require less effort to process, at least initially. This pattern of less activation for characters with radicals in dominant position, however, changes at the later N400 component when character recognition proceeds from radical level facilitation to lexical level competition. Note that radical position effects however, continue to show a smaller N400 component for target characters sharing a similar radical position to their primes. At the N400 component, characters with dominant position radicals elicited greater negativity compared with subordinate position. We argue that radical processing at the N400 reflects competition at the lexical level (see also Lee et al., 2007) as dominant radical position characters naturally have more neighbors that are simultaneously activated, and thus may require greater effort and/or lateral inhibition to suppress irrelevant neighboring competitors to select the appropriate lexical entry. Characters with subordinate position radicals, on the other hand, co-activate a smaller set of neighbors and would therefore involve less conflict resolution. An issue, however, arises that the N400 is generally considered

to be sensitive to lexico-semantic features and assumed to represent post-lexical semantic integration or access to semantic representations in the long term memory (e.g., Kutas and Hillyard, 1980, 1984; Nobre and McCarthy, 1994; Barber and Kutas, 2007). However, the N400 time window of 300–450 ms in this study is earlier than the typical N400 associated with post-lexical semantic integration; thus, the authors argue that the N400 here may capture the earlier phase of the N400 (see Grainger and Holcomb, 2009, for a similar interpretation of ERP masked repetition priming effects). All models of Chinese character recognition take the view that radicals serve as perceptual input only (Taft et al., 1999; Perfetti et al., 2005; Perfetti and Liu, 2006; Taft, 2006), whether the N400 component continues to reflect lexical competition at the word form level or interference at the semantic level from non-target neighbors requires further investigation. We are, however, inclined to argue that the interference reflected in the earlier N400 in our findings occurs at the word form level as equal numbers of semantic and phonetic radicals served as target radicals, and all prime-target pairs were semantically unrelated. Another pertinent observation is that the main effect at the P200 and N400 show similar topographic difference distribution suggesting that the P200 effect argued to be at the sub-lexical level could be related to the later N400. However, we favor the view that the N400 effect is a lexical event associated with the more conventional interpretation of lexical level competition. Nonetheless, what is clear is that position specified radicals play a crucial role in character recognition in Chinese, and affect this process in a complex manner that behavioral experiments alone cannot

reveal. Specifically, changes from facilitation to competition of the position-specific radicals from the radical level to the lexical level may offset each other, and result in a null effect of radical dominance in RT.

The presence of radical dominance effects at the occipital N170 could be seen as reflecting neighborhood size effects (Chen and Weekes, 2004; Chen et al., 2006; Hauk et al., 2009; Hsu et al., 2009) or a neighborhood-by-position interaction (Grainger et al., 2006). The observation of greater processing effort for characters with subordinate radicals is consistent with findings of neighborhood size effects reported from behavioral, functional imaging, and ERP studies in the form of facilitative effects for words with larger neighborhoods (Hsiao et al., 2007; Hsu et al., 2009, 2011; Li et al., 2010). Hsiao et al.'s (2007) study showing position effects at occipital N170 and frontal P200 components can also be described as a type of neighborhood-by-position bias effects between two types of characters having opposite distribution of semantic and phonetic radicals of left-right configuration characters. The dominant radical positions in this study can be analogous to their SP condition where the phonetic radical occurs on the right side of character and left for semantic radical (S and P denote the semantic and phonetic radical, respectively). The less preferred or subordinate position has the opposite alignment (PS). In this case, greater negativity at the occipital N170 or positivity at the frontal P200 for the less preferred positions of radicals is similar to our findings. However, our study differs significantly from Hsiao et al. as it takes a more parsimonious approach of not conflating radical position with functional specificity (see also Zhou and Marslen-Wilson, 1999). As radicals assuming particular functions are more likely to be located in particular positions, the relationship between radical position and functional units still requires delineating. If indeed the radical dominance effects can be conceptualized to be associated with the neighborhood size effect, we suggest that the N170 could reflect a lexico-orthographic component in which the representation of the word form neighborhood is processed. Based on Grainger and Holcomb (2010), the time course of the N170 component in the current study, corresponding to their N250, suggests that activation from orthographic units (e.g., radicals) to words (e.g., character) may take place in this time window with its neural generator located at the left fusiform gyrus region associated with the visual word from area (VWFA). In relating the N170 to N250 in Grainger and Holcomb, it is notable that the time course of our N170 (as well as P1) peaks at a similar time to Grainger and Holcomb (i.e., P150 and N250) which use a similar masked priming design. This may account for the delay of our N170 compared to previous ERP studies with Chinese radicals which typically used covert naming (e.g., Lee et al., 2007) or semantic judgment tasks (e.g., Liu et al., 2003) without a mask.

Finally, the early N1 component observed in this study may be a consequence of the prime's N1 component overlapping with the target character, and reflect a delayed N1 activity of the prime. This may explain the null effects and more crucially its early peak latency of the N1 at 75 ms (prime + backward mask = 112 ms). The typical visual N1 component peaks at around 120–150 ms.

To integrate the present results within the context of Taft's (2006) interactive-activation model in terms of the temporal dynamics of radical processing, we propose that position-general radicals are initially activated at the occipital P1 component at approximately 150 ms, and subsequently spread activation to their position-coded radicals. Around 280 ms in the occipital N170/frontal P200 component, character representations containing position-specific radicals are activated and lexical selection of the word form (or semantics) occurs at approximately 380 ms at the N400, with greater negativity reflecting greater lexical competition.

In conclusion, our results support the role of position-specific radicals in orthographic processing as proposed by Taft (2006) and Taft et al. (1999). This study importantly separates facilitative effects due to visual overlap from position information of radicals and proposes a temporal framework for radical processing. Distinct position-specific representations are conceptualized in terms of dominance of radical position via relative type-token frequency, and we demonstrated that independent position-specific radicals are activated within the first 250 ms of character recognition. Such early access is revealed via modulation of the occipital N170/frontal P200 component, followed by the retrieval of lexico-orthographic or lexico-semantic information at the N400. Our evidence for position-specific radical representations, therefore, highlights the over simplification of the Lexical Constituency Model (Perfetti et al., 2005) as an account of Chinese character processing. While character configuration may be relevant, it alone may not be adequate to access word form representation.

ACKNOWLEDGMENTS

REFERENCES

Barber, H. A., and Kutas, M. (2007). Interplay between computational models and cognitive electrophysiology in visual word recognition. *Brain Res. Rev.* 53, 98–123.

Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F., and Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. *J. Cogn. Neurosci.* 11, 235–260.

Chen, M. J., and Weekes, B. S. (2004). Effects of semantic radicals on Chinese character categorisation and character decision. *Chin. J. Psychol.* 46, 181–196.

Chen, M. J., Weekes, B. S., Peng, D. L., and Lei, Q. (2006). "Effects of semantic radical consistency and combinability on Chinese character processing," in *The Handbook of East Asian Psycholinguistics*, 1st Edn, Vol. 1: Chinese, eds P. Li, E. Bates, L. H. Tan, and O. J. L. Tzeng (Cambridge: Cambridge University Press), 175–186.

Chen, Y. P., Allport, D. A., and Marshall, J. C. (1996). What are the functional orthographic units in Chinese word recognition: the stroke or the stroke pattern? *Q. J. Exp. Psychol. A* 49, 1024–1043.

Ding, G., Peng, D., and Taft, M. (2004). The nature of the mental representation of radicals in Chinese: a priming study. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 530–539.

Feldman, L. B., and Siok, W. W. T. (1997). The role of component function in visual recognition of Chinese character. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 776–781.

Feldman, L. B., and Siok, W. W. T. (1999). Semantic radicals contribute to the visual identification of Chinese characters. *J. Mem. Lang.* 40, 559–576.

Fu, Y.-H. (1993). "The structure and construction of Chinese characters," in *Computational Analysis of*

Modern Chinese Characters, ed. Y. Chen (Shanghai: Shanghai Educational Press), 108–169. [in Chinese].

Grainger, J., Granier, J.-P, Farioli, F., van Assche, E., and van Heuven, W. J. B. (2006). Letter position information and printed word perception: the relative-position priming constraint. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 865–884.

Grainger, J., and Holcomb, P. J. (2009). Watching the word go by: on the time-course of component processes in visual word recognition. *Lang. Linguist. Compass* 3, 128–156.

Grainger, J., and Holcomb, P. J. (2010). "Neural constraints on a functional architecture for word recognition," in *The Neural Basis of Reading*, eds P. Cornelissen, P. Hansen, M. Kringlebach, and K. Pugh (Oxford: Oxford University Press), 3–32.

Grainger, J., and Van Heuven, W. J. B. (2003). "Modeling letter position coding in printed word perception," in *The Mental Lexicon*, ed. P. Bonin (New York: Nova Science Publishers), 1–24.

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

Hauk, O., Pulvermüller, F., Ford, M., Marslen-Wilson, W. D., and Davis, M. H. (2009). Can I have a quick word? Early electrophysiological manifestations of psycholinguistic processes revealed by event-related regression analysis of the EEG. *Biol. Psychol.* 80, 64–74.

Holcomb, P. J., and Grainger, J. (2006). On the time course of visual word recognition: an event-related potential investigation using masked repetition priming. *J. Cogn. Neurosci.* 18, 1631–1643.

Holcomb, P. J., and Grainger, J. (2009). ERP effects of short interval masked associative and repetition priming. *J. Neurolinguistics* 22, 301–312.

Holcomb, P. J., Grainger, J., and O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *J. Cogn. Neurosci.* 14, 938–950.

Hoosain, R. (1991). *Psycholinguistic Implications for Linguistic Relativity: A Case Study of Chinese*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hsiao, J. H., and Shillcock, R. (2006). Analysis of a Chinese phonetic compound database: implications for orthographic processing. *J. Psycholinguist. Res.* 35, 405–426.

Hsiao, J. H., Shillcock, R., and Lee, C. (2007). Neural correlates of foveal splitting in reading: evidence from an ERP study of Chinese character recognition. *Neuropsychologia* 45, 1280–1292.

Hsu, C. H., Lee, C. Y., and Marantz, A. (2011). Effects of visual complexity and sublexical information in the occipitotemporal cortex in the reading of Chinese phonograms: a single-trial analysis with MEG. *Brain Lang.* 117, 1–11.

Hsu, C. H., Tsai, J. L., Lee, C. Y., and Tzeng, O. J. L. (2009). Orthographic combinability and phonological consistency effects in reading Chinese phonograms: an event-related potential study. *Brain Lang.* 108, 56–66.

Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–208.

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

Lee, C.-Y., Tsai, J.-L., Chan, W.-H., Hsu, C.-H., Hung, D. L., and Tzeng, O. J. L. (2007). Temporal dynamics of the consistency effect in reading Chinese: an event-related potentials study. *Neuroreport* 18, 147–151.

Lee, C.-Y., Tsai, J.-L., Chiu, Y.-C., Tzeng, O. J. L., and Hung, D. L. (2006a). The early extraction of sublexical phonology in reading Chinese pseudocharacters: an event-related potentials study. *Lang. Linguist.* 7, 619–635.

Lee, C.-Y., Tsai, J.-L., Huang, H.-W., Hung, D. L., and Tzeng, O. J. L. (2006b). The temporal signatures of semantic and phonological activations for Chinese sublexical processing: an event-related potential study. *Brain Res.* 1121, 150–159.

Lee, C.-Y., Tsai, J.-L., Su, E. C.-I., Tzeng, O. J. L., and Hung, D. L. (2005). Consistency, regularity and frequency effects in naming Chinese characters. *Lang. Linguist.* 6, 75–107.

Leung, M. T., and Lau, K. Y. D. (2010). *Hong Kong Corpus of Chinese Newspaper*. The University of Hong Kong. [Unpublished database].

Li, H., and Chen, H.-C. (1997). "Processing of radicals in Chinese character recognition," in *The Cognitive Processing of Chinese and Related Asian Languages*, ed. H.-C. Chen (Hong Kong: Chinese University Press), 141–160.

Li, Q. L., Bi, H. Y., and Zhang, J. X. (2010). Neural correlates of the orthographic neighborhood size effect in Chinese. *Eur. J. Neurosci.* 32, 866–872.

Liu, Y., Perfetti, C. A., and Hart, L. (2003). ERP evidence for the time course of graphic, phonological, and semantic information in Chinese meaning and pronunciation decisions. *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 1231–1247.

Nobre, A. C., and McCarthy, G. (1994). Language-related ERPs: scalp distributions and modulations by word type and semantic priming. *J. Cogn. Neurosci.* 6, 233–255.

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

Perfetti, C. A., and Liu, Y. (2006). "Reading Chinese characters: orthography, phonology, meaning, and the Lexical Constituency Model," in *The Handbook of East Asian Psycholinguistics*, 1st Edn, Vol. 1: Chinese, eds P. Li, E. Bates, L. H. Tan, and O. J. L. Tzeng (Cambridge: Cambridge University Press), 225–236.

Perfetti, C. A., Liu, Y., and Tan, L. H. (2005). The lexical constituency model: some implications of research on Chinese for general theories of reading. *Psychol. Bull.* 112, 43–59.

Perfetti, C. A., and Tan, L. H. (1998). The time course of graphic, phonological, and semantic activation in Chinese character identification. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 24–33.

Rugg, M. D. (1984). Event-related potentials and the phonological processing of words and non-words. *Brain Lang.* 23, 225–240.

Sereno, S. C., Rayner, K., and Posner, M. I. (1998). Establishing a timeline of word recognition: evidence from eye movements and event-related potentials. *Neuroreport* 9, 2195–2200.

Shillcock, R., Ellison, T. M., and Monoghan, P. (2001). Eye-fixation behavior, lexical storage, and visual word recognition in a split processing model. *Psychol. Rev.* 107, 824–851.

Taft, M. (2006). "Processing of characters by native Chinese readers," in *Handbook of East Asian Psycholinguistics*, 1st Edn, Vol. 1: Chinese, eds P. Li, L. H. Tan, E. Bates, and O. J. L. Tzeng (Cambridge: Cambridge University Press), 237–249.

Taft, M., and Zhu, X. (1997). Submorphemic processing in reading Chinese. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 761–775.

Taft, M., Zhu, X., and Ding, G. (2000). The relationship between character and radical representation in Chinese. *Xin Li Xue Bao* 32(Suppl.), 3–12.

Taft, M., Zhu, X., and Peng, D. (1999). Positional specificity of radicals in Chinese character recognition. *J. Mem. Lang.* 40, 498–519.

Tsang, Y.-K., and Chen, H.-C. (2009). Do position-general radicals have a role of play in processing Chinese characters? *Lang. Cogn. Process.* 24, 947–966.

Yang, J., McCandliss, B., Shu, H., and Zevin, J. (2009). Simulating language-specific, and language-general effects in a statistical learning model of Chinese reading. *J. Mem. Lang.* 61, 238–257.

Yeh, S.-L., and Li, J. L. (2002). Role of structure and component in judgments of visual similarity of Chinese characters. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 933–947.

Zhou, X., and Marslen-Wilson, W. (1999). The nature of sublexical processing in reading Chinese characters. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 819–837.

Early and sustained supramarginal gyrus contributions to phonological processing

*Magdalena W. Sliwinska, Manali Khadilkar, Jonathon Campbell-Ratcliffe, Frances Quevenco and Joseph T. Devlin**

Cognitive, Perceptual and Brain Sciences, University College London, London, UK

Reading is a difficult task that, at a minimum, requires recognizing a visual stimulus and linking it with its corresponding sound and meaning. Neurologically, this involves an anatomically distributed set of brain regions cooperating to solve the problem. It has been hypothesized that the supramarginal gyrus (SMG) contributes preferentially to phonological aspects of word processing and thus plays an important role in visual word recognition. Here, we used chronometric transcranial magnetic stimulation (TMS) to investigate the functional specificity and timing of SMG involvement in reading visually presented words. Participants performed tasks designed to focus on either the phonological, semantic, or visual aspects of written words while double pulses of TMS (delivered 40 ms apart) were used to temporarily interfere with neural information processing in the left SMG at five different time windows. Stimulation at 80/120, 120/160, and 160/200 ms post-stimulus onset significantly slowed subjects' reaction times in the phonological task. This inhibitory effect was specific to the phonological condition, with no effect of TMS in the semantic or visual tasks, consistent with claims that SMG contributes preferentially to phonological aspects of word processing. The fact that the effect began within 80–120 ms of the onset of the stimulus and continued for approximately 100 ms, indicates that phonological processing initiates early and is sustained over time. These findings are consistent with accounts of visual word recognition that posit parallel activation of orthographic, phonological, and semantic information that interact over time to settle into a distributed, but stable, representation of a word.

Keywords: reading, phonology, semantics, chronometric TMS, inferior parietal lobe

INTRODUCTION

From texts to twitter, e-mails to blogs, we live in a society that is dominated by *written* communication. The ease with which we read masks a complex set of processes necessary to link visual symbols with their sounds and meaning. At a neural level, these processes engage an anatomically distributed set of brain regions that, at a minimum, include broad areas of the ventral occipito-temporal (vOT) cortex, the inferior parietal lobule (IPL), and inferior frontal cortex (Pugh et al., 2001; Shaywitz et al., 2002; Price and Mechelli, 2005). Here we focused on a specific sub-field of the IPL, namely the supramarginal gyrus (SMG), and investigated both its functional contribution to reading and also its time course using transcranial magnetic stimulation (TMS).

The IPL is an anatomically heterogeneous area consisting of several distinct cytoarchitectonic fields (Brodmann, 1909; Von Bonin and Bailey, 1947), each with their own pattern of connectivity (Rushworth et al., 2006; Caspers et al., 2011). The most anterior field corresponds to the SMG, an area strongly linked to phonological processing (Petersen et al., 1988; Booth et al., 2004; Seghier et al., 2004; Zevin and McCandliss, 2005; Prabhakaran et al., 2006; Raizada and Poldrack, 2007; Buchsbaum and D'Esposito, 2008; Obleser and Kotz, 2009; Sharp et al., 2010; Yoncheva et al., 2010). Indeed, neuroimaging evidence demonstrates that SMG

responds more strongly during phonological than semantic processing (Demonet et al., 1994; Price et al., 1997; Mummery et al., 1998; Devlin et al., 2003), suggesting a level functional specificity during word recognition. Thus it was surprising that a recent TMS experiment (Stoeckel et al., 2009) found stimulation of the left SMG facilitated both phonological *and* semantic processing, calling into question the specificity of SMG's contribution to reading. It is certainly possible that differences between the fMRI and TMS methodologies could yield conflicting results (e.g., Hamidi et al., 2009), thus we were motivated to further investigate the functional specificity of the SMG during word recognition using a more robust stimulation technique than used in the earlier study.

Our second aim was to investigate the temporal dynamics of SMG contributions to language processing. Traditionally, event-related potential (ERP) and magnetoencephalography (MEG) are most commonly used to measure the time course of processing, taking advantage of their outstanding temporal resolution. Several such studies have reported phonological effects occurring 250–350 ms after the appearance of the visual word (Niznikiewicz and Squires, 1996; Bentin et al., 1999; Newman and Connolly, 2004; Grainger et al., 2006; Ashby and Martin, 2008). Sereno and colleagues, however, have argued compellingly that the phonological (and semantic) processing must happen more rapidly, based on

the rapidity of eye movements during text reading (Sereno et al., 1998; Sereno and Rayner, 2003). In addition, they have used ERPs to demonstrate that higher order properties of words are accessed as early as 100–200 ms after stimulus onset (Sereno et al., 1998). These findings receive additional support from recent ERP and MEG studies suggesting that phonological processing may begin within the first 100 ms of visual word recognition (Ashby et al., 2009; Wheat et al., 2010). Here, we used chronometric TMS to take advantage of its combined temporal (tens of ms) and spatial resolution (approximately 10 mm) to investigate the timing of SMG involvement in reading.

MATERIALS AND METHODS

PARTICIPANTS

Forty right-handed, monolingual native English speakers volunteered to participate in this study, and of these 32 (19 women, 13 men; aged 18–41, mean = 25) were included in the main experiment. For the other eight the functional location procedure failed to identify a region of SMG for testing in the main experiment (see Experimental Procedures below). All participants were neurologically normal, with no personal or family history of epilepsy. In addition, none had any form of dyslexia according to self-reports. Each person provided informed consent after the experimental procedures were explained and subjects were paid for their participation. The experiment was approved by the University College London (UCL) Research Ethics Committee.

EXPERIMENTAL PROCEDURES

There were two testing sessions. The first involved a 30-min visit to the Birkbeck-UCL Neuroimaging Centre (BUCNI) to acquire a T1-weighted structural magnetic resonance imaging (MRI) scan [FLASH sequence, repetition time (TR) = 12 ms, echotime (TE) = 5.6 ms, flip angle = 19, resolution = 1 mm × 1 mm × 1 mm] used to anatomically identify the left SMG in each participant. The second session occurred 2–10 days later and involved the main TMS testing which lasted approximately 1 h.

Before a participant arrived for TMS testing, three potential stimulation targets were identified and marked on their MRI scan

using the Brainsight frameless stereotaxy system (Rogue Research, Montreal, Canada). The first was located just superior to the termination of the posterior ascending ramus of the Sylvian fissure. The second was placed at the ventral end of the anterior SMG, superior to the Sylvian fissure, posterior to the postcentral sulcus, and anterior to the posterior ascending ramus of the Sylvian fissure. The third was approximately halfway between these sites and at least 10–15 mm from the other two (see **Figure 1A**). These three sites were chosen within the anterior region of the left SMG since this area has been shown to be sensitive to phonological processing in neuroimaging studies (e.g., Petersen et al., 1988; Price et al., 1997; Devlin et al., 2003; Seghier et al., 2004; Zevin and McCandliss, 2005; Raizada and Poldrack, 2007). Each site was then tested to functionally localize the specific target site where stimulation interfered with phonological processing.

Target site localization

Participants performed a visual rhyme judgment task to focus attention on the sounds of the words. Each trial began with a fixation cross centrally presented on the screen for 1000 ms immediately followed by two words that appeared simultaneously above and below the cross and remained on the screen for 500 ms. Subjects were asked to judge whether the two words rhymed or not (e.g., *kite-white*) during a 2500-ms inter-trial interval (ITI). Responses were indicated by button press using the left and right index fingers. The pairing of yes/no responses with fingers was counter-balanced across participants. Each run included 34 trials and lasted 1:35 min. Repetitive TMS (10 Hz, 500 ms) was delivered randomly on half of the trials with the caveat that they occurred equally often on yes and no trials. Stimulation began 100 ms after the onset of the word pair. The data from the first two trials per run were discarded to allow participants to get past anticipating the first rTMS trial. When TMS consistently slowed median reaction times (RTs) relative to non-TMS trials, that site was used for testing in the main experiment.

At the beginning of testing, participants performed a practice run of the rhyme judgment task where no TMS was delivered to become familiar with the task. Once they felt comfortable, the first

FIGURE 1 | Stimulation sites. (A) Three possible stimulation targets marked within each participant's left SMG using a frameless stereotaxy system. The first one located just superior to the termination of the posterior ascending ramus of the Sylvian fissure; the second one at the ventral end of the anterior SMG; and the third one approximately halfway between the other two sites. **(B)** The final testing sites for all 32 participants (white filled circles) and the mean group location (black filled circle) on the averaged brain of all participants normalized to the standard MNI152 space with an affine registration (Jenkinson and Smith, 2001) shown on a parasagittal plane.

testing site was chosen and the participant was introduced to the sensation of rTMS at that site. TMS was introduced by placing the coil on the scalp such that the line of maximum magnetic flux intersected the target site. After familiarization with the sensation, each participant performed two more practice runs with concurrent rTMS. Localization then began at the first testing site and each site was tested using one of five matched stimulus sets. When rTMS facilitated RTs, the next site was tested. When rTMS produced numerically longer (i.e., slower) responses, the site was re-tested using a different stimulus set to determine whether the observed slowdown was consistent. Any site that produced two or more RT slowdowns during the localizer task was selected for stimulation in the main experiment. Note that any numeric increase in RTs, including a single millisecond, was considered a "slowdown," which is why it was important to show that slowdowns were consistent rather than a result of idiosyncratic factors. In general, the fact that the "wrong" SMG sites typically led to small speedup effects (presumably due to intersensory facilitation), made even small slowdowns convincing as long as they were reproducible. Once a testing site was identified, the localization procedure stopped in order to limit unnecessary stimulation received by subjects. The order of testing the target sites was counter-balanced across participants. If after 10 runs, no site resulted in consistent TMS-induced slowdowns, then the experiment was terminated.

Experimental tasks

In the main experiment, participants performed three different tasks: (i) a homophone judgment task where they decided whether two words sounded the same; (ii) a synonym judgment task where they decided whether two words meant the same thing; and (iii) a visual judgment control task where they decided whether two consonant letters strings were identical. The first two tasks were designed to emphasize phonological or semantic processing, respectively. The third task was included as a control condition in which stimulation was not expected to affect performance. This task shared visual, decision, and response features of the lexical tasks but no linguistic components. The number of "yes" and "no" responses was equal in all cases.

There were 105 trials per task. The tasks were presented in blocks of 21 trials to minimize task-switching costs. Following a short instruction screen to remind the participant of the task, the first trial in each block was a dummy item and discarded from the analyses to exclude the RT cost of switching tasks. The remaining 20 items in the block constituted the data used for further analysis. A trial commenced with a fixation cross displayed for 500 ms, followed by two letter strings presented above and below the fixation cross for another 500 ms. A blank screen was then presented for a random interval between 1300 and 2300 ms, giving an average duration of 2500 ms per trial. Participants indicated their response with the same button press they used in the functional localizer task. The experiment was divided into three runs of five blocks each lasting approximately 5 min. In between runs, subjects took a self-paced break. The order of tasks was counter-balanced across participants.

Chronometric TMS

A double pulse of TMS was delivered on every trial, at one of five different timing conditions. Pulses occurred at either 40 and 80, 80 and 120, 120 and 160, 160 and 200, or 200 and 240 ms

post-stimulus onset. The TMS timings were not randomly distributed; instead, they were ordered in either an ascending or descending staircase in sets of four trials (**Figure 2**). For instance, the first four trials might have pulses delivered at 40/80 ms, while the next four were at 80/120, etc., such that all 20 trials in the block had TMS delivered at one of the five timing conditions. For the following block (i.e., the next task), the timing went in the opposite direction (i.e., $4 \times 200/240$ followed by $4 \times 160/200$, etc.). The aim of this procedure was to avoid any late stimulation trials (e.g., 160/200) randomly following early trials (40/80) because during pilot studies there was some concern that participants were implicitly waiting for the TMS pulse before responding, and thus artificially inflating RTs on those trials. With the current staircase method there was no evidence that participants waited for the TMS before responding. Indeed, subjects reported that they were not aware that stimulation onsets differed. In contrast, when chronometric timings are delivered randomly subjects are typically aware of the different timings.

Testing began with a practice run performed without TMS in order to familiarize subjects with the task requirements. It included all three tasks and provided practice in switching between them. Subjects were then familiarized with the sensation of double-pulse TMS at the SMG testing site. Finally, they completed the actual experiment for the given site using one of five different stimulus versions. The order of the versions was counter-balanced among participants. None of stimuli used in practice or in the localization procedure were repeated in the main task.
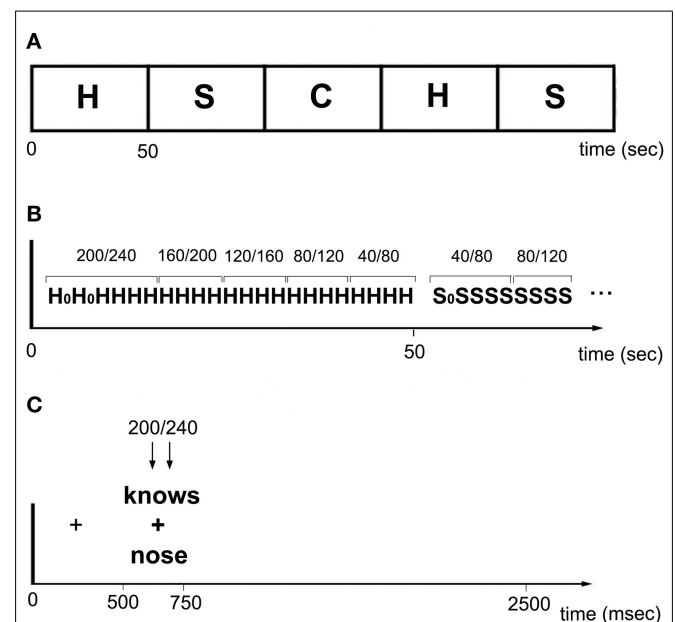
FIGURE 2 | (A) Within a run, homophones (H), synonyms (S), and consonant strings (C) alternated in 50 s blocks. (B) Each block consisted of 20 trials. Pulses occurred at either 40/80, 80/120, 120/160, 160/240, or 200/240 ms post-stimulus onset. TMS timings were ordered in either an ascending or descending staircase in sets of four trials. H_0 and S_0 indicate dummy trials. (C) Each trial began with a fixation cross presented for 500 ms. A stimulus was then presented for 250 m, followed by a blank screen displayed for random interval between 1300 and 2300 ms. Stimulation occurred at one of five time windows.

STIMULI

For the localizer task, word stimuli ($n = 160$ plus 10 dummy trials) ranged in length from three to eight letters and were divided into five separate lists, matched for concreteness, familiarity, written word frequency, number of letters, and number of syllables [one-way ANOVA, all $F(1, 158) < 1.1$, $p > 0.31$]. Concreteness and familiarity ratings were taken from the MRC Psycholinguistic database (Coltheart, 1981), and British English word frequencies came from the Celex database (Baayen and Pipenbrook, 1995). In addition, within each list trials were divided into TMS and no-TMS items that were also matched across these five factors [all $t(30) < 1.8$, $p > 0.1$]. It is worth noting that the orthography of the paired words was manipulated such that participants could not perform rhyme judgment based solely on the word's spelling. The words in rhyming and non-rhyming trials had different spellings in half of the cases (e.g., rhyming: *wall-call* vs. *style-pile*; non-rhyming: *work-pork* vs. *egg-pen*).

For the main experiment, the word stimuli (200 trials plus 12 dummies trials) ranged in length from 3 to 10 letters and were matched across the homophone and synonym tasks for concreteness, familiarity, written word frequency, number of letters, and number of syllables [all $t(198) < 1.66$, $p > 0.11$]. In addition, the consonant strings in the non-lexical task were matched in length to the lexical stimuli. Within each task, the items were divided into five lists, again matched for all factors [all $F(4, 95) < 2.1$, $p > 0.1$]. Then, the lists were paired with each of the five time windows such that the lists occurred with equal frequency within each time window across participants.

TRANSCRANIAL MAGNETIC STIMULATION

Stimulation was performed using a Magstim Rapid[2] stimulator (Magstim, Carmarthenshire, UK) and a 70-mm diameter figure-of-eight coil. The stimulation intensity was set to 55% of the maximum stimulator output and held constant for all subjects. During the localizer task, trains of five pulses (i.e., 10 Hz for 500 ms) were pseudorandomly delivered at 100, 200, 300, 400, 500 ms post-stimulus onset in half of all trials. During the main task, double pulses were delivered 40 ms apart at five different time windows: 40/80, 80/120, 120/160, 160/200, and 200/240 ms following stimulus onset in each trial. The TMS frequency, intensity, and duration were well within established international safety limits (Wassermann, 1998; Rossi et al., 2009). During testing, a Polaris Vicra infrared camera (Northern Digital, Waterloo, ON, Canada) was used in conjunction with the Brainsight frameless stereotaxy system (Rogue Research, Montreal, Canada) to register the participant's head to their own MRI scan in order to accurately target stimulation throughout the experiment. All participants used an earplug in their left ear to attenuate the sound of the coil discharge and avoid damage to the ear (Counter et al., 1991). All participants tolerated TMS well. In some cases, stimulation affected the temporalis muscle and produced a small, unilateral facial twitch. Participants described the sensations as "unusual" but not uncomfortable.

ANALYSES

Reaction times were recorded from the onset of the stimulus and only correct responses were analyzed. TMS was expected to affect RTs rather than accuracy, as previous studies utilizing similar language tasks and stimuli indicate that TMS rarely affects accuracy (Devlin and Watkins, 2007). For the localizer task, the group analysis compared responses to TMS and no-TMS trials when TMS was delivered to the main testing site vs. when it was delivered to the other SMG targets. For the main task the earliest timing window (i.e., pulses delivered at 40/80 ms) was considered the baseline condition as previous ERP, MEG, and TMS findings (e.g., Khateb et al., 1999; Pammer et al., 2004; Stoeckel et al., 2009) indicate that this is too early for TMS to have an effect on SMG during phonological processing. As a result, within each of the three tasks, each of the four later time windows was compared to the baseline, using two-tailed, planned paired t-tests. Anticipatory responses were defined as RTs ≤ 300 ms and were trimmed from the data (0.04% of responses). In all analyses, median RTs for correct responses were used in the statistical analyses to minimize the effect of outliers (Ulrich and Miller, 1994).

In order to identify testing sites in terms of standard space coordinates, each participant's structural scan was registered to the Montreal Neurological Institute-152 template using an affine registration (Jenkinson and Smith, 2001). Note that all stimulation was done in native anatomical space – the standard space coordinates were computed solely for reporting purposes. In addition, for illustrative purposes a group mean structural scan was created in standard space and used as a background image when presenting the stimulation sites in order to accurately reflect the anatomical variability across subjects (Devlin and Poldrack, 2007).

RESULTS

FUNCTIONAL LOCALIZATION

In 8 out of 40 participants, the functional localization process failed and testing ceased after 10 runs. In the remaining 32 participants, an average of five localizer runs per subject (range: 2–10, mean = 6) were required to successfully identify the main SMG testing site. In these participants, rTMS produced a significant inhibitory effect of 44 ms relative to the no-TMS trials [paired t-test; $t(31) = 9.8$, $p < 0.001$]. When normalized to reflect between-subject variability in overall RT, this equated to a 6% slowdown in individuals. In contrast, stimulation of the other SMG sites produced a significant facilitation effect of 32 ms [paired t-test; $t(31) = 4.9$, $p < 0.001$]. When normalized, this constituted a 4% speed-up in RTs. In other words, there was a clear difference between the final test site and other locations, even though they were only 1–2 cm away and still within anterior SMG. The precise location where stimulation interfered with phonological processing varied across individuals and is illustrated in **Figure 1B**. Here, white filled circles show where stimulation led to a slowdown for rhyme judgments in each participant. The mean coordinate in standard space was [−52, −37, +32], a region previously implicated in phonological processing (e.g., Price et al., 1997; Devlin et al., 2003; Seghier et al., 2004; Zevin and McCandliss, 2005; Raizada and Poldrack, 2007).

CHRONOMETRIC TASK

Overall accuracy levels were reasonably high (88%) indicating that participants did not have any difficulty performing the tasks. When accuracy was analyzed with an omnibus 3×5 ANOVA with

Task (Phonological, Semantic, Visual) and TMS (40/80, 80/120, 120/160, 160/200, 200/240) as independent factors, it revealed a significant main effect of Task [$F(2, 63) = 30.4$, $p < 0.001$] indicating that the semantic task (83%) was significantly more difficult than either the phonological task (90%) or the visual task (91%). Neither the main effect of TMS nor its interaction with Task were significant (both $F < 1$). In other words, there was no evidence that TMS affected accuracy in performing any of the three tasks.

The RT results are shown in **Figure 3**. From the figure, it is apparent that there was a main effect of Task [$F(2, 62) = 98$, $p < 0.001$], with slowest responses on the semantic task (893 ms), followed by the phonological task (803 ms) and then the visual task (665 ms), each of which was significant different from the others (all $p < 0.001$, after Bonferroni correction for multiple comparisons). Neither the main effect of TMS [$F(4, 124) = 1.2$, $p = 0.31$] nor the Task × TMS interaction reached significance [$F(8, 248) = 1.26$, $p = 0.27$] in the omnibus ANOVA. Even so, a set of planned comparisons were performed to specifically evaluate whether TMS modified RTs in the phonological and/or semantic task.

For the phonological task, a comparison of each time condition to the baseline condition (40/80 ms) indicated inhibitory effects at all four time windows relative to baseline (plotted in **Figure 4**). We observed RT increases of 30, 30, 25, and 21 ms, although only the first three were significant [80/120: $t(31) = 3.9$, $p = 0.001$; 120/160: $t(31) = 2.4$; $p = 0.02$; 160/200: $t(31) = 2.3$, $p = 0.03$; 200/240: $t(31) = 1.6$, $p = 0.11$]. Despite a similar size inhibitory effect, the final time window did not reach statistical significance because of greater inter-subject variability. Specifically, only 20 out of 32 participants were slowed by TMS during the 200/240 time window. In contrast, 26 subjects showed a slowdown in the 80/120 window, 22 subjects in 120/160 window, and 24 subjects in the 160/200 window. In summary, double pulses of TMS delivered to the same site that slowed performance in the rhyme judgment localizer task resulted in significantly longer RTs between 80 and 200 ms post-stimulus onset.

In contrast, SMG stimulation had no significant effect on either the semantic or visual judgment task. For the semantic task, there were net slowdowns in each of the time windows relative to the baseline condition (40/80 ms), but none of these were significant [all $t(31) < 0.96$, $p > 0.34$]. This was due to considerable inter-subject variability. Specifically, only 18, 15, 19, and 14 participants (out of 32) showed increased RTs in the four respective time windows. For the visual judgment control task, the effects of TMS were variable and none were significant [all $t(31) < 1.1$, $p > 0.3$].

To investigate the functional specificity of the slowdowns observed in the phonological test, we compared them to the TMS effects in the semantic and visual tasks. **Figure 4** illustrates the difference in RTs between TMS and no-TMS trials per time window. Dark gray, light gray, and white bars show TMS effects for phonological, semantic, and visual tasks, respectively. It is clear from the figure that slowdown in the phonological task was significantly greater than both the semantic [paired t-test: $t(31) = 2$, $p = 0.03$] and visual task [$t(31) = 3.1$, $p = 0.002$] in the 80/120 time window. In the later time window, however, the phonological TMS effect did not differ statistically from the semantic TMS effect, despite the fact that there were significant slowdowns relative to baseline in the phonological, but not the semantic, task. Relative to the TMS effects in the visual task, TMS produced significantly larger slowdowns in the phonological task in the 120/160 [$t(31) = 2.2$, $p = 0.02$] and 160/200 [$t(31) = 2.6$, $p = 0.01$] time windows. Finally, there were no significant differences between the TMS effects in the semantic and visual tasks in any time windows [all $t(31) < 0.83$, $p > 0.41$].

DISCUSSION

In the present study TMS was used to investigate functional specificity and timing of phonological processing within the left SMG during reading. There were two main findings. First, the effects of TMS were present for phonological judgments but were not observed for either semantic or visual judgments. Moreover, the
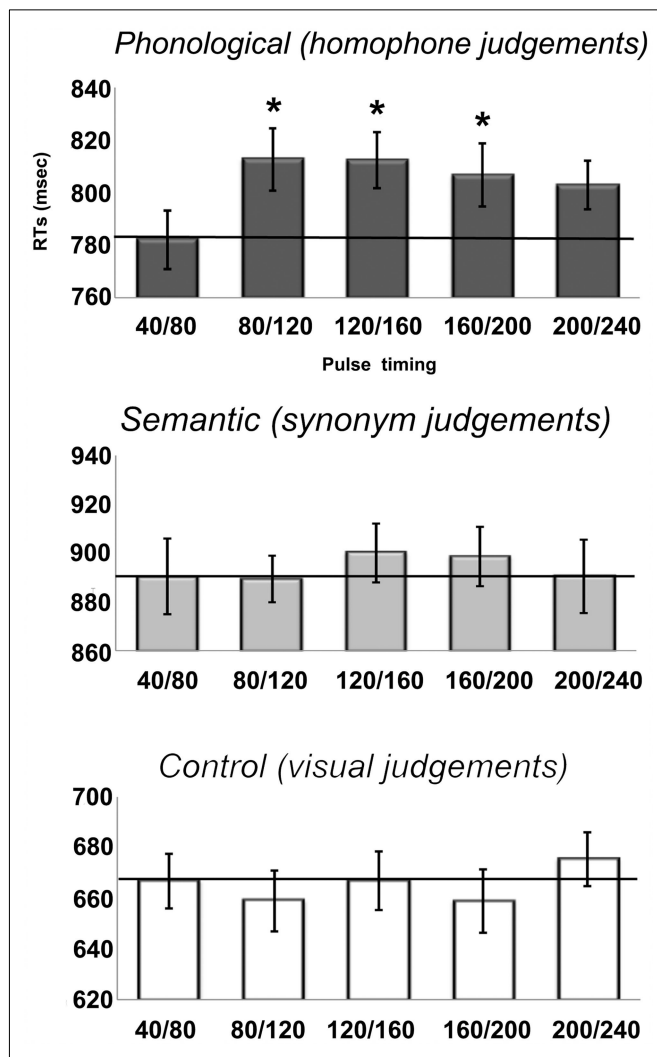
FIGURE 3 | Reaction times (RTs) from the onset of the visual stimulus for each of the five stimulation timings for all three tasks in the main experiment. Note the scales of the y-axes are not identical due to different RTs across the three tasks with visual < phonological < semantic. The solid line represents the baseline RTs. Error bars reflect standard error of the mean adjusted to correctly reflect the variance in the within-subject design (Loftus and Masson, 1994).

FIGURE 4 | The difference between reaction times for each time window relative to its baseline condition (i.e., the 40/80 time window) is plotted for all three tasks. Dark gray bars represent the phonological task, light gray the semantic task and white the visual control task. *$p < 0.05$.

effect of TMS was significantly greater for phonological judgments than either semantic or visual judgments in the 80/120 time window. Second, the inhibitory effects of TMS were apparent as early as 80–120 ms following stimulus presentation and were sustained for approximately another 100 ms. Both of these findings are discussed as they pertain to the neural information processing underlying visual word recognition.

The first aim of this study was to investigate the functional specificity of SMG contributions to word recognition. Previous functional imaging studies involving explicit phonological decisions have consistently revealed SMG activation (Petersen et al., 1988; Booth et al., 2004; Seghier et al., 2004; Zevin and McCandliss, 2005; Raizada and Poldrack, 2007; Buchsbaum and D'Esposito, 2008; Yoncheva et al., 2010). Moreover, the region is activated when participants focus on the sounds of words relative to their meaning (Demonet et al., 1994; Price et al., 1997; Mummery et al., 1998; Devlin et al., 2003; McDermott et al., 2003) suggesting that SMG is preferentially engaged by phonological, rather than semantic, processes. Indeed, the current TMS results are consistent with the imaging findings, confirming a causal link between SMG and phonological processing (Hartwigsen et al., 2010a). SMG stimulation increased response latencies in the phonological task but not in the semantic or visual control tasks. I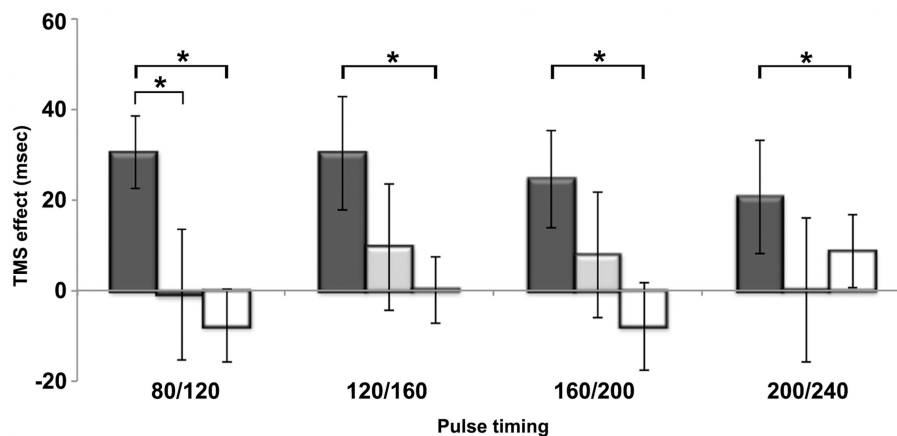ndeed, at the earliest time window (80/120) the effect of TMS on the phonological task (+30 ms) was significantly greater than in the semantic (−1 ms) or the visual (−8 ms) task, suggesting a degree of functional specificity for phonology early in the time course of processing visual words. Moreover, the results imply that the region is not necessary for other types of linguistic processing such as visual word recognition or semantic processing, nor for more domain-general processes such as sustained attention, decision making, action selection, and initiation, etc. A previous study, however, found a different pattern of results where SMG stimulation affected both phonological *and* semantic processing (Stoeckel et al., 2009). We are cautious about these previous findings for three reasons. First, Stoeckel et al. (2009) reported that TMS facilitated, rather than inhibited, response times – an effect that has no clear physiological

basis (Walsh and Pascual-Leone, 2003; Devlin and Watkins, 2007). Second, this facilitation was only present following single pulse stimulation; trains of repetitive TMS delivered to the same site inhibited phonological processing (Stoeckel et al., 2009). Finally, their findings stand in contrast to several previous studies (as well as the current results) that demonstrate stimulation of SMG preferentially interferes with phonological processing (Romero et al., 2006; Hartwigsen et al., 2010a; Pattamadilok et al., 2010). As a result, the weight of evidence from TMS seems to support the imaging findings and suggests that SMG provides a necessary contribution to phonological, but not semantic, processing.

Precisely what aspects of phonological processing are being computed in SMG are open to debate. Studies of speech comprehension, for instance, typically do not show supramarginal activation (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009), even though phonology plays a central role in speech perception. Instead, the region seems to be engaged by more demanding phonological tasks such as rhyme (Petersen et al., 1988; Yoncheva et al., 2010), syllable (Price et al., 1997; Devlin et al., 2003), or phoneme judgments (Zevin and McCandliss, 2005; Raizada and Poldrack, 2007). Pattamadilok et al. (2010) hypothesized that this may be because each of these tasks involves a form of covert articulation where the participant monitors their own inner speech. The SMG is anatomically well situated for this role with reciprocal connections linking it to ventral premotor (PMv) cortex and pars opercularis (POp; Catani et al., 2005; Rushworth et al., 2006; Petrides and Pandya, 2009), two regions involved in articulatory motor planning (Price, 2010). These reciprocal connections between PMv/POp and SMG may form a processing loop for acting on reproducible sound patterns that would provide a simple resonance circuit for temporarily storing these patterns (McClelland and Elman, 1986; Botvinick and Plaut, 2006). Indeed, studies of verbal working memory commonly implicate these regions (Paulesu et al., 1993; Buchsbaum and D'Esposito, 2008; Koelsch et al., 2009) and TMS delivered to PMv/POp also disrupts phonological processing (Nixon et al., 2004; Gough et al., 2005; Romero et al., 2006; Hartwigsen et al., 2010b). In other words, SMG may

play an integral role in representing and processing representations for phono-articulatory patterns that contribute to "phonological processing."

It is important to note, however, that phonological processing is only one of several functions that the SMG contributes to. For instance, the region is also involved in making visually guided hand actions (Rushworth et al., 2001; Binkofski et al., 2004; Price, 2010) and in spatially localizing auditory stimuli (Lewald and Ehrenstein, 2001; Renier et al., 2009). In other words, the apparent functional specificity of the SMG for phonological processing is limited to a very restricted context – namely when processing linguistic information.

The second aim of this study was to investigate the temporal dynamics of SMG contributions to each task by disrupting processing at different time intervals during the first 250 ms of stimulus processing. In the phonological task, a TMS-induced inhibitory effect was present from 80/120 ms post-stimulus onset. Although the detailed mechanisms of action on the cerebral cortex remain unknown (Wagner et al., 2009), it is clear that TMS induces ionic currents in a percentage of neurons in all cortical layers within the stimulated area, leading to inhibitory and excitatory currents within local microcircuits (Esser et al., 2005). These can cause spiking of pyramidal neurons that in turn send a volley of spikes to distal, but anatomically connected regions. Affected neurons then enter a brief refractory state, such that the local physiological effect of a single TMS pulse within the stimulated area lasts approximately 10 ms (Esser et al., 2005), although the distal effects may last for tens of milliseconds. Indeed, chronometric TMS experiments have shown functionally distinct effects of TMS for pulses separated by as little as 40 ms (Amassian et al., 1993; Corthout et al., 1999; Juan and Walsh, 2003; Pitcher et al., 2007). Consequently, it is reasonable to assume that the inhibitory effects of 80/120 stimulation did not last beyond 160 ms post-stimulus onset – earlier than expected based on many ERP findings. For instance, Bentin et al. (1999) used ERPs to measure the time course of phonological processing during a rhyme monitoring task. Both written words and pseudowords produced a negative-going potential beginning as early as 290 ms after the onset of the stimulus, consistent with many similar studies showing phonological effects in the 250- to 300-ms time range (Niznikiewicz and Squires, 1996; Newman and Connolly, 2004; Grainger et al., 2006). Other studies have reported even later phonological effects ranging from 350- to 550-ms (Rugg, 1984; Carreiras et al., 2009). In other words, many studies indicate that the time course of phonological processing in word recognition begins roughly 100 ms later than reported here.

One possible explanation for this apparent discrepancy may have to do with the nature of the different methodologies. ERP and MEG signals reflect the aggregate electromagnetic activity of synchronous neuronal firing and as a result may be less sensitive to the earliest processing dynamics within a region before synchrony has time to develop (Schroeder et al., 1998). In contrast, the effect of TMS occurs immediately with the stimulation pulse and can interfere with neuronal activity that contributes to the build up of the ERP/MEG signal (Walsh and Cowey, 2000). As a result, TMS effects tend to precede those seen in ERP/MEG

and correspond more closely to the timings seen in intracellular recording studies (Corthout et al., 2000; Duncan et al., 2010; Schuhmann et al., 2012). In other words, despite its poorer temporal resolution (tens of ms as opposed to ms), TMS may provide more precise information regarding the onset of regional neuronal activity.

Another possible explanation for the relatively late ERP recordings is that the ERP components such as the N250 or N400 index processes based on recurrent feedback rather than the initial information passing through the system (Sereno and Rayner, 2003). When reading text, the eyes fixate on a word for an average of 250–300 ms (Just and Carpenter, 1980; Rayner et al., 1996), indicating that lexical processing must be underway well before the next saccade. Indeed, Sereno et al. (1998) found that during reading, early ERP components such as the P1 and N1 are influenced by factors such as lexicality and frequency, demonstrating that higher order properties of the word are accessed as early as 100–200 ms post-stimulus onset (see also Hauk and Pulvermuller, 2004). In other words, there is growing evidence that non-visual properties of a word become available as early as 100–200 ms from the onset of the visual word (Ashby et al., 2009; Wheat et al., 2010; Reichle et al., 2011; Hauk et al., 2012).

In addition to this rapid onset, we observed that the effects of TMS were sustained through the 160/200 ms time windows. In contrast, most previous chronometric TMS studies of visual processing have demonstrated separate early and late effects of stimulation, suggesting temporally distinct feed-forward and feedback phases of processing (e.g., Corthout et al., 1999). In our data, however, TMS to each of the time windows between 80/120 and 160/200 ms significantly slowed responses, suggesting on-going phonological processing, presumably due to dynamic interactions with regions processing other aspects of the word including visual and semantic information (Cao et al., 2008; Carreiras et al., 2009; Frye et al., 2010). Indeed, the same temporal pattern of disruption was observed in a chronometric TMS study of left vOTcortex – a region critically involved in processing the visual forms of words (Duncan et al., 2010). Taken together, the results suggest continuous and simultaneous communication between vOT and SMG occurring between approximately 100 and 200 ms after the presentation of a visual word. This type of interactive processing (as opposed to strictly feed-forward processing) is a fundamental principle of virtually all computationally explicit cognitive accounts of visual word recognition (McClelland and Rumelhart, 1981; Seidenberg and McClelland, 1989; Plaut et al., 1996; Coltheart et al., 2001; Jacobs et al., 2003; Harm and Seidenberg, 2004; Perry et al., 2007) and is increasingly important for neuroanatomical models of reading as well (Price and Devlin, 2011; Twomey et al., 2011; Wang et al., 2011; Woodhead et al., 2011). In other words, these data are not only consistent with accounts of visual word recognition that suggest parallel processing of orthographic, phonological (and presumably semantic) information over time and their integration as a result of constant regional interaction in order to achieve stable word representations, but they also provide a tentative time frame for this processing (i.e., 80–200 ms), consistent with estimates of the time available based on both eye movement and ERP data (Sereno and Rayner, 2003).

REFERENCES

Amassian, V. E., Cracco, R. Q., Mac-Cabee, P. J., Cracco, J. B., Rudell, A. P., and Eberle, L. (1993). Unmasking human visual perception with the magnetic coil and its relationship to hemispheric asymmetry. *Brain Res.* 605, 312–316.

Ashby, J., and Martin, A. E. (2008). Prosodic phonological representations early in visual word recognition. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 224–236.

Ashby, J., Sanders, L. D., and Kingston, J. (2009). Skilled readers begin processing sub-phonemic features by 80 ms during visual word recognition: evidence from ERPs. *Biol. Psychol.* 80, 84–94.

Baayen, R. H., and Pipenbrook, R. (1995). *Linguistic Data Consortium. The Celex Lexical Database.* Philadelphia, PA: University of Pennsylvania.

Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F., and Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. *J. Cogn. Neurosci.* 11, 235–260.

Binkofski, F., Buccino, G., Zilles, K., and Fink, G. R. (2004). Supramodal representation of objects and actions in the human inferior temporal and ventral premotor cortex. *Cortex* 40, 159–161.

Booth, J. R., Burman, D. D., Meyer, J. R., Gitelman, D. R., Parrish, T. B., and Mesulam, M. M. (2004). Development of brain mechanisms for processing orthographic and phonologic representations. *J. Cogn. Neurosci.* 16, 1234–1249.

Botvinick, M. M., and Plaut, D. C. (2006). Short-term memory for serial order: a recurrent neural network model. *Psychol. Rev.* 113, 201–233.

Brodmann, K. (1909). *Vegleichende Lokalisationslehre der Grosshirnde.* Leipzig: Barth.

Buchsbaum, B. R., and D'Esposito, M. (2008). The search for the phonological store: from loop to convolution. *J. Cogn. Neurosci.* 20, 762–778.

Cao, F., Bitan, T., and Booth, J. R. (2008). Effective brain connectivity in children with reading difficulties during phonological processing. *Brain Lang.* 107, 91–101.

Carreiras, M., Perea, M., Vergara, M., and Pollatsek, A. (2009). The time course of orthography and phonology: ERP correlates of masked priming effects in Spanish. *Psychophysiology* 46, 1113–1122.

Caspers, S., Eickhoff, S. B., Rick, T., Von Kapri, A., Kuhlen, T., Huang, R., Shah, N. J., and Zilles, K. (2011). Probabilistic fibre tract analysis of cytoarchitectonically defined human inferior parietal lobule areas reveals similarities to macaques. *Neuroimage* 58, 362–380.

Catani, M., Jones, D. K., and Ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Ann. Neurol.* 57, 8–16.

Coltheart, M. (1981). The MRC psycholinguistics database. *Q. J. Exp. Psychol. (Hove)* 33A, 497–505.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychol. Rev.* 108, 204–256.

Corthout, E., Uttl, B., Walsh, V., Hallett, M., and Cowey, A. (1999). Timing of activity in early visual cortex as revealed by transcranial magnetic stimulation. *Neuroreport* 10, 2631–2634.

Corthout, E., Uttl, B., Walsh, V., Hallett, M., and Cowey, A. (2000). Plasticity revealed by transcranial magnetic stimulation of early visual cortex. *Neuroreport* 11, 1565–1569.

Counter, S. A., Borg, E., and Lofqvist, L. (1991). Acoustic trauma in extracranial magnetic brain stimulation. *Electroencephalogr. Clin. Neurophysiol.* 78, 173–184.

Demonet, J. F., Price, C., Wise, R., and Frackowiak, R. S. (1994). Differential activation of right and left posterior sylvian regions by semantic and phonological tasks: a positron-emission tomography study in normal human subjects. *Neurosci. Lett.* 182, 25–28.

Devlin, J. T., Matthews, P. M., and Rushworth, M. F. (2003). Semantic processing in the left inferior prefrontal cortex: a combined functional magnetic resonance imaging and transcranial magnetic stimulation study. *J. Cogn. Neurosci.* 15, 71–84.

Devlin, J. T., and Poldrack, R. A. (2007). In praise of tedious anatomy. *Neuroimage* 37, 1033–1041; discussion 1050–1038.

Devlin, J. T., and Watkins, K. E. (2007). Stimulating language: insights from TMS. *Brain* 130, 610–622.

Duncan, K. J., Pattamadilok, C., and Devlin, J. T. (2010). Investigating occipito-temporal contributions to reading with TMS. *J. Cogn. Neurosci.* 22, 739–750.

Esser, S. K., Hill, S. L., and Tononi, G. (2005). Modeling the effects of transcranial magnetic stimulation on cortical circuits. *J. Neurophysiol.* 94, 622–639.

Frye, R. E., Liederman, J., Hasan, K. M., Lincoln, A., Malmberg, B., McLean, J. III, and Papanicolaou, A. (2010). Diffusion tensor quantification of the relations between microstructural and macrostructural indices of white matter and reading. *Hum. Brain Mapp.* 32, 1220–1235.

Gough, P. M., Nobre, A. C., and Devlin, J. T. (2005). Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. *J. Neurosci.* 25, 8010–8016.

Grainger, J., Kiyonaga, K., and Holcomb, P. J. (2006). The time course of orthographic and phonological code activation. *Psychol. Sci.* 17, 1021–1026.

Hamidi, M., Tononi, G., and Postle, B. R. (2009). Evaluating the role of prefrontal and parietal cortices in memory-guided response with repetitive transcranial magnetic stimulation. *Neuropsychologia* 47, 295–302.

Harm, M. W., and Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychol. Rev.* 111, 662–720.

Hartwigsen, G., Baumgaertner, A., Price, C. J., Koehnke, M., Ulmer, S., and Siebner, H. R. (2010a). Phonological decisions require both the left and right supramarginal gyri. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16494–16499.

Hartwigsen, G., Price, C. J., Baumgaertner, A., Geiss, G., Koehnke, M., Ulmer, S., and Siebner, H. R. (2010b). The right posterior inferior frontal gyrus contributes to phonological word decisions in the healthy brain: evidence from dual-site TMS. *Neuropsychologia* 48, 3155–3163.

Hauk, O., Coutout, C., Holden, A., and Chen, Y. (2012). The time-course of single-word reading: evidence from fast behavioral and brain responses. *Neuroimage* 60, 1462–1477.

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

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.

Jacobs, A. M., Graf, R., and Kinder, A. (2003). Receiver operating characteristics in the lexical decision task: evidence for a simple signal-detection process simulated by the multiple read-out model. *J. Exp. Psychol. Learn Mem. Cogn.* 29, 481–488.

Jenkinson, M., and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* 5, 143–156.

Juan, C. H., and Walsh, V. (2003). Feedback to V1: a reverse hierarchy in vision. *Exp. Brain Res.* 150, 259–263.

Just, M. A., and Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychol. Rev.* 87, 329–354.

Khateb, A., Annoni, J. M., Landis, T., Pegna, A. J., Custodi, M. C., Fonteneau, E., Morand, S. M., and Michel, C. M. (1999). Spatio-temporal analysis of electric brain activity during semantic and phonological word processing. *Int. J. Psychophysiol.* 32, 215–231.

Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Muller, K., and Gruber, O. (2009). Functional architecture of verbal and tonal working memory: an fMRI study. *Hum. Brain Mapp.* 30, 859–873.

Lewald, J., and Ehrenstein, W. H. (2001). Spatial coordinates of human auditory working memory. *Brain Res. Cogn. Brain Res.* 12, 153–159.

Loftus, G. R., and Masson, M. E. J. (1994). Using confidence-intervals in within-subject designs. *Psychon. Bull. Rev.* 1, 476–490.

McClelland, J. L., and Elman, J. L. (1986). The TRACE model of speech perception. *Cogn. Psychol.* 18, 1–86.

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

McDermott, K. B., Petersen, S. E., Watson, J. M., and Ojemann, J. G. (2003). A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. *Neuropsychologia* 41, 293–303.

Mummery, C. J., Patterson, K., Hodges, J. R., and Price, C. J. (1998). Functional neuroanatomy of the semantic system: divisible by what? *J. Cogn. Neurosci.* 10, 766–777.

Newman, R. L., and Connolly, J. F. (2004). Determining the role of phonology in silent reading using event-related brain potentials. *Brain Res. Cogn. Brain Res.* 21, 94–105.

Nixon, P., Lazarova, J., Hodinott-Hill, I., Gough, P., and Passingham, R. (2004). The inferior frontal gyrus and phonological processing: an investigation using rTMS. *J. Cogn. Neurosci.* 16, 289–300.

Niznikiewicz, M., and Squires, N. K. (1996). Phonological processing and the role of strategy in silent reading: behavioral and electrophysiological evidence. *Brain Lang.* 52, 342–364.

Obleser, J., and Kotz, S. A. (2009). Expectancy constraints in degraded speech modulate the language comprehension network. *Cereb. Cortex* 20, 633–640.

Pammer, K., Hansen, P. C., Kringelbach, M. L., Holliday, I., Barnes, G., Hillebrand, A., Singh, K. D., and Cornelissen, P. L. (2004). Visual word recognition: the first half second. *Neuroimage* 22, 1819–1825.

Pattamadilok, C., Knierim, I. N., Kawabata Duncan, K. J., and Devlin, J. T. (2010). How does learning to read affect speech perception? *J. Neurosci.* 30, 8435–8444.

Paulesu, E., Frith, C. D., and Frackowiak, R. S. (1993). The neural correlates of the verbal component of working memory. *Nature* 362, 342–345.

Perry, C., Ziegler, J. C., and Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: the CDP+ model of reading aloud. *Psychol. Rev.* 114, 273–315.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., and Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature* 331, 585–589.

Petrides, M., and Pandya, D. N. (2009). Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biol.* 7, e1000170. doi:10.1371/journal.pbio.1000170

Pitcher, D., Walsh, V., Yovel, G., and Duchaine, B. (2007). TMS evidence for the involvement of the right occipital face area in early face processing. *Curr. Biol.* 17, 1568–1573.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., and Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular

domains. *Psychol. Rev.* 103, 56–115.

Prabhakaran, R., Blumstein, S. E., Myers, E. B., Hutchison, E., and Britton, B. (2006). An event-related fMRI investigation of phonological-lexical competition. *Neuropsychologia* 44, 2209–2221.

Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann. N. Y. Acad. Sci.* 1191, 62–88.

Price, C. J., and Devlin, J. T. (2011). The interactive account of ventral occipitotemporal contributions to reading. *Trends Cogn. Sci. (Regul. Ed.)* 15, 246–253.

Price, C. J., and Mechelli, A. (2005). Reading and reading disturbance. *Curr. Opin. Neurobiol.* 15, 231–238.

Price, C. J., Moore, C. J., Humphreys, G. W., and Wise, R. J. S. (1997). Segregating semantic from phonological processes during reading. *J. Cogn. Neurosci.* 9, 727–733.

Pugh, K. R., Mencl, W. E., Jenner, A. R., Katz, L., Frost, S. J., Lee, J. R., Shaywitz, S. E., and Shaywitz, B. A. (2001). Neurobiological studies of reading and reading disability. *J. Commun. Disord.* 34, 479–492.

Raizada, R. D., and Poldrack, R. A. (2007). Selective amplification of stimulus differences during categorical processing of speech. *Neuron* 56, 726–740.

Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724.

Rayner, K., Sereno, S. C., and Raney, G. E. (1996). Eye movement control in reading: a comparison of two types of models. *J. Exp. Psychol. Hum. Percept. Perform.* 22, 1188–1200.

Reichle, E. D., Tokowicz, N., Liu, Y., and Perfetti, C. A. (2011). Testing an assumption of the E-Z reader model of eye-movement control during reading: using event-related potentials to examine the familiarity check. *Psychophysiology* 48, 993–1003.

Renier, L. A., Anurova, I., De Volder, A. G., Carlson, S., Vanmeter, J., and Rauschecker, J. P. (2009). Multisensory integration of sounds and vibrotactile stimuli in processing streams for "what" and "where". *J. Neurosci.* 29, 10950–10960.

Romero, L., Walsh, V., and Papagno, C. (2006). The neural correlates of

phonological short-term memory: a repetitive transcranial magnetic stimulation study. *J. Cogn. Neurosci.* 18, 1147–1155.

Rossi, S., Hallett, M., Rossini, P. M., and Pascual-Leone, A. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clin. Neurophysiol.* 120, 2008–2039.

Rugg, M. D. (1984). Event-related potentials in phonological matching tasks. *Brain Lang.* 23, 225–240.

Rushworth, M. F., Behrens, T. E., and Johansen-Berg, H. (2006). Connection patterns distinguish 3 regions of human parietal cortex. *Cereb. Cortex* 16, 1418–1430.

Rushworth, M. F., Krams, M., and Passingham, R. E. (2001). The attentional role of the left parietal cortex: the distinct lateralization and localization of motor attention in the human brain. *J. Cogn. Neurosci.* 13, 698–710.

Schroeder, C. E., Mehta, A. D., and Givre, S. J. (1998). A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. *Cereb. Cortex* 8, 575–592.

Schuhmann, T., Schiller, N. O., Goebel, R., and Sack, A. T. (2012). Speaking of which: dissecting the neurocognitive network of language production in picture naming. *Cereb. Cortex.* 22, 701–709.

Seghier, M. L., Lazeyras, F., Pegna, A. J., Annoni, J. M., Zimine, I., Mayer, E., Michel, C. M., and Khateb, A. (2004). Variability of fMRI activation during a phonological and semantic language task in healthy subjects. *Hum. Brain Mapp.* 23, 140–155.

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

Sereno, S. C., and Rayner, K. (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends Cogn. Sci. (Regul. Ed.)* 7, 489–493.

Sereno, S. C., Rayner, K., and Posner, M. I. (1998). Establishing a timeline of word recognition: evidence from eye movements and event-related potentials. *Neuroreport* 9, 2195–2200.

Sharp, D. J., Awad, M., Warren, J. E., Wise, R. J. S., Vigliocco, G.,

and Scott, S. K. (2010). The neural response to changing semantic and perceptual complexity during language processing. *Hum. Brain Mapp.* 31, 365–377.

Shaywitz, B. A., Shaywitz, S. E., Pugh, K. R., Mencl, W. E., Fulbright, R. K., Skudlarski, P., Constable, R. T., Marchione, K. E., Fletcher, J. M., Lyon, G. R., and Gore, J. C. (2002). Disruption of posterior brain systems for reading in children with developmental dyslexia. *Biol. Psychiatry* 52, 101–110.

Stoeckel, C., Gough, P. M., Watkins, K. E., and Devlin, J. T. (2009). Supramarginal gyrus involvement in visual word recognition. *Cortex* 45, 1091–1096.

Twomey, T., Kawabata Duncan, K. J., Price, C. J., and Devlin, J. T. (2011). Top-down modulation of ventral occipito-temporal responses during visual word recognition. *Neuroimage* 55, 1242–1251.

Ulrich, R., and Miller, J. (1994). Effects of truncation on reaction time analysis. *J. Exp. Psychol. Gen.* 123, 34–80.

Von Bonin, G., and Bailey, P. (1947). *The Neocortex of Macaca Mulatta.* Urbana, IL: University of Illinois Press.

Wagner, T., Rushmore, J., Eden, U., and Valero-Cabre, A. (2009). Biophysical foundations underlying TMS: setting the stage for an effective use of neurostimulation in the cognitive neurosciences. *Cortex* 45, 1025–1034.

Walsh, V., and Cowey, A. (2000). Transcranial magnetic stimulation and cognitive neuroscience. *Nat. Rev. Neurosci.* 1, 73–79.

Walsh, V., and Pascual-Leone, A. (2003). *Transcranial Magnetic Stimulation. A Neurochronometrics of Mind.* London: The MIT Press.

Wang, X., Yang, J., Shu, H., and Zevin, J. D. (2011). Left fusiform BOLD responses are inversely related to word-likeness in a one-back task. *Neuroimage* 55, 1346–1356.

Wassermann, E. M. (1998). Risk and safety of repetitive transcranial magnetic stimulation: report and suggested guidelines from the International Workshop on the safety of repetitive transcranial magnetic stimulation, June 5-7, 1996. *Electroencephalogr. Clin. Neurophysiol.* 108, 1–16.

Wheat, K. L., Cornelissen, P. L., Frost, S. J., and Hansen, P. C. (2010). During visual word recognition, phonology is accessed within 100 ms and

may be mediated by a speech production code: evidence from magnetoencephalography. *J. Neurosci.* 30, 5229–5233.

Woodhead, Z. V., Brownsett, S. L., Dhanjal, N. S., Beckmann, C., and Wise, R. J. (2011). The visual word form system in context. *J. Neurosci.* 31, 193–199.

Yoncheva, Y. N., Zevin, J. D., Maurer, U., and McCandliss, B. D. (2010). Auditory selective attention to speech modulates activity in the visual word form area. *Cereb. Cortex* 20, 622–632.

Zevin, J. D., and McCandliss, B. D. (2005). Dishabituation of the BOLD response to speech sounds. *Behav. Brain Funct.* 1, 4.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Electrophysiological cross-language neighborhood density effects in late and early English-Welsh bilinguals

Giordana Grossi[1]*, Nicola Savill[2], Enlli Thomas[3] and Guillaume Thierry[3]

[1] State University of New York at New Paltz, New Paltz, NY, USA
[2] University of York, York, UK
[3] Bangor University, Bangor, UK

Behavioral studies with proficient late bilinguals have revealed the existence of orthographic neighborhood density (ND) effects across languages when participants read either in their first (L1) or second (L2) language. Words with many cross-language (CL) neighbors have been found to elicit more negative event-related potentials (ERPs) than words with few CL neighbors (Midgley et al., 2008); the effect started earlier, and was larger, for L2 words. Here, 14 late and 14 early English-Welsh bilinguals performed a semantic categorization task on English and Welsh words presented in separate blocks. The pattern of CL activation was different for the two groups of bilinguals. In late bilinguals, words with high CLND elicited more negative ERP amplitudes than words with low CLND starting around 175 ms after word onset and lasting until 500 ms. This effect interacted with language in the 300–500 ms time window. A more complex pattern of early effects was revealed in early bilinguals and there were no effects in the N400 window. These results suggest that CL activation of orthographic neighbors is highly sensitive to the bilinguals' learning experience of the two languages.

Keywords: bilingualism, ERPs, neighborhood density, reading, orthography

INTRODUCTION

Research over the last 20 years has shown that, within a language, the number of neighbors (i.e., words created by changing a single letter of a target word – Coltheart et al., 1977) of a target stimulus influences the processing of the target. This effect, named the neighborhood density (ND) effect, is modulated by several factors. For example, whereas words with a high number of neighbors are generally recognized faster than words with a low number of neighbors in lexical decision tasks, an inhibitory effect has generally been found with non-words (e.g., Coltheart et al., 1977; Andrews, 1989; Holcomb et al., 2002). With words, the effect is also modulated by the frequency of the target (e.g., Andrews, 1989, 1992) and the relative frequency of the neighboring words compared to the frequency of the target words (longer RT when neighbors have a higher frequency than the target; see Perea, 1998, for a review). Finally, different ND effects have been observed in different tasks. For example, Carreiras et al. (1997) found that ND effects were inhibitory in a progressive demasking task (where participants had to identify the stimuli), null in a lexical decision task, and facilitatory in a naming task.

Electrophysiological studies have investigated neural indices of such effects. Holcomb et al. (2002) showed that the N400, a marker of lexical and semantic processing usually observed between 350 and 500 ms (e.g., Kutas et al., 2006), was larger when targets had a high compared to low ND. This effect was found in both a lexical decision (for both words and non-words) and semantic categorization task, which suggests that similar mechanisms are at work in the two tasks, and was recently replicated by Müller et al. (2010) and Laszlo and Federmeier (2011). The larger N400 to targets

with high, compared to low, ND has been interpreted in terms of increased lexico-semantic activation of, and competition among, neighbors, according to Holcomb and colleagues, and increased semantic activation of neighbors according to Laszlo and Federmeier (2011). Because ND effects in the N400 time window have been found for both words and pseudowords, Laszlo and Federmeier have concluded that access to meaning is attempted regardless of the orthographic status of the target. According to the authors, these data therefore argue against staged models of word recognition (e.g., Forster, 1999) and support cascade models (e.g., Harm and Seidenberg, 2004).

Both behavioral and electrophysiological studies have shown that ND effects can also be observed cross-linguistically. For example, in van Heuven et al.'s (1998) first experiment, proficient Dutch-English bilinguals performed a progressive demasking task on both Dutch (L1) and English (L2) words. Identification speed in both languages was negatively influenced by the number of orthographic neighbors in the other language (i.e., the higher the ND, the longer the RT). In Experiment 4, a different group of proficient Dutch-English bilinguals performed a lexical decision task on English (L2) words. Again, RTs were longer for English words that had a high number of neighbors in Dutch (L1). These and other data (e.g., Alternberg and Cairns, 1983; Frenck-Mestre, 1993; Bijeljac-Babic et al., 1997) suggest that orthographic representations for the first and the second languages might be organized together in highly proficient bilinguals and trigger a complex series of activation and inhibition processes among words belonging to different languages (Dijkstra and Van Heuven, 2002).

The N400 modulation by ND has also been observed cross-linguistically (Midgley et al., 2008). In a categorization experiment, late French-English bilinguals, all proficient in L2 (English), were asked to perform a go/no-go task and press a button when an animal name was presented on the screen. Participants were presented with two separate lists (French and English words) whose order was counterbalanced across subjects. Cross-language (CL) ND was manipulated in the following way: 50% of the French words had a high number of neighbors in English and 50% had a low CLND. Similarly, 50% of the English words had a high number of neighbors in French and 50% had a low CLND. In general, event-related potentials (ERPs) were more negative for targets with high, compared to low, CLND. However, the pattern of effects depended on the target language. The N400 (300–500 ms) effect peaked later and was less widely distributed for L1 than L2 targets. Furthermore, early effects (P2/N2, 175–275 ms) were present only for L2 targets. These effects were absent in a group of monolingual English speakers.

Midgley et al. (2008) interpreted the difference in CLND effects between the two languages in terms of frequency of exposure: the participants were more proficient in French, French being their first language; therefore, the connection strength between lexical representations was stronger for L1 than L2. As a consequence, French neighbors were more easily activated by English targets than English neighbors by French targets. A similar interpretation was proposed to explain the presence of early effects (P2/N2) for L2 targets (which were present in Holcomb et al., 2002, but only in the categorization task). According to the authors, differences in frequency between the targets and their neighbors in the two studies would explain the discrepancy in results. In Holcomb et al. (2002), both target and neighboring words had a high subjective written frequency, whereas, in Midgley et al. (2008), L2 targets had a lower subjective frequency than their L1 neighbors. Therefore, in the second study, the activation and competition from high-frequency neighbors would have started earlier.

GOALS OF THE PRESENT STUDY

The behavioral and electrophysiological data reviewed so far support the non-selective access hypothesis, according to which, during presentation of single words, multiple lexical representations are activated (mainly bottom-up); especially those representations from L1 that have some sort of orthographic, phonological, or semantic overlap with L2 input (e.g., van Heuven et al., 1998; Dijkstra et al., 1999; Haigh and Jared, 2007; for activation through translation, see Thierry and Wu, 2007; Wu and Thierry, 2010; Zhang et al., 2011). According to the Bilingual Interaction-Activation (BIA+) model (Dijkstra and Van Heuven, 2002), these multiple representations compete with each other through lateral inhibition. As a result, both within-language and CL lexical interference effects can arise. A similar interpretation was proposed by Holcomb and Grainger (2007) to explain Holcomb et al.'s (2002) data on within-language ND effects. Midgley et al. (2008) also interpreted their CL effects in terms of lexical competition between word form representations (orthographic and/or phonological) from the two languages.

The goal of the present study was to replicate and extend Midgley et al.'s (2008) experiment by employing a different language pair (English and Welsh) and two groups of bilingual individuals: late bilinguals, who started learning Welsh during or after puberty, and early bilinguals, who learn both English and Welsh early in life. The comparison between late and early bilinguals will provide invaluable information on whether the pattern of CL activation differs depending on when the second language is learned: consecutively to, or concurrently with, the first language. Studying Welsh and English as a language pair allows testing potential interactions between orthographic transparency and language non-selective lexical access.

Welsh orthography is rather different from English orthography. First, it is transparent and, in contrast to English, has essentially one-to-one mapping between graphemes and phonemes (Frost et al., 1987; Ellis and Hooper, 2001). Also, it is characterized by letter combinations fairly uncommon in English. For example, many words start with double consonants such as "ll"/ɬ/and "ff"/f/. Diphthongs like "wy"/ʊɨ/and "ae"/ɑːɨ/or/ɑːɪ/ are quite common; and "w"/u/and "y"/ə/ are vowels. Therefore, Welsh word forms can look quite different from English word forms. Indeed, native English speakers who are not familiar with Welsh show no word and pseudoword superiority effects (considered to be measures of familiarity with the words and the orthography of a language, respectively; McClelland, 1976; Carr and Pollatsek, 1985; Grainger et al., 2003) in a forced-choice letter identification task (Grossi et al., 2008).

Participants performed a semantic categorization task with Welsh and English words presented in separate blocks. Based on previous literature on within-language and CLND, it was predicted that high, compared to low, CLND words would generate more negative ERPs starting at around 175 ms post-stimulus onset. Based on Midgley et al. (2008), this effect was predicted to be asymmetric in late bilinguals, with stronger effects for L2 compared to L1 targets, assuming that different pattern of early and late effects for L1 and L2 in late proficient bilinguals reflects frequency of exposure. In early bilinguals, based on the frequency of exposure hypothesis, we predicted similar effects for L1 and L2 targets, as these participants had extensive exposure to both languages.

MATERIALS AND METHODS
PARTICIPANTS

A detailed description of participants' characteristics can be found in Grossi et al. (2010); see also Table 1, p. 126)[1]. Analyses were carried out on 14 early Welsh/English bilinguals (six females, mean age of 38.4 years, range 22–52 years) and 14 late learners of Welsh (10 females, mean age of 40.3 years, range 25–52 years). Based on self-report, all participants had normal or corrected-to-normal vision (20/20), and none had a history of neurological disorders. Based on self-report and the Edinburgh Handedness Inventory (Oldfield, 1971), all late bilinguals were right-handed; in the early bilingual group, 12 participants were right-handed, one was left-handed, and one was ambidextrous. All participants were paid £7/h for their participation.

---

[1]The data discussed in this paper are from the same study described in Grossi et al. (2010); in that paper, we focused on the N1 lateralization for the two languages; here, we focused on cross-language neighborhood effects, investigated in different time windows.

Based on self-reports, early bilinguals learned Welsh from birth ($n = 10$) or early in life (three from age 3, and one from age 5); as for English, seven learned it from birth, two before age 3, three from age 4, and two from age 5. The primary language spoken at home until 2 years of age was Welsh for six participants, a mix of Welsh and English for four participants, and English for four participants. Elementary education was in Welsh for five participants, balanced for one participant, predominantly in Welsh for six participants, and predominantly in English for two participants. Middle school and high school instruction was in both Welsh and English for all early bilinguals. In terms of language proficiency, all early bilingual participants rated themselves as native-like speakers in both languages. All participants rated themselves as native-like in reading English; for Welsh, eleven participants rated themselves as native-like, and three participants as somewhat proficient. Early participants reported speaking Welsh almost half of the time ($M = 47.5\%$, $SD = 25.8$) and reading Welsh for recreational reading 28% of the time ($SD = 26.7$).

For late bilinguals, the mean age of acquisition for Welsh was 28.3 years ($SD = 8.7$), and the average number of years of Welsh was 11.9 ($SD = 6.9$). Four participants held a college degree, and 10 held a post-graduate degree. The primary language spoken at home until 2 years of age was English for 13 participants, and Polish for 1 participant. Elementary education was in English for all participants. Most participants had English as the only language of instruction in both middle school ($n = 11$) and high school ($n = 12$; the other participants were exposed to some Welsh). When asked to indicate how well they felt they spoke Welsh and English, all participants rated themselves as native-like in English; nine participants rated themselves as native-like in Welsh, four as somewhat proficient, and one between these two levels. In terms of proficiency in reading, all participants rated themselves as native-like in English; eight participants rated themselves as native-like in Welsh, five as somewhat proficient, and one as low proficient. Participants reported to speak Welsh 30% of the time ($SD = 22.3$) and to read Welsh for recreational reading 22.5% of the time ($SD = 14.8$).

Proficiency in Welsh was also measured objectively with a translation task including all Welsh words used in the semantic categorization task ($n = 96$). The task was administered at the end of the experimental session before the debriefing. Participants were asked to circle all the familiar Welsh words and, when possible, provide the correct English translation. As expected, early bilinguals translated Welsh words with a higher degree of accuracy (91.15 vs. 80.73%) than late bilinguals and indicated fewer Welsh words as being completely unfamiliar (3.57 vs. 11.24%; see Grossi et al., 2010, Table 2, p. 126 for more information).

STIMULI AND MATERIALS
Two lists of 80 Welsh and 80 English words were created: 50% with high CLND and 50% with low CLND. Therefore, there were 40 words in each of the following categories: high CLND Welsh, low CLND Welsh, high CLND English, low CLND English. In addition, animal names were used as probe stimuli (20% per block, $n = 16$ for each language block). Welsh words were selected from the Cronfa Electroneg o Gymraeg (Ellis et al., 2001); English words were selected from the CELEX database (Baayen et al., 1995).

Words were four- or five-letter words, either mono- or bi-syllabic. Words with at least one occurrence per million were selected and used to calculate the number of orthographic neighbors of words within and across languages. The final set of stimuli for the study were 80 English (mean frequency = 80.32, SD = 93.92) and 80 Welsh words (mean frequency = 74.85, SD = 70.81; the difference in frequency was not significant, $p = 0.69$) between four and five letters in length with half of the items in each language having many orthographic neighbors in the other language and the other half having few neighbors in the other language. English items with high Welsh ND had a mean number of Welsh neighbors of 7.9 (range = 4–12, SD = 2.1). English items with low Welsh ND had 0.23 (range = 0–2, SD = 0.58) neighbors on average. The difference between the two means was significant ($p < 0.0001$, two-tailed). Stimuli were matched on within-language neighborhood size. The list of stimuli and information about orthographic and lexical characteristics can be found in Grossi et al. (2010).

The 16 Welsh and 16 English animal names were matched in length (Welsh, $M = 4.5$, SD = 0.52; English, $M = 4.43$, SD = 0.51; $p = 0.73$, two-tailed) and frequency (Welsh, $M = 26.56$, SD = 41.64; English, $M = 15.63$, SD = 29; $p = 0.4$, two-tailed).

PROCEDURE
Participants gave written consent and filled out the handedness and biographical questionnaires. Next, they performed the semantic categorization task. All participants were tested in a sound-attenuating and electrically shielded booth, and seated 100 cm directly in front of a 19-inch monitor. The sequence of events was the following: a fixation point appeared at the center of the screen and served as a warning signal that a trial was about to begin; the fixation point was followed by a random and variable interval between 500 and 700 ms, after which words were presented for 1000 ms and followed by 1000 ms of blank screen. Each trial ended with a screen indicating that participants could blink. Participants were instructed to press a button, as quickly and as accurately as they could, every time an animal name would appear on the screen. Practice trials presented at the beginning allowed participants to familiarize themselves with trial structure. The session was self-paced: participants controlled when the next trial would begin by pressing a button on a response box. The entire experimental session lasted between 2 and 3 h.

ERP DATA COLLECTION
Electrophysiological data were recorded in reference to Cz at a rate of 1000 Hz from 64 Ag/AgCl electrodes placed according to the extended 10–20 convention (Neuroscan system). Impedances were kept below 7 kΩ. EEG activity was filtered on-line band pass between 0.1 and 200 Hz and re-filtered off-line with a 30 Hz low pass zero phase shift digital filter. Eye-blinks were detected using the vertical electrooculogram bipolar channel. Potential variations exceeding a threshold of 20% of maximum EEG amplitude over the duration of a complete individual recording session were automatically registered as artifacts and contributed to the computing of a model blink artifact (derived from more than 100 individual blink artifacts in each participant). Artifacts were then individually corrected by subtracting point-by-point amplitudes of the model from signals measured at each channel proportionally

to local maximum signal amplitude. Eye movements, drifts, and other artifacts were removed by an algorithm that eliminated all events associated with brain waves that were larger than 75 μV or smaller than −75 μV. The percentage of accepted trials was 89%. Epochs ranged from −500 to 1000 ms after the onset of the critical word. Baseline correction was performed in reference to pre-stimulus activity (500 ms baseline) and individual averages were digitally re-referenced to the left and right mastoid channels offline. Behavioral data were collected simultaneously to ERP data.

MEASURES AND ANALYSES

Analyses were conducted in the following time windows: 175–300 and 300–500 ms (classical N400 window). Omnibus analyses were conducted on the following factors: Group (between-subjects), Language (English, Welsh), and CLND (high, low). In order to describe the scalp distribution of Language and CLND effects, the following repeated measures factors were also included: Hemisphere (left, right), Laterality (lateral, medial), and Anteriority (central, centroparietal, parietal). Analyses were informed by regions of interest highlighted by Midgley et al. (2008) and conducted at the sites where CLND effects were largest, based on visual inspection. The following electrodes were included in the main analyses: C3/4, C1/2, CZ (central), CP3/4, CP1/2, CPZ (centroparietal), and P3/4, P1/2, PZ (centroparietal). Analyses on midline sites were run separately from hemisphere analyses.

In late bilinguals, CLND effects started at around 175 ms for Welsh stimuli over central, centroparietal, and parietal sites and continued until approximately 500 ms. In early bilinguals, the largest differences were more frontal. Therefore, for this group, analyses were also carried out over frontal (F5/6, F3/4, Fz) and frontocentral (FC3/4, FC1/2, FCZ) sites. The dependent variable was mean ERP amplitude in each of the intervals of interest. Words rated as unfamiliar by the participants were excluded from analysis. Significant interactions involving condition effects were followed up by simple effects analyses. Adjusted p-values (Geisser–Greenhouse correction) are reported for all within-subject measures with more than one degree of freedom.

RESULTS

BEHAVIORAL RESULTS

A detailed discussion of the results can be found in Grossi et al., 2010; see Table 4 on p. 129). Late bilinguals were faster and more accurate in detecting target words in English than Welsh. Mean accuracy was 99.11% (SD = 1.91) for English and 84.15% (SD = 16.65) for Welsh. Mean RTs were 575.96 ms (SD = 74.54) for English and 666.40 ms (SD = 76.14) for Welsh. The difference between language conditions was significant for both RT and accuracy (both p's < 0.01). Early bilinguals showed no differences in accuracy for the two languages (English, M = 98.21, SD = 2.66; Welsh, M = 94.20, SD = 9.48; p = 0.16), and were faster in recognizing English (M = 565.07 ms, SD = 64.47) than Welsh (M = 619.26 ms, SD = 72.92) targets (p = 0.008).

EVENT-RELATED POTENTIALS

Figure 1 depicts the ERPs elicited by English and Welsh words for the two groups of participants. Welsh targets elicited more negative ERPs than English targets from around 300 ms and until approximately 650 ms for late but not in early bilinguals. The distribution of the Language effect (in

FIGURE 1 | Mean grand-averages ERPs for Welsh and English stimuli over medial and midline sites in late and early bilinguals. Negative is plotted up.
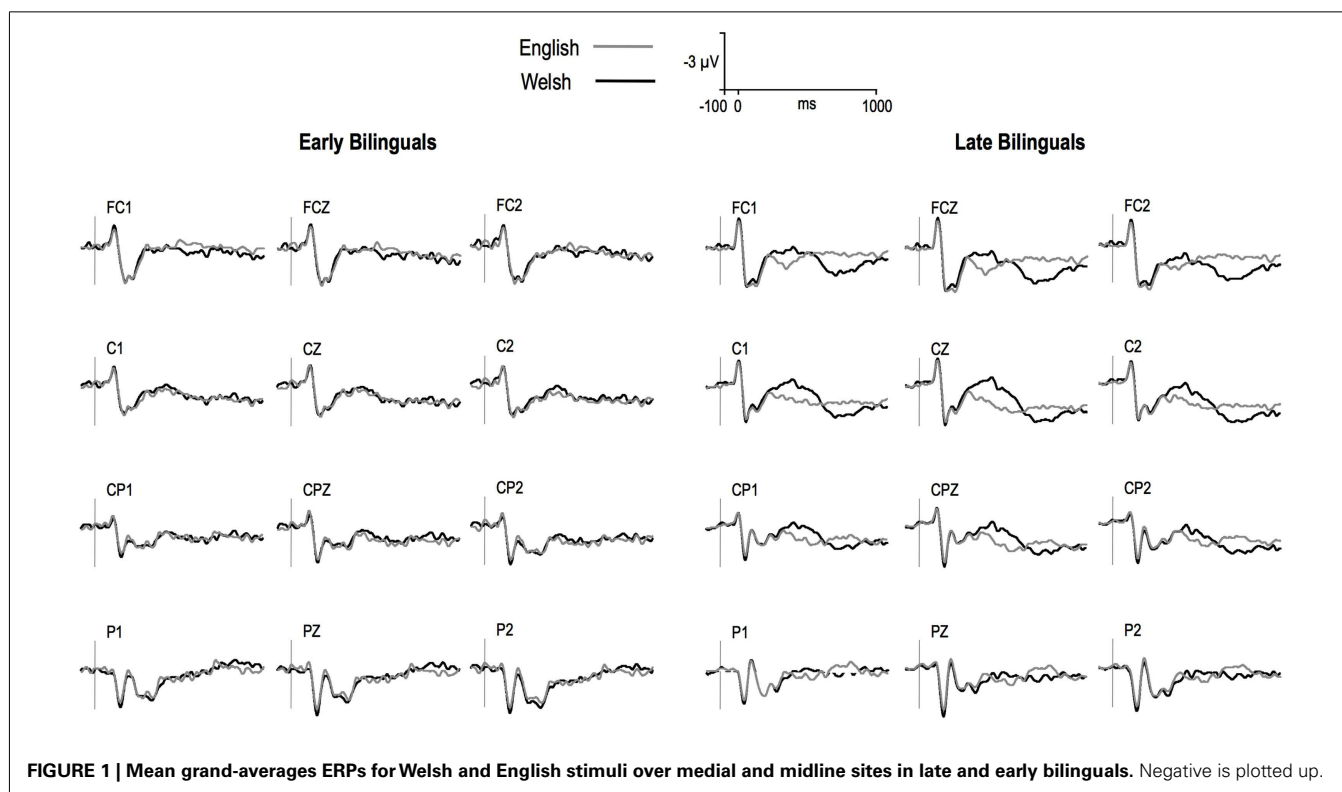
FIGURE 2 | Difference voltage maps representing the 300–650 ms language effect (English–Welsh) in the two groups of bilinguals and mean grand-averages ERPs at the site where the effect was largest (negative is plotted up).

Table 1 | Relevant findings for late and early bilinguals in analyses at central, centroparietal, and parietal sites.

Late bilinguals	175–300 ms	Hemisphere analysis ND, $F(1,13) = 4.68$, $p = 0.05$ ND × Hemisphere × Laterality, $F(1,13) = 4.67$, $p = 0.05$ Language × Laterality × Anteriority, $F(2,26) = 9.73$, $p = 0.001$ Midline analysis ND, $F(1,13) = 9.25$, $p = 0.009$ Language × Anteriority, $F(2,26) = 3.3$, $p < 0.08$
	300–500 ms	Hemisphere analysis ND, $F(1,13) = 4.07$, $p = 0.065$ Language × ND × Hemisphere × Anteriority, $F(2,26) = 3.25$, $p = 0.056$ Language × Laterality, $F(1,13) = 10.17$, $p = 0.007$ Language × Laterality × Anteriority, $F(2,26) = 13.37$, $p < 0.0001$ Midline analysis ND, $F(1,13) = 5.13$, $p = 0.04$ Language, $F(1,13) = 4.13$, $p = 0.06$ Language × Anteriority, $F(2,26) = 8.62$, $p = 0.01$
Early bilinguals	175–300 ms	Hemisphere analysis ND × Hemisphere × Laterality, $F(1,13) = 8.84$, $p = 0.01$ Language × Laterality, $F(2,26) = 4.86$, $p = 0.04$
	300–500 ms	Hemisphere analysis ND × Hemisphere × Laterality, $F(1,13) = 4.2$, $p = 0.06$

The results pertain to omnibus ANOVAs.

terms of difference voltage maps) is shown in **Figure 2**. Omnibus hemisphere analyses for the two time windows showed that the CLND effect and Language effect differed between groups [175–300 ms: Language × ND × Group, $F(1,26) = 5.52$, $p < 0.03$; ND × Hemisphere × Group, $F(1,26) = 5.63$, $p < 0.03$; ND × Hemisphere × Laterality × Group, $F(1,26) = 3.94$, $p < 0.06$; ND × Hemisphere × Laterality × Group, $F(1,26) = 12.99$, $p = 0.001$; 300–500 ms: Language × Laterality × Group, $F(1,26) = 3.74$, $p = 0.06$; ND × Hemisphere × Laterality × Group, $F(1,26) = 5.18$, $p = 0.03$; Language × Laterality × Anteriority × Group, $F(2,52) = 5.95$, $p = 0.006$]. **Table 1** presents a summary of relevant findings

at centroparietal sites in an omnibus ANOVAs for the two groups. Only the main results and follow-up analyses will be discussed in the next section.

LATE BILINGUALS
175–300 ms
Analyses conducted on lateral and medial electrodes showed that ERP amplitudes were more negative for high compared to low CLND targets in this time window [$F(1,13) = 4.68$, $p = 0.05$]; this effect interacted with Hemisphere and Laterality. Follow-up analyses showed that CLND was significant as a main effect

FIGURE 3 | Difference voltage maps representing the cross-language ND effect (high ND – low ND) in late bilinguals and mean grand-averages ERPs at the site where the effect was largest (negative is plotted up).

over the left hemisphere sites [$F(1,13) = 5.45$, $p < 0.04$]; over the right hemisphere sites, ND was significant only over medial sites [ND × Laterality, $F(1,13) = 9.5$, $p = 0.009$; medial sites, $p < 0.05$; over right lateral sites, a significant interaction between ND and Language was observed at the 0.05 level, but analyses carried out separately for the two languages did not reveal any significant ND effect]. Therefore, overall, the CLND effect (more negative ERPs to high than low CLND targets) was more robust over the left hemisphere sites and over the medial sites (**Figure 3**). This main effect did not interact with Language (all p's > 0.1). These results were confirmed by midline analyses (ND, $F(1,13) = 9.25$, $p = 0.009$; no significant interactions between Language and CLND were observed, all p's > 0.11).

In hemisphere analyses, Language interacted with Laterality and Anteriority, revealing some distributional differences between targets in the two languages. However, follow-up analyses did not reveal any reliable Language effect in this time window (all p's > 0.08). Similarly, midline analyses only revealed a trend for significance for the Language × Anteriority interaction, but no significant Language effects were found when analyses were run at each level of Anteriority (all p's > 0.34).

300–500 ms
In hemisphere analyses, ERPs tended to be more negative for high compared to low CLND targets [$F(1,13) = 4.07$, $p = 0.065$]; this effect was qualified by a four-way interaction with Language, Hemisphere, and Anteriority. For English targets, CLND was not significant as a main effect ($p = 0.5$), but interacted with Hemisphere and Laterality [$F(1,13) = 7.56$, $p < 0.02$]. However, no ND effects were significant in follow-up analyses by Hemisphere

and Laterality (all p's > 0.21). In contrast, the ND main effect was significant for Welsh targets [$F(1,13) = 4.77$, $p < 0.05$]. No other interaction between Language and ND reached significance. Analyses over the midline sites revealed a main effect for CLND [$F(1,13) = 5.13$, $p = 0.04$]. This effect did not interact with Language (all p's > 0.2).

Event-related potentials were more negative for Welsh than English targets in this time window. In hemisphere analyses, Language interacted with Laterality and with Laterality and Anteriority. Follow-up analyses showed that the Language effect was significant over medial sites [Language × Anteriority, $F(2,26) = 6.35$, $p = 0.02$; central, $p = 0.003$; centroparietal, Language × Hemisphere, $p = 0.05$; parietal, all p's > 0.14] but not lateral sites (all p's > 0.13). Midline analyses revealed a similar pattern of results: the main effect of Language approached significance [$F(1,13) = 4.13$, $p = 0.06$] and was qualified by a Language × Anteriority interaction [$F(2,26) = 8.62$, $p = 0.01$]: Welsh targets elicited more negative ERP amplitudes than English targets at CZ and CPZ sites ($p = 0.003$, $p < 0.05$, respectively; PZ, $p = 0.78$).

EARLY BILINGUALS
175–300 ms
Hemisphere analyses over central, centroparietal, and parietal sites revealed an interaction between CLND, Hemisphere, and Laterality [$F(1,13) = 8.84$, $p = 0.01$]. Follow-up analyses carried out on each hemisphere separately did not reveal any significant ND effects. No ND effects were observed over the midline sites (all p's > 0.33). ERP amplitudes were more negative for English than Welsh targets at centroparietal and parietal sites
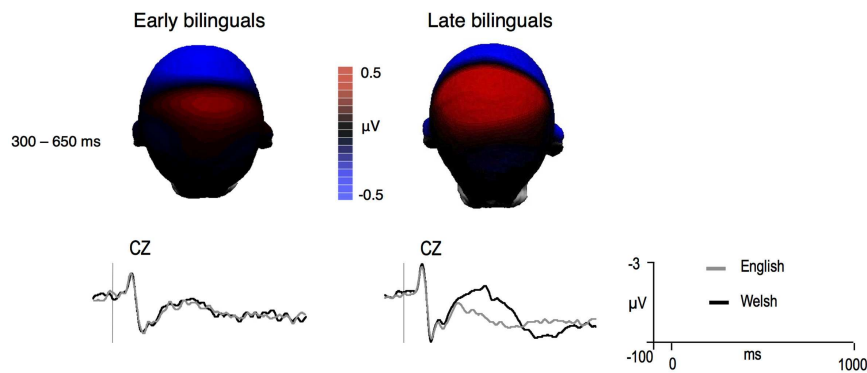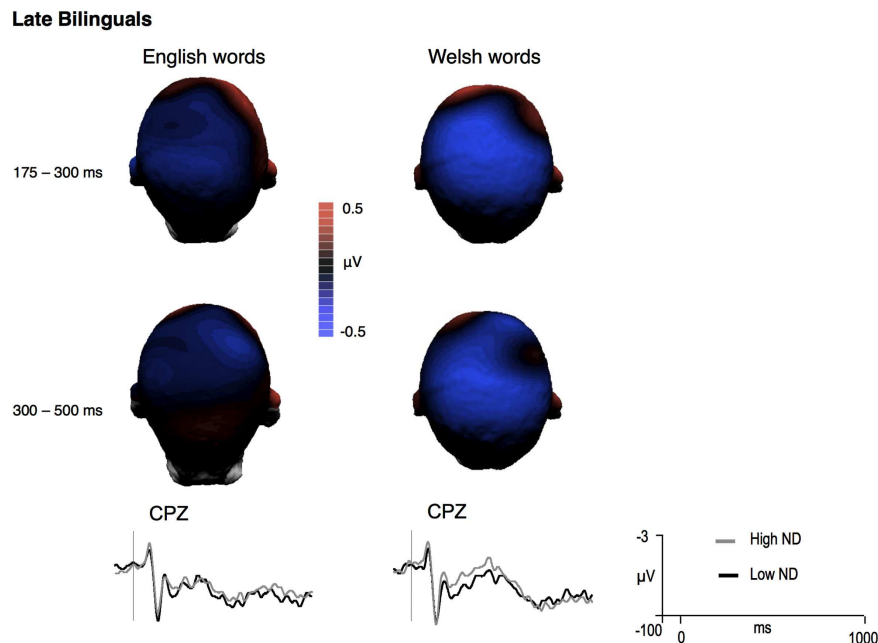
FIGURE 4 | Difference voltage maps representing the 175–300 ms language effect (English – Welsh) in early bilinguals and mean grand-averages ERPs at the site where the effect was largest (negative is plotted up).

FIGURE 5 | Difference voltage maps representing the cross-language ND effect (high ND – low ND) in early bilinguals and mean grand-averages ERPs at the site where the effect was largest (only for **the 175–300 ms time window, as no significant cross-language ND effects were observed during the 300–500 ms time window; negative is plotted up).**

[Language × Anteriority interaction, $F(2,26) = 4.86$, $p = 0.04$; central, $p = 0.79$; centroparietal, $p < 0.03$; parietal, $p < 0.05$]. The distribution of the effect is shown in **Figure 4**. No differences between Welsh and English targets were detected during this time window in midline analyses (all p's > 0.22).

Hemisphere analyses over frontal and frontocentral sites revealed a trend for the interaction between Language and CLND [$F(1,13) = 3.8$, $p = 0.07$]. The ND effect tended to be significant for English targets [$F(1,13) = 4.1$, $p = 0.06$]. For Welsh targets, ND interacted with Hemisphere [$F(1,13) = 4.68$, $p = 0.05$]. Over the left hemisphere, the ND effect was reversed, in that Welsh words with high CLND tended to elicit more positive ERP amplitudes than Welsh words with low CLND [$F(1,13) = 4.25$, $p = 0.06$]. No ND effects were observed over the right hemisphere sites (all p's > 0.29). The distribution of the effects is shown in **Figure 5**.

Midline analyses revealed a significant interaction between Language and CLND [$F(1,13) = 4.86$, $p < 0.05$]. The ND effect was significant for English [$F(1,13) = 5.00$, $p < 0.05$] but not for Welsh ($p = 0.39$) targets.

300–500 ms

In hemisphere analyses, a trend was found for the interaction between ND, Hemisphere, and Laterality [$F(1,13) = 4.2$, $p = 0.06$], suggesting distributional differences between high and low CLND targets. However, no ND effects resulted significant in follow-up analyses. No ND or Language effects were detected in this time window in midline analyses (all p's > 0.21).

DISCUSSION

The present study was aimed at replicating and extending Midgley et al.'s (2008) data on the effects of CLND in visual word recognition by comparing late and early bilinguals. Late bilinguals learned Welsh later in life, whereas early bilinguals were exposed to both English and Welsh either at birth or during early childhood. Both behavioral and electrophysiological data revealed differences between the two languages in late bilinguals: they were less accurate and slower in detecting Welsh targets compared to English targets in the categorization task; furthermore, Welsh words elicited more negative ERPs than English words starting at around 300 ms, suggesting that L2 words required more processing resources than L1 words. These large effects were absent in early bilinguals, who only showed slower RT to Welsh than English targets in the categorization task, likely reflecting the fact that English remained, in terms of reading, the dominant language. Electrophysiologically, only a small (0.2 μV) effect was found over centroparietal and parietal sites, where English targets elicited more negative ERP amplitudes compared to Welsh targets in the 175–300 ms time window.

As expected, based on Midgley et al. (2008), targets with high CLND elicited more negative ERPs as compared to low CLND targets over central, centroparietal, and parietal sites from 175 to 500 ms in late bilinguals. In contrast to Midgley and colleagues, this effect did not interact with Language, implying that both English and Welsh targets contributed to it. Therefore, in proficient late bilinguals, words in one language activate the orthographic representation of words in the other language before 250 ms, supporting the non-selective access account of single word recognition (e.g., Dijkstra and Van Heuven, 2002, but see Wu and Thierry, 2010, for the case of low proficient bilinguals with languages very different in terms of script). According to this model, the two languages are integrated in a single lexicon; presentation of a word in a language causes the activation of words in the other language that overlap in form (orthographic and phonological) and/or meaning. Therefore, it is the similarity between the stimulus and internal representations that drives activation, not the language to which words belong (Dijkstra and Van Heuven, 2002). Indeed, the Language effect started later than the CLND effect in late bilinguals. Furthermore, given that L2 was opaque in Midgley et al. (2008) and transparent in the present experiment, we can conclude that CL orthographic neighborhood effects are not modulated by orthographic transparency, in line with data from studies on within-language ND.

The lack of interaction between Language and ND in the 175–300 ms time window might have been due to the small number of participants. Inspection of **Figure 3** suggests that the effect was not completely symmetrical (analyses run on each language separately confirmed that the ND effect was significant for Welsh but not English targets: Welsh, $p < 0.04$, English, $p = 0.67$ in hemisphere analyses; Welsh, $p = 0.005$, English, $p = 0.16$ in midline analyses). We asked whether differences in experience and proficiency with L2 among our participants might have contributed to this pattern of results. Our participants were, as a group, highly proficient, considering their performance in the translation and categorization tasks (a few scored nearly at, or at, ceiling). However, differences in proficiency and experience existed among them (for example, accuracy in the translation task ranged from 48 to 100%). Furthermore, they reported using Welsh for recreational reading 22.5% of the time (only two participants reported reading Welsh 50% of the time). It is therefore possible that even many years of experience with a second language do not translate in completely symmetrical effects in reading experiments if the first language remains dominant, particularly here in the domain of reading (which is certainly the case for most English-Welsh bilinguals, given that Welsh is a "minority" language in Wales; Lyon, 1996). The Language effect, along with the behavioral results, supports this picture.

In order to assess whether the difference in ERP amplitude between high and low CLND English (L1) targets was modulated by the participants' experience with L2, *post hoc* analyses were carried out based on a median split with Years of Experience with Welsh and Translation Accuracy as measures of experience. The results (see **Table 2**) revealed the presence of larger CLND effects for L1 words in more, compared to less, proficient bilinguals for both the early and late time window, as expected: more experienced bilinguals were supposed to have a broader Welsh vocabulary, likely including many Welsh words that were neighbors of English targets in the present study. These findings suggest that CLND had some effect on the processing of L1 words, depending on the experience with the second language. This pattern is in agreement with non-selective access models, given that CL neighbors are hypothesized to be activated differentially based on a variety of factors that affect the level of activation of single items, such as subjective frequency and proficiency in the second language.

As in Midgley et al. (2008), the interaction between CLND and Language was significant in the 300–500 ms time window in late bilinguals. The effect was significant for L2 targets but not for L1 targets, revealing asymmetric effects for the two languages. Therefore, the early activation of Welsh neighbors when participants read English words might have dissipated rapidly and did not carry out to the N400 time window. Median split *post hoc* analyses based on language proficiency suggest the presence of larger CLND effects for L1 words in more, compared to less, proficient bilinguals, as for the 175–300 ms time window. Overall, these results suggest that, in late bilinguals, electrophysiological CLND effects tend to be asymmetrical, although the level of asymmetry was modulated by experience with the second language, in agreement with behavioral data (e.g., Bijeljac-Babic et al., 1997).

Table 2 | Cross-language ND effects for English targets in late bilinguals in terms of effect size (differences are in μV).

Years of experience	Time window (ms)	Less experienced bilinguals	More experienced bilinguals	Cohen's *d*
	175–300	−0.08 (1.37)	−0.59 (0.77)	0.60*
	300–500	0.03 (1.27)	−0.32 (0.70)	0.34††
Translation accuracy	**Time window (ms)**	**Lower accuracy translators**	**Higher accuracy translators**	**Cohen's *d***
	175–300	−0.14 (1.43)	−0.38 (0.80)	0.21†
	300–500	−0.03 (1.28)	−0.26 (0.70)	028†

*Medium effect; ††, small-to-medium effect; †small effect. The effects were calculated at CPZ, where the cross-language ND effect was largest (SD are shown in parentheses). The median split for Years of Experience with Welsh was 12 years (the less experienced group had an average of 6.7 years, whereas the more experienced group had an average of 17 years). The median split for Translation Accuracy was 83% (the lower accuracy translators had an average translation accuracy of 69%; the higher accuracy translators had an average translation accuracy of 92.6%).

Based on the non-selective access hypothesis, it was hypothesized that symmetrical effects would be present for the two languages in early bilinguals. However, this hypothesis was not supported. A frontocentral CLND effect was found at midline sites for English targets in the 175–300 ms time window. For Welsh targets, the effect was mainly localized over the frontal left hemisphere sites and it was reversed. Perhaps high CLND Welsh words (e.g., *bara*, *coes*, *nain*) automatically activated competing English phonological representations, which would cause inhibition (e.g., Dijkstra et al., 1999); but it is unclear why this would occur only with Welsh targets and only in early bilinguals. Additionally, no CLND effects were found in the 300–500 ms time windows in early bilinguals. Therefore, CLND effects were weaker and more transient in early bilinguals. Furthermore, their pattern only partially resembled the one observed in late bilinguals in terms of distribution and direction. Although the meaning of these differences is unclear, the presence of effects in the 175–300 ms time window in early bilinguals reveals the existence of CL activation during the early stages of reading. It might be safe to conclude, based on the present data, that this activation is quickly suppressed or dissipated, potentially because the inhibitory control operating in early bilinguals is more efficient and has a faster turn around than that developed by late bilinguals.

These results are not entirely consistent with a non-selective model of lexical access, as they seem to contradict behavioral accounts of CL activation in bilinguals. However, most of the available data on CLND effects was gathered in proficient late bilinguals (e.g., Midgley et al., 2008) or participants whose age of acquisition for L2 was not specified (e.g., Grainger and Dijkstra, 1992; van Heuven et al., 1998), with a few exceptions. For example, Bijeljac-Babic et al. (1997) found CL activation of orthographically related words in early French-English bilinguals who learned both languages during early childhood and who used them daily. However, the masked priming paradigm employed by the authors is fairly different from the categorization task used in the present experiment, since, in the latter, the "context" language was known to the participants (while it was masked in Bijeljac-Babic and colleagues' study). Therefore, early bilinguals might be skillful at applying top-down inhibition to block interference from words from the other

language if the linguistic context is clear (e.g., Rodriguez-Fornells et al., 2002).

Electrophysiological evidence is mixed. In a letter detection task, Rodriguez-Fornells et al. (2002) asked early Spanish-Catalan bilinguals to respond to Spanish words presented singularly on a computer screen along with Spanish pseudowords and Catalan words and non-words (different response hands were used depending on the word's initial letter). The authors found a N400 modulation by lexical frequency only for Spanish words and therefore hypothesized that proficient bilinguals are able to block semantic processing in the unattended language (for a critique of this work, see Grosjean et al., 2003). This conclusion contradicts more recent evidence of CL automatic semantic priming in early bilinguals. Martin et al. (2009) asked participants to indicate whether words presented on a computer monitor at regular intervals in a visual stream had more than five letters or five or fewer letters. This task was aimed at forcing participants to focus on the stimuli's low-level features, instead of their meaning. Participants saw two blocks of trials, depending on whether they had to respond only to Welsh or English stimuli. They were not informed that words were presented in pairs, belonging to the same or different languages and being semantically related or unrelated. The results revealed that the N400 was modulated by the semantic relationship between primes and targets, regardless of whether the words belonged to the same language and regardless of whether they were in the language under the focus of attention. Martin and colleagues concluded that word meaning is accessed automatically for both languages in early bilinguals because it occurred even when participants were explicitly instructed to neglect words in a given language. According to them, the task was successful in driving the participants' attention away from semantic processing, as no behavioral semantic priming effect was found in either experiment for reaction times. However, it is unclear why the activation of meaning would have any priming effect on a letter-counting task. Furthermore, the authors did not perform a manipulation check to establish that participants were indeed unaware of the semantic relationship between some of the words. Finally, as Martin and colleagues acknowledged (p. 330), in order to decide whether a stimulus required a manual response, attention

needs to be paid to either its word form or meaning. Therefore, the very goal of having participants disregard words in one language might have caused them to engage in lexical and semantic processing of every word. This being said, Martin et al. (2012) recently showed that the same task in monolingual speakers of English failed to elicit any semantic modulation of the N400, even when participant focused on English words. Obviously, the critique of Martin et al. (2009) applies equally to Rodriguez-Fornells et al.'s (2002). Further research is needed to settle the question. In the meantime, the present results suggest that late and early bilinguals might exercise different levels of control on one language when processing words in their other language, at least as regards CL activation of orthographic neighbors.

The functional meaning of the differences in CLND effects between early and late bilinguals is not clear. Differences in proficiency alone are unlikely to explain this pattern, as targets in both languages contributed to the ND main effect in late bilinguals. Therefore, based on proficiency, a more symmetric pattern would be expected in early bilinguals[2]. Early and late bilingualism differ on a variety of dimensions. Early or childhood bilingualism (which itself can be distinguished in various forms, e.g., simultaneous and sequential) tends to occur in more naturalistic settings, while late bilingualism is usually fostered through direct instruction and often without a relevant pragmatic context (Baker, 2011). Furthermore, because early bilinguals usually learn to speak their languages in different contexts and with different people, they develop an awareness of the distinct use of different languages and two separate language systems very early (Baker, 2011). This original and reciprocal independence might help set up control mechanisms that are not present in late bilinguals. Whilst speculative, this hypothesis highlights the fact that current models of non-selective access (e.g., the BIA+) do not take into consideration differences in learning experiences that often characterize language acquisition in early and late bilinguals.

The results observed in late bilinguals support the recent literature on the modulation of the N400 amplitude by ND (Holcomb et al., 2002; Midgley et al., 2008; Müller et al., 2010; Laszlo and Federmeier, 2011). They also support the presence of early ND effects in CL experiments starting at around 175–200 ms. Interestingly, early effects have not been reliably described in studies of within-language ND, with the exception of Holcomb et al.'s (2002) second experiment. Midgley et al. (2008) explained this apparent discrepancy in terms of differences in the relative frequency of targets and their neighbors in within- and between language ND studies: in the latter, L2 targets might have L1 neighbors with higher subjective frequencies, compared to L1 neighbors of L1 targets. This relative frequency would translate in an earlier influence of the neighbors on the processing of the target word. This reasonable explanation, however, does not account for the presence of early effects in Holcomb et al.'s (2002) second experiment. Furthermore, the ND effects in Laszlo and Federmeier (2011) seemed to start earlier than 250 ms, based on their Figure 3, although the authors limited their analysis to the 250–450 ms time window. Similarly, Müller et al.'s (2010) Figure 2 suggests the presence of an early effect; however, the authors concentrated their analyses on the 350–550 and later time windows. Clearly, the presence of early within-language ND effects will need to be substantiated in future experiments. In the meantime, we would recommend that analyses be carried out on earlier time windows, as both Midgley et al.'s study and the present findings suggest that ND effects are detectable before 300 ms.

One of the limitations of the present study is the relatively small sample size. Future studies should investigate individual differences more systematically, as the present data suggest that the presence of CLND effects for L1 depend on proficiency in L2, at least in late bilinguals. Furthermore, future studies should investigate how different language learning experiences shape aspects of cognition and brain organization in terms of CL interaction, as the differences between early and late bilinguals are not trivially explained by non-selective models of lexical access.

---

[2] However, some authors have remarked that cross-language effects tend to be larger when the target words in one language have a lower frequency than related words in the other language (e.g., Dijkstra et al., 1999; p. 497; see also Beauvillain and Grainger, 1987). Therefore, based on written frequency or familiarity, larger cross-language effects would be expected for L2 targets in late bilinguals, and smaller effects would be obtained in early bilinguals, for whom words in the two languages have more similar frequencies.

REFERENCES

Alternberg, E. P., and Cairns, H. S. (1983). The effects of phonotactic constraints on lexical processing in bilingual and monolingual subjects. *J. Verb. Learn. Verb. Behav.* 22, 174–188.

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: activation or search? *J. Exp. Psychol. Learn. Mem. Cogn.* 15, 802–814.

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: lexical similarity or orthographic redundancy? *J. Exp. Psychol. Learn. Mem. Cogn.* 18, 234–254.

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX Lexical Database (Release 2)*. Philadelphia: Linguistic Data Consortium.

Baker, C. (2011). *Foundations of Bilingual Education and Bilingualism*, 5th Edn. Bristol: Multilingual Matters.

Beauvillain, C., and Grainger, J. (1987). Accessing interlexical homographs: some limitation of a language-selective access. *J. Mem. Lang.* 26, 658–672.

Bijeljac-Babic, R., Biardeau, A., and Grainger, J. (1997). Masked orthographic priming in bilingual word recognition. *Mem. Cognit.* 25, 447–457.

Carr, T. H., and Pollatsek, A. (1985). "Recognizing printed words: a look at current models," in *Reading Research: Advances in Theory and Practice*, Vol. 5, eds D. Besner, T. G. Waller, and G. E. MacKinnon (New York, NY: Academic Press), 1–82.

Carreiras, M., Perea, M., and Grainger, J. (1997). Effects of orthographic neighborhood in visual word recognition: cross-task comparisons. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 857–871.

Coltheart, M., Davelaar, E., Jonasson, J. T., and Besner, D. (1977). "Access to the internal lexicon," in *Attention and Performance IV*, ed. S. Dornic (Hillsdale, NJ: Erlbaum), 535–555.

Dijkstra, A. F. J., Grainger, J., and Van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: the neglected role of phonology. *J. Mem. Lang.* 41, 496–518.

Dijkstra, A. F. J., and Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: from identification to decision. *Biling. Lang. Cogn.* 5, 175–197.

Ellis, N. C., and Hooper, A. M. (2001). Why learning to read is easier in Welsh than in English: orthographic transparency effects evinced with frequency-matched tests. *Appl. Psycholinguist.* 22, 571–599.

Ellis, N. C., O'Dochartaigh, C., Hicks, W., Morgan, M., and Laporte, N. (2001). *Cronfa Electroneg o Gymraeg/A 1 Million Word Lexical Database and Frequency Count for Welsh.* Bangor: University of Wales. Available at: http://www.bangor.ac.uk/canolfanbedwyr/ceg.php.en

Forster, K. I. (1999). The microgenesis of priming effects in lexical access. *Brain Lang.* 68, 5–15.

Frenck-Mestre, C. (1993). Use of orthographic redundancies and word identification speed in bilinguals. *J. Psycholinguist. Res.* 22, 397–409.

Frost, R., Katz, L., and Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: a multilingual comparison. *J. Exp. Psychol. Hum. Percept. Perform.* 13, 104–115.

Grainger, J., Bouttevin, S., Truc, C., Bastien, M., and Ziegler, J. (2003). Word superiority, pseudoword superiority, and learning to read: a comparison of dyslexic and normal readers. *Brain Lang.* 87, 432–440.

Grainger, J., and Dijkstra, T. (1992). "On the representation and use of language information in bilinguals," in *Cognitive Processing in Bilinguals,* ed. R. J. Harris (Amsterdam: North Holland), 207–220.

Grosjean, F., Li, P., Münte, T. F., and Rodriguez-Fornells, A. (2003). Imaging bilinguals: when the neurosciences meet the language sciences. *Biling. Lang. Cogn.* 6, 159–165.

Grossi, G., Savill, N., Thomas, E., and Thierry, G. (2010). Posterior N1 asymmetry to English and Welsh words in early and late English-Welsh bilinguals. *Biol. Psychol.* 85, 124–133.

Grossi, G., Thierry, G., Thomas, E., and DiPietro, J. (2008). Word and pseudoword superiority effects in early/native Welsh-English bilinguals. *Paper Presented at the 49th Meeting of the Psychonomic Society,* Chicago.

Haigh, C. A., and Jared, D. (2007). The activation of phonological representations by bilinguals while reading silently: Evidence from interlingual homophones. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 623–644.

Harm, M. W., and Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychol. Rev.* 111, 662–720.

Holcomb, P. J., and Grainger, J. (2007). Exploring the temporal dynamics of visual word recognition in the masked repetition priming paradigm using event-related potentials. *Brain Res.* 1180, 39–58.

Holcomb, P. J., Grainger, J., and O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *J. Cogn. Neurosci.* 14, 938–950.

Kutas, M., Van Petten, C., and Kluender, R. (2006). "Psycholinguistics electrified II: 1994-2005," in *Handbook of Psycholinguistics,* 2nd Edn, eds M. Traxler and M. A. Gernsbacher (New York: Elsevier), 659–724.

Laszlo, S., and Federmeier, K. D. (2011). The N400 as a snapshot of interactive processing: evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology* 48, 176–186.

Lyon, J. (1996). *Becoming Bilingual. Language Acquisition in a Bilingual Community.* Clevedon: Multilingual Matters.

Martin, C. D., Costa, A., Dering, B., Hoshino, N., Wu, Y. J., and Thierry, G. (2012). Effects of speed of word processing on semantic access: the case of bilingualism. *Brain Lang.* 120, 61–65.

Martin, C. D., Dering, B., Thomas, E. M., and Thierry, G. (2009). Brain potentials reveal semantic priming in both the 'active' and the 'non-attended' language of early bilinguals. *Neuroimage* 47, 326–333.

McClelland, J. L. (1976). Preliminary letter identification in the perception of words and nonwords. *J. Exp. Psychol. Hum. Percept. Perform.* 2, 80–91.

Midgley, K. J., Holcomb, P. J., van Heuven, W. J. B., and Grainger, J. (2008). An electrophysiological investigation of cross-language effects of orthographic neighborhood. *Brain Res.* 1246, 123–135.

Müller, O., Duñabeitia, J. A., and Carreiras, M. (2010). Orthographic and associative neighborhood density effects: what is shared, what is different? *Psychophysiology* 47, 455–466.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia* 9, 97–113.

Perea, M. (1998). Orthographic neighbours are not all equal: evidence using an identification technique. *Lang. Cogn. Process.* 13, 77–90.

Rodriguez-Fornells, A., Rotte, M., Heinze, H.-J., Nösselt, T., and Münte, T. F. (2002). Brain potential and functional MRI evidence for how to handle two languages with one brain. *Nature* 415, 1026–1029.

Thierry, G., and Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proc. Natl. Acad. Sci. U.S.A.* 104, 12530–12535.

van Heuven, W. J. B., Dijkstra, T., and Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *J. Mem. Lang.* 39, 458–483.

Wu, Y. J., and Thierry, G. (2010). Chinese-English bilinguals reading English hear Chinese. *J. Neurosci.* 30, 7646–7651.

Zhang, T., van Heuven, W. J. B., and Conklin, K. (2011). Fast automatic translation and morphological decomposition in Chinese-English bilinguals. *Psychol. Sci.* 22, 1237–1242.

Word-initial letters influence fixation durations during fluent reading

Christopher J. Hand[1], Patrick J. O'Donnell[2] and Sara C. Sereno[3,2]*

[1] Department of Psychology, University of Bedfordshire, Luton, UK
[2] School of Psychology, University of Glasgow, Glasgow, UK
[3] Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK

The present study examined how word-initial letters influence lexical access during reading. Eye movements were monitored as participants read sentences containing target words. Three factors were independently manipulated. First, target words had either high or low constraining word-initial letter sequences (e.g., *dwarf* or *clown*, respectively). Second, targets were either high or low in frequency of occurrence (e.g., *train* or *stain*, respectively). Third, targets were embedded in either biasing or neutral contexts (i.e., targets were high or low in their predictability). This 2 (constraint) × 2 (frequency) × 2 (context) design allowed us to examine the conditions under which a word's initial letter sequence could facilitate processing. Analyses of fixation duration data revealed significant main effects of constraint, frequency, and context. Moreover, in measures taken to reflect "early" lexical processing (i.e., first and single fixation duration), there was a significant interaction between constraint and context. The overall pattern of findings suggests lexical access is facilitated by highly constraining word-initial letters. Results are discussed in comparison to recent studies of lexical features involved in word recognition during reading.

Keywords: reading, eye movements, word-initial letter constraint, word frequency, contextual predictability

INTRODUCTION

The greatest advancements in understanding fluent reading over the past few decades have come from investigations that measure eye movement behavior (for reviews, see Rayner, 1998, 2009). Such studies have identified several oculomotor, perceptual, and cognitive factors that modulate the reader's decisions of where and when to move the eyes while processing text. For example, words in text which are shorter in length, higher in frequency of occurrence, or more predictable from a prior context are fixated for less time and are skipped more often than words that are longer, lower in frequency, or less predictable. The present study investigates the role of word-initial letters in reading.

One of the key findings of eye movement reading research is that the information available on a single fixation is not limited to the currently fixated (foveal) word. Readers are able to acquire information from the upcoming parafoveal word before its subsequent fixation. The importance of parafoveal vision in reading was substantiated in classic eye movement reading studies using the "moving window" (McConkie and Rayner, 1975) and "boundary" (Rayner, 1975) paradigms. In these paradigms, changes are made in the text contingent on the reader's eye position.

In "moving window" studies, text *outside* a window defined around the fixated letter is altered in some way (e.g., valid text is replaced by strings of Xs). Under such conditions, when parafoveal preview is invalid, reading time is slowed, demonstrating the use of both foveal and parafoveal information during normal reading. The *perceptual span* – the region of text from which useful information can be extracted – has been functionally approximated from "moving window" studies. For English, it is estimated to

extend from three characters to the left of fixation (approximately the beginning of the fixated word) to around 14 characters to the right of fixation (McConkie and Rayner, 1975; Miellet et al., 2009). Although the span encompasses a significant number of letters to the right of fixation, the level of analysis drops off substantially from the fovea – from recognizing words to identifying letters to merely determining the length of the upcoming parafoveal word(s).

In "boundary" studies, only a single word of the text changes. While reading, participants parafoveally view either a valid or invalid preview in the target location, which then changes to the target when the reader saccades across a pre-specified (invisible) boundary located just before the target word. "Boundary" experiments have varied the visual, phonological, and semantic similarity between the foveated target and its initial parafoveal preview and have generally shown that orthographic and phonological, but not semantic, information is extracted parafoveally (e.g., McConkie and Zola, 1979; Rayner et al., 1980; Balota et al., 1985; Pollatsek et al., 1992). The fixation time advantage on a target word (fixation n) when parafoveal information associated with that target (obtained from fixation $n - 1$) is valid vs. invalid is termed *parafoveal preview benefit*. Rayner et al. (1982) found that the when the first three letters (i.e., word-initial trigram) of the parafoveal preview were identical to those of the (eventual) target word and when the remaining letters of the preview were replaced by letters that were visually similar to the target, reading rate was only slightly impaired compared to when the preview was completely identical to the target (i.e., the valid preview condition). The implication is that the identification of word-initial letters is fundamental to

obtaining a parafoveal preview benefit (c.f. Inhoff et al., 1989; Inhoff, 1990; Johnson et al., 2007). Given that the first few letters of the parafoveal word are nearest the fovea and that the space before the parafoveal word serves to decrease lateral masking of its beginning letters, such findings are not unexpected.

If the identification of the word-initial trigram facilitates reading, as evidenced by parafoveal preview benefit, the question arises whether the level of lexical constraint conferred by the trigram can affect word identification. Within the auditory word recognition literature, the homologous issue of word beginnings and their role in spoken word identification has been the topic of innumerable studies. Marslen-Wilson and Welsh (1978; see also Marslen-Wilson, 1987) proposed the *cohort* model of spoken word recognition. In this model, the initial acoustic information activates a large number of candidate words (i.e., a cohort) in parallel, but as further evidence accumulates, the activation of words that are no longer compatible with the input decays until a single candidate remains (the point in a spoken word which delivers a single candidate is called the *uniqueness point*). Although the signal is produced and processed in a more continuous and sequential way in the auditory compared to the visual domain, parafoveal preview nevertheless gives emphasis to the initial letters of an upcoming word. Thus, it is reasonable to expect similar activation and selection processes to occur in visual word recognition during fluent reading. High constraint (HC) initial trigrams rarely appear in words whereas low constraint (LC) initial trigrams often do. For example, the HC trigram *dwa-* includes very few words in its cohort (e.g., *dwarf, dwarves, dwam*); in contrast, the LC trigram *clo-* has many words in its cohort (e.g., *clown, close, clock, cloud, cloth, cloak, clone, clout, clove, clog, cloy, clothes, clover, closet, cloister, clobber*; N.B., this excludes morphologically related suffixed words).

To determine whether such cohort effects operated in the visual domain, Lima and Inhoff (1985; Experiment 1), in an eye movement reading study, tested whether the constraint of a word-initial trigram affected reading behavior. They hypothesized that lexical access would be facilitated when a word's candidate set was limited by its initial letters. Target words were either HC (e.g., *dwarf*) or LC (e.g., *clown*) words of similar length and frequency presented in single-line neutral sentences. Lima and Inhoff additionally varied parafoveal preview across three conditions: one- and two-word moving window conditions (with strings of Xs replacing text outside the window), and a full-line condition (i.e., normal reading). In the one-word condition, readers were prevented from obtaining a valid parafoveal preview of the target; in both the two-word and full-line conditions, a valid parafoveal preview of the target was available. In accordance with prior findings, Lima and Inhoff found a preview benefit whereby targets were read faster with a valid (two-word and full-line conditions) vs. invalid (one-word condition) parafoveal preview. In contrast to their predictions, however, preview benefit did not interact with target constraint. They had expected to find greater preview benefit for HC than LC words. In terms of target fixation time, they did find an effect of constraint. The effect, however, was in the opposite direction of their prediction – HC words were fixated *longer* than LC words. It is important to note that this effect was only significant in the more immediate first fixation duration (FFD) measure (i.e., the duration

of the initial fixation on a target word, regardless of whether that word is refixated); the effect did not reach significance in the relatively delayed gaze duration (GD) measure (i.e., the sum of all consecutive fixations, including the first, before moving to another word). Lima and Inhoff concluded that higher trigram familiarity (LC words) could benefit lexical access by increasing the efficiency of foveal processing.

Although past eye movement research has explored the effects of whole-word orthographic (and phonological) regularity (Inhoff and Topolski, 1994; Sereno and Rayner, 2000), more recent studies have examined the effects of word-initial orthographic regularity on eye movement behavior. In particular, the focus of these studies has been on whether the orthographic regularity of a target word's beginning letters, viewed parafoveally from the prior fixation, can affect the location of the ensuing fixation on the target (i.e., landing position). Evidence for the influence of word-initial orthographic regularity on fixation location (with more regular word beginnings giving rise to more rightward landing positions), however, has been equivocal (for a review, see White and Liversedge, 2004).

White and Liversedge (2004) suggested that prior studies had confounded two variables associated with word-initial orthographic regularity, namely, "orthographic familiarity," and "informativeness." These two variables represent different ways of measuring the frequency of a word's beginning letter sequence. Orthographic familiarity is calculated by summing the frequency of all words (tokens) beginning with that letter sequence, while informativeness is calculated by summing the number of words (types) beginning with that letter sequence. White and Liversedge (2004) conducted two experiments that manipulated these variables by misspelling the beginning letter sequences of words. They found that landing position was closer to the beginning of misspelled words (i.e., nearer the location of the misspelling; e.g., *aoricultural, akricultural, ngricultural*) compared to correct words (e.g., *agricultural*), even when the misspelling employed a highly frequent word-initial trigram (e.g., *acricultural*). They also found no difference in landing position between correctly spelled words having informative word-initial trigrams (e.g., *escalator*) and misspelled informative (e.g., *eacalator*) or uninformative (e.g., *encalator*) controls. Although these manipulations permit a high degree of control over certain orthographic characteristics of the stimuli, the use of misspelled words, however, limits the generalizability of such results to normal reading.

The purpose of the present experiment was to further investigate the effect of word-initial letter constraint in reading. Like Lima and Inhoff (1985), we compared fixation time on HC (e.g., *dwarf*) and LC (e.g., *clown*) words in text. Unlike Lima and Inhoff (1985), however, we additionally manipulated two key variables known to affect word recognition, namely, word frequency, and contextual predictability. When lexical variables such as word length are controlled, high frequency (HF) words are read faster than low frequency (LF) words, and words preceded by a contextually biasing context are read faster than those in a neutral context (see, e.g., Hand et al., 2010; for reviews, see Rayner, 1998, 2009). In Lima and Inhoff's study, target words were mainly LF words embedded in neutral contexts. Prior research using gaze-contingent display change paradigms, however, has demonstrated

increased parafoveal preview benefit to HF vs. LF words (Inhoff and Rayner, 1986) as well as to contextually predictable vs. less predictable words (Balota et al., 1985). Thus, we implemented a 2 (Constraint: HC, LC) × 2 (Frequency: HF, LF) × 2 (Context: Biasing, Neutral) design. Because parafoveal preview benefit is modulated both by frequency and contextual predictability, it is possible that HC words will, in fact, show a processing advantage over LC words when favorable parafoveal preview conditions are present. Accordingly, we expected to find an interaction between Constraint and Frequency and/or Constraint and Context. In line with Lima and Inhoff's (1985) findings, we anticipated *longer* fixations on HC vs. LC words for LF targets in Neutral contexts. However, we predicted *shorter* fixations on HC vs. LC words for HF targets, for targets in Biasing contexts, or, minimally, for HF targets in Biasing contexts.

MATERIALS AND METHODS

PARTICIPANTS

Forty-eight members of the University of Glasgow community (30 females; mean age 23) were paid £6 or given course credit for their participation. All were native English speakers with normal or corrected-to-normal vision and had not been diagnosed with any reading disorder.

APPARATUS

Eye movements were monitored via an SR Research Desktop-Mount EyeLink 2K eyetracker, with a chin/forehead rest. The eyetracker has a spatial resolution of 0.01° and eye position was sampled at 1000 Hz using corneal reflection and pupil tracking. Text (black letters on a white background, using 14-point Bitstream Vera Sans Mono, a non-proportional font) was presented on a Dell P1130 19″ flat screen CRT (1024 × 768 resolution; 100 Hz). At a viewing distance of 72 cm, approximately four characters of text subtended 1° of visual angle. Viewing was binocular with eye movements recorded from the right eye.

DESIGN AND MATERIALS

A 2 (Constraint: HC, LC) × 2 (Frequency: HF, LF) × 2 (Context: Biasing, Neutral) design was used. All target words were five letters long. With a total of 88 experimental items, there were 11 items in each of the eight conditions. All experimental items are listed in the Appendix. An example set of materials, showing all eight target conditions, is presented in **Table 1**. Target words were always positioned near the middle of a line of text. Because each participant only read a given target word in one of its Context conditions (Neutral or Biasing), two participant groups were used. One group read half of the materials in Neutral and the other half in Biasing contexts; the second group read the materials in their opposing context conditions. In addition, experimental items were blocked by Context condition, with all Neutral materials presented first followed by all Biasing materials. Within each block, experimental items were presented in a different random order to each participant. Stimulus specifications across conditions are presented in **Table 2**.

Constraint

Half of the target words had HC and half had LC initial trigrams. We calculated several measures to characterize the constraint of

Table 1 | Example materials.

Condition		Example
NEUTRAL CONTEXT		
LF	LC	He had enjoyed being a <u>clown</u> but it was time to retire.
	HC	In gym class, he felt like a <u>dwarf</u> next to his classmates.
HF	LC	He bought tickets for the <u>train</u> to Waterloo on the internet.
	HC	She wanted to talk to the <u>girls</u> about the incident.
BIASING CONTEXT		
LF	LC	Pierre had entertained kids at the circus for 50 years.
		He had enjoyed being a <u>clown</u> but it was time to retire.
	HC	Jamie loved basketball but he was very short for his age.
		In gym class, he felt like a <u>dwarf</u> next to his classmates.
HF	LC	Stuart did not want to travel to London by bus or plane.
		He bought tickets for the <u>train</u> to Waterloo on the internet.
	HC	At school, Miss Jones told only the boys to leave early.
		She wanted to talk to the <u>girls</u> about the incident.

Target words are underlined.
LF, low frequency; HF, high frequency; LC, low constraint; HC, high constraint.

the trigram neighborhood for each (five-letter) target. These were performed on both length-invariant (i.e., only five-letter words) and length-variant (i.e., words of any length or *x*-letter words) trigram neighborhoods. All measures included the target word. Similar to White and Liversedge (2004), we computed the number of words (type frequency, per million) and the summed frequency of words (token frequency, per million) that shared the initial trigram. We also calculated the percentage that each target represented of its trigram neighborhood, dividing each target word's frequency of occurrence by the summed frequency of all five- or *x*-letter words (including the target) that shared a given trigram. Finally, we obtained the rank position of the target within the trigram neighborhood based on its frequency relative to the frequency of its trigram neighbors. To determine these neighborhood profiles for *x*-letter words, we used the Brigham Young University on-line resource[1] (Davies, 2004) for the British National Corpus (BNC). Average values for each of these measures across conditions are presented in **Table 2**. Overall, in both five- and *x*-letter trigram neighborhoods, HC words, in comparison to LC words, had far fewer neighbors, had much smaller neighborhood frequencies, accounted for a much higher percentage of their neighborhood, and were ranked much closer to the top of their neighborhood.

Frequency

In addition, half of the targets were HF and half were LF words. Word frequencies were obtained using BNC[2], a corpus of 90 million written word tokens. Mean frequencies were 88 occurrences per million for HF targets and nine occurrences per million for LF targets (see **Table 2**).

Predictability

Finally, half of the targets were presented in a Neutral context and half in a Biasing context. As illustrated in **Table 1**, Neutral

---

[1]http://corpus.byu.edu/bnc
[2]http://www.natcorp.ox.ac.uk

Table 2 | Means (with SDs) of target specifications across experimental conditions.

Measure	LF		HF	
	LC	**HC**	**LC**	**HC**
N	22	22	22	22
Length	5 (0)	5 (0)	5 (0)	5 (0)
Frequency	8 (5)	10 (6)	86 (69)	90 (84)
NUMBER OF NEIGHBORS				
Five-letter	20 (8)	2 (2)	19 (7)	5 (3)
x-Letter	209 (135)	17 (14)	269 (142)	50 (46)
FREQUENCY OF NEIGHBORHOOD				
Five-letter	429 (470)	11 (7)	1035 (1630)	93 (88)
x-Letter	1615 (1759)	31 (29)	5396 (15867)	358 (281)
% OF NEIGHBORHOOD				
Five-letter	4 (5)	95 (5)	15 (7)	96 (4)
x-Letter	1 (2)	38 (21)	5 (3)	33 (25)
RANK IN NEIGHBORHOOD				
Five-letter	7 (4)	1 (0)	3 (1)	1 (0)
x-Letter	28 (22)	1 (1)	8 (10)	2 (2)
CLOZE				
Neutral	0.03 (0.09)	0.04 (0.07)	0.04 (0.11)	0.03 (0.09)
Biasing	0.64 (0.23)	0.60 (0.19)	0.64 (0.25)	0.60 (0.32)
PREDICTABILITY RATING				
Neutral	3.70 (0.89)	3.43 (1.04)	4.21 (1.19)	3.98 (1.15)
Biasing	5.87 (0.52)	5.76 (0.60)	6.01 (0.40)	5.92 (0.61)

LF, low frequency; HF, high frequency; LC, low constraint; HC, high constraint; N, number of items; Length, word length (number of letters); Frequency, frequency of occurrence (per million); Number of Neighbors, number of trigram neighbors for five-letter or x-letter (any length) words; Frequency of Neighborhood, summed frequency (per million) of trigram neighborhood; % of Neighborhood, word frequency percentage that each target represents of its five-letter or x-letter trigram neighborhood; Rank in Neighborhood, rank of target in neighborhood, based on its frequency; Cloze, Cloze value of target, on a scale of 0 (target word not guessed) to 1 (target word correctly guessed); Predictability Rating, predictability rating of target in text, on a scale of 1 (highly unpredictable) to 7 (highly predictable); Neutral, neutral context condition (target sentence only); Biasing, biasing context condition (context plus target sentence).

conditions comprised one single-line sentence. Biasing conditions, however, comprised two single-line sentences: for a given target, the first sentence contained contextually biasing information for that word; the second sentence was the Neutral sentence in which the target was embedded. In this way, biasing information was established in and confined to the first of two sentences. In addition, the identical sentence containing the target could be used across the Neutral and Biasing context conditions (between participant groups).

The level of contextual predictability was determined by two norming tasks – a Cloze probability task and a predictability rating task. For both tasks, the materials were divided into two sets with equal numbers of Neutral and Biasing sentences and were presented to two participant groups to avoid repetition of the target sentence across conditions. In the Cloze task, two groups of 13 participants (none of whom participated in the main experiment

or the predictability rating task) were given each experimental item up to but not including the target word. Their task was to generate the next word in the sentence. Items were scored as "1" for correct responses and "0" for all other guesses. A 2 (Constraint: HC, LC) × 2 (Frequency: HF, LF) × 2 (Context: Biasing, Neutral) analysis of variance (ANOVA) on Cloze probabilities by items (F_2) revealed, as expected, a main effect of Context, with more targets generated in Biasing (0.62) than in Neutral (0.04) contexts (see **Table 2**) [$F_2(1,21) = 991.25$, MSE = 0.02, $p < 0.001$]. No other main effects or interactions were significant [all F_2s < 1].

In the predictability rating task, two groups of 13 participants (again, none of whom participated in the main experiment or Cloze task) were presented with each item in its entirety with the target word underlined. Ten percent of the materials were non-experimental filler items (one- and two-line texts) that were clearly anomalous. The participants' task was to indicate how predictable they considered the target word to be on a scale of 1 (highly unpredictable) to 7 (highly predictable). A 2 (Constraint: HC, LC) × 2 (Frequency: HF, LF) × 2 (Context: Biasing, Neutral) ANOVA on predictability ratings by items (F_2) revealed, as expected, a main effect of Context, with targets rated more predictable in Biasing (5.89) than in Neutral (3.83) contexts (see **Table 2**) [$F_2(1,21) = 590.73$, MSE = 0.32, $p < 0.001$]. The relatively high ratings of targets in Neutral contexts reflected the fact that they were designed to be *less* predictable (and not implausible or anomalous) compared to targets in Biasing contexts. The main effect of Frequency, although numerically small, was also significant, with higher ratings for HF (5.03) than for LF (4.69) targets [$F_2(1,21) = 4.64$, MSE = 1.08, $p < 0.001$]. Most likely, this reflects the underlying fact that HF words are, by definition, more likely to occur than LF words within any context (see, e.g., Scott et al., 2012). The main effect of Constraint was not significant [$F_2(1,21) = 2.45$, MSE = 0.55, $p = 0.132$], nor were any of the interactions [Frequency × Predictability: $F_2(1,21) = 1.51$, MSE = 1.04, $p > 0.20$; Constraint × Frequency, Constraint × Predictability, and Constraint × Frequency × Predictability: all F_2s < 1].

PROCEDURE

Participants were given written and verbal instructions about the eyetracking task. They were told to read for comprehension, as they would normally, and that questions would appear after half of the trials to ensure they were paying attention.

The experiment involved the initial calibration of the eyetracking system, reading five practice one-line (Neutral) sentences, recalibration, reading the 44 Neutral experimental sentences, recalibration, reading five practice two-line (Biasing) passages, recalibration, and reading the 44 Biasing experimental passages. The nine-point calibration display comprised a series of calibration points extending over the maximal horizontal and vertical range of the display. After participants fixated each point in a random order, the accuracy of the calibration was checked (validation). The experiment proceeded only when the calibration was highly accurate (average error <0.30°; maximal error on any one point <0.50°). If necessary, participants could be recalibrated at any time during the experiment.

Each trial began with a black square which corresponded to the position of the first letter of the experimental item. An accurately

calibrated fixation at this location triggered the presentation of the item. After reading each item, participants moved their eyes to the lower, right corner of the screen and pressed a button to clear the screen. On half of the trials, a *yes–no* comprehension question followed. Participants had no difficulty in answering these questions (average over 92% correct). Prior to each new trial, participants were required to fixate a central point allowing the experimenter to implement a drift-correction routine.

RESULTS

The target region comprised the space before the target word and the target itself. Lower and upper cut-off values for individual fixations were 100 and 750 ms, respectively. Data were additionally eliminated if there was a blink or track loss on the target, or if the fixation on the target was either the first or last fixation on a line. Overall, 2% of the data were excluded for these reasons. In reading, most content words are generally fixated once – sometimes words are immediately refixated, sometimes they are skipped altogether. In the present study, the probabilities for target word single fixation, immediate refixation, and skipping were 0.67, 0.07, and 0.24, respectively.

The resulting data were analyzed over a number of standard fixation time measures on the target word: (1) FFD; (2) single fixation duration (SFD; fixation time when the word is only fixated once); (3) GD; and (4) total fixation time (TT; the sum of all fixations, including later regressions made to that word). We also examined several other commonly used measures: (5) the duration of the next forward-going fixation from the target ($T + 1$) as a measure of processing spillover; (6) the duration of the pre-target fixation ($T - 1$; the last fixation before the target) as a measure of parafoveal pre-processing of the target; (7) the probability of making a first-pass fixation on the target (PrF); and (8) the landing position (LandPos) or location of the first fixation on the target. The average values across all measures (with SDs) are presented in **Table 3**.

The different measures can be viewed as a series of snapshots over the temporal course of processing the target – from pre-target, to target, and then to post-target measures. The earliest measures are $T - 1$, PrF, and, to some extent, LandPos, which can reflect varying degrees of target pre-processing. These measures should, however, be interpreted with some caution as the pre-target text differed across conditions in our study (N.B. most pre-target words were HF function words). With respect to PrF, although the decision to skip a word occurs on the pre-target fixation, target processing can occur on both pre- and post-target fixations (e.g., Reichle et al., 2003; Kliegl and Engbert, 2005). With respect to LandPos, although it represents fixation location on the target, itself, the saccade target is determined from the pre-target fixation (e.g., McConkie et al., 1988; Rayner et al., 1996). Target measures include FFD, SFD, and GD during which the target is foveated. These measures tend to be highly correlated because the majority of data points contributing to each measure are shared – that is, most FFDs are SFDs, and most GDs are FFDs. As GD includes cases when an additional (consecutive) fixation is made on the target, in this respect, it is not as immediate as FFD or SFD. Finally, $T + 1$ and TT represent relatively delayed, later stages of target word processing, since these measures comprise fixations

Table 3 | Means (with SDs) of fixation time measures, fixation probability, and landing position across conditions.

Measure	Context	LF		HF	
		LC	**HC**	**LC**	**HC**
FFD	Neutral	204 (32)	194 (32)	195 (26)	187 (23)
	Biasing	189 (29)	189 (29)	182 (29)	180 (25)
SFD	Neutral	207 (38)	196 (41)	197 (28)	189 (24)
	Biasing	190 (31)	189 (29)	183 (29)	179 (25)
GD	Neutral	223 (47)	208 (46)	206 (30)	201 (35)
	Biasing	201 (36)	196 (32)	188 (29)	184 (26)
TT	Neutral	261 (64)	234 (75)	233 (43)	223 (47)
	Biasing	212 (42)	204 (31)	207 (35)	196 (32)
$T + 1$	Neutral	216 (33)	201 (34)	207 (35)	202 (32)
	Biasing	199 (27)	198 (28)	199 (30)	191 (27)
$T - 1$	Neutral	202 (31)	191 (31)	193 (28)	191 (25)
	Biasing	191 (27)	190 (27)	188 (24)	181 (20)
PrF	Neutral	0.82 (0.14)	0.79 (0.16)	0.86 (0.14)	0.80 (0.15)
	Biasing	0.71 (0.19)	0.69 (0.19)	0.66 (0.21)	0.67 (0.20)
LandPos	Neutral	2.81 (0.62)	2.64 (0.54)	2.78 (0.59)	2.63 (0.57)
	Biasing	2.82 (0.68)	2.88 (0.63)	2.86 (0.58)	3.02 (0.56)

FFD, first fixation duration; SFD, single fixation duration; GD, gaze duration; TT, total fixation time; T + 1, next forward-going fixation from target; T − 1, pre-target fixation duration; PrF, probability of target fixation; LandPos, landing position on target; LF, low frequency; HF, high frequency; LC, low constraint; HC, high constraint.

FIGURE 1 | Average single fixation duration (ms), with SE bars, on target words as a function of Constraint (LC, HC), Frequency (LF, HF), and Context (Neutral, Biasing). SFD, single fixation duration; LC, low constraint; HC, high constraint; LF, low frequency; HF, high frequency.
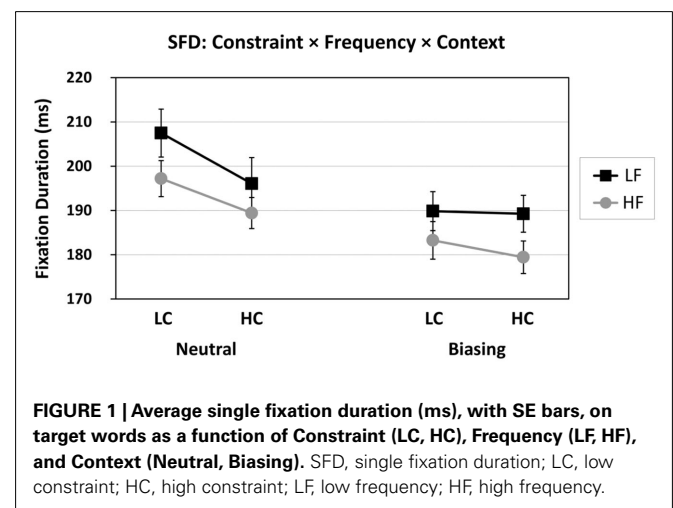
occurring after the initial fixation(s) on the target. Nonetheless, TT tends to be correlated with GD as there is a high degree of data overlap.

As the majority of target word fixations were single fixations, SFD condition means, including SE bars, are displayed in **Figure 1**. For all measures, 2 (Constraint: HC, LC) × 2 (Frequency: HF, LF) × 2 (Context: Biasing, Neutral) ANOVAs were conducted both by participants (F_1) and by items (F_2). A summary of all main effects and interactions across all measures is presented in **Table 4**.

Table 4 | Main effects and interactions by participants (F_1) and by items (F_2) on all measures.

Factor	F_1	MSE	p	F_2	MSE	p
CONSTRAINT						
FFD	7.56	357	<0.01	5.31	242	<0.05
SFD	7.38	453	<0.01	4.46	321	<0.05
GD	12.87	411	<0.001	5.25	500	<0.05
TT	15.51	1267	<0.001	15.74	675	<0.001
$T+1$	20.73	248	<0.001	11.37	211	<0.01
$T-1$	12.46	240	<0.001	7.86	173	<0.05
PrF	3.98	0.02	=0.052	1.94	0.02	>0.15
LandPos	<1			<1		
FREQUENCY						
FFD	12.75	451	<0.001	11.49	276	<0.01
SFD	12.57	532	<0.001	9.13	304	<0.01
GD	21.25	682	<0.001	13.34	537	<0.01
TT	11.15	1454	<0.01	12.82	608	<0.01
$T+1$	3.97	331	=0.052	3.48	191	=0.076
$T-1$	8.12	291	<0.01	6.46	180	<0.05
PrF	<1			<1		
LandPos	<1			1.03	0.18	>0.30
CONTEXT						
FFD	16.89	610	<0.001	30.53	139	<0.001
SFD	20.42	688	<0.001	32.48	132	<0.001
GD	24.30	1191	<0.001	38.65	304	<0.001
TT	41.51	2531	<0.001	77.98	612	<0.001
$T+1$	16.74	527	<0.001	19.71	216	<0.001
$T-1$	9.30	480	<0.01	12.72	157	<0.01
PrF	96.04	0.02	<0.001	65.15	0.01	<0.001
LandPos	10.18	0.30	<0.01	11.62	0.12	<0.01
CONSTRAINT × FREQUENCY						
FFD	<1			<1		
SFD	<1			<1		
GD	1.40	465	>0.20	1.00	570	>0.30
TT	1.19	1170	>0.25	<1		
$T+1$	<1			<1		
$T-1$	<1			<1		
PrF	<1			<1		
LandPos	<1			<1		
CONSTRAINT × CONTEXT						
FFD	4.63	321	<0.05	4.56	154	<0.05
SFD	3.35	391	=0.074	4.74	174	<0.05
GD	1.69	557	>0.15	<1		
TT	1.65	1062	>0.20	1.20	730	>0.25
$T+1$	1.66	456	>0.20	1.87	198	>0.15
$T-1$	<1			<1		
PrF	2.04	0.02	>0.15	1.02	0.01	>0.30
LandPos	4.96	0.35	<0.05	2.77	0.18	=0.111
FREQUENCY × CONTEXT						
FFD	<1			<1		
SFD	<1			<1		
GD	<1			<1		
TT	4.98	820	<0.05	1.65	1177	>0.20
$T+1$	<1			<1		
$T-1$	<1			<1		
PrF	4.87	0.02	<0.05	4.75	0.01	<0.05
LandPos	<1			2.05	0.12	>0.15
CONSTRAINT × FREQUENCY × CONTEXT						
FFD	<1			<1		
SFD	<1			<1		
GD	<1			<1		
TT	1.99	1387	>0.15	<1		
$T+1$	5.21	374	<0.05	4.19	228	=0.053
$T-1$	4.31	316	<0.05	3.19	204	=0.089
PrF	1.19	0.02	>0.25	1.25	0.01	>0.25
LandPos	<1			<1		

FFD, first fixation duration; SFD, single fixation duration; GD, gaze duration; TT, total fixation time; $T+1$, next forward-going fixation from target; PrF, probability of target fixation; LandPos, landing position on target.
Degrees of freedom are $F_1(1,47)$ and $F_2(1,21)$. MSE, mean squared error.

MAIN EFFECTS

Constraint

In each of the fixation time measures (FFD, SFD, GD, TT, $T+1$, and $T-1$), there was a significant main effect of Constraint (see **Table 4**). In contrast to Lima and Inhoff's (1985) earlier findings, HC words were fixated for *less* time than LC words (HC vs. LC: 187 vs. 193 ms for FFD, 189 vs. 194 ms for SFD, 197 vs. 205 ms for GD, 214 vs. 228 ms for TT, 198 vs. 205 ms for $T+1$; and 188 vs. 194 ms for $T-1$). For PrF, although the effect was not significant (marginal by participants, non-significant by items), the direction of the numerical effect was consistent with the fixation time results, with HC words (0.74) less likely to be fixated than LC words (0.76). Finally, the main effect of Constraint for LandPos was not significant.

Frequency

The main effect of Frequency (see **Table 4**) was significant across all target fixation time measures (FFD, SFD, GD, TT) and the pre-target $T-1$ measure, but only marginally significant (both by participants and items) in the post-target $T+1$ measure. In line with numerous eye movement studies on word frequency, HF words were associated with shorter fixations than LF words (HF vs. LF: 186 vs. 194 ms for FFD, 187 vs. 196 ms for SFD, 195 vs. 207 ms for GD, 215 vs. 228 ms for TT, 200 vs. 204 ms for $T+1$, and 188 vs. 193 ms for $T-1$). There was no reliable effect in the PrF and LandPos measures.

Context

The main effect of Context (see **Table 4**) was significant across all measures, including fixation probability and landing position (FFD, SFD, GD, TT, $T+1$, $T-1$, PrF, and LandPos). Again, similar to several eye movement studies investigating predictability, targets in Biasing contexts were fixated for less time than those in Neutral contexts (Biasing vs. Neutral: 185 vs. 195 ms for FFD, 185 vs. 198 ms for SFD, 192 vs. 210 ms for GD, and 205 vs. 238 ms for TT,), were less likely to be fixated (Biasing vs. Neutral:0.68 vs. 0.82 for PrF), and were associated with shorter pre- and post-target

(Continued)

fixations (197 vs. 207 ms for $T+1$, and 188 vs. 194 ms for $T-1$). For LandPos, readers fixated further into targets when they were more predictable (Biasing vs. Neutral: 2.89 vs. 2.72 characters).

INTERACTIONS
Although the interactions, in general, tended to be non-significant, there were a few exceptions (see **Table 4**).

Constraint × Frequency
The Constraint × Frequency interaction was not significant across any measure.
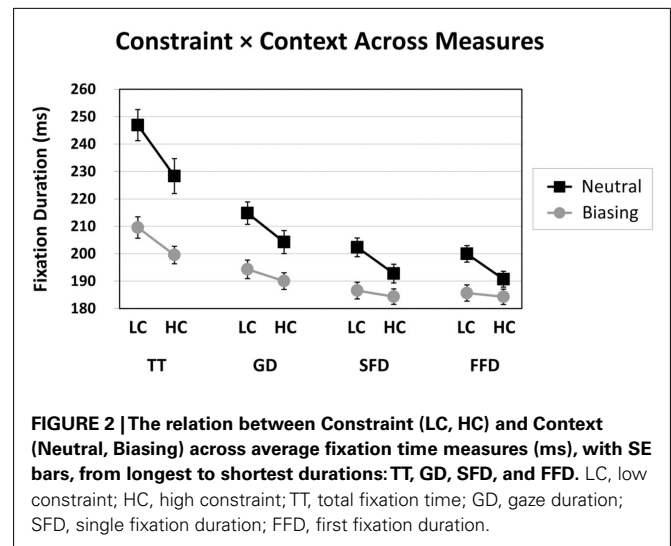
Constraint × Context
Constraint × Context, however, did reach significance in the more immediate fixation time measures of FFD and SFD (although this effect was marginal by participants in SFD) and, for Land-Pos, was significant by participants but only a statistical trend by items. In all other measures (GD, TT, $T+1$, $T-1$, and PrF), Constraint × Context failed to reach significance.

For LandPos, although the numerical means suggested an opposing pattern of results, with landing positions for HC words nearer word beginnings in Neutral contexts (HC = 2.64 and LC = 2.80 characters), but nearer word endings in Biasing contexts (HC = 2.95 and LC = 2.84 characters), this pattern was not maintained statistically. Rather, the follow-up contrasts in general were more supportive of an interpretation in which the landing position for HC-Neutral targets (2.64 characters) was nearer the beginning of the word compared to the other three conditions (2.80, 2.95, and 2.84 characters for LC-Neutral, HC-Biasing, and LC-Biasing, respectively) (HC-Neutral vs. LC-Neutral: $F_1 = 3.55$, $p = 0.066$, $F_2 = 2.14$, $p = 0.158$; HC-Neutral vs. HC-Biasing: $F_1 = 13.42$, $p < 0.001$, $F_2 = 9.68$, $p < 0.01$; HC-Biasing vs. LC-Biasing: $F_1 = 1.60$, $p > 0.20$, $F_2 < 1$; LC-Neutral vs. LC-Biasing: all $Fs < 1$).

For FFD and SFD, follow-up contrasts revealed significant differences between LC-Neutral and LC-Biasing conditions (FFD: $F_1 = 6.19$, $p < 0.05$, $F_2 = 4.84$, $p < 0.05$; SFD: $F_1 = 8.68$, $p < 0.01$, $F_2 = 3.88$, $p = 0.062$), HC-Neutral and HC-Biasing conditions (FFD: $F_1 = 30.59$, $p < 0.001$, $F_2 = 27.25$, $p < 0.001$; SFD: $F_1 = 30.62$, $p < 0.001$, $F_2 = 25.48$, $p < 0.001$), LC-Neutral and HC-Neutral conditions (FFD: $F_1 = 12.76$, $p < 0.001$, $F_2 = 12.62$, $p < 0.01$; SFD: $F_1 = 11.29$, $p < 0.01$, $F_2 = 12.73$, $p < 0.01$), but not between LC-Biasing and HC-Biasing conditions (all $Fs < 1$). In sum, while the effect of Context was maintained for both LC and HC words, the effect of Constraint was only upheld in Neutral contexts. In **Figure 2**, we plotted the Constraint × Context data (collapsed across Frequency) over the different fixation time measures, from the longest to the shortest duration (TT, GD, SFD, FFD). It seems that the interaction in the early SFD and FFD measures may actually arise from floor effects. That is, fixation times in HC-Biasing conditions just cannot get any shorter.

Frequency × Context
With respect to Frequency × Context, our results confirmed those of past eye movement studies that have typically demonstrated a lack of an interaction in fixation times but the presence of one in PrF (e.g., Rayner et al., 2004a; Hand et al., 2010). With the

FIGURE 2 | The relation between Constraint (LC, HC) and Context (Neutral, Biasing) across average fixation time measures (ms), with SE bars, from longest to shortest durations: TT, GD, SFD, and FFD. LC, low constraint; HC, high constraint; TT, total fixation time; GD, gaze duration; SFD, single fixation duration; FFD, first fixation duration.

exception of TT, in which the interaction was only significant by participants, all other measures (FFD, SFD, GD, $T+1$, $T-1$, and LandPos) failed to show an interaction. For the reliable interaction in PrF, follow-up contrasts were significant for LF-Neutral vs. LF-Biasing ($F_1 = 32.67$, $p < 0.001$, $F_2 = 32.02$, $p < 0.001$) and HF-Neutral vs. HF-Biasing ($F_1 = 78.07$, $p < 0.001$, $F_2 = 76.42$, $p < 0.001$), were not significant for LF-Neutral vs. HF-Neutral ($F_1 = 1.67$, $p > 0.20$, $F_2 = 1.59$, $p > 0.20$), and were only marginally significant for LF-Biasing vs. HF-Biasing ($F_1 = 3.34$, $p = 0.074$, $F_2 = 3.32$, $p = 0.083$). Thus, Biasing contexts gave rise to a lower likelihood of fixating the target (or an increased probability of skipping it), and when the target was additionally HF, these effects were enhanced. This pattern of differences stands in partial contrast to prior research which has found fewer fixations (or increased skipping) only in the combined condition of high predictability and HF (Rayner et al., 2004a; Hand et al., 2010).

Constraint × Frequency × Context
Finally, the three-way interaction was significant (although marginal by items) in the pre- and post-target measures, $T-1$ and $T+1$ (see **Table 4**). Recall that these measures are considered to reflect parafoveal pre-processing and post-target processing spillover. All other measures (FFD, SFD, GD, TT, PrF, and Land-Pos) failed to demonstrate an interaction. Follow-up contrasts for $T-1$ and $T+1$ revealed similar effects, with Neutral and Biasing contexts producing distinct patterns (for condition means, see **Table 3**). In general, in Neutral contexts, pre- and post-target fixations were longer with LF–LC words (e.g., *clown*) compared to any other condition; in Biasing contexts, pre-and post-target fixations were shorter with HF–HC words (e.g., *girls*) relative to the other conditions.

For $T-1$ in Neutral contexts, the three contrasts involving the LF–LC condition were significant by participants and items (LF–LC vs. LF–HC/HF–LC/HF–HC: all $Fs > 4.50$, $ps < 0.05$). The remaining Neutral conditions did not differ from each other (LF–HC vs. HF–LC vs. HF–HC: all $Fs < 1$). For $T-1$ in Biasing contexts, the three contrasts involving the HF–HC condition were significant by participants but marginal in two of the items contrasts

(HF–HC vs. LF–LC/LF–HC/HF–LC: all F_1s > 4.45, p_1s < 0.05; all F_2s > 3.00, p_2s < 0.10). The remaining Biasing conditions did not differ from each other (LF–LC vs. LF–HC vs. HF–LC: all Fs < 1).

An identical pattern of means was obtained in $T + 1$, although the results tended to be less reliable. For $T + 1$ in Neutral contexts, the three contrasts involving the LF–LC condition were significant by participants and items (LF–LC vs. LF–HC/HF–LC/HF–HC: all Fs > 4.75, ps < 0.05). The remaining Neutral conditions did not differ from each other (LF–HC vs. HF–LC vs. HF–HC: all Fs < 1.80, ps > 0.15, except LF–HC vs. HF–LC with F_1 = 2.58, p = 0.115). For $T + 1$ in Biasing contexts, the contrasts involving the HF–HC condition were largely significant by participants (significant in two, marginal in one), but marginal at best by items (marginal in two, trend in one) (HF–HC vs. LF–LC/LF–HC/HF–LC: all F_1s > 3.35, p_1s < 0.08; all F_2s > 2.47, p_2s < 0.13). The remaining Biasing conditions did not differ from each other (LF–LC vs. LF–HC vs. HF–LC: all Fs < 1).

SUMMARY

The overall pattern of results across all measures (FFD, SFD, GD, TT, $T + 1$, $T - 1$, PrF, and LandPos), with a few notable exceptions detailed below, generally showed main effects of Constraint, Frequency, and Context with no interactions. For the main effects of Constraint and Frequency, with the exception of PrF and Land-Pos, all measures showed reliable facilitation for HC over LC and for HF over LF words, respectively. For the main effect of Context, all measures, including PrF and LandPos, showed significant facilitation in Biasing vs. Neutral conditions. In terms of the interactions, Constraint × Frequency was statistically unreliable. Constraint × Context generally reached significance (exceptions noted) in only three measures – LandPos (trend by items), FFD, and SFD (marginal by participants). However, the interaction in the early FFD and SFD measures seemed to be the result of a floor effect impeding HC-Biasing conditions. The Frequency × Context interaction was only reliable in the PrF measure (TT was significant by participants but non-significant by items), replicating prior eye movement studies. Target words were more likely to be skipped when they were in Biasing contexts with an additional (marginal) advantage when the target was HF vs. LF. Finally, the Constraint × Frequency × Context was significant (marginal by items) only in $T - 1$ and $T + 1$. Although some of the follow-up contrasts were marginal, in general, the longest pre- and post-target fixations occurred with LF–LC words in Neutral contexts and the shortest with HF–HC words in Biasing contexts, a pattern that substantiated the underlying main effects of Constraint, Frequency, and Context.

DISCUSSION

The present study was carried out in order to investigate whether there was a difference in processing words beginning with LC initial trigrams (e.g., *clown*), having numerous trigram neighbors, vs. those with HC initial trigrams (e.g., *dwarf*), having few trigram neighbors. Previous work by Lima and Inhoff (1985) had found, contrary to their original predictions, that LC words received shorter fixations than HC words, but only in the FFD measure. In their study, however, LC and HC words were LF words embedded in Neutral contexts. Our study additionally manipulated the

word frequency (LF vs. HF) of LC and HC targets as well as their predictability (Neutral vs. Biasing preceding context). We had expected to replicate Lima and Inhoff's findings in our LF-Neutral condition, with LC words fixated for less time than HC words. However, in HF, Biasing, and/or HF-Biasing conditions, we had expected that HC words might demonstrate a processing advantage over LC words. If, as prior research has demonstrated, parafoveal processing is facilitated for words that are HF (Inhoff and Rayner, 1986) or predictable (Balota et al., 1985), then it seemed probable that HC words in these conditions would show a processing benefit relative to LC words. In general, our findings showed that, regardless of target frequency or predictability, HC words were reliably fixated for *less* time than LC words.

We first review our findings within the context of a time-course framework, delineating the effects in terms of pre-target ($T - 1$, PrF, and LandPos), target (FFD, SFD, and GD), and post-target ($T + 1$ and TT) measures. We then present some further analyses in an attempt to address possible methodological concerns with our experiment. We return to Lima and Inhoff's (1985) study and discuss differences in methods that may have led to their different pattern of results. Finally, we examine recent eye movement studies investigating issues related to word-initial letter constraint whose results are more consistent with our findings.

PATTERNS OF EFFECTS

Pre-target effects

Pre-target fixation duration effects have been a focus of several recent eye movement studies, with both positive and null effects reported (e.g., Rayner et al., 2004b; Drieghe et al., 2005, 2008; Inhoff et al., 2005; Kennedy and Pynte, 2005; Kliegl et al., 2006; Kennedy, 2008; Miellet et al., 2009; Hand et al., 2010). Such effects are termed "parafoveal-on-foveal" effects because characteristics of the (parafoveal) target can begin to emerge in fixation time on the pre-target (foveal) word, before the target is directly fixated. There is no question that information about the upcoming parafoveal word is obtained prior to its fixation – moving window and boundary experiments have demonstrated that normal reading behavior is impaired when parafoveal text is altered. The issues of debate, however, concern (1) the level of parafoveal pre-processing (whether it is limited to lower-level, perceptual analysis or can extend to higher-level, semantic activation); and (2) the implications for models of eye movement control in reading (whether visual attention is allocated in a serial or parallel manner which, consequently, determines if parafoveal information can affect the duration of the current fixation). In our study, pre-target fixations ($T - 1$) demonstrated sensitivity to the target word's constraint, frequency, and predictability, with shorter durations when the parafoveal target was HC, HF, or in a Biasing context. The three-way interaction (marginal by items) showed, in Neutral contexts, a relative disadvantage to LF–LC parafoveal targets and, in Biasing contexts, a relative advantage to HF–HC parafoveal targets. Although such effects apparently support the notion of parafoveal-on-foveal processing at a deep level, we are reluctant to draw any firm conclusions. The aim of our study was not to investigate parafoveal-on-foveal processing. As such, unlike most investigations of parafoveal-on-foveal processing, we did not insure that targets were preceded by longer, content words. We will

return to this issue when we additionally examine whether launch site (i.e., the location of the pre-target fixation) affected target fixation duration.

For PrF, readers were more likely to skip targets that were HC (vs. LC) or were embedded in a Biasing (vs. Neutral) context. Although there was no main effect of Frequency, there was a Frequency × Context interaction. The pattern of effects, in general, replicated past studies (Rayner et al., 2004a; Hand et al., 2010) in which HF-Biasing targets were skipped more often than targets in the other conditions. No other PrF effects were significant.

For LandPos, readers' fixation location on the target (determined from the pre-target fixation) was further into the word in Biasing (vs. Neutral) contexts. Although some eye movement studies show similar findings (e.g., Lavigne et al., 2000; McDonald and Shillock, 2003; Kennedy et al., 2004), others do not (e.g., Rayner et al., 2001; Vainio et al., 2009). The only other effect was a Constraint × Context interaction (significant by participants, trend by items), which generally showed that landing position within HC-Neutral words were further to the left than those in the other conditions (see, e.g., Hyönä, 1995).

Target effects

The three target fixation time measures (FFD, SFD, and GD) all exhibited a significant effect of Constraint, with shorter fixation times associated with HC (vs. LC) targets. The other main effects of Frequency and Context were also significant, replicating past eye movement studies that demonstrate an advantage for HF vs. LF words and for words in Biasing vs. Neutral contexts, respectively. The lack of a Frequency × Context interaction also replicated past studies. The only significant interaction was Constraint × Context in the earlier FFD and SFD measures (although marginal by participants in SFD), showing a null effect of Constraint selectively in Biasing contexts. We suggested, however, that the lack of any difference here was most likely due to a floor effect in which individual fixation times on words in the HC-Biasing condition had reached their lower limit.

Post-target effects

Refixations on the target made after first leaving the target only contributed to 6% of the total possible data. Thus, TT effects tended to be similar to those of GD, demonstrating main effects of Constraint, Frequency, and Context. The only difference was a Frequency × Context interaction that was significant by subjects but not by items, a result similar to that reported in Hand et al. (2010).

$T + 1$ also showed main effects of Constraint, Frequency (marginal by participants and items), and Context. As with $T - 1$, there was a three-way interaction (significant by participants, marginal by items). The pattern of results from the follow-up contrasts (several of which were statistically marginal) revealed increased processing spillover in the LC–LF-Neutral condition and decreased spillover in the HC–HF-Biasing condition, the "hardest" and "easiest" conditions, respectively, as defined by the direction of main effects.

FURTHER ANALYSES

There are two issues with our current experiment that demand further attention. The first is related to our experimental method, the second to our interpretation. A potential confound of our study was that Neutral, single-line sentences were always presented as a first block, followed by a second block of Biasing, two-line materials. We adopted this approach for several reasons. We thought that having the Neutral materials first would enable a more cautious comparison to Lima and Inhoff's (1985) original study which involved only single-line sentences. We also thought it would be less confusing to the participants if similar materials were presented together. Finally, we reasoned that presenting the Biasing materials first may have induced participants to engage in different strategies when subsequently presented with Neutral materials. At the outset, we had originally started to construct "empty" contexts to be presented as the first sentence for our Neutral materials and had intended to randomized all materials within a single block. However, the "empty" contexts generally served to introduce a certain degree of incoherence. Nevertheless, the issue remains that if participants tend to speed up over the course of the experiment, it is possible that our effect of Context may be due to practice effects and not our manipulation.

In general, we do not think that our Context effect is an order effect – past eye movement studies that have manipulated the predictability of targets in fully randomized designs have found similar effects (e.g., Rayner et al., 2004a; Hand et al., 2010; see also, Rayner, 1998, 2009). Additionally, effects from fatigue could offset those of practice over the course of an experiment. To address this concern, however, we performed separate Constraint × Frequency ANOVAs on FFD and SFD for Neutral and Biasing conditions. FFD and SFD represent the earliest measures of processing. If participants sped up from Neutral to Biasing blocks, then it is possible that effects of Constraint or Frequency would likewise be attenuated. Recall, however, that Constraint interacted with Context for the early measures, with Biasing contexts functionally eliminating effects of Constraint. The separate ANOVAs confirmed this [Constraint: neutral-FFD $F_1(1,47) = 11.11$, MSE $= 368$, $p < 0.01$, $F_2(1,21) = 12.91$, MSE $= 150$, $p < 0.01$; Neutral-SFD $F_1(1,47) = 8.00$, MSE $= 552$, $p < 0.01$, $F_2(1,21) = 9.55$, MSE $= 232$, $p < 0.01$; Biasing-FFD and Biasing-SFD all Fs < 1]. These results cannot distinguish between an interaction (possibly due to floor effects) and a general acceleration of fixation times over the experiment. However, Frequency did not interact with Context and such effects were maintained in both halves of the experiment [Frequency: neutral-FFD $F_1(1,47) = 6.49$, MSE $= 471$, $p < 0.05$, $F_2(1,21) = 6.38$, MSE $= 272$, $p < 0.05$; Neutral-SFD $F_1(1,47) = 5.40$, MSE $= 638$, $p < 0.05$, $F_2(1,21) = 4.37$, MSE $= 260$, $p < 0.05$; Biasing-FFD $F_1(1,47) = 6.22$, MSE $= 435$, $p < 0.05$, $F_2(1,21) = 6.00$, MSE $= 241$, $p < 0.05$; Biasing-SFD $F_1(1,47) = 7.47$, MSE $= 434$, $p < 0.01$, $F_2(1,21) = 5.90$, MSE $= 283$, $p < 0.05$].

We also examined the first-pass reading time on each region of the target sentence (i.e., the only sentence in the Neutral condition; the second sentence in the Biasing condition) across Context conditions. Sentences were divided into four regions: the target, itself, including the space preceding it (always six characters); a pre-target region before the target (always 10 characters); a beginning region of text occurring before the pre-target region (13 characters on average); and a post-target region of all text occurring after the target (27 characters on average). For each region, the

first-pass reading time was divided by the number of characters in that region to yield a reading time per character (ms/char) measure. The averages for beginning, pre-target, target, and post-target regions were 33.1, 26.9, 35.2, and 32.7 ms/char for the Neutral condition and 37.5, 23.4, 32.7, and 24.1 ms/char for the Biasing condition, with corresponding differences (Neutral–Biasing) of −4.4, 3.5, 2.5, and 8.6 ms/char. While most regions were read faster in the Biasing compared to the Neutral condition, the first region was read slower. The greatest numerical advantage for the Biasing condition arose from the final region, where discourse integration processes would be most facilitated. While the current data cannot unequivocally demonstrate that our Context effect is *solely* due to the target's predictability (and not the by-product of an order effect), the overall weight of evidence, including that from prior eye movement studies investigating contextual effects, seems to favor an interpretation in which reading behavior across several measures is facilitated by more predictable contexts.

A final point regarding Neutral vs. Biasing conditions is related to anaphor resolution. The concern is that in the Neutral condition, pre-target anaphoric references (e.g., pronouns) have no antecedents, whereas in the Biasing condition, some do. Unresolved anaphors could serve to increase processing time selectively in the Neutral condition, and thus masquerade as a context effect. The conditions under which anaphor resolution proceeds with relative ease or difficulty is, itself, not fully resolved, nor is the issue of how isolated pronouns are processed in context-free circumstances. Nevertheless, the data do not seem to support the contention that the context effect is the result of unresolved anaphors. We examined the Neutral sentences containing unresolved anaphors. Some of these anaphors were located in the

beginning region, others were located in the pre-target region, and some spanned these two regions. Our comparison of reading times in these early target sentence regions (above), however, revealed no evidence of systematic differences. Given that the unresolved anaphors were fairly equally distributed across these two regions, it seems unlikely that they are responsible for the pattern of effects.

The second issue concerns how much we can conclude about parafoveal processing in the absence of employing a boundary paradigm. An invalid parafoveal preview (a letter string different from the target that changes to the target when eyes cross a pre-target boundary) can be used to insure foveal-only processing. By its nature, however, an invalid preview does not simply deny parafoveal processing; it permits parafoveal processing of an incorrect stimulus. Nevertheless, the complexity of our existing design ($2 \times 2 \times 2$) made an additional parafoveal preview manipulation impractical. We can, however, make some tentative conclusions about parafoveal processing based in part on our pre-target ($T − 1$) findings of parafoveal-on-foveal effects as well as on further analyses of our data.

Launch distance (i.e., the number of characters from the pre-target fixation to the beginning of the target region) can be used as a proxy measure of the degree of parafoveal processing of the target (see, e.g., Hand et al., 2010). This argument assumes that nearer launch sites allow for better parafoveal pre-processing than further ones. We first calculated descriptive statistics for our launch site analysis. **Figure 3** shows the landing position as well as the number of data points on target words as a function of launch distance across all conditions. The pattern of target landing position data shows that closer launch sites resulted in saccades further into the target. The pattern of data points shows that launch
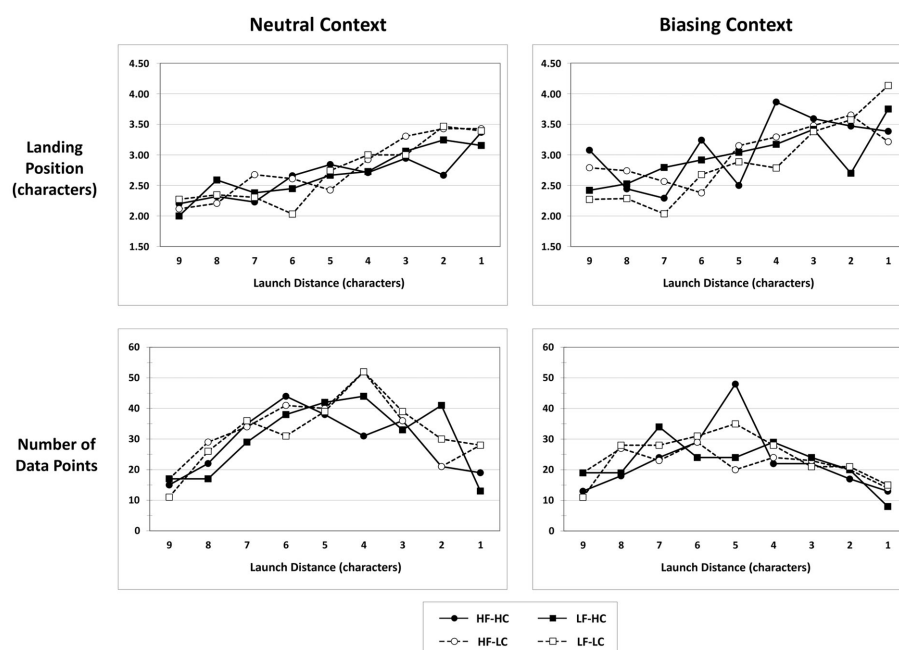

FIGURE 3 | Target landing position (characters) and number of data points as a function of launch distance (characters) from the target for HF–HC, HF–LC, LF–HC, and LF–LC words in Neutral and Biasing contexts. HF, high frequency; LF, low frequency; HC, high constraint; LC, low constraint.

distance was relatively normally distributed. These patterns are confirmed by past eye movement research (e.g., McConkie et al., 1988; Rayner et al., 1996). There are more data points in Neutral context conditions as the target was more likely to be skipped in Biasing context conditions. While the data are somewhat noisy, there do not seem to be any systematic differences between the experimental conditions.

We performed a 2 (Launch Distance: Near, Far) × 2 (Constraint) × 2 (Frequency) × 2 (Context) ANOVA on the FFD data by participants $[F_1(1,47)]$ and by items $[F_2(1,21)]$. For Launch Distance, we defined Near as saccades originating from one to three characters and Far as saccades originating from seven to nine characters. For missing data (less than 2% overall; 11 of 768 participant and 7 of 352 item cells), appropriate condition means adjusted by participant or item were substituted. As in our original analyses, the main effects of Constraint, Frequency, and Context were all significant (Constraint: $F_1 = 8.89$, $p < 0.01$, $F_2 = 5.75$, $p < 0.05$; Frequency: $F_1 = 13.40$, $p < 0.001$, $F_2 = 13.49$, $p < 0.01$; Context: $F_1 = 15.24$, $p < 0.001$, $F_2 = 33.17$, $p < 0.001$). FFDs were shorter on HC vs. LC targets (181 vs. 188 ms), on HF vs. LF targets (180 vs. 188 ms), and on targets in Biasing vs. Neutral contexts (178 vs. 191 ms). Launch Distance was also significant, with shorter FFDs associated with Near vs. Far launch sites (175 vs. 193 ms) ($F_1 = 33.61$, $p < 0.001$, $F_2 = 40.99$, $p < 0.001$). Two interactions were significant by participants but not by items (Launch Distance × Constraint and Frequency × Context: F_1s > 9.10, p_1s < 0.01, F_2s < 1). No other interactions approached significance. Thus, it seems that Launch Distance (within a range of nine characters) did not modulate any of the reported main effects. However, these effects should be considered with caution as they only represent a relatively small sample of the data (see **Figure 3**).

RECONCILING DIFFERENCES

Recall that Lima and Inhoff (1985) only found an advantage for LC words in the FFD measure. Our finding of a processing advantage for HC words was demonstrated across several eye movement measures. The issue remains, however, as to how we can best account for the pattern of our results, both in light of Lima and Inhoff's study as well as in the broader theoretical context of recent related research. It is possible that differences in results between the current experiment and Lima and Inhoff's were due to differences in aspects of materials and methods.

First, the specifications for the number of five- and x-letter neighbors across conditions in their study was 9 and 80 for LC, and 1 and 5 for HC, respectively; in our study, the corresponding values (for comparable LF targets) were 20 and 209 for LC, and 2 and 17 for HC, respectively. Thus, it seems that our LC words were more "unconstrained" than theirs, having denser neighborhoods. In terms of the *lexical constraint hypothesis* – Lima and Inhoff's (1985) initial position, in which word-initial letter information acquired parafoveally is used to constrain the number of possible candidates – LC words having bigger trigram neighborhoods should be additionally disadvantaged. Our findings lend support to this account. According to Lima and Inhoff's revised view, however, larger trigram neighborhoods should lead to even greater subsequent foveal processing efficiency. While both accounts seem

plausible, we believe that the weight of evidence, as discussed below, favors an interpretation in which a higher constraining parafoveal trigram, when clearly visible, acts to facilitate that word's recognition.

Second, in terms of methods, a combination of an expanded experimental design and a greater number of participants in our experiment ($N = 48$) compared to Lima and Inhoff's (1985) ($N = 18$) resulted in over five times more data points available for analysis in our study compared to theirs (4224 vs. 756 observations, respectively). Although the difference between studies in the number of data points per participant per condition was moderate (11 in ours vs. 7 in theirs), it does represent a 57% increase which, nonetheless, serves to enhance the reliability of our results.

Third, Lima and Inhoff (1985) always preceded their target word by a content word that had an average length of seven characters. In our study, the pre-target word tended to be a HF function word. The average length of our pre-target words was four letters (which did not differ across conditions). Although our analysis of launch distance and landing position (**Figure 3**) shows that fixations were made on the pre-target word (launch sites of one to four characters), the median launch site in our sample was five characters. It seems reasonable, then, to assume that our pre-target words were skipped more often than those used in Lima and Inhoff's experiment. The consequences, however, are not straightforward. On the one hand, a single fixation on a longer, content, pre-target word would result in less parafoveal pre-processing of the subsequent target (e.g., Henderson and Ferreira, 1990). However, if a second fixation were made on that pre-target word (the probability of which increases with word length), then a greater degree of target pre-processing could occur (e.g., Sereno, 1992). On the other hand, a higher degree of skipping a shorter, function, pre-target word entails that, although launch distance to the target word is maintained, the parafoveal preview of the target would include an intervening word. Without knowing the frequencies of the different fixation scenarios in Lima and Inhoff's study, it is difficult to speculate further about how the variation in pre-target words between our experiments differentially affected target processing. Nevertheless, when launch distance is taken into consideration, our target word data provide evidence to suggest that pre-target skipping did not interact with the variables of interest. Although the mean pre-target word length was four characters, the median value was three characters. Consequently, our launch distances of Near (one to three characters) vs. Far (seven to nine characters) correspond, to a large extent, to having fixated or skipped the pre-target word, respectively. In the Launch Distance × Constraint × Frequency × Context analysis detailed above, only the main effects reached significance (with shorter target fixation times associated with Near launch distances or with words that were HC, HF, or in Biasing contexts). We can tentatively conclude that, with respect to the experimental manipulations, skipping the word before the target in general only additively modulated subsequent FFDs on the target.

Fourth, Lima and Inhoff's (1985) materials were presented on a Hewlett–Packard 1300A CRT with letters plotted in a dot-matrix font (cyan letters on a black background) in a darkened room. Under these conditions, the text can appear quite pixelated and is more difficult to read. Our materials were presented in a situation

more akin to natural reading – a high quality font (black letters on a white background) in a well-lit room. The difficulty reading a dot-matrix font is substantiated by the longer fixation times in Lima and Inhoff's study. The average FFD and GD in their full-line (i.e., normal reading) condition was 225 and 253 ms, whereas the average FFD and GD in our LF-Neutral condition (i.e., the condition most comparable to their stimuli) was 199 and 216 ms, a reduction of 26 and 37 ms, respectively. Assuming that both experiments sampled typical university students with similar abilities in reading relatively simple short lines of text, it seems that the most plausible explanation for the slower reading times in the Lima and Inhoff study is related to the intelligibility of the font used.

In terms of the speed of identifying parafoveal letters in a dot-matrix font, it is possible that LC trigrams would show an advantage over HC trigrams for reasons related to differential lower-level visual processing. Recently, Kveraga et al. (2007) used low resolution (blurred) and high resolution (clear) stimuli to bias processing toward the magnocellular (M) and parvocellular (P) pathways, respectively. They found that M-stimuli were projected rapidly from early visual areas to the orbitofrontal cortex (OFC) which, in turn, sent rapid feedback in the form of predictions to inferotemporal (object identification) areas. P-stimuli, however, were only projected from occipital cortex to the fusiform gyrus, without the rapid mediation via the OFC. In the current context, a blurred (dot-matrix) parafoveal stimulus, in comparison to a clear one, paradoxically would lead to faster top-down processing. That is, top-down processing predicting a parafoveal word-initial trigram would be easier for common or prototypical (LC) trigrams than for rare (HC) ones.

Finally, a recent eye movement experiment by White (2008) examined the effects of word-initial orthographic familiarity, using HF-familiar, LF-familiar, and LF-unfamiliar words as targets in sentences. The comparison of interest for the current study is that between LF-familiar and LF-unfamiliar words. White (2008) measured orthographic familiarity in terms of *n*-gram token frequencies (i.e., the summed frequency of all words containing a particular letter sequence). White (2008) obtained trigram token values from CELEX (Baayen et al., 1995). In particular, the token-initial trigram frequency was significantly larger for LF-familiar than LF-unfamiliar words. In this respect, these conditions are similar to our LF–LC and LF–HC conditions, respectively. White found that SFD was significantly longer for LF-unfamiliar words (FFD was significant by participants but trend by items; GD was significant by participants and marginal by items; TT was not significant). As with the Lima and Inhoff (1985) study, although the effect is less well expressed in fixation time measures in comparison to our study, the direction of the effect is, nevertheless, inconsistent with our findings.

In order to appropriately evaluate White's words, using the BNC (Davies, 2004), we calculated the same measures we had used to characterize the trigram (*x*-letter) neighborhoods, namely, the number of trigram neighbors (type frequency), the summed frequency of the trigram neighborhood (token frequency, per million), the percentage of the trigram neighborhood accounted for by the target based on its frequency, and the rank of the target within the trigram neighborhood, again, based on its frequency (see **Table 2**). Specifically, our LF–LC words (vs. White's LF-familiar words) had substantially more trigram neighbors (209 vs. 121) and a slightly higher trigram neighborhood summed frequency (1615 vs. 1144 per million), while accounting for a similar percentage of the trigram neighborhood (1 vs. 2%) and relative rank within the trigram neighborhood (28 vs. 30). Our LF–HC words (vs. White's LF-unfamiliar words) had fewer trigram neighbors (17 vs. 31), had a lower summed frequency of trigram neighbors (31 vs. 192 per million), accounted for a greater percent of the trigram neighborhood (38 vs. 22%), and were higher ranking within the trigram neighborhood (1 vs. 9). In neighborhood terms, in comparison to White's words, our LF–LC words were unknown members lost in larger crowds and our LF–HC words were unique members conspicuous within smaller gatherings. In general, there was a greater difference between our LF–LC and LF–HC words than White's LF-familiar and LF-unfamiliar words which could have contributed to the different pattern of results.

Another possible reason for the different pattern of results between White's (2008) and our study, as with Lima and Inhoff (1985), may be related to the quality of the display used. Although White's LF words were slightly lower in frequency in comparison to ours (3 vs. 9 per million, as per the BNC), they were shorter (half four- and half five-letter words vs. all five-letter words). Nevertheless, fixation times were substantially longer (FFDs, SFDs, and GDs were 280, 284, and 309 ms for LF-familiar and 286, 294, and 324 ms for LF-unfamiliar, respectively) than those in our study (see **Table 3**, LF–LC/Neutral and LF–HC/Neutral conditions). As mentioned earlier, it seems that the intelligibility of the font used is the most likely driving force behind differences in reading speed between participant groups. If this were the case, then the pattern of results in White's study may have arisen in part for reasons of diminished visual clarity as discussed earlier.

RELATED FINDINGS

Within the eye movement reading literature, two recent studies have examined issues related to word-initial letter constraint. In the first, Williams et al. (2006) investigated the role of orthographic neighbors as parafoveal previews to targets in a reading study using the boundary paradigm. A word's orthographic neighbors are words of the same length that differ by only a single letter from that word (Coltheart et al., 1977). For example, the neighbors of *sleet* are *fleet*, *sheet*, *sweet*, *slept*, *sleek*, and *sleep*. Williams et al. (2006) compared fixation time on targets when the parafoveal preview was identical to the target (e.g., *sleet*), an orthographic neighbor of the target (e.g., *sweet*), or an orthographically matched non-word (e.g., *speet*). In their first experiment, targets were LF and orthographic neighbor previews were HF words; in their second experiment, targets were HF and orthographic neighbor previews were LF words. They found that the amount of preview benefit depended on the frequency of the preview. When orthographic neighbor previews were HF, the preview benefit was equivalent to identical (LF) previews, with both conditions showing facilitation relative to the non-word preview condition. When orthographic neighbor previews were LF, only the identical (HF) preview condition was facilitated. These results, in partial contrast to those of Lima and Inhoff (1985), demonstrate that when parafoveal information is orthographically similar as well as lexical

(word vs. non-word) and salient (HF vs. LF), lexical processing, as reflected in the subsequent fixation time on the parafoveal word, is facilitated.
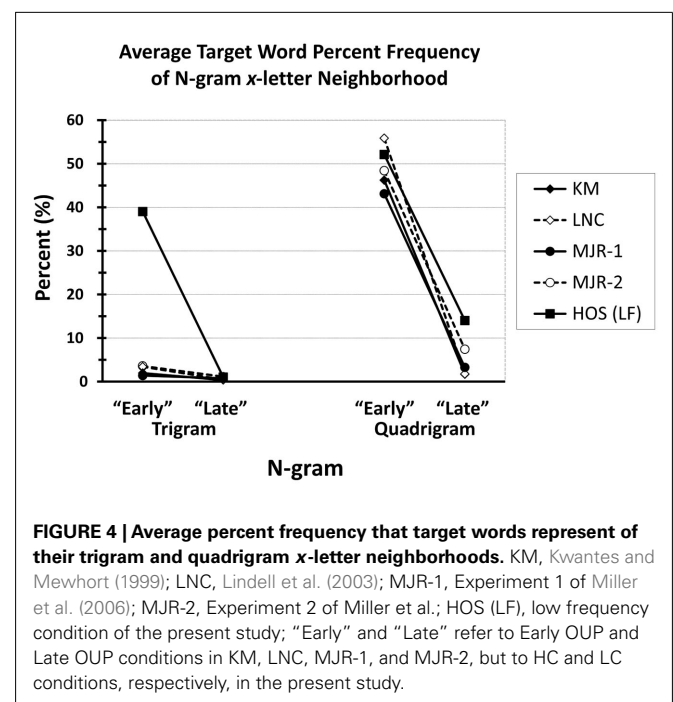
The second study examined the *orthographic uniqueness point* (OUP) in fluent reading (Miller et al., 2006). The OUP is the visual analog of the spoken word uniqueness point, that is, the letter position in a word that differentiates that word from other words based on orthography. For example, a typical early OUP word has its uniqueness point at letter position 4 (e.g., *actress*) whereas a late OUP word cannot be specified until letter 6 or 7 (e.g., *cartoon* or *curtail*). Prior research had used foveally presented words for naming (Kwantes and Mewhort, 1999) and lateralized presentation for a lexical decision task (Lindell et al., 2003) to investigate the OUP. Both studies found an RT advantage for early compared to late OUP words, providing evidence that a word's letters are at some point processed serially, in a left-to-right manner (in English). Specifically, according to Kwantes and Mewhort (1999), the seriality in processing occurs when a reader begins searching for the word in memory, not at the earlier stage of letter identification. Miller et al. (2006), however, raised several methodological concerns with these studies which they addressed in two experiments. First, they used early and late OUP words in the context of a normal reading task while recording participants' eye movements. Second, they generally used different words than those that had been previously tested (Lindell et al.'s words were a subset of those used by Kwantes and Mewhort). In Experiment 1, Miller et al. expanded and altered the stimulus list from the earlier studies. In Experiment 2, Miller et al. further refined their stimuli to address Lamberts' (2005) prior criticism that early OUP words tended to have fewer orthographic neighbors than late OUP words. Finally, using the boundary paradigm, Miller et al. manipulated the parafoveal preview of early and late OUP words across three conditions. The preview could be identical to the subsequent target, have the same first four letters as the target with the remaining letters visually different, or be entirely visually different from the target. Across both experiments, Miller et al. found no evidence to support the notion of serial processing. Late OUP words were read as fast as early OUP words, regardless of the amount of preview available. They attributed the lack of an OUP effect to differences in methodology and stimuli employed in the prior studies.

In the context of our current findings, a positive OUP effect could be interpreted as a relative advantage for words beginning with HC four-letter (quadrigram) combinations (i.e., early OUP words, whose OUP is at letter position 4) vs. words beginning with LC quadrigrams (i.e., late OUP words, whose OUP is at letter position 6 or 7). Because the eye movement experiments (Miller et al., 2006) which did not find an OUP effect used different stimuli than the naming (Kwantes and Mewhort, 1999) and lexical decision (Lindell et al., 2003) studies which did, the differing results may have arisen from the level of constraint conferred by the word-initial quadrigram. One of our measures of constraint was the percentage that each word represented of its entire (*x*-letter) trigram neighborhood (see **Table 2**). For this measure, we divided the frequency of each target word by the summed frequency of all words (including the target) of any length that shared that word-initial trigram. Using this same procedure, we

calculated (as per Davies, 2004) the average percentage that a given target represented of its quadrigram neighborhood in early and late OUP conditions. We found that, across all three of the above studies, early OUP words represented a far greater proportion of their quadrigram neighborhoods (average 48%, range 43–55%) than late OUP words (average 3%, range 2–7%). The percentages for each study are presented in **Figure 4**. While early OUP words, by definition, should comprise a larger percentage of their quadrigram neighborhoods than late OUP words, there was no apparent difference in these means across the different studies.

The possibility remains, however, that the experiments reporting an advantage for early over late OUP words (Kwantes and Mewhort, 1999; Lindell et al., 2003) may have used early OUP words that had higher constraining *trigram* neighborhoods than the experiments that found no such difference (Miller et al., 2006). For each study, we calculated (using Davies, 2004) the percentage that each early and late OUP word represented of its trigram neighborhood. These percentages are presented in **Figure 4**. In terms of trigrams, both early and late OUP words represented only a negligible percentage of their neighborhoods, with a minimal difference between early OUP (average 2.6%, range 1.4–3.6%) and late OUP (average 0.7%, range 0.4–1.1%) words. As with the quadrigram neighborhoods, these proportions did not differ between studies. Thus, although the results of RT and eye movement experiments were in conflict, the profiles of quadrigram and trigram neighborhoods for early and late OUP words were similar.

Assuming that the presence of an OUP effect in naming and lexical decision is due to task effects and that the lack of one in fluent reading more accurately reflects processes associated with recognizing words in text (for an extended discussion, see Miller et al., 2006), the question remains why we found a fixation time

FIGURE 4 | Average percent frequency that target words represent of their trigram and quadrigram *x*-letter neighborhoods. KM, Kwantes and Mewhort (1999); LNC, Lindell et al. (2003); MJR-1, Experiment 1 of Miller et al. (2006); MJR-2, Experiment 2 of Miller et al.; HOS (LF), low frequency condition of the present study; "Early" and "Late" refer to Early OUP and Late OUP conditions in KM, LNC, MJR-1, and MJR-2, but to HC and LC conditions, respectively, in the present study.

advantage for words with HC trigrams while Miller et al. found no such advantage for words with HC quadrigrams. As noted previously, the stimuli used in the prior OUP studies were generally LF words; thus, any comparisons to our study will be limited to our LF–HC and LF–LC conditions. With respect to trigrams, our (LF) HC words represented a much larger proportion of their neighborhoods than did our LC words (see **Table 2**; **Figure 4**). In contrast, Miller et al.'s early OUP words were equally as unrepresentative as their late OUP words in corresponding neighborhoods. With respect to quadrigrams, we first calculated (using Davies, 2004) the percentage that our HC and LC words represented of their quadrigram neighborhoods. Similar to Miller et al.'s early and late OUP stimuli, respectively, our HC words comprised a large proportion (52%) and our LC words a relatively small proportion (14%) of their quadrigram neighborhoods (see **Figure 4**). In short, our stimulus conditions became differentiated one letter position prior to those used in Miller et al. These differences in *n*-gram profiles and in the empirical findings, taken together, would seem to suggest that word-initial letter constraint is only effective if it occurs within the first three (and not four) letters of a word.

Although this is a rather bold claim, eye movement research on the use of parafoveal information does provide support for the attentional relevance of word beginnings (e.g., Rayner et al., 1982; McConkie and Zola, 1987). Nonetheless, we do not want to imply that *no more than* the first three letters of a word are processed in a certain way. Rather, we would suggest that the rate of gain of parafoveal information levels out the further the distance (in letters) from the beginning of the parafoveal word (see, e.g., Engbert et al., 2005; Kliegl et al., 2006; Miellet et al., 2009). Other issues, however, would also come into play. First, fixations to a target can originate from closer or further launch distances which would affect the amount of parafoveal preview obtained (e.g., Hand et al., 2010). Also, on any given fixation, more or less parafoveal preview can be acquired as a function of the difficulty of the currently fixated, foveal word (e.g., Henderson and Ferreira, 1990). One way to test the limits of parafoveal information capture of word-initial quadrigrams in early and late OUP words would be – as we suggested at the outset regarding Lima and Inhoff's (1985) findings – to additionally manipulate word frequency and contextual predictability. That is, an early OUP word may be facilitated if it were both an HF and highly predictable word. As mentioned previously, OUP stimuli tend to be LF words. In the Miller et al. (2006) study, OUP targets appeared in contextually neutral sentences (average Cloze values were less than 0.01). If increased frequency and predictability of the parafoveal word enhances the parafoveal preview benefit of that word, as prior research has demonstrated (e.g., Balota et al., 1985; Inhoff and Rayner, 1986), then it is possible that the highly constraining quadrigrams of such early OUP words would facilitate that word's recognition.

Theoretically, our results have implications for models of eye movement control in reading (e.g., E–Z Reader of Reichle et al., 2003; SWIFT of Engbert et al., 2005). It is beyond the scope of this paper, however, to detail the different mechanisms which may account for our findings (see, e.g., White, 2008). Likewise, our results have implications for a range of word recognition models. Nevertheless, caution must be exercised in making generalizations beyond the specific reading task employed. Effects do not always generalize from lexical decision, or even self-paced reading, to fluent reading conditions. With respect to orthographic neighborhood size (i.e., the number of words differing from the target by exactly one letter), Pollatsek et al. (1999) reported a pattern of results homologous to our own findings. They showed that a large neighborhood size facilitated lexical decision but had an inhibitory effect on reading, even when using the same experimental target words. Such differences in findings are sometimes explained by different mechanisms which are engaged by the different tasks. Norris (2006), on the other hand, adopts a more parsimonious approach in arguing that readers behave like optimal Bayesian decision-makers and exploit whatever statistical patterns that are available in order to deliver the most efficient result. In these terms, a word-initial HC trigram viewed parafoveally greatly raises the *post hoc* probability of the occurrence of that target. Proponents of Bayesian reading models would therefore suggest that the choice of a reading mechanism should be secondary to assuming that readers will learn to recognize visual words in an optimal manner.

CONCLUSION

We examined the word-initial letter constraint of target words in an eye movement reading study that additionally manipulated the word frequency and contextual predictability of these targets. Several results replicated prior research – for example, demonstrating frequency and predictability effects in fixation times and an interaction of these effects in word skipping rates. In direct contrast to Lima and Inhoff (1985), however, we found an effect of trigram constraint in which HC words (e.g., *dwarf*) were consistently fixated for *less* time than LC words (e.g., *clown*). Although Constraint interacted with Context, it did so only in early fixation time measures and was most likely the result of a floor effect. We suggested that the differences in our findings in relation to those of Lima and Inhoff were due to differences in materials and methods. Finally, we evaluated recent related eye movement research in light of our findings. Although this research does not fully corroborate our results, neither does it refute our claims. Additionally, our findings are consistent with a Bayesian account (Norris, 2006) in which readers respond to the statistical information available to perform in an optimal fashion. In sum, this study reports evidence that supports the notion that the level of orthographic constraint conferred by the first few letters of an upcoming word is advantageously processed by the reader.

REFERENCES

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX Lexical Database [CD-ROM]*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Balota, D. A., Pollatsek, R., and Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cogn. Psychol.* 17, 364–390.

Coltheart, M., Davelaar, E., Jonasson, J. F., and Besner, D. (1977). "Access to the internal lexicon," in *Attention and Performance VI*, ed. S. Dornic (Hillsdale: Erlbaum), 535–555.

Davies, M. (2004). *BYU-BNC: The British National Corpus*. Available at: http://corpus.byu.edu/bnc

Drieghe, D., Brysbaert, M., and Desmet, T. (2005). Parafoveal-on-foveal effects on eye movements in text reading: does an extra space make a difference? *Vision Res.* 45, 1693–1706.

Drieghe, D., Rayner, K., and Pollatsek, A. (2008). Mislocated fixations can account for parafoveal-on-foveal effects in eye movements during reading. *Q. J. Exp. Psychol.* 61, 1239–1249.

Engbert, R., Nuthmann, A., Richter, E. M., and Kliegl, R. (2005). SWIFT: a dynamical model of saccade generation during reading. *Psychol. Rev.* 112, 777–813.

Hand, C. J., Miellet, S., O'Donnell, P. J., and Sereno, S. C. (2010). Word frequency and contextual predictability effects in reading: it depends where you're coming from. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 1294–1313.

Henderson, J. M., and Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: implications for attention and eye movement control. *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 417–429.

Hyönä, J. (1995). Do irregular letter combinations attract readers' attention? Evidence from fixation locations in words. *J. Exp. Psychol. Hum. Percept. Perform.* 21, 68–81.

Inhoff, A. W. (1990). Integrating information across eye fixations in reading: the role of letter and word units. *Acta Psychol. (Amst.)* 73, 281–297.

Inhoff, A. W., Eiter, B. M., and Radach, R. (2005). Time course of linguistic information extraction from consecutive words during eye fixations in reading. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 979–995.

Inhoff, A. W., Pollatsek, A., Posner, M. I., and Rayner, K. (1989). Covert attention and eye movements during reading. *Q. J. Exp. Psychol.* 41A, 63–89.

Inhoff, A. W., and Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: effects of word frequency. *Percept. Psychophys.* 40, 431–439.

Inhoff, A. W., and Topolski, R. (1994). Use of phonological codes during eye fixations in reading and in on-line and delayed naming tasks. *J. Mem. Lang.* 33, 689–713.

Johnson, R. L., Perea, M., and Rayner, K. (2007). Transposed-letter effects in reading: evidence from eye movements and parafoveal preview. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 209–229.

Kennedy, A. (2008). Parafoveal-on-foveal effects are not an artifact of mis-located saccades. *J. Eye Mov. Res.* 2, 1–10.

Kennedy, A., Murray, W. S., and Boissiere, C. (2004). Parafoveal pragmatics revisited. *Eur. J. Cogn. Psychol.* 16, 128–153.

Kennedy, A., and Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Res.* 45, 153–168.

Kliegl, R., and Engbert, R. (2005). Fixation durations before word skipping in reading. *Psychon. Bull. Rev.* 12, 132–138.

Kliegl, R., Nuthmann, A., and Engbert, R. (2006). Tracking the mind during reading: the influence of past, present, and future words on fixation durations. *J. Exp. Psychol. Gen.* 135, 12–35.

Kveraga, K., Boshyan, J., and Bar, M. (2007). Magnocelllular projections as the trigger for top-down facilitation in recognition. *J. Neurosci.* 27, 13232–13240.

Kwantes, P., and Mewhort, D. (1999). Evidence for sequential processing in visual word recognition. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 376–381.

Lamberts, K. (2005). Left-to-right sequential processing in visual word recognition: Alternative interpretation of orthographic uniqueness point effects. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 14–19.

Lavigne, F., Vitu, F., and d'Ydewalle, G. (2000). The influence of semantic context on initial landing sites in words. *Acta Psychol. (Amst.)* 104, 191–214.

Lima, S. D., and Inhoff, A. W. (1985). Lexical access during eye fixations in reading: effects of word-initial letter sequence. *J. Exp. Psychol. Hum. Percept. Perform.* 11, 272–285.

Lindell, A., Nicholls, M., and Castles, A. E. (2003). The effect of orthographic uniqueness and deviation points on lexical decisions: evidence from unilateral and bilateral-redundant presentations. *Q. J. Exp. Psychol.* 56A, 287–307.

Marslen-Wilson, W., and Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cogn. Psychol.* 10, 29–63.

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

McConkie, G. W., Kerr, P. W., Reddix, M. D., and Zola, D. (1988). Eye movement control during reading: I. The location of initial eye fixations in words. *Vision Res.* 28, 1107–1118.

McConkie, G. W., and Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Percept. Psychophys.* 17, 578–586.

McConkie, G. W., and Zola, D. (1979). Is visual information integrated across successive fixations in reading? *Percept. Psychophys.* 25, 221–224.

McConkie, G. W., and Zola, D. (1987). "Visual attention during eye fixations while reading," in *Attention and Performance XII*, ed. M. Coltheart (Hillsdale: Erlbaum), 385–401.

McDonald, S. A., and Shillcock, R. C. (2003). Low-level predictive inference in reading: the influence of transitional probabilities on eye movements. *Vision Res.* 43, 1735–1751.

Miellet, S., O'Donnell, P. J., and Sereno, S. C. (2009). Parafoveal magnification: visual acuity does not modulate the perceptual span in reading. *Psychol. Sci.* 20, 721–728.

Miller, B., Juhasz, B. J., and Rayner, K. (2006). The orthographic uniqueness point and eye movements during reading. *Br. J. Psychol.* 97, 191–216.

Norris, D. (2006). The Bayesian reader: explaining word recognition as an optimal Bayesian decision process. *Psychol. Rev.* 113, 327–357.

Pollatsek, A., Lesch, M., Morris, R. K., and Rayner, K. (1992). Phonological codes are used in integrating information across saccades in word identification and reading. *J. Exp. Psychol. Hum. Percept. Perform.* 18, 148–162.

Pollatsek, A., Perea, M., and Binder, K. S. (1999). The effects of "neighborhood size" in reading and lexical decision. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 1142–1158.

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cogn. Psychol.* 7, 65–81.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 372–422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Q. J. Exp. Psychol.* 62, 1457–1506.

Rayner, K., Ashby, J., Pollatsek, A., and Reichle, E. D. (2004a). The effects of frequency and predictability on eye fixations in reading: implications for the E-Z reader model. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 720–730.

Rayner, K., Warren, T., Juhasz, B. J., and Liversedge, S. P. (2004b). The effect of plausibility on eye movements in reading. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 1290–1301.

Rayner, K., Binder, K. S., Ashby, J., and Pollatsek, A. (2001). Eye movement control in reading: word predictability has little influence on initial landing positions in words. *Vision Res.* 41, 943–954.

Rayner, K., McConkie, G. W., and Zola, D. (1980). Integrating information across eye movements. *Cogn. Psychol.* 12, 206–226.

Rayner, K., Sereno, S. C., and Raney, G. E. (1996). Eye movement control in reading: a comparison of two types of models. *J. Exp. Psychol. Hum. Percept. Perform.* 22, 1188–1200.

Rayner, K., Well, A. D., Pollatsek, A., and Bertera, J. H. (1982). The availability of useful information to the right of fixation in reading. *Percept. Psychophys.* 31, 537–550.

Reichle, E. D., Rayner, K., and Pollatsek, A. (2003). The E-Z reader model of eye-movement control in reading: comparisons to other models. *Behav. Brain Sci.* 26, 445–476.

Scott, G. G., O'Donnell, P. J., and Sereno, S. C. (2012). Emotion words affect eye fixations during reading. *J. Exp. Psychol. Learn. Mem. Cogn.* doi: 10.1037/a0027209.

Sereno, S. C. (1992). "Early lexical effects when fixating a word in reading," in *Eye Movements and Visual Cognition: Scene Perception and Reading*, ed. K. Rayner (New York: Springer-Verlag), 304–316.

Sereno, S. C., and Rayner, K. (2000). Spelling-sound regularity effects on eye fixations in reading. *Percept. Psychophys.* 62, 402–409.

Vainio, S., Hyönä, J., and Pajunen, A. (2009). Lexical predictability exerts robust effects on fixation duration, but not on initial landing position during reading. *Exp. Psychol.* 56, 66–74.

White, S. J. (2008). Eye movement control during reading: effects of word frequency *and orthographic familiarity. J. Exp. Psychol. Hum. Percept. Perform.* 34, 205–223.

White, S. J., and Liversedge, S. P. (2004). Orthographic familiarity influences initial eye fixation positions in reading. *Eur. J. Cogn. Psychol.* 16, 52–78.

Williams, C. C., Perea, M., Pollatsek, A., and Rayner, K. (2006). Previewing the neighborhood: the role of orthographic neighbors as parafoveal previews in reading. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 1072–1082.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX

EXPERIMENTAL MATERIALS

The materials are listed as they would appear in the Biasing context condition. The Neutral condition is simply the second sentence of each item, containing the target word (underlined). Target words were low or high frequency (LF, HF) words with low or high constraint (LC, HC) word-initial trigrams. Items are sorted by these four conditions, with 22 items per condition. One participant group read half the items of each condition in a Biasing and half in a Neutral context. The other participant group read the same items in their opposite context condition.

LF–LC

1. Leon was unhappy with the tough bread he got with his soup.
He complained that it was <u>stale</u> and the waitress apologised.

2. Jill's friends were drinking red wine all night in her flat.
In the morning, she noticed an enormous <u>stain</u> on the carpet.

3. Robert was polishing his shoes before his big job interview.
He wanted them to be <u>shiny</u> enough to see his face in them.

4. Maude added two brown sugars to her cappuccino.
She put her spoon through the <u>froth</u> and stirred them in.

5. Sidney had tried a new shampoo for his terrible dandruff.
He massaged it into his <u>scalp</u> before rinsing it out well.

6. Eve's cat had begun to scratch her new furniture.
She would need to get his <u>claws</u> cut to prevent more damage.

7. Ray lived for six months with groups of pygmies in Africa.
He studied each <u>tribe</u> and learned about their customs.

8. Albert thought he looked good with his new facial hair.
His friends disagreed and thought his <u>beard</u> looked awful.

9. Lorna had gone on a five-mile run in the midday sun.
You could see the <u>sweat</u> running down her face by the end.

10. Luke's first job was working at the supermarket.
His responsibility was to <u>stack</u> the shelves.

11. When Geoffrey got a nosebleed, Dawn nearly keeled over.
We thought she was going to <u>faint</u> at the sight of his blood.

12. The child couldn't sleep after watching the monster movie.
It had been really <u>scary</u> and she was afraid to be alone.

13. Gavin placed the expensive necklace around his wife's neck.
It was a string of <u>pearl</u> beads and she adored him for it.

14. There were fingerprints all over the handrail at the bar.
They took away the shine from the <u>brass</u> and looked grubby.

15. Rory was going to dig all day in the potato fields.
He picked up his <u>spade</u> and headed off to work.

16. Pierre had entertained kids at the circus for fifty years.
He had enjoyed being a <u>clown</u> but it was time to retire.

17. The Big Ranch restaurant's specialty was high quality beef.
Bill ordered their biggest <u>steak</u> and a pitcher of beer.

18. Emily had never seen such an enormous bowl of ice cream.
She excitedly grabbed a <u>spoon</u> and began to stuff herself.

19. The shopkeeper suspiciously eyed the girl in the hooded top.
He knew she was a <u>thief</u> and hoped to catch her red-handed.

20. The teacher scrawled sentences onto the blackboard.
The noise of the <u>chalk</u> sent shivers up everyone's spine.

21. Tania first prepared the tomatoes, cucumber and lettuce.
She finished making the <u>salad</u> with oil and vinegar dressing.

22. The letter Lucas had posted was returned to him.
He had forgotten to put a <u>stamp</u> on it before posting it.

LF–HC

1. The heavy rain had washed the dirt and soil into the stream.
This made the water <u>muddy</u> and unsafe to drink.

2. I couldn't stop sneezing as I cleaned out the storage room.
Everything was <u>dusty</u> and it got up my nose as I worked.

3. After many washes, Karl's shirt had lost most of its colour.
It was so badly <u>faded</u> that he would need to buy a new one.

4. Betty only needed the egg whites to make her meringue nests.
Later, she used the <u>yolks</u> to make a separate dish.

5. Hounds used for hunting are trained in special kennels.
They are taught to chase <u>foxes</u> out of their burrows.

6. Heroin addicts often tie a belt tightly around their arms.
This makes it easier to find <u>veins</u> that they inject into.

7. Everyone was excited about going to see big cats at the zoo.
The children wanted to see the <u>lions</u> and tigers most of all.

8. Nadia had been practising her tennis stroke for six hours.
She now had a pain in her <u>elbow</u> and went to get an ice pack.

9. The cause of death was a hammer blow to the head.
The damage to the victim's <u>skull</u> was quite sickening.

10. Valerie's neighbour's Alsatian kept coming into her garden.
She got her son to build a <u>fence</u> to keep the dog out.

11. The boys got into a fist fight in the playground.
They began to furiously <u>punch</u> each other in the face.

12. Andrea constantly suffered from severe eczema.
Her skin was always <u>itchy</u> and she constantly scratched it.

13. The forecast warned drivers of poor visibility on the roads.
As Will drove home, it became <u>foggy</u> and he could barely see.

14. At the ceilidh, Steven vigorously spun Emma round and round.
This made her very <u>dizzy</u> but she still had a good time.

15. The grey squirrel was foraging at the foot of the oak tree.
He recovered the <u>acorn</u> that he had buried last winter.

16. Jamie loved basketball but he was very short for his age.
In gym class, he felt like a <u>dwarf</u> next to his classmates.

17. Poachers still illegally hunt elephants for their tusks.
    It is possible to buy <u>ivory</u> items on the black market.

18. Karen had jumped and landed awkwardly while ice skating.
    She badly hurt her <u>ankle</u> and would need to have an x-ray.

19. Leanne was thirsty so she ordered a diet coke from the bar.
    It came with a slice of <u>lemon</u> and lots of ice and a straw.

20. Maintaining a healthy digestive system requires roughage.
    Foods that are high in <u>fibre</u> are recommended by experts.

21. The music teacher hired removal men when he moved house.
    He couldn't move his <u>piano</u> on his own as it was too heavy.

22. Tara had taken heaps of photos of her Egyptian holiday.
    She would have to begin a new <u>album</u> to keep them together.

HF–LC

1. Maria's only son was graduating today from Oxford.
   As she watched, she felt so <u>proud</u> of his achievements.

2. Marcus almost hurt himself badly lifting weights at the gym.
   He had picked ones that were too <u>heavy</u> for him to lift.

3. During apartheid in South Africa, most races could not vote.
   Only people who were <u>white</u> could take part in the elections.

4. Susan was bored in the lecture and time passed slowly.
   She kept glancing at the <u>clock</u> and counted down the minutes.

5. The pirates located the spot where the treasure was buried.
   They opened up the <u>chest</u> and marvelled at the booty inside.

6. Mary's young son gave her a kick as she washed the dishes.
   She was so surprised, she dropped a <u>plate</u> and it smashed.

7. Tiger Woods was angry when he was distracted playing a shot.
   Apparently, someone in the <u>crowd</u> cheered as he hit the ball.

8. Stuart did not want to travel to London by bus or plane.
   He bought tickets for the <u>train</u> to Waterloo on the internet.

9. Terry went to the new gardening centre.
   He bought a rare <u>plant</u> for his garden.

10. Harry was slightly late for the play in the theatre.
    He missed the <u>start</u> but caught up with the plot quickly.

11. The toddler held onto the furniture to keep himself upright.
    On his own, he was unable to <u>stand</u> without falling down.

12. The joiner hadn't smoothed the edges of the cabinets yet.
    They were still quite <u>rough</u> and not ready to be varnished.

13. Nigel was struggling to cut the turkey with a blunt knife.
    He asked his wife for a <u>sharp</u> one and he continued to carve.

14. During the War, German submarines targeted supply convoys.
    They would attack the <u>ships</u> that carried weapons and food.

15. Every morning, Jeff would walk past the baker's shop.
    He enjoyed the smell of <u>bread</u> and frequently bought a loaf.

16. Everyone knew that "EastEnders" was just beginning.
    We recognised the familiar <u>theme</u> tune and sat down to watch.

17. The park-keepers took good care of the lawns.
    They made sure that the <u>grass</u> was cut every day.

18. There had been a terrible crash at the weekend's Grand Prix.
    Oil had leaked onto the <u>track</u> and caused a massive pile-up.

19. The yacht crew were pleased with the favourable strong wind.
    They used it to gain <u>speed</u> and were sure to win the race.

20. I could feel something in my shoe which dug into my heel.
    It was a small <u>stone</u> which had come from the gravel path.

21. Johnny liked his father to read to him before bedtime.
    There was one particular <u>story</u> he liked about a tiger.

22. David increased his vocabulary by reading lots of books.
    His knowledge of difficult <u>words</u> was far better than others.

HF–HC

1. Meg was driving and spotted a badly injured hedgehog.
   She tried to prevent it from <u>dying</u> but it was too late.

2. Special police units rushed to the bank robbery in progress.
   The men inside were <u>armed</u> and had taken customers hostage.

3. The couple finally got pregnant after trying for months.
   They were extremely <u>happy</u> when they eventually succeeded.

4. Derek asked for a bacon double cheeseburger at Burger King.
   He also ordered an extra <u>large</u> drink to wash it all down.

5. Sheena had to shop for many things in many different stores.
   She made up several <u>lists</u> so that she remembered everything.

6. Henry had been injured in a scrum at school.
   He was unable to play <u>rugby</u> for several weeks.

7. Ted was diabetic and had to monitor what he ate.
   If he ate too much <u>sugar</u> he could become unwell.

8. Dan was traumatised by seeing the mutilated body as a child.
   He could never get rid of the <u>image</u> from his mind's eye.

9. At school, Miss Jones told only the boys to leave early.
   She wanted to talk to the <u>girls</u> about the incident.

10. Keith liked to listen to Mozart, the Beatles, and techno.
    He liked all kinds of <u>music</u> with no particular preference.

11. The Sultan kept his gold bullion hidden in his palace.
    There was always someone there to <u>guard</u> it around the clock.

12. It had rained all night and the footpath was very muddy.
    Hannah's shoes were <u>dirty</u> and she trailed mud in the house.

13. The Queen has never voted in a General Election.
    Members of the <u>royal</u> family are not allowed to.

14. Seth could easily carry six plastic chairs at a time.
    They were incredibly <u>light</u> and could be stacked together.

15. Craig knew the law about carrying illegal weapons in public.
    He still carried a <u>knife</u> despite the risk of being caught.

16. Jack's aunt was supposed to pick him up after school.
    Instead, it was his <u>uncle</u> who was waiting for him.

17. The Ministry of Defence discovered a spy in their operation.
It was a Russian <u>agent</u> who was relaying details to Moscow.

18. Sarah had saved money to have veneers fitted at the dentist.
When they were finished, her <u>teeth</u> looked fabulous.

19. The DVD is now the most common form of movie entertainment.
It seems that the <u>video</u> will soon be a thing of the past.

20. Claire's knee was causing her a lot of pain after exercise.
The specialist said the <u>joint</u> was inflamed and needed rest.

21. It was a cold day and Barbara had forgotten her gloves.
She decided to keep her <u>hands</u> in her pockets for warmth.

22. Jennifer tried a cigarette for the first time and loved it.
She started to regularly <u>smoke</u> when she went out.

The time course of contextual effects on visual word recognition

Chia-Ying Lee[1,2,3,4]*, Yo-Ning Liu[1] and Jie-Li Tsai[4]

[1] Brain and Language Laboratory, The Institute of Linguistics, Academia Sinica, Taipei, Taiwan
[2] Laboratory for Cognitive Neuropsychology, National Yang-Ming University, Taipei, Taiwan
[3] Institute of Cognitive Neuroscience, National Central University, Jhongli, Taiwan
[4] Department of Psychology, National Chengchi University, Taipei, Taiwan

Sentence comprehension depends on continuous prediction of upcoming words. However, when and how contextual information affects the bottom-up streams of visual word recognition is unknown. This study examined the effects of word frequency and contextual predictability (cloze probability of a target word embedded in the sentence) on N1, P200, and N400 components, which are related to various cognitive operations in early visual processing, perceptual decoding, and semantic processing. The data exhibited a significant interaction between predictability and frequency at the anterior N1 component. The predictability effect, in which the low predictability words elicited a more negative N1 than high predictability words, was only observed when reading a high frequency word. A significant predictability effect occurred during the P200 time window, in which the low predictability words elicited a less positive P200 than high predictability words. There is also a significant predictability effect on the N400 component; low predictability words elicited a greater N400 than high predictability words, although this effect did not interact with frequency. The temporal dynamics of the manner in which contextual information affects the visual word recognition is discussed. These findings support the interactive account, suggesting that contextual information facilitates visual-feature and orthographic processing in the early stage of visual word processing and semantic integration in the later stage.

Keywords: anterior N1, contextual effect, event-related potentials, lexical access, P200

INTRODUCTION

Studies have used various measures to demonstrate how the processing of a word can be influenced by its preceding context. In behavioral studies, readers are usually faster or more accurate in responding to words that are congruent with its preceding context (Stanovich and West, 1981; Duffy et al., 1989). By recording eye movements during natural reading, the fixation, and gaze durations are usually shorter for highly expected words than for unexpected words that are embedded in the sentences (Kliegl et al., 2004; Rayner et al., 2004; Sereno et al., 2006; Dambacher et al., 2008; Hand et al., 2010). These results suggest that sentence comprehension depends on continuous prediction of upcoming words. However, *when* and *how* contextual information affects bottom-up streams of visual word recognition is unknown. The temporal resolution of the event-related potentials (ERPs) technique has been used to detail the time course of language comprehension by using a series of ERP components to index various stages of lexical processing (Van Petten and Kutas, 1990; Dambacher et al., 2006; Sereno et al., 2006; Federmeier, 2007; Hand et al., 2010; Molinaro et al., 2010). This study attempted to delineate the manner in which contextual information modulates word recognition during sentence comprehension, especially in the early stage of word processing.

Word recognition models usually assume that a mental lexicon is associated with a "pool" of mentally stored information. Lexical access describes retrieval of such information or access to a discrete lexical entry, either through a search procedure (Forster and Chambers, 1973) or by activating a threshold based on features extracted from the stimulus (Morton, 1969). In general, visual word recognition can be subdivided into three stages, as follows: *prelexical*, *lexical*, and *post-lexical processing* (Forster, 1981; Fodor, 1983). There have been different perspectives on whether contextual information affects word recognition at an early stage, at the moment of lexical access, or at the post-lexical stage of lexical processing. The **modular view** proposes that processing at one level of representation must be completed before the output of this processing can be combined with information from other processing levels (Forster, 1981; Fodor, 1983). The word processing in the sentences must be bottom-up driven (i.e., initiated only after the physical properties of the stimuli are received). The context can only exert its effect at post-lexical stage of the word processing for semantic integration. The facilitation effect of the word in the context is simply because it is easier to integrate upon receipt. Consequently, the modular view does not predict interactions between frequency and predictability, especially in the early stage of lexical processing.

An alternative view is the **interactive account** (Morton, 1969; McClelland and Rumelhart, 1981), which lacks informational encapsulation, and predicts the immediate and mutual influence at various levels of lexical processing. The contextual information can be used in an anticipatory or predictive manner, and exert its effect from the early stage of word recognition, such as the early perceptual features analysis, to the later stage of lexical activation and selection (Federmeier, 2007). This account allows the features of upcoming words to be pre-activated during online sentence processing as a result of top-down contextual processing. Thus, both frequency and context may affect early stages in word recognition. The facilitation effect of a word in the context may be attributed to the contextual information, which can be used to activate words prior to receiving them. Considering the fundamental difference between the modular and interactive accounts regarding the timing of the influence of information from one linguistic level of representation on the processing of another level, the ERPs technique is particularly suitable in evaluating these distinct claims between modular (integrative) and interactive (predictive) views of language comprehension.

In the ERP literature, the effect of context was usually evaluated by manipulating the degree of fit or semantic congruency between the context and its upcoming word (Kutas and Hillyard, 1980a,b; 1984), predictability (Van Petten and Kutas, 1990; Dambacher et al., 2006; Dambacher and Kliegl, 2007), or sentential constraint (Hoeks et al., 2004; Meyer and Federmeier, 2007) in various studies. Despite the various ways to term the contextual influences, empirically, these are usually determined by the cloze procedure, in which participants were asked to complete a sentence fragment with a word that first comes to their mind. The cloze probability of a word refers to the percentage of people who completed a sentence frame with that particular word (Taylor, 1953). A well-replicated finding in the ERP literatures is that N400 amplitudes are inversely proportional to the cloze probability. For example, in the following sentence from the study by Kutas and Hillyard (1984): "*He liked lemon and sugar in his tea/coffee*," the higher the cloze probability of a word (*tea*) in a context, the more reduced the amplitude of the N400 compared to an unexpected word (*coffee*). In general, the reduction of N400 amplitudes is found with words that can be easily integrated into the preceding word, sentence, or discourse context (Kutas and Hillyard, 1980a,b; Van Petten and Kutas, 1990, 1991; van Berkum et al., 1999). These types of findings suggest that the N400 component is sensitive to the processing of lexical integration. The facilitation of processing words in a sentence reflects the ease of integrating the word into context, or the extent to which the context pre-activates specific properties of those words.

To further examine the manner in which contextual information modulates word level processing, studies manipulated the contextual constraint, and the lexical properties of an upcoming word, such as word frequency. Among all types of lexical properties, the word frequency effect has been recognized for its robust influence on the process of word recognition. Relative to high frequency words, readers tend to require a longer period of time to respond to low frequency words in naming and lexical decision tasks (Forster and Chambers, 1973), and require longer fixation and gaze time on low frequency words in natural sentence reading (Inhoff and Rayner, 1986; Kliegl et al., 2004). Some studies

manipulating word frequency and word predictability propose the possible mechanism of contextual influence on parafoveal preview. For instance, Hand et al. (2010) showed the word frequency and word predictability interaction on duration measures when considering the launch distance. The word predictability effect was stronger for low frequency words than for high frequency words at the near launch site, but the effect was stronger for high frequency words than for low frequency words. Tracking the word frequency effect across behavioral and electrophysiological paradigms is particularly relevant because its presence is considered a marker for successful lexical access (Embick et al., 2001; Sereno and Rayner, 2003; Hand et al., 2010). Since the word frequency effect has been used to determine the point in time of lexical access, the earliest word frequency effect on ERPs provides an upper limit for the latency of lexical access.

The ERP data has clearly demonstrated that, when all other factors are constant, N400 amplitude is an inverse function of the eliciting frequency of a word (Bentin et al., 1985; Rugg, 1990). In addition, the N400 frequency effect interacts with a variety of other factors that influence the ease of semantic processing, such as repetition, word position in the sentences, and the predictability of the word in the sentences. For example, the repeated presentation of a word in the word list can reduce or eliminate the N400 frequency effect (Rugg, 1990). Van Petten and Kutas (1990) revealed that the effect of frequency on N400, in which low frequency words elicited larger N400 than high frequency words, was found only when the word occurred early in the sentence, but not at the end of the sentence. Given that the word position may reflect the build-up of context "online," the interaction between word frequency and word position may imply that the frequency effect of lexical processes can be superimposed by the contextual constraint of the sentence. However, it is important to note that a later-occurring word position in a sentence does not necessarily imply that there is increased contextual constraint. Dambacher et al. (2006) further examined the effects of frequency, predictability, and position of words during word-by-word sentence reading. Congruent with Van Petten and Kutas (1990), this study found interactions of predictability and frequency, as well as of position and frequency on N400. The N400 amplitude exhibited a larger predictability effect for low frequency than for high frequency words, and suggested that semantic contextual constraints can override N400 frequency effects (Embick et al., 2001; Dambacher et al., 2006). In addition, a strong frequency effect was observed on the frontocentral P200, in which the P200 amplitude was smaller for high frequency words than for low frequency words. By treating frequency as an index for lexical access, the authors claimed that lexical access was presumably completed for high frequency words within the first 200 ms after stimulus presentation during sentence reading, whereas low frequency words were being processed. This also explained the larger predictability effect on the N400 for low frequency words than for high frequency words. This occurred because the lexical access of low frequency words benefits from contextual information during the N400 time window, and this benefit is strongly reduced in the processing of high frequency words that were previously recognized.

Recent studies also claimed that early processing makes contact with lexical entries for words that include semantic and

phonological properties; therefore, lexical frequency, semantic features, and lexicality affect neural computation within 200 ms post-stimulus onset in a visual word recognition task (Sereno et al., 1998, 2003; Hauk et al., 2006a,b; Penolazzi et al., 2007; Scott et al., 2009). Sereno et al. (2003) examined the temporal locus of contextual influence on word frequency and word ambiguity, and revealed that the contextual effect, coincident with frequency effect, was found on the N1 component from 132 to 192 ms post-stimulus. The ERP literature usually considers the N1 as an index of the visual signal associated with the early stage of word recognition. The findings of Sereno et al. (2003) suggest that the context affects the selection of the appropriate meaning of an ambiguous word in the early stage of lexical processing, which supports the interactive view. Penolazzi et al. (2007) demonstrated the word frequency and probability effects at 120 and 180 ms after written word onset. Other studies also found frequency by predictability interaction (Dambacher et al., 2006; Hauk et al., 2006b; Penolazzi et al., 2007) and semantic coherence effect (Hauk et al., 2006a) as early as approximately 130 ms in the early stage of lexical processing. Federmeier and Kutas (2001) also demonstrated that context begins to have its effects very early in frontal N1, which peaked at around 150 ms, and that this influence continues into the early and late N400 time windows. The effect of constraint on the response to expected exemplars begins in the N1 time window, with a reduced N1 to expected exemplars in high- as opposed to low-constraint sentences. These results indicate that semantic context integration may occur at an early stage, and almost simultaneously with the processing of information regarding the form and lexical properties of a word.

Most studies claim that the early semantic effects on lexical access are mainly based on early effects of lexicality and word frequency (Sereno et al., 2003; Dambacher et al., 2006; Hauk et al., 2006b; Penolazzi et al., 2007; Scott et al., 2009). However, both lexicality and word frequency are highly correlated with the word-form properties (such as bigram, trigram, and word-form frequencies). These early effects maybe attributed to word-form recognition, rather than actual lexical access. Penolazzi et al. (2007) orthogonally manipulated the length, the lexical frequency, and the cloze probability of a word that occurred in a specified semantic context, and found that frequency and probability effects were modulated by word length at 120 and 180 ms after written word onset. Particularly, the long and short words exhibited opposite word frequency effects in the early time window, which may explain the lack of early word-related ERP effects in earlier studies if the physical properties, such as word length, were not controlled effectively. Penolazzi et al. (2007) found that the word length interacts with both frequency and cloze probability during the early time window, but not on the N400 component. The main effect of cloze probability was found on the N400 and post-N400 time windows, and these late ERP indexes are insensitive to stimulus variance. Although the contextual influence starts at the early time window, these early neuropsychological markers depend mainly on the perceptual or other prelexical features of the stimuli. Thus, the contextual effect on early ERP components, such as N1 or P200, may not imply that access to lexico-semantic information occurs within the first stages of lexical access, but acts in an anticipatory or predictive manner for the early perceptual features analysis.

This is further supported by Solomyak and Marantz (2009), who examined the visual recognition of heteronyms to distinguish the abstract word-form process from actual lexical access in the brain. Heteronyms (e.g., "wind," which has two distinct meanings depending on the pronunciation) are phonological and semantically distinct words that share a common orthography, which provides a unique opportunity to distinguish between the processing of lexical property (the frequency ration of one meaning to the other) and word-form properties (open bigram, trigram, and whole-word-form frequencies) in the early stages of processing. Their data revealed a considerable effect of the form properties of the heteronym in the left hemisphere on the M170 and of heteronym frequency ration on the M350. The true lexical properties of heteronyms did not affect processing until after 300 ms post-stimulus, which supports the late access theory. This finding also suggests that the early frequency effect, as reported in previous literature, may only reflect abstract word-form identification rather than actual lexical access (Solomyak and Marantz, 2009).

Related literature has consistently demonstrated that the N400 amplitude is sensitive to the expectancy of a word in a semantic context. However, it remains unclear whether the effects of contextual influence or lexico-semantic processing can be found in early components, such as N1 or P200. A few studies have demonstrated these early ERP effects under the influence of physical or prelexical variables, such as word length, bigram, and trigram frequencies (Hauk and Pulvermüller, 2004; Hauk et al., 2006a). In alphabetic writing systems, word frequency is usually confounded with word length. Most studies that manipulate word frequency have carefully controlled the word length. However, in some cases, the mixed usage of words with various lengths or other physical factors is unavoidable (such as using a corpus of sentences as the stimulus set in Dambacher's series of studies) which may affect or attenuate the short-lived early ERP effects, since these early components are typically focal and brief. Thus, physical properties of the stimulus must be efficiently controlled or explicitly considered to clarify the functional characteristics of these early ERP effects in sentence comprehension.

The English and Chinese writing systems differ in their orthographic features, and the manner in which these features map onto the phonological structure of words. English is an alphabetical language that uses letters and letter combinations to represent the sounds of words. By contrast, the Chinese writing system uses square-shaped characters as the basic reading unit that links directly to monosyllabic sounds, but not to phonemes. Most importantly, according to the Chinese word corpus of Academia Sinica Balanced Corpus (2004), over 76% of the words (type) consist of two or three characters. This study used the advantages of Chinese two-character compounds, which allowed us to bypass the natural confound between word frequency and word length in alphabetic writing systems, to delineate the nature of predictive processing mechanisms in sentence comprehension, especially in the early stage of lexical processing. The effects of contextual predictability (cloze probability) and word frequency of the two-character words in the middle of the sentence are measured during the time windows of the anterior N1, P200, and N400. The target words will be presented in two sessions in order to counterbalance their appearances at high and low predictable

contexts. To reduce the effect from repetition, participants were required to come back for the second session at 2 weeks later. The repetition effects will also be examined to evaluate if the repeated presentation of targets would cause any effect on these ERPs components. This allows researchers to determine the functional stage of word recognition in which the contextual information begins to interact with bottom-up processing of visually presented sentence completions.

MATERIALS AND METHODS
PARTICIPANTS
Twenty-one right-handed native Chinese speakers (eight males) were paid to participate in this experiment (mean age = 23.6 years, range: 18–29 years), and had no history of neurological or psychiatric disorders. All participants were native Chinese speakers with normal or corrected-to-normal vision. Written consent was obtained from all participants.

EXPERIMENTAL DESIGN AND MATERIALS
The contextual predictability (high versus low) and word frequency of the target word were manipulated in a two-by-two factorial design (see **Table 1**). One hundred two-character words were chosen as target words from a Chinese corpus (2004). There were an equal number of high- and low frequency words among the target words, 50 words for each condition. The mean word frequency per million was 91.25 (Max: 294.07; Min: 13.84) for high frequency target words, and 1.59 (Max: 7.27; Min: 0.23) for low frequency target words. The words in high- and low frequency conditions were further matched for the visual complexity and

orthographic neighborhood size. Two types of sentences were constructed for each target word. Therefore, 200 sentences containing 25 or 26 characters were generated, and a target word was embedded at the 11th to 16th character positions of each sentence. When constructing the sentences, care was taken to avoid the lexical associate word of the target appearing in the context preceding the target. The cloze probability of the target was assessed by 19 participants who did not participate in the ERP experiment. In the cloze task, the participants were presented with sentence fragments preceding the target words, and were asked to fill in a word that first came to their mind to complete each sentence fragment. The predictability value was calculated based on the proportion of raters (19 participants) who filled in the target words as their first answers. Each list consisted of 100 sentences, in which half of the target words were highly predictable, and the other half were less predictable. For the high predictability condition, the mean cloze probability value was 0.80 for high frequency predictable target words, and 0.75 for low frequency predictable target words. For the low predictability condition, the mean cloze probability value was 0.05 for high frequency predictable target words, and 0.07 for low frequency predictable target words. For each participant, two lists were created to counterbalance the predictability and frequency of the target word in each sentence. For each type of predictability, half of the targets were high frequency words, and half were low frequency words. The target words in high- and low frequency conditions were matched for the number of strokes ($t = 1.31$, $p = 0.17$). For the word preceding the target word, the statistical analyses revealed that there were no significant difference in their word frequency (raw and log frequency, $Fs < 1$) and

Table 1 | Means of word frequency and predictability of target words and example sentences (Chinese, word-by-word translation, and whole sentence translation) for each condition.

Condition		Word frequency	Log frequency	Predictability	Example sentences
HF	HP	91.25	2.84	0.80	身為/政府/高層/官員/, /理應/遵守/*利益*/迴避/原則/, /以免/知/法/犯法/
					As/government/senior/officials/⸰/should/obey/***interest***/avoid/principle/⸰ /lest/know/laws/violate the laws/
					As senior officials in government, one should avoid conflict of ***interests*** lest we consciously violate the laws.
	LP	91.25	2.84	0.01	人類/若/漠視/環保/, /過分/追求/*利益*/將/導致/無法/挽回/的/生態/浩劫/
					Human/if/ignore/environment/⸰/excessively/pursue after/***interest***/will/cause/can't/reverse/DE/ecological/catastrophe/
					If human keep ignoring the environmental issues and pursuing of ***interests*** excessively will, it would lead to irreversible ecological catastrophe.
LF	HP	1.59	0	0.75	經過/古董店/發現/這/個/漂亮/的/*沙漏*/讓/她/愛不釋手/想要/買回家/
					Come over/antique shop/find/this/GE/beautiful/DE/***sandglass***/make/her/be foud of it/want to/buy it home/
					She passed by an antique shoe and was fond of the beautiful ***sandglass***, which make her wants to buy it home.
	LP	1.59	0	0.01	古代/西方人/在/時鐘/發明/之前/, /使用/*沙漏*/當做/計算/時間/的/工具/
					Ancient time/Western/at/clock/invent/before/⸰/use/***sandglass***/as/measure/time/DE/tool/
					Before the clock was invented, the ancient Western uses ***sandglasses*** for timing.

Target words are highlighted with bolds and underlines in the example sentences. HF, high frequency; LF, low frequency; HP, high predictability; LP, low predictability.

word classes ($X2 = 2.994$, $p = 0.81$) among four conditions. Each participant read the two lists in two separate experimental sessions. To prevent the repetition effect on the target words, participants were required to return to the second session at least 2 weeks after the first session.

PROCEDURE

Participants were individually seated at a distance of approximately 70 cm in front of a monitor, in an electrically shielded room. Each participant received 12 trials for practice and 100 randomized experimental trials in four test blocks. Participants were allowed to rest between test blocks for as long as they required. For each trial, a fixation cross was presented in the center of the screen for 500 ms as a warning that a sentence was about to begin. Sentences were subsequently presented one word at a time at the center of the screen. The size of each character presented on the screen was 32×32 pixels, and there was a space of 4 pixels between characters. The width of a character and the space before it subtended 0.9° of visual angle. Each word appeared for 250 ms, and was followed by a blank screen for 450 ms. Participants were asked to read for comprehension, and tried not to blink during this period of time. A total of 29% of sentences were followed by a comprehension question. Participants were asked to answer by clicking the left or right button on the mouse for Yes and No responses. Otherwise, participants started the next trial by pressing the left mouse button. Across participants, an average of 98.3% of comprehension questions were answered correctly (Max: 100%; Min: 87.3%). The entire session lasted for approximately 40 min.

EEG RECORDING AND PREPROCESSING

The electroencephalogram (EEG) was recorded from 64 sintered Ag/AgCl electrodes (QuickCap, Neuromedical Supplies, Sterling, TX, USA) with a common vertex reference located between Cz and CPz. The EEG was continuously recorded and digitized at a rate of 500 Hz. The signal was amplified by SYNAMPS2 (Neuroscan, Inc., El Paso, Texas, USA) with a low-pass filter of 100 Hz for off-line analysis. The data were re-referenced off-line to the average of the right and left mastoids for further analysis. Vertical eye movements were recorded by a pair of electrodes placed on the supraorbital and infraorbital ridges of the left eye, and horizontal eye movements were recorded by electrodes placed lateral to the outer canthus of the right and left eyes. A ground electrode was placed on the forehead anterior to the FZ electrode. Electrode impedance remained below 5 kΩ.

For off-line analysis, the continuous EEG was epoched with 100 ms before the onset of the target word, and 700 ms post-stimulus intervals. The pre-stimulus interval (-100 to 0 ms) was used for baseline correction. Trials contaminated by eye movement or with voltage variations larger than 60 μV were rejected. The band-pass filter of 0.1 and 30 Hz (zero phase shift mode, 12 dB) was used. The ERPs were calculated for each participant and each condition for every electrode.

RESULT

Figure 1 shows the grand averaged ERPs to the high- and low frequency target words in high- and low predictability contexts across

two sessions from representative electrodes. Visual inspection of the data revealed three main components in all conditions for further analysis. The first distinct negative peak was the anterior N1, which peaked at approximately 100 ms at frontocentral sites. It was followed by the P200, which was a positive-going wave that reached its peak at approximately 220 ms, and was most prominent at the frontocentral electrodes. The third component was the N400, a negative deflection following the N1-P200 complex, which peaked at approximately 350 ms with central-parietal distribution. Effects of word frequency and predictability were accessed by comparisons of mean amplitudes in the following three time windows of interest: anterior N1 (120–150 ms), P200 (200–250 ms), and N400 (300–500 ms).

The repeated measures of ANOVA were performed on these ERPs components, with factors of predictability (high versus low), word frequency (high versus low), repetition (session 1 versus session 2), and electrodes in the region of interest. The examination of repetition effect was used to determine if the repeated presentation of target word in two sessions would cause any repetition effect, and in particular, if any interaction would be caused by the repetition. Thus, the data of two sessions would be merged for each ERP component, if the repetition effect did not interact with condition effects. Otherwise, only the data of the first session would be analyzed if the repetition effect interacted with condition effects. For each ANOVA, the Greenhouse–Geisser adjustment to the degrees of freedom was applied to correct violations of sphericity associated with repeated measures. Consequently, the corrected p-value was reported for all F tests with more than one degree of freedom in the numerator. The *post hoc* tests were conducted by using Tukey's procedure.

ANTERIOR N1

The mean amplitude of the N1 was analyzed by a four-way ANOVA with predictability (high and low), word frequency (high versus low), repetition (Session 1 versus Session 2), and electrode (FZ, FCZ, CZ, F3/4, FC3/4, C3/4) as within-subject factors. The mean amplitudes for all conditions are presented in **Table 2**. The choice of electrodes was motivated by previous studies that reported the contextual effects on the frontal N1 (Federmeier and Kutas, 2001; Dambacher et al., 2006). The main effect of repetition was not significant ($F < 1$). The repetition effect failed to demonstrate any significant interaction with predictability or with word frequency (Fs < 1). The data from Session 1 and Session 2 were merged for further analysis by a three-way ANOVA with predictability (high and low), frequency (high and low), and electrode (FZ, FCZ, CZ, F3/4, FC3/4, C3/4) as within-subject factors. The data revealed that the main effects of both predictability and word frequency were insignificant (Fs < 1). The two-way interaction between predictability and word frequency [$F(1, 20) = 5.45$, $p = 0.030$] was significant, whereas three-way interaction among predictability, word frequency, and electrode was insignificant [$F(8, 160) = 1.63$, $p = 0.17$]. The *post hoc* test revealed that low predictability words elicited a larger negativity than high predictability words in the reading of high frequency words [$F(1, 20) = 8.77$, $p = 0.007$], but not in the reading of low frequency words [$F(1, 20) = 2.33$, $p = 0.14$] (see **Figure 2**).
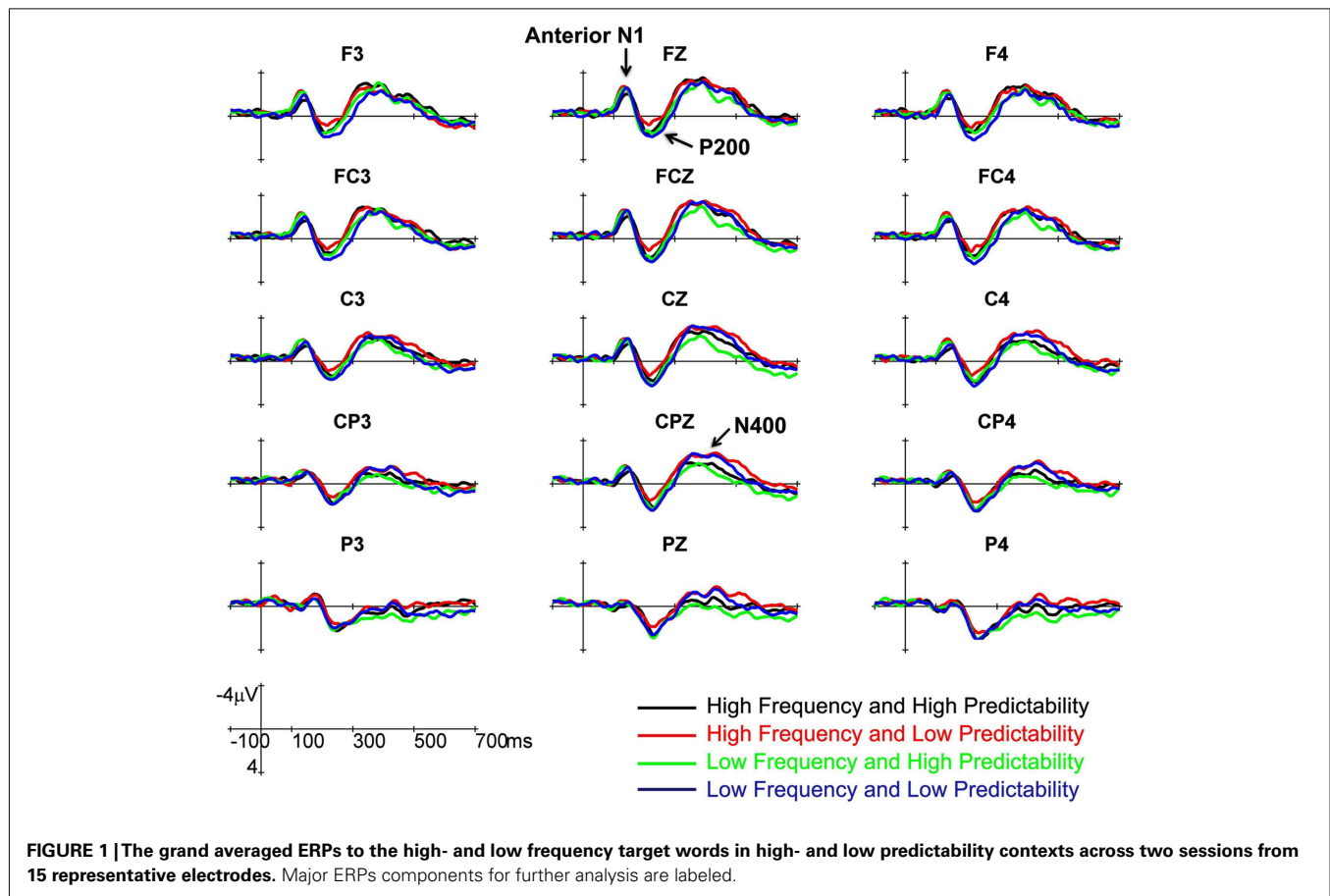
FIGURE 1 | The grand averaged ERPs to the high- and low frequency target words in high- and low predictability contexts across two sessions from 15 representative electrodes. Major ERPs components for further analysis are labeled.

Table 2 | Mean amplitudes of anterior N1 for each condition and each electrode of interest in each session.

N1 (120–150 ms)		Session 1						Session 2					
		High frequency			Low frequency			High frequency			Low frequency		
		Left	Midline	Right	Left	Midline	Right	Left	Midline	Right	Left	Midline	Right
High predictability	F	−1.07	−1.90	−1.50	−2.62	−2.96	−2.37	−1.22	−1.68	−1.35	−1.54	−2.03	−1.80
	FC	−1.15	−1.60	−1.21	−2.58	−2.92	−2.12	−1.58	−1.91	−1.55	−1.73	−2.02	−1.87
	C	−0.80	−0.93	−0.44	−2.23	−2.29	−1.76	−1.29	−1.45	−1.02	−1.36	−1.35	−1.51
Low predictability	F	−1.99	−2.50	−2.35	−1.80	−2.70	−1.92	−2.33	−2.77	−2.15	−1.46	−1.96	−1.38
	FC	−1.86	−2.31	−2.19	−1.95	−2.67	−1.84	−2.61	−2.80	−2.42	−1.53	−1.99	−1.27
	C	−1.43	−1.58	−1.72	−1.79	−2.17	−1.51	−2.24	−2.32	−2.05	−1.07	−1.27	−0.73

P200

Previous studies have suggested a frontocentral distributed contextual effect on the P200 (Federmeier et al., 2005; Dambacher et al., 2006). Accordingly, nine left anterior electrodes (FZ, FCZ, CZ, F3/4, FC3/4, C3/4) were chosen for analyzing the P200. The mean amplitude of the P200 was analyzed by a four-way ANOVA with predictability (high and low), word frequency (high versus low), repetition (Session 1 versus Session 2), and electrode (FZ, FCZ, CZ, F3/4, FC3/4, C3/4) as within-subject factors. The data revealed a significant three-way interaction among repetition, predictability, and frequency [$F(1, 20) = 4.73$, $p < 0.05$]. Therefore, only the data

from Session 1 was further analyzed by a three-way ANOVA with predictability (high and low), frequency (high and low), and electrode (FZ, FCZ, CZ, F3/4, FC3/4, C3/4) as within-subject factors. The mean amplitudes for all conditions are presented in **Table 3**. The data failed to demonstrate significant main effects of predictability [$F(1, 20) = 1.89$, $p = 0.18$] and word frequency [$F(1, 20) = 2.32$, $p = 0.14$]. The only significant interaction was the two-way interaction between predictability and electrode [$F(8, 160) = 2.74$, $p = 0.03$]. The *post hoc* analysis revealed that the high predictability words elicited more positive P200 than low predictability words did in most of the frontocentral electrode
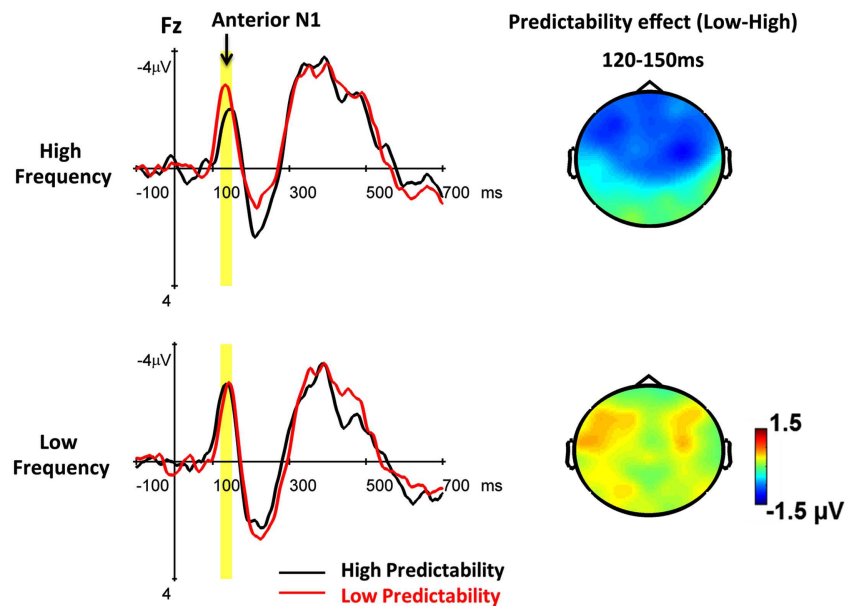
FIGURE 2 | The grand average ERPs words embedded in high and low predictability contexts under high and low frequency conditions across two sessions are shown at Fz. The yellow area indicates the time window (120–150 ms) of the anterior N1. The topographic maps show the predictability effects (low minus high) for high and low frequency words, respectively.

Table 3 | Mean amplitudes of P200 for each condition and each electrode of interest in each session.

P200 (200–250 ms)		Session 1						Session 2					
		High frequency			Low frequency			High frequency			Low frequency		
		Left	Midline	Right	Left	Midline	Right	Left	Midline	Right	Left	Midline	Right
High predictability	F	2.32	2.36	2.23	2.32	2.83	2.63	1.49	1.57	1.51	1.70	1.37	1.38
	FC	2.37	2.67	2.56	2.66	3.20	3.23	1.28	1.66	1.61	1.42	1.42	1.23
	C	2.14	2.70	2.97	2.48	3.42	3.15	1.31	1.70	1.82	1.37	1.78	1.28
Low predictability	F	1.92	1.94	1.86	2.51	2.20	2.64	0.29	0.10	0.83	2.58	2.73	2.84
	FC	1.99	1.94	1.80	2.71	2.56	2.82	0.38	0.58	0.94	2.38	2.80	2.92
	C	1.77	1.94	1.72	2.17	2.59	2.55	0.57	1.11	1.41	2.24	2.98	3.00

($ps < 0.001$), except for F3 [$F(1, 160) = 0.55, p = 0.458$], FC3 [$F(1, 160) = 1.21, p = 0.274$], and F4 [$F(1, 160) = 1.72, p = 0.192$] (see **Figure 3**).

N400

Based on previous ERP studies that have examined the semantic processing of Chinese words (Lee et al., 2007; Hsu et al., 2009; Huang et al., 2011), the analysis of mean N400 amplitude was conducted separately for data derived from the midline and lateral sites using the four-way ANOVAs with predictability (high and low), word frequency (high versus low), repetition (Session 1 versus Session 2), and electrode in the region of interest as within-subject factors. Five electrodes (FZ, FCZ, CZ, CPZ, and PZ) were selected for midline N400 analysis. For N400s in lateral electrode sites, 10 electrodes (F3/4, FC3/4, C3/4, CP3/4, and P3/4) were chosen as the electrode variable. The mean

amplitudes for all conditions are presented in **Table 4**. The midline analysis revealed a significant main effect of repetition [$F(1, 20) = 5.29, p < 0.05$]. The repetition by predictability interaction [$F(1, 20) = 6.58, p < 0.05$] and the repetition by frequency interaction [$F(1, 20) = 5.11, p < 0.05$] were also significant. A similar pattern was also found in the lateral analysis. Therefore, only the data from the first session was used for further analysis with the three-way ANOVA with predictability (high and low), frequency (high and low), and electrode in the region of interest as within-subject factors.

The midline analysis revealed a significant predictability effect [$F(1, 20) = 15.82, p = 0007$] (see **Figure 4**). Low predictability words elicited more negative N400 responses than high predictability words. However, the frequency effect was not significant ($F < 1$). A significant predictability by electrode interaction was observed [$F(1, 20) = 5.33, p = 0.009$]. The predictability effects
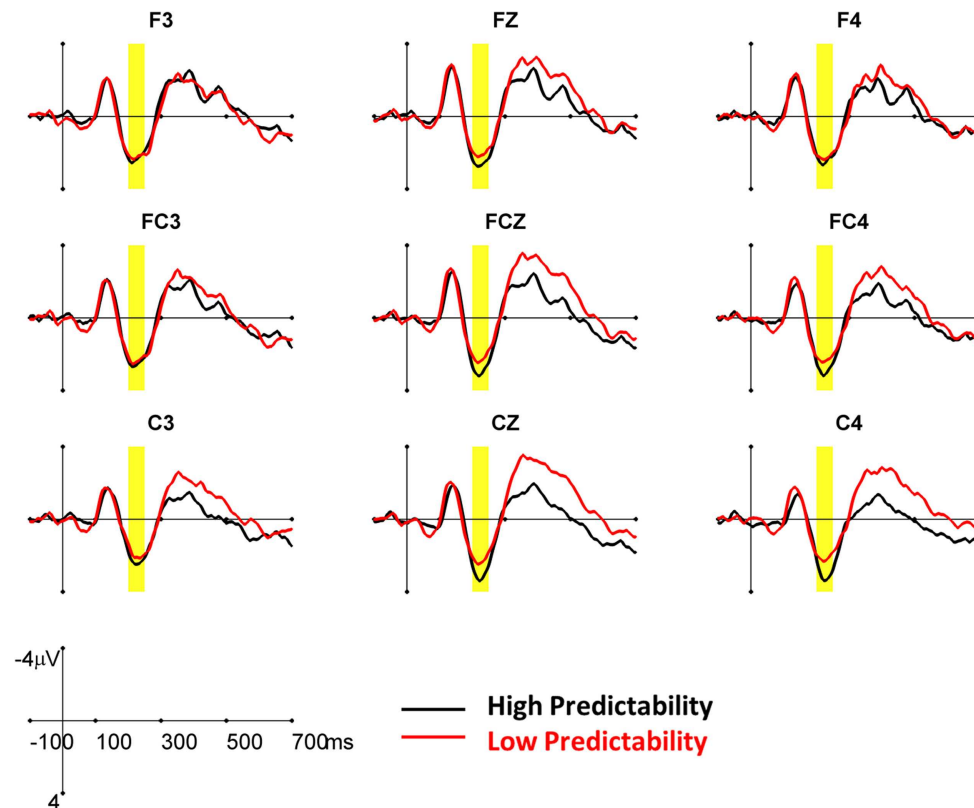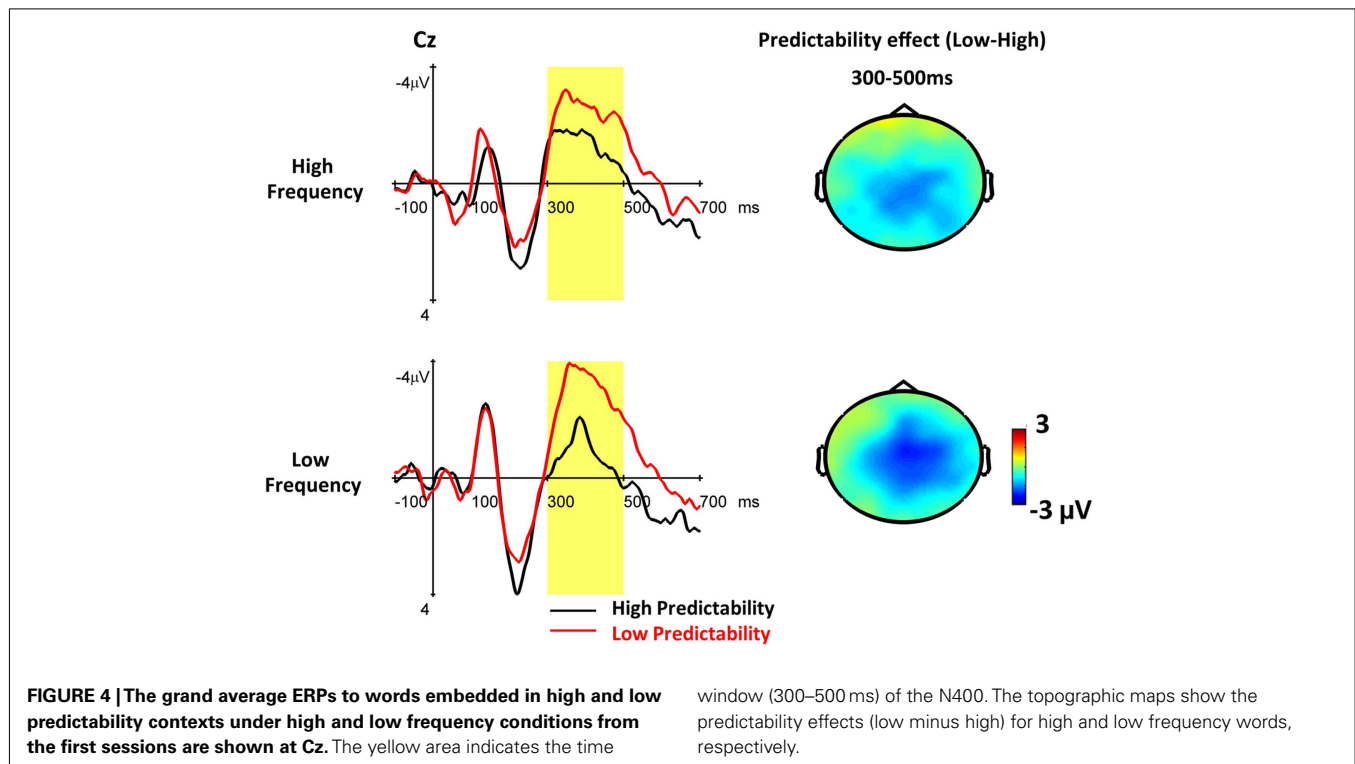
FIGURE 3 | The grand average ERPs for target words embedded in high and low predictability conditions from the first sessions are shown at nine frontocentral electrodes. The yellow area indicates the time window (200–250 ms) of the P200 component.

Table 4 | Mean amplitudes of N400 for each condition and each electrode of interest in each session.

N400 (300–500 ms)		Session 1						Session 2					
		High frequency			Low frequency			High frequency			Low frequency		
		Left	Midline	Right	Left	Midline	Right	Left	Midline	Right	Left	Midline	Right
High predictability	F	−1.61	−2.00	−1.62	−1.61	−1.37	−0.93	−3.25	−3.90	−3.00	−2.23	−2.66	−2.03
	FC	−1.31	−1.71	−1.46	−1.07	−1.18	−0.61	−3.19	−3.93	−2.88	−2.18	−2.47	−2.02
	C	−0.78	−1.39	−0.84	−0.64	−0.79	−0.32	−2.42	−3.21	−1.86	−1.49	−1.75	−1.42
	CP	0.33	−0.18	0.14	0.27	−0.29	0.48	−1.07	−2.64	−0.94	−0.17	−1.46	−0.24
	P	1.68	0.72	1.37	2.01	0.85	1.78	0.54	−0.60	0.68	1.45	0.65	1.77
Low predictability	F	−1.65	−2.30	−1.98	−1.51	−2.76	−1.72	−2.35	−3.57	−2.35	−1.48	−2.13	−1.44
	FC	−1.83	−2.57	−2.07	−1.55	−2.92	−1.85	−2.65	−3.73	−2.72	−1.84	−2.49	−1.93
	C	−1.66	−2.51	−2.14	−1.81	−2.96	−2.16	−2.56	−3.40	−2.48	−1.46	−2.36	−1.56
	CP	−0.78	−1.72	−0.75	−0.97	−2.17	−1.13	−1.21	−3.13	−1.64	−0.55	−2.12	−0.89
	P	0.85	−0.68	0.26	1.09	−0.89	0.35	0.29	−1.13	0.23	0.58	−0.39	0.72

were significant at FCZ, CZ, CPZ, and PZ ($ps < 0.0001$), but only marginally significant at Fz ($p = 0.06$). The two-way interaction between frequency and predictability [$F(1, 20) = 1.03$, $p = 0.32$] and the three-way interaction among frequency, predictability, and electrode [$F(4, 80) = 1.02$, $p = 0.37$] were not significant. Planned comparisons revealed that the predictability effect was

significant at low frequency words [$F(1, 20) = 11.42$, $p = 0.003$], but it was only marginally significant at high frequency words [$F(1, 20) = 3.78$, $p = 0.066$]. For the lateral analysis, both the main effects of frequency and predictability ($Fs < 1$) and their interaction [$F(1, 20) = 1.04$, $p = 0.32$] were not significant. All other interactions also failed to reach significance ($Fs < 1$).

FIGURE 4 | The grand average ERPs to words embedded in high and low predictability contexts under high and low frequency conditions from the first sessions are shown at Cz. The yellow area indicates the time window (300–500 ms) of the N400. The topographic maps show the predictability effects (low minus high) for high and low frequency words, respectively.

DISCUSSION

This study investigated *when* and *how* context influences word recognition, especially in the early stage of lexical processing. The cloze probability and the word frequency effects of the two-character Chinese compound embedded in the middle of the sentences were measured in relation to a set of ERP components (anterior N1, P200, and N400) to index various stages of lexical processing. The data revealed predictability effects on the anterior N1, P200, and N400 components, but demonstrated a differing modulation effect to word frequency and long-term repetition (the same set of target words were embedded in various sentence frames with a 2-week interval). Repetition did not modulate the predictability effect on the anterior N1, whereas it significantly reduced and changed the predictability effects on the P200 and N400. In the literature, the anterior N1 usually indexes the early stage of perceptual analysis, whereas the N400 reflects the retrieval of the lexical item from semantic memory. The differing sensitivity to long-term repetition is consistent with previous studies, which demonstrated that the perceptual properties were less affected by the long-term repetition, whereas the representational changes involved in semantic decisions about previously encountered stimuli may last for several days within the semantic network (Meister et al., 2007). Moreover, the interactions between predictability and word frequency were only found on the anterior N1, but not on the P200 and N400. The different patterns on the anterior N1, P200, and N400 suggest that contextual effects may differ at various stages of lexical processing.

The data demonstrated a significant predictability-by-frequency interaction on the anterior N1 and suggests that contextual information exerts its effect within 100–200 ms after

perceiving the upcoming word. Unlike the typical finding on the N400, the predictability effect, in which the low predictability words elicited a more negative N1 than high predictability words, was only found in reading high frequency words, but not in reading low frequency words. Our findings are in general congruent with the early frequency or contextual effects (within 200 ms) that have been reported in a number of studies (Federmeier and Kutas, 2001; Sereno et al., 2003; Dambacher et al., 2006, 2009; Dikker et al., 2009; Kim and Lai, 2011) and suggest a top-down influence on pre-activated form-based representations. Studies have used the RMS analysis to identify at least two brain responses for the word-evoked potential that occurs within 200 ms after perceiving a word, peaking at 100–120 ms and at 160–180 ms (Hauk et al., 2006a; Penolazzi et al., 2007). The grand average scalp topographies of these peak activations usually exhibit a frontocentral negativity, the anterior N1, and a posterior positivity, the posterior P1, for the first peak, and approximately the opposite polarity pattern for the second peak, the N170. These early contextual or semantic ERP effects that occur within 200 ms are usually short-lived and topographically specific, and thus are much more vulnerable than the widely distributed long-lasting late ERP effects. Therefore, there are some inconsistencies among these early ERP effects in their time windows and spatial distribution, including the anterior N1, the posterior P1, and the N170. Several studies have suggested that the early top-down modulation might originate from the visual cortex. For instance, Sereno et al. (2003) reported that sentence context modulated the ERP in the posterior regions elicited by ambiguous words 132–192 ms after stimulus onset. Dikker et al. (2009) reported the effect of syntactic expectedness on visual M100 at occipital cortex, in which the unexpected item elicited an

enhanced M100 relative to the expected controls, but only when word category was overtly marked by a functional morpheme, supporting the hypothesis that the early visual responses to word-forms can be influenced by prior syntactic context. Kim and Lai (2011) examined the time course of interactions between lexical semantic and word-form analysis during reading of sentences, in which the target word might be replaced by pseudowords which either did or did not orthographically resemble a contextually supported real word, or could be replaced by non-word consonant strings. The pseudowords resembling the contextually supported real words elicited an enhanced occipital distributed P130 relative to real words. The pseudowords that did not resemble the contextually supported real words elicited an enhanced N170 relative to non-word consonant strings. These findings support the view of a top-down excitation of form features in the information flow within visual cortex.

Other studies, however, suggested that the early contextual effects may occur in regions other than the visual cortex. For example, Federmeier and Kutas (2001) reported a contextual effect on the anterior N1 for picture processing, in which the expected example showed a reduced anterior N1 relative to unexpected examples, but only for high constraint sentences. Hauk et al. (2006a) examined the cortical activation elicited by words and pseudowords that varied in orthographic typicality (the frequency of their component letter pairs (bigrams) and triples (trigrams). The typicality effect was found within 100 ms after stimulus onset, in which words and pseudowords with atypical orthography elicited a stronger activity in left peri-sylvian areas (regions extending from Wernicke's area to the posterior/inferior parietal cortex and prefrontal cortex) than those with typical orthographic patterns. However, the lexicality (words versus pseudowords) did not interact significantly with the orthographic typicality until 160 ms. The findings suggest a series of distinct but interactive processing stages in word recognition, from the early form-based analysis to the later lexico-semantic processes. This is further supported by the study of Dambacher et al. (2009), which demonstrated an early predictability effect that was found at approximately 100 ms at right anterior and left posterior sites. In sum, these findings support a top-down influence on early feature processing, in which the context may afford form-specific predictions for the upcoming stimuli.

The other possibility for the early contextual effect might be that the context plays a role in directing attention to specific sensory features early in the information processing stream. In the literature, the attentional effect over the N1 commonly shows a posterior distribution (central, parietal, and occipital), but an anterior distributed N1 (central and frontal) has also been reported (Luck and Hillyard, 1995; Luck et al., 2000; Vogel and Luck, 2000; Tollner et al., 2009). In the visual domain, two types of N1 responses can be found: an early anterior N1 (which occurred over frontocentral electrodes and peaked approximately 120 ms post-stimulus) and a somewhat later posterior N1 (which occurred over lateral posterior electrodes and peaked approximately 175 ms post-stimulus for contralateral stimuli; Mangun and Hillyard, 1991; Mangun et al., 1993; Luck and Hillyard, 1995). In general, both the anterior N1 and the posterior N1 reflect a benefit of correctly allocated attentional resources, and are manifestations of a crucial sensory

attention-gating mechanism. For example, in the visual cueing paradigm, the N1 amplitude is largest for perceptual features in attended (versus unattended) locations and on attended (versus unattended) objects. It suggests that perceptual features are only selected for further perceptual processing if they are in attended locations or on attended objects (Anllo-Vento and Hillyard, 1996; Martinez et al., 2006). In addition, the effect of modality change was most pronounced on the anterior N1 and almost disappeared on the central-posterior N1. This suggests that the anterior N1 enhancement may reflect the detection of a modality change and the initiation of the attentional readjustment, in order to optimize target detection (Tollner et al., 2009).

Lee et al. (submitted) examined the contextual predictability (cloze probability of the final word in the sentence) and orthographic similarity (identical words, orthographically similar homophones, and orthographically dissimilar homophones) of the final words, in an online sentence comprehension task. The data revealed an interaction between predictability and orthographic similarity on the anterior N1. Orthographic similarity only had an effect on the anterior N1 with high predictability sentences, in which an identical character elicited a greater N1 than both orthographically similar and dissimilar homophones. In other words, a larger N1 was evident when the expected character was presented. However, this is only true in reading high predictability sentences. This can be further supported by the current findings, in which the predictability effect can be obtained when the upcoming words are of high frequency. Based on the logogen model (Morton, 1969), a large number of passive word-detector elements (logogens) can accrue information or be activated from a number of sources in parallel. Word frequency and contextual information both may act to reduce the amount of stimulus information that was required to exceed the threshold frequency of a logogen, by lowering the threshold and by raising the level of activation. Assuming a high frequency words would maintain a relatively high resting state in the system, it is easier for the high frequency word to reach the threshold and to except its effect (such as becoming available to capture attention or to achieve lexical access) with the help of context. Our findings are compatible with the early selection model of attention, which contends that attention acts as a sensory gain mechanism that enhances perception of the expected stimuli.

The current data revealed a significant predictability effect on the P200, in which the low predictability words elicited a less positive P200 than high predictability words. Recent studies have demonstrated that the P200 is larger (more positive) for strongly constrained sentence endings, regardless of whether the actual word was the expected word, especially for right, but not left, visual field presentations (Federmeier et al., 2005; Wlotko and Federmeier, 2007). These findings suggest that the contextual effect occurred in the early time window. Dambacher et al. (2006) demonstrated a significant predictability effect on the P200. However, contradictory to our findings and those of other studies, they found a more positive P200 for low predictability words in the sentences. It is important to note that, in their data, the word position also strongly modulates the P200, which was not included as factor in this initial analysis. In fact, there was a high correlation between word predictability and word position ($r = 0.41$) in the

sentence corpus that they used. When effects of predictability and position were estimated within one model, neither the word predictability nor the predictability-by-frequency interaction affected the P200 amplitudes. The only significant factor influencing the P200 was the word position, in which the P200 was larger for words at the beginning and end of sentences than for words in the middle of sentences (Dambacher et al., 2006). This was unexpected because our findings and those of other studies have suggested that the P200 varies with the level of expectancy for a particular item in a sentence. According to Dambacher et al. (2006), the increased working memory load or alertness in the middle of the sentence might be possible reasons for the decreasing P200 amplitude toward the center of a sentence and for subsequently increasing amplitudes. However, further studies are needed to examine this explanation. Indeed, in the literature on the visual search paradigm, the P200 has been used to index the mechanisms for selective attention, feature detection (including color, orientation, and shape), and the early stage of item encoding (Luck and Hillyard, 1994). In general, decreased amplitude of the P200 results from increased attention, which decreases the amount of search space and facilitates feature classification in visual search during the perceptual processing. While recording eye movements during natural reading, the fixation, and gaze durations are usually shorter for highly predictable words than for lowly predictable words that are embedded in the sentences (Kliegl et al., 2004; Rayner et al., 2004; Sereno et al., 2006; Dambacher et al., 2008; Hand et al., 2010). These might reflect the amount of attention to be allocated to the words in the sentences. The lowly predictable words require more attention for further processing, thus eliciting less positive P200s than the highly predictable ones. Taken together, the P200 may reflect the matching of input with expectation. The contextual information can be used to predict or to pre-activate the expected word, thereby facilitating the perceptual matching process in the early stage.

In the N400 time window, the interaction between frequency and predictability was not significant. However, the *post hoc* comparison revealed a significant predictability effect for low frequency words and a marginally significant predictability effect ($p = 0.06$) for high frequency words. The overall pattern is consistent with previous studies (Van Petten and Kutas, 1990; Dambacher and Kliegl, 2007) in which low predictability words elicited a larger N400 than high predictability words, whereas the predictability effect was stronger for low frequency words than high frequency words. The N400 reflects the brain activity associated with semantic access, and the N400 reduction is regarded as ease of semantic integration. Our data revealed that both high and low frequency words can benefit from the contextual information. However, this benefit is substantially reduced for high frequency words because

these are recognized or processed faster and more efficiently in lexical access.

In summary, this study demonstrates contextual predictability effects on the anterior N1, P200, and N400 components. The findings support the interactive account, and suggest that contextual information facilitates visual-feature and orthographic processing in the early stage of word recognition, and semantic integration in the later stage. Similar conclusions were reached by demonstrating the contextual constraining effect on phonological regularity of English indefinite articles ("an" precedes nouns beginning with vowel sounds, whereas "a" precedes nouns beginning with consonant sounds; DeLong et al., 2005) and lexical status (Laszlo and Federmeier, 2009). However, these effects were mainly found in the classical N400 time window. There have been debates on whether the N400 reflects an automatic lexical access or the post-lexical semantic integration. The contextual modulation of the N400 is difficult to differentiate if readers use context to generate expectancies for upcoming items (prediction view) or if they are forced by the words to devote more or fewer resources in integrating words into sentence representations (integration view). Recently, Molinaro et al. (2010) demonstrated a larger N400 for words with neighbors of higher frequency compared to words without such neighbors only when the critical word was embedded in low-constraining sentences. Most importantly, the cloze probability manipulation affects ERPs about 100 ms before the effect of neighbor frequency manipulation (Molinaro et al., 2010). The context facilitates the word recognition even before the lexical competition among a set of word neighbors begins, which thus supports the predictive view. In addition, this study used the anterior N1 and P200 to index the modulation of attention and perceptual analysis in the bottom-up stream of word processing. The interaction between predictability and frequency on the anterior N1 and the main effect of predictability on the P200 provide strong supports for the hypothesis that contextual information can be used to predict and pre-activate the features of upcoming words.

SUPPLEMENTARY MATERIAL
The Supplementary Material for this article can be found online at http://www.frontiersin.org/Language_Sciences/10.3389/fpsyg.2012.00285/abstract

REFERENCES
Academia Sinica Balanced Corpus (2004). [CDROM]: *Academia Sinica*. Taipei.

Anllo-Vento, L., and Hillyard, S. A. (1996). Selective attention to the color and direction of moving stimuli: electrophysiological correlates of hierarchical feature selection. *Percept. Psychophys.* 58, 191–206.

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

Dambacher, M., Goellner, K., Nuthmann, A., Jacobs, A., and Kliegl, R. (2008). Frequency and predictability effects on

event-related potentials and eye-movements. *Int. J. Psychol.* 43, 46–47.

Dambacher, M., and Kliegl, R. (2007). Synchronizing timelines: relations between fixation durations and N400 amplitudes during sentence reading. *Brain Res.* 1155, 147–162.

Dambacher, M., Kliegl, R., Hofmann, M., and Jacobs, A. M.

(2006). Frequency and predictability effects on event-related potentials during reading. *Brain Res.* 1084, 89–103.

Dambacher, M., Rolfs, M., Göllner, K., Kliegl, R., and Jacobs, A. M. (2009). Event-related potentials reveal rapid verification of predicted visual input. *PLoS ONE* 4, e5047. doi:10.1371/journal.pone.0005047

DeLong, K. A., Urbach, T. P., and Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8, 1117–1121.

Dikker, S., Rabagliati, H., and Pylkkanen, L. (2009). Sensitivity to syntax in visual cortex. *Cognition* 110, 293–321.

Duffy, S. A., Henderson, J. M., and Morris, R. K. (1989). Semantic facilitation of lexical access during sentence processing. *J. Exp. Psychol. Learn Mem. Cogn.* 15, 791–801.

Embick, D., Hackl, M., Schaeffer, J., Kelepir, M., and Marantz, A. (2001). A magnetoencephalographic component whose latency reflects lexical frequency. *Cogn. Brain Res.* 10, 345–348.

Federmeier, K. D. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology* 44, 491–505.

Federmeier, K. D., and Kutas, M. (2001). Meaning and modality: influences of context, semantic memory organization, and perceptual predictability on picture processing. *J. Exp. Psychol. Learn Mem. Cogn.* 27, 202–224.

Federmeier, K. D., Mai, H., and Kutas, M. (2005). Both sides get the point: hemispheric sensitivities to sentential constraint. *Mem. Cognit.* 33, 871–886.

Fodor, J. A. (1983). *The Modularity of Mind.* Cambridge: MIT Press.

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time – evidence for autonomous lexical processing. *Q. J. Exp. Psychol. A* 33, 465–495.

Forster, K. I., and Chambers, S. M. (1973). Lexical access and naming time. *J. Verb. Learn. Verb. Behav.* 12, 627–635.

Hand, C. J., Miellet, S., O'Donnell, P. J., and Sereno, S. C. (2010). The frequency-predictability interaction in reading: it depends where you're coming from. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 1294–1313.

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

Hauk, O., Patterson, K., Woollams, A., Watling, L., Pulvermuller, F., and Rogers, T. T. (2006b). [Q:] When would you prefer a SOSSAGE to a SAUSAGE? [A:] At about 100 msec.

ERP correlates of orthographic typicality and lexicality in written word recognition. *J. Cogn. Neurosci.* 18, 818–832.

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

Hoeks, J. C., Stowe, L. A., and Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. *Brain Res. Cogn. Brain Res.* 19, 59–73.

Hsu, C. H., Tsai, J. L., Lee, C. Y., and Tzeng, O. J. (2009). Orthographic combinability and phonological consistency effects in reading Chinese phonograms: an event-related potential study. *Brain Lang.* 108, 56–66.

Huang, H. W., Lee, C. Y., Tsai, J. L., and Tzeng, O. J. (2011). Sublexical ambiguity effect in reading Chinese disyllabic compounds. *Brain Lang.* 117, 77–87.

Inhoff, A. W., and Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: effects of word frequency. *Percept. Psychophys.* 40, 431–439.

Kim, A., and Lai, V. (2011). Rapid interactions between lexical semantic and word form analysis during word recognition in context: evidence from ERPs. *J. Cogn. Neurosci.* 24, 1104–1112.

Kliegl, R., Grabner, E., Rolfs, M., and Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *Eur. J. Cogn. Psychol.* 16, 262–284.

Kutas, M., and Hillyard, S. A. (1980a). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biol. Psychol.* 11, 99–116.

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

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

Laszlo, S., and Federmeier, K. D. (2009). A beautiful day in the neighborhood: an event-related potential study of lexical relationships and prediction in context. *J. Mem. Lang.* 61, 326–338.

Lee, C. Y., Tsai, J. L., Chan, W. H., Hsu, C. H., Hung, D. L., and Tzeng, O. J. (2007). Temporal dynamics of the consistency effect in reading

Chinese: an event-related potentials study. *Neuroreport* 18, 147–151.

Luck, S. J., and Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology* 31, 291–308.

Luck, S. J., and Hillyard, S. A. (1995). The role of attention in feature detection and conjunction discrimination: an electrophysiological analysis. *Int. J. Neurosci.* 80, 281–297.

Luck, S. J., Woodman, G. F., and Vogel, E. K. (2000). Event-related potential studies of attention. *Trends Cogn. Sci. (Regul. Ed.)* 4, 432–440.

Mangun, G. R., and Hillyard, S. A. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. *J. Exp. Psychol. Hum. Percept. Perform.* 17, 1057–1074.

Mangun, G. R., Hillyard, S. A., and Luck, S. J. (1993). "Electrocortical substrates of visual selective attention," in *Attention and Performance XIV (Silver Jubilee Volume)*, eds E. M. David and K. Sylvan (MIT Press), 219–243.

Martinez, A., Teder-Salejarvi, W., Vazquez, M., Molholm, S., Foxe, J. J., Javitt, D. C., Di Russo, F., Worden, M. S., and Hillyard, S. A. (2006). Objects are highlighted by spatial attention. *J. Cogn. Neurosci.* 18, 298–310.

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

Meister, I., Buelte, D., Sparing, R., and Boroojerdi, B. (2007). A repetition suppression effect lasting several days within the semantic network. *Exp. Brain Res.* 183, 371–376.

Meyer, A. M., and Federmeier, K. D. (2007). The effects of context, meaning frequency, and associative strength on semantic selection: distinct contributions from each cerebral hemisphere. *Brain Res.* 1183, 91–108.

Molinaro, N., Conrad, M., Barber, H. A., and Carreiras, M. (2010). On the functional nature of the N400: contrasting effects related to visual word recognition and contextual semantic integration. *Cogn. Neurosci.* 1, 1–7.

Morton, J. (1969). Interaction of information in word recognition. *Psychol. Rev.* 76, 165–178.

Penolazzi, B., Hauk, O., and Pulvermüller, F. (2007). Early semantic

context integration and lexical access as revealed by event-related brain potentials. *Biol. Psychol.* 74, 374–388.

Rayner, K., Ashby, J., Pollatsek, A., and Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: implications for the E-Z reader model. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 720–732.

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

Scott, G. G., O'Donnell, P. J., Leuthold, H., and Sereno, S. C. (2009). Early emotion word processing: evidence from event-related potentials. *Biol. Psychol.* 80, 95–104.

Sereno, S. C., Brewer, C. C., and O'Donnell, P. J. (2003). Context effects in word recognition: evidence for early interactive processing. *Psychol. Sci.* 14, 328–333.

Sereno, S. C., O'Donnell, P. J., and Rayner, K. (2006). Eye movements and lexical ambiguity resolution: investigating the subordinate-bias effect. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 335–350.

Sereno, S. C., and Rayner, K. (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends Cogn. Sci. (Regul. Ed.)* 7, 489–493.

Sereno, S. C., Rayner, K., and Posner, M. I. (1998). Establishing a timeline of word recognition: evidence from eye movements and event-related potentials. *Neuroreport* 9, 2195–2200.

Solomyak, O., and Marantz, A. (2009). Lexical access in early stages of visual word processing: a single-trial correlational MEG study of heteronym recognition. *Brain Lang.* 108, 191–196.

Stanovich, K. E., and West, R. F. (1981). The effect of sentence context on ongoing word recognition: tests of a two-process theory. *J. Exp. Psychol. Hum. Percept. Perform.* 7, 658–672.

Taylor, W. L. (1953). Cloze procedure: a new tool for measuring readability. *Journal. Q.* 30, 415–433.

Tollner, T., Gramann, K., Muller, H. J., and Eimer, M. (2009). The anterior N1 component as an index of modality shifting. *J. Cogn. Neurosci.* 21, 1653–1669.

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

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

Van Petten, C., and Kutas, M. (1991). Influences of semantic and syntactic context on open- and closed-class words. *Mem. Cognit.* 19, 95–112.

Vogel, E. K., and Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology* 37, 190–203.

Wlotko, E. W., and Federmeier, K. D. (2007). Finding the right word: hemispheric asymmetries in the use of sentence context information. *Neuropsychologia* 45, 3001–3014.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Task-dependent masked priming effects in visual word recognition

Sachiko Kinoshita[1] and Dennis Norris[2]*

[1] *Department of Psychology, ARC Centre of Excellence in Cognition and its Disorders (CCD), Macquarie University, Sydney, NSW, Australia*
[2] *MRC Cognition and Brain Sciences Unit, Cambridge, Cambridgeshire, UK*

A method used widely to study the first 250 ms of visual word recognition is masked priming: These studies have yielded a rich set of data concerning the processes involved in recognizing letters and words. In these studies, there is an implicit assumption that the early processes in word recognition tapped by masked priming are automatic, and masked priming effects should therefore be invariant across tasks. Contrary to this assumption, masked priming effects are modulated by the task goal: For example, only word targets show priming in the lexical decision task, but both words and non-words do in the same-different task; semantic priming effects are generally weak in the lexical decision task but are robust in the semantic categorization task. We explain how such task dependence arises within the Bayesian Reader account of masked priming (Norris and Kinoshita, 2008), and how the task dissociations can be used to understand the early processes in lexical access.

Keywords: masked priming, visual word recognition, Bayesian reader

Human readers are remarkably efficient at recognizing words: As noted in the introduction to this special issue, the time window in which a letter string passes from being a mere sequence of printed curves and strokes to being perceived as a word takes no longer than one-third of a second. Speed and minimal effort are hallmarks of automatic processes, and a procedure that has been valuable in studying the automatic aspects of visual word recognition is masked priming.

Forster and Davis (1984) pioneered the masked priming procedure that has come to be the standard in studies of visual word recognition. In this procedure, a trial consists of a sequence of three events: a forward mask (typically a series of # symbols) presented for 500 ms, a prime presented briefly (usually about 50 ms), followed immediately by the target to which a response is required – usually lexical decision. The target is presented either for a fixed duration (e.g., 500 ms) or until subject's response. Typically the prime is presented in lowercase letters and the target in uppercase, so that the prime related in form does not overlap the target physically, and the target functions as a backward mask for the prime. Despite the prime being presented so briefly that subjects have little phenomenological awareness of it; a prime related to the target in some way – for example, by identity (e.g., chair-CHAIR) or form (e.g., cheir-CHAIR) facilitates the response to the target, relative to an unrelated control.

The standard interpretation of masked priming has been driven by two consistent findings. First, as noted above, the prime and target are presented in different cases. Combined with the results of experiments that have specifically manipulated the perceptual overlap between primes and targets (Bowers et al., 1998), this implies that priming is driven by representations at the level of abstract letter identity rather than visual form. Second, in the lexical decision task, identity priming is found consistently for

words, but not for non-words. At first blush this seems to be clear and convincing evidence that priming is purely lexical. If priming were not lexical, why would it be observed only for words? Indeed, the idea that masked priming is lexically mediated has become the "conventional wisdom." For example, Forster (2004) suggested that masked priming is an "index of lexical access" (p. 277).

However, despite the consistency of these findings when using lexical decision, a completely different picture emerges when the task is changed. With careful choice of task, priming can be obtained for non-words, and even made to disappear for words (Forster, 1985; Norris and Kinoshita, 2008). The same pairing of primes and targets can produce different patterns of priming in different tasks. Priming is not an automatic function of the relation between prime and target but depends critically on the nature of the experimental task. Note that we are not the first to make this claim: Earlier, Dehaene et al. (1998) had proposed that subliminal processing can be found under conditions where "subjects unconsciously apply the task instruction to the prime" (p. 598)[1]. This would seem to imply that masked priming effects are "strategic," in which case much of the appeal of the procedure is lost. However, the fact that priming is task-dependent is not a cause

---

[1] Dehaene et al. (1998) put forward the view in the context of subliminal perception. To date, there has been relatively little contact between researchers who use masked priming to study subliminal perception, and those who use it to study the visual word recognition. In part, this is because the former have typically used a small set of stimuli (e.g., arrows pointing left and right, single digits) that are presented repeatedly, and there has been a concern that the mechanism supporting masked priming in this case is a simple stimulus-response mapping process which has little to do with visual word recognition (e.g., Damian, 2001). We will discuss this issue later, with respect to whether masked priming effects can be semantic.

for gloom and despondency, nor does it make a case for abandoning the paradigm. On the contrary, we will suggest that the lawful way in which the influence of a prime presented for only 50 ms can be modulated by the task provides significant insights into the process involved in the first 250 ms of word recognition.

In this paper, we first explain why it is that masked priming should be influenced by the task. We will do this in the context of the framework of the Bayesian Reader model Norris, 2006, 2009; Norris and Kinoshita, 2008). We will then illustrate this by reviewing data[2] that show that almost all of the main phenomena that have been studied with masked priming are modulated by task demands. In fact, some patterns of data are turned on their heads completely when the task is changed. The data challenge all of the following common assumptions about masked priming:

- Masked priming is lexically mediated.
- Orthographic processing is modulated by morphology.
- Orthographic processing is language-specific.
- Transposed-letter priming effects reflect the orthographic representations in the lexicon.
- There is no semantic priming.
- Mapping of orthography to phonology occurs serially across a letter string.

We conclude by noting how the task dependence in masked priming does not imply that it is "strategic."

DECISION PROCESS IN THE BAYESIAN READER

An assumption common to almost all models of reading, whether verbal or computational, is that there is a fixed processing architecture. Word recognition always proceeds in the same automatic fashion and all that varies between tasks is how the output of the system is used. According to this view priming takes place within the fixed component of the system and should therefore be unaffected by task. Even in models that explicitly recognize task differences (e.g., Bimodal Interactive Activation model, BIAM; Grainger and Ferrand, 1994; Grainger and Ziegler, 2011), these differences are explained in terms of the different representations and pathways involved in different tasks (or different memory systems – lexical vs. episodic – in the case of Forster, 1985), that is, in terms of different architectures supporting different tasks.

Here we present a very different view: we suggest that it is the task that drives perception. More specifically, we suggest that perception can be characterized as embodying a process of optimal Bayesian decision making (cf. Knill et al., 1996). According to this view, all perception involves making decisions. A necessary implication of this is that behavior will vary with task demands, i.e., the nature of the decision required. This admits the possibility that the pattern of priming may vary quite radically as a function of the subject's task, even if they involve the same representations. This possibility has huge ramifications for the kind of inferences that can be drawn from masked priming data.

TASK DEPENDENCE IN MASKED PRIMING

We now turn to the review of findings demonstrating task dependence in masked priming. The review is necessarily selective. Several different tasks have been used with the masked priming procedure. In studies of visual word recognition, the most popular task is lexical decision, in which subjects are asked to classify a letter string as either a word, or a non-word. The read-aloud task (also called the naming task or the pronunciation task) is another task that is frequently used by visual word recognition researchers. The task here is to read aloud the word as quickly and as accurately as possible. In a semantic categorization task, subjects are asked to decide whether a word is an exemplar of a category (e.g., "animals"). In addition to these tasks, more recently, Norris and Kinoshita (2008) adapted the masked priming procedure to be used with the same-different task, which we describe below. In all of these tasks, response latency (reaction time, RT) is the main dependent variable, as the tasks are generally designed to produce a high level of accuracy and it is less sensitive to masked priming[3].

IS PRIMING LEXICAL?

As noted earlier, the dominant view of masked priming is that masked priming is lexical (Forster and Davis, 1984; Forster, 1998; Forster et al., 2003). The main support for this view comes from the fact that in the lexical decision task word targets show robust masked priming effects but non-word targets do not. Forster (1998) reported that across 40 lexical decision experiments, the mean size of identity priming effect for non-word targets was 8.7 ms, and only in three cases it was statistically significant at the 0.05 level. This effect of lexical status is easily explained within the lexical view: A masked prime automatically activates its corresponding representation (or within Forster's "entry-opening" account, "opens the entry") in the lexicon, but non-words have no representations to activate (or have no entries to open), hence non-words do not show priming.

Norris and Kinoshita (2008) developed an account of masked priming based on the Bayesian Reader model of word recognition (Norris, 2006). The Bayesian Reader is a stimulus sampling model. The model accumulates samples of evidence from the perceptual input and makes near-optimal decisions as dictated by the experimental task. This simple assumption correctly accounts for a wide range of phenomena in visual word recognition including the logarithmic function relating ease of recognition to word frequency, how neighborhood effects are modulated by task, and how reaction-time distributions change as a function of both frequency and the type of non-words used in lexical decision (Norris, 2009).

---

[2]Our review is limited to behavioral data. For a review of ERP data in masked priming studies, see, e.g., Grainger and Holcombe (2009).

[3]In addition, the perceptual (tachistoscopic) identification task has also been used with the masked priming procedure (e.g., Evett and Humphreys, 1981; Humphreys et al., 1988). Here, a trial sequence consists of a forward mask, prime, target, and a backward mask. The prime and the target are both presented briefly, and the subject's task is to identify the target ("the item presented in uppercase letters"), with accuracy being the dependent measure. Unlike the RT tasks in which the target is presented clearly, subjects sometimes report the prime instead of the target, and intrude the letters from the prime. Furthermore, Davis and Forster (1994) have shown that priming in this task may be entirely attributable to target legibility which results from the physical fusion between the prime and target. As these pattern of priming effects are quite different from the other tasks in which the target is presented undegraded, we will limit our coverage of masked priming to RT tasks in which the target perception is accurate, and RT is the main dependent variable.

In order to extend the model to simulate masked priming Norris and Kinoshita made one additional assumption: they assumed that the perceptual system does not treat the prime and target as separate perceptual events, and therefore evidence from both the prime and targets are integrated in reaching a decision.

As we will explain below, this accounts for the fact that priming in lexical decision is seen only for words but not non-words, but it also makes an interesting and rather counter-intuitive prediction: If the task is changed then it should be possible to observe priming for non-words, and to eliminate priming for words. The task chosen was a same-different task. This has the same basic procedure as masked priming in lexical decision, but now an additional word or non-word is presented for 1000 ms before the prime. The subject's task is to decide whether the target is the same as or different from this referent stimulus. In contrast to the lexical decision task, the same-different task shows equally robust priming effects for word and non-word targets requiring a Same decision, but no priming for either words or non-words for *Different* decision (Norris and Kinoshita, 2008, Experiment 1; Kinoshita and Norris, 2009, Experiment 1, 4). That is, targets requiring a *Same* response pattern like *words* in lexical decision regardless of their lexical status, whereas targets requiring a *Different* response pattern like *non-words* in lexical decision regardless of their lexical status. These results are exactly as predicted by the Bayesian Reader.

In order to understand how the model predicts this pattern of data we need to remind ourselves exactly what the decision is that we ask subjects to make. In a typical lexical decision experiment subjects are told to respond "Yes" if the stimulus is a word, and "No" if it is not a word. They are not told to respond only when they know exactly what the word is, or only when they know exactly what the non-word is. Similarly, in the same-different task, subjects are not required to uniquely identify a stimulus that is *Different*, they just need to know that it is not the same as the referent. The significance of this rather pedantic analysis is that it highlights an important parallel between the two tasks. In both cases subjects have a set of stimuli in memory. In lexical decision this set corresponds to the entire lexicon. In the same-different task the set contains only the referent. In both cases the task is to determine whether the target is a member of the specified set. The data have a simple pattern: targets in the set show priming, targets not in the set show no priming.

But how is this pattern predicted by the Bayesian Reader? Although we explain this more formally and back it up with simulations in Norris and Kinoshita (2008; see also Norris et al., 2010) here we will try to give a more intuitive account of how this pattern emerges from the combination of stimulus sampling and optimal decision making. We begin by examining why in lexical decision only the words show priming. We will use the standard statistical metaphor of drawing balls from an urn. On each trial the urn contains balls which are a mixture of colors. On half of the trials the balls are mainly black, and on the other half they are some other color. The task is to draw balls from the urn to determine whether the balls are mainly black or not. We can think of black balls as words, and colored balls as non-words. The assumption that the balls in the urn are never purely one color reflects the stimulus sampling component of the Bayesian Reader. The mix of colors corresponds to noise in the sampling process and means

that many samples are required to make a confident decision. See Norris (2006) for full details of the mathematics of the decision process.

Consider what happens when we add a "prime" where some extra balls are sampled before the balls representing the target. We assume that these "prime" balls are mainly a single color (black, blue, red, etc.). In the case where the prime and target balls are both mainly black, you will obviously need that many fewer balls to reach a decision that most of the balls are black. If you prime the black balls with a sample of balls of a different color you will clearly need to sample more balls form the target to appreciate that most of the balls are in fact black. There will therefore be an overall priming effect with identity primes producing faster responses (needing fewer samples) than unrelated primes. Now consider what happens when the target balls are some other color. A prime of the same color will provide exactly the same kind of evidence as the samples from the target, so this will provide a head start equal to the number of balls in the prime. But, a prime of a different color will provide exactly the same head start. Any ball that is not black contributes in exactly the same way to the decision that the target balls are not black. As noted above, when performing lexical decision it does not matter exactly what the non-word is, all that matters is that it is not a word. Similarly, the exact color of the balls is immaterial because there is no need to know what color the balls are, so long as they are not black. All that counts in reaching the decision is whether or not the balls are black. There will therefore be a "priming effect" for black balls, but no priming effect for balls of any other color. Balls of any other color are simply balls that are not black.

As with all analogies, the urn analogy fails to capture the more subtle details of the full model. For example, it might seem to imply that priming a colored-ball target with a black-ball prime should bias the decision process toward "black" and give rise to inhibition. However, in lexical decision, RTs to non-words are unaffected by the lexical status of the prime. This is because the main effect of the prime is to generate evidence that the target is in a particular area of orthographic/lexical space, but is insufficient to provide specific evidence that the target is a word or not.

Below we describe examples where this framework has been useful in understanding the nature of task dependence. Before doing so, we put to rest an alternative account of why there is no priming for non-words in lexical decision and *Different* decisions in the same-different task.

FAMILIARITY BIAS

It has sometimes been suggested that masked priming effects reflect a combination of lexical activation and "familiarity bias." If this were true it would not only undermine our explanation of masked priming, but also undermine the value of the task as a tool for providing insights into the first 250 ms of reading. It is worth emphasizing that the changes in pattern of priming with task reported by Norris and Kinoshita (2008) were exactly as predicted by the Bayesian Reader. No other account of masked priming would have lead one to expect that priming would be completely different in lexical decision and the same-different task. Nevertheless, Bowers (2010) suggested that the data can be

explained in terms of a "familiarity bias," originally suggested by Bodner and Masson (1997). The idea is that a masked non-word prime facilitates the identification of a repeated non-word target by "preactivating the relevant sublexical representations," but a repeated non-word target is "perceived as more familiar (due to its improved perception)" (Bowers, 2010, p. 786), and this familiarity bias counteracts the benefit due to the preactivation of sublexical representation for non-word targets. Similarly, in the same-different task, the absence of priming effect for *Different* responses is explained in terms of the familiarity bias due to increased fluency of perceiving the target producing a bias toward responding "*Same*" (because "the increased fluency can be taken as evidence that the target has been repeated," Bowers, 2010, p. 787), which inhibits a *Different* response.

The problem with the familiarity bias hypothesis is that it is *ad hoc*, and it would seem possible to rationalize any conceivable pattern of data within this loosely formulated view. Furthermore, attempts to test it empirically have not succeeded in producing support for it. Kinoshita and Norris (2011) attempted to replicate a finding by Bodner and Masson (1997), which has been taken as providing evidence for the familiarity bias hypothesis. The finding concerns the emergence of priming for non-word targets in a lexical decision task presented in a cAsE-AlTeRnAtEd format. According to the familiarity bias hypothesis, case-alternated targets are visually unfamiliar, and this should have the effect of reducing reliance on "perceived fluency" which generates a bias toward responding "Word" in the lexical decision task. Consequently, when targets are presented in a visually unfamiliar, case-alternated format, masked priming effects should emerge for non-word targets, reflecting the preactivation of sublexical representations, and this is what Bodner and Masson (1997) found. Kinoshita and Norris' (2011) replication of Bodner and Masson's lexical decision experiment, using the stimuli used by Bodner and Masson, showed that the priming for case-alternated non-word targets did emerge. However, further analysis showed that this was limited to non-words containing letters that were ambiguous when presented in a case-alternated format, namely, a lowercase "l" or an uppercase "I" (e.g., lOvInk, lOmIT); non-words that did not contain ambiguous letters (e.g., jAsEnT, nOrBaT) did not show any priming. These results suggest that the priming of non-words emerged with case-alternated targets because of priming of letters made ambiguous by case-alternation, not because subjects abandoned familiarity bias when all targets were visually unfamiliar. Follow-up experiments using case-alternated targets that did not contain ambiguous letters showed the standard pattern of priming, namely, robust masked priming effects for words but not non-words in lexical decision, and masked priming effects for the *Same* decision but no priming for the *Different* decision in the same-different task. These results provide little support for the familiarity bias hypothesis, but are as expected from the Bayesian Reader account[4].

DO TRANSPOSED-LETTER PRIMING EFFECTS DEPEND ON LEXICAL REPRESENTATIONS?

One of the most widespread uses of masked priming has been to investigate the nature of orthographic representations. The standard procedure is to manipulate the degree of orthographic overlap between prime and target with the aim of "cracking the orthographic code" (Grainger, 2008). The paradigmatic example of this enterprise is the transposed-letter (hereafter TL) priming effect. This effect refers to the finding that a prime generated by transposing the positions of letters in the target (usually adjacent, internal letters, e.g., *jugde-JUDGE*) facilitates the recognition of the target (often almost as much as the identity prime) more than a prime generated by replacing the corresponding letters with letters not contained in the target (e.g., *junpe-JUDGE*). First reported by Forster et al. (1987), this effect has been replicated many times in the lexical decision task across different languages (e.g., Perea and Lupker, 2003; Lupker et al., 2008, in English; Perea and Lupker, 2004 in Spanish; Schoonbaert and Grainger, 2004, in French; Perea and Perez, 2009, in Japanese kana)[5].

As with other priming effects, in the lexical decision task, TL priming is readily observed with words but not with non-words. In studies examining TL priming with non-words (e.g., Perea and Lupker, 2003; Schoonbaert and Grainger, 2004; Perea and Carreiras, 2008) the effect is absent or unreliable at best. This invites the inference that TL priming is telling us about specifically lexical representations rather than representations at a purely orthographic level (Grainger and van Heuven, 2003; Whitney and Cornelissen, 2005, 2008). However, in the light of the task differences we have already described, we need to ask whether these effects are indeed truly lexical. That is, might we see TL effects for non-words in another task? It should now be clear that the obvious way to investigate orthographic effects in non-word processing is to use the same-different task. Kinoshita and Norris (2009) did exactly this and found that TL priming effects were equally robust for words and non-words, indicating that the effect is pre-lexical in origin. Also using the same-different task, García-Orza et al. (2010) extended the finding of TL priming effects to digit- and symbol-strings, demonstrating that the effect is not even limited to letter stimuli. These results are consistent with the assumption of the Overlap model (Gomez et al., 2008) and the noisy-position Bayesian Reader model (Norris et al., 2010). Both of these models suggest that TL priming effects arise from perceptual uncertainty in the location of visual objects during the brief period in which masked primes are presented.

The comparison between the lexical decision task and the same-different task has also been useful in further elucidating how TL priming effects interact with linguistic factors. In lexical decision, TL priming effects have been reported to be modulated by morphological structure. Duñabeitia et al. (2007) used the lexical decision task and reported finding robust TL priming effects in both Basque and Spanish if the letters in the prime are transposed within a morpheme (e.g., spekaer) but not if the letters were transposed across a morphemic boundary (e.g., speaekr). (See however

---

[4]To be complete, Kinoshita and Norris (2011) noted that there are conditions under which familiarity bias does seem to operate, but these are not the conditions under which masked priming is used to study early processes in visual word recognition. Readers are referred to Kinoshita and Norris (2011) for details.

[5]But not in Hebrew (e.g., Velan and Frost, 2009, 2011). We will turn to this finding shortly.

Rueckl and Rimzhim, 2011, for a failure to replicate these "morphological boundary effects" in English.) Previous studies (e.g., Rastle et al., 2004) have shown that priming is observed with masked primes that merely appear to be morphologically complex (e.g., "corner," which appears to contain the suffix "er") primes the stem target (e.g., CORN) and that this effect is not due to mere orthographic overlap ("brothel," where the ending "el" is not a suffix does not prime "BROTH"). This has been taken as evidence for an orthographically driven morphological decomposition process. Within this context, Duñabeitia et al. interpreted their own results as suggesting that morphologically complex words are decomposed into morpheme constituents at the same stage as when letter position coding takes place. Duñabeitia et al. (2010) investigated whether morphological decomposition is an obligatory part of orthographic processing, or occurs only in the service of lexical access. They used the Spanish stimuli used by Duñabeitia et al. (2007) and found that in this task, unlike the lexical decision task, TL priming effects were unaffected by the morphological structure: TL priming effects were equally robust when the letter transposition was within a morpheme, or when it straddled across a morpheme boundary. From these results, Duñabeitia et al. (2010) concluded that the presence of an orthographically defined morpheme (prefix or suffix) is not sufficient to drive the morphological decomposition process (as would be expected, for example, from the view that orthographic representations become structured as a result of learning of structural regularities such as low bigram frequency associated with morpheme boundaries); the morpho-orthographic segmentation process only comes into play when lexical access is attempted.

We (Kinoshita et al., 2012) recently reported a similar task dissociation with Hebrew, a language for which morphology is believed to play a more important role. In contrast to the Indo-European languages with linear concatenative morphology (where prefix/suffix is simply appended to the stem), Semitic morphology is comprised of tri-consonantal *roots* which are embedded in *phonological word patterns*. For example, the Hebrew word TIZ-MORET ("orchestra") consists of the root ZMR, which alludes to the concept of singing, and the phonological word pattern TI-O-ET, which is used to form feminine nouns. Frost et al. (1997) showed that in lexical decision roots but not word patterns prime the whole word and argued that in Hebrew roots are "lexical units." Frost et al. (2005) further showed that unlike in English, primes that are one-letter-different from the target do not produce priming in Hebrew. Based on these results, Frost (2009) argued that whereas in English and other Indo-European languages the lexical space is structured in terms of the constituent letters and their positions, Hebrew lexical space is structured according to the morphological roots. Velan and Frost (2009, 2011) further pointed out that in many Hebrew words, transposing letters in a root produces another root, and showed that in these words, TL priming effects are not found. With this as background, Kinoshita et al. (2012) tested whether Hebrew morphology also modulates TL priming in the same-different task. The results were clear: Robust TL priming effects were found with Hebrew words and non-words, irrespective of morphological structure, even for the words for which Velan and Frost (2009, 2011) did not find TL priming effects in the lexical decision task. Norris and Kinoshita (2012) took these results

to argue that the basic perceptual processes supporting the identification of written symbols are universals: They are governed by exactly the same principles as all other forms of visual object recognition, and that it is what the reader does with those symbols that depend on the properties of the language. These dynamic, task-dependent patterns of TL priming effects would be hard to explain within models of word recognition which assume that orthographic representation with fixed properties – properties that are built in to the orthographic representations to reflect the structure of the language – get activated automatically whenever the word is presented.

The way that TL priming is modulated by morphology might appear to suggest that orthographic processing is different in the two tasks. For example, one might assume that in lexical decision there is some kind of feedback from morphology that alters orthographic processing. However, there is a much simpler explanation that is in line with our suggestion that the primary difference between the tasks is in the way information is used in making decisions.

A system performing optimal decisions should obviously make use of all of the information available. Lexical decision could be based on whole-word forms, but any morphological representations that become available in during the access process should also be taken into account. Whereas transposing letters within a word-sized unit, whether in lexical decision or in the same-different task, may produce a representation that is still a close match to the target, transposing letters between the much smaller morphemic units in a word may well cause much greater disruption, simply because the units are shorter. Indeed, letter transpositions are more apparent with short words (compare for example ALBE/ABLE vs. TRANLSATE/TRANSLATE). So, to the extent that word recognition and lexical decision take advantage of morphology, TL priming effects will be modulated by morphology in tasks that require lexical access, even if morphology has no direct influence on orthographic or letter-level processing. This is another advantage of comparing data from different tasks. One tasks informs us about the importance of morphology, the other tells us that lower level orthographic processing can be completely independent of morphology.

IS PRIMING SEMANTIC? SEMANTIC CATEGORIZATION

The semantic priming effects found with masked primes in the early studies of subliminal perception (e.g., Marcel, 1983) were treated with a great deal of skepticism, and the results generally did not stand up to close methodological scrutiny (e.g., Holender, 1986, see Kouider and Dehaene, 2007, for a historical review of the literature). In visual word recognition studies also, semantic priming effects with masked primes are generally weak and unreliable (e.g., Frost et al., 1997; Rastle et al., 2000). These studies used tasks such as lexical decision and perceptual identification that do not require semantic processing. In contrast, the semantic categorization task necessarily requires semantic processing, as the decision is whether the target word has the semantic features of a category exemplar (e.g., McRae and Boisvert, 1998; Grondin et al., 2009). Thus, masked primes that share semantic features with the target ought to produce priming in a semantic categorization task, and indeed this has now been shown in many studies

(e.g., Dehaene et al., 1998; Bueno and Frenck-Mestre, 1999, 2002; Kunde et al., 2003; Quinn and Kinoshita, 2008). Bueno and Frenck-Mestre (2008) further showed that semantic priming effects with masked primes can be demonstrated at a shorter prime-target SOA in semantic categorization than in lexical decision.

It is important to consider what sort of decision process is involved in semantic categorization when using categories like "animals," and "living things" typically used in studies of word recognition. Natural categories like these are characterized by family resemblance: While category exemplars generally resemble each other to some extent and share features, there is no necessary and sufficient set of features that all exemplars possess (Rosch and Mervis, 1975). That is, contrary to the suggestion made by some researchers (e.g., Carreiras et al., 1997; Forster and Hector, 2002) in semantic categorization, there is no single dimension (e.g., "animalness") that can be monitored to make a decision. One implication of this is that the prime need not be a category member to produce priming; sharing many semantic features with the exemplars should be sufficient to produce priming. Quinn and Kinoshita (2008, Experiment 3, 4) demonstrated this with what they called "impostor" priming. Impostors were non-members of a category which nevertheless shared many semantic features with the exemplars, e.g., for the category "Planets," "comet," "asteroid"; for the category "Human body parts," "mind," "claw." Although subjects correctly rejected these items as non-members when the items were presented as targets in a non-speeded condition, they were slower to reject them in a speeded condition, indicating that these items were similar to the exemplars. When used as primes, these impostors facilitated categorization of targets relative to unrelated primes.

Another point to note with regards natural categories is that the degree of family resemblance varies from category to category. In general, small categories (categories with a small number of members) such as "single-digit numbers" (e.g., one, seven, four), "precious stones" (e.g., diamond, ruby, sapphire), or "planets" (e.g., Mars, Venus, Jupiter) are homogeneous, whereas categories containing a large number of exemplars like "animals" tend to be a superordinate category that comprise heterogeneous sub-categories such as birds, mammals, fish, etc. Quinn and Kinoshita (2008) noted that consequently, the prime-target pairs drawn from a small category are more likely to share semantic features relevant to the category classification (e.g., category − "planet," mars-VENUS) than when the exemplars are drawn from a large category (e.g., category − "animal," parrot-RABBIT). Quinn and Kinoshita attributed the failures to find facilitation of target categorization by category-congruent primes with large categories (e.g., Forster et al., 2003; Forster, 2004, Experiment 3) to the lack of semantic feature overlap, and showed that provided that feature overlap is high (e.g., hawk-EAGLE, frog-TOAD)[6], category congruence effects with masked primes are also found with large categories.

---

[6]Quinn and Kinoshita (2008) quantified the amount of feature overlap using the feature production norm of McRae et al. (2005). Other measures of semantic overlap using co-occurrence statistics of words based on text corpus analysis [e.g., Latent Semantic Analysis (LSA): Landauer and Dumais, 1997; Correlated Occurrence Analog to Lexical Semantics (COALS): Rohde et al., 2004] are also possible.

RESPONSE CONGRUENCE EFFECT AND STIMULUS-RESPONSE MAPPING

One of the issues that has been of debated vigorously in research on subliminal perception (see Kouider and Dehaene, 2007, for a review) is whether the category congruence effects observed with masked primes are semantic, or reflect a response conflict that has a different origin. In semantic categorization tasks requiring a binary categorization decision (e.g., is the target bigger/smaller than 5?) and a key-press response to indicate the decision, category congruence is confounded with decision congruence and response congruence. Consequently, these terms are often used interchangeably even though they are conceptually distinct. As noted above, our view is that it is the congruence in the information used to make the decision required by the task that produces priming, not merely response congruence. Nevertheless, when a small set of stimuli are responded to repeatedly in the same task, repetition benefits not only the semantic classification process but also other levels of response representation such as action (which finger is used; see Horner and Henson, 2009). This is a point to consider when interpreting masked priming effects in categorization tasks that used a small set of items repeatedly.

Dehaene et al. (1998) reported a highly influential study using single digits (Arabic numerals, e.g., 1, 8) and number words (e.g., ONE, EIGHT) as primes and targets in a "bigger-than-5?" task. They showed that primes that belonged to the same category as the target (e.g., prime = 3, target = ONE) facilitated the response to the target relative to primes that belonged to the opposite category (e.g., prime = 7, target = ONE). In addition to the behavioral data, they showed congruence effects in the hemodynamic (fMRI) data and the electrophysiological (ERP) measures of brain activity related to the preparation of motor responses, and took the results to argue that masked primes were semantically categorized and then processed all the way to the level of a motor response.

Damian (2001) questioned this conclusion on the basis that Dehaene et al. (1998) used a small set of stimuli repeatedly as both primes and targets and it was therefore possible that the primes were activating a motor response directly, on the basis of a learned stimulus-response mapping. In support of this claim, Damian showed that in a categorization task that required size judgment against an arbitrary reference ("Is the real-world object corresponding to the word larger or smaller than 20 cm × 20 cm?") using a small set of Dutch words (e.g., appel/apple, huis/house), the congruence effect emerged only from the second block, after the prime had been used as a target. Moreover, when the primes were used in a task that did not require the same categorization decision and key-press response (a read-aloud task), they did not produce a congruence effect. Damian concluded from these results that Dehaene et al.'s (1998) findings also reflected stimulus-response mappings rather than congruence in semantic category.

Damian's (2001) claim has in turn been challenged by the finding that with other stimuli (e.g., numbers), primes that have not been responded to ("novel" primes) do produce congruence effects (Naccache and Dehaene, 2001; Kunde et al., 2003; Forster, 2004; Reynvoet et al., 2005; Kinoshita and Hunt, 2008; Quinn and Kinoshita, 2008). Naccache and Dehaene (2001) have further reported that with their number stimuli in the "bigger-than-5" task, the priming effect was greater the closer in numerical distance

the prime and target were (e.g., for the target 1, the prime 2 produced more facilitation than the prime "3"), indicating the semantic nature of the effect. The semantic categorization studies reviewed earlier (e.g., Forster et al., 2003; Forster, 2004, Experiment 3; Quinn and Kinoshita, 2008) used novel primes and a large set of targets presented only once and obtained robust category congruence effects. Stimulus-response mapping alone therefore cannot explain the masked priming effects found in these studies.

Kinoshita and Hunt (2008) used RT distribution analysis to tease apart the contribution of these two levels of congruence – stimulus-response mapping and semantic features – in the "bigger-than-5" task, using single digits as stimuli. They found different RT distributions for novel primes and used primes, with a disproportionate slowdown of congruent trials (e.g., 3-1; 6-9) in the slow RT bins when the prime was a used prime (see Ansorge et al., 2010, for a similar finding). Borrowing ideas developed in the context of response conflict literature (e.g., De Jong et al., 1994; Hommel, 1994), Kinoshita and Hunt suggested that the slowdown specific to the used primes reflects the decay over time, or active suppression of the response code activated directly by the prime. The congruence effect for novel primes, in contrast, was suggested to be semantic in origin, and time locked to the processing of the target. More specifically, the congruence effect for the novel primes was suggested to be a semantic priming effect, and is based on the overlap in semantic features (quantity information) between the prime and the target that are relevant to the categorization.

In sum, when categorizing a small set of stimuli repeatedly, the priming effect could be semantic in origin, or could be due to stimulus-response mapping (see Finkbeiner and Friedman, 2011, for converging evidence based on the trajectory measure using a reaching response). Because visual word recognition researchers have typically used a large set of stimuli presented only once, this issue has not been a concern, however, it is a factor to consider when the set of potential stimuli is small as in single digits or letters of the alphabet, as we will see below.

STIMULUS-RESPONSE MAPPING IN LETTER PROCESSING

As noted in the introduction, one of the most consistent findings in masked priming studies using words as stimuli is that the perceptual overlap between the prime and target does not modulate the size of priming, indicating that priming is driven by representations at the level of abstract letter identity, that is letters that are abstract with regards font, size, and case (i.e., $_A = A = A = a$). Bowers et al. (1998) made an important observation that in contrast to studies using word stimuli, studies using single letters as stimuli failed to find evidence for masked priming of abstract letter identities. Bowers et al. reasoned that if the letter representations are abstract, then the size of identity priming effect should not differ whether the prime and target differing in case are visually similar (e.g., c/C, k/K) or dissimilar (e.g., a/A, b/B). While this was the pattern found with the lexical decision task and noun-verb decision task for words made up of visually similar prime-target pairs (e.g., kiss–KISS) and dissimilar pairs (e.g., edge–EDGE), for individual letter stimuli used in an alphabet decision task and consonant-vowel decision task, priming effects were found only for visually similar pairs. A similar interaction between priming and visual

similarity of prime-target pairs was reported by Arguin and Bub (1995) and Ziegler et al. (2000) in an alphabet decision task. Bowers et al. (1998) concluded that from these data that "abstract letter codes and abstract word codes exist in the orthographic system but for some reason, only orthographic word codes support priming" (p. 1718).

Kinoshita and Kaplan (2008) noted that in studies using single letters as stimuli, the stimulus set was necessarily small (Bowers et al., 1998, used eight visually similar pairs and eight dissimilar pairs), and the letters were used repeatedly both as primes and targets. Thus, just as Damian (2001) suggested with regards the "bigger-than-5?" task with numbers as stimuli, the priming effect observed in these studies may have reflected stimulus-response mapping, based on a partial analysis of the visual features of the prime. To circumvent this response strategy, Kinoshita and Kaplan used the cross-case same-different task. Here, the referent is always in the opposite case to the target, and subjects are asked to decide whether the target is the same letter as the referent, ignoring case. Kinoshita and Kaplan reasoned that here because a letter can be used both in the SAME and DIFFERENT conditions (e.g., SAME: referent = A, target = a; DIFFERENT: referent = B, target = a) equally often, stimulus-response mapping cannot be learned. Because the decision requires abstract letter identity and not physical identity, the decision supporting priming in this task was assumed to be based on abstract letter representations (e.g., one that corresponds to both uppercase A and lowercase a). In line with this assumption, the results showed robust priming effects which were equal in magnitude for the prime-target letter pairs which were visually similar (e.g., c-C, x-X) and visually dissimilar pairs (e.g., a-A, b-B).

To sum up, as summarized in a review of letter perception by Grainger et al. (2008), in the binary categorization tasks (alphabet decision, consonant-vowel decision), letter priming effects are largely driven by the visual similarity of prime-target pairs (for priming effects found with the letter naming task, see Is the Assembly of Phonology Serial? Onset Priming Effect in Reading Aloud). This pattern could not been taken as evidence for priming of abstract letter identities. Only in the cross-case same-different match task, robust priming effects which were insensitive to the visual similarity could be demonstrated. These results highlight the usefulness of task analysis in guiding the design of masked priming experiment.

IS THE ASSEMBLY OF PHONOLOGY SERIAL? ONSET PRIMING EFFECT IN READING ALOUD

Although the Bayesian Reader is not a model of reading aloud, the pattern of data seen in reading aloud should still be modulated by the goal of the task. In reading aloud, the goal is to generate a speech response, and to initiate articulation as quickly as possible while minimizing errors. One feature of the masked priming effect in this task is that it is highly sensitive to the overlap in phonemic onset. Forster and Davis (1991) were the first to note that relative to the all-letter-different unrelated control (e.g., fame-SINK), overlap in the onset alone (e.g., same-SINK) facilitates the naming of the target. This onset priming effect has been replicated in a number of languages that use alphabetic scripts

(in English, Kinoshita, 2000; Kinoshita and Woollams, 2002; in Dutch, Schiller, 2004; in French, Grainger and Ferrand, 1996; in Spanish, Dimitropoulou et al., 2010; in Korean alphabetic Hangul, Kim and Davis, 2002) but not in syllabic scripts like Korean Hanja (e.g., Kim and Davis, 2002) or in mora-based Japanese (Verdonschot et al., 2011). The onset priming effect has been found with the naming of single letter targets (Bowers et al., 1998) and picture targets (Schiller, 2008), so it is not specific to reading. Taken together with the fact that it is absent in the lexical decision task (Forster and Davis, 1991, Experiment 5; Carreiras et al., 2005, Experiment 1; Grainger and Ferrand, 1996, Experiment 4; Kim and Davis, 2002, Experiment 1b), the results indicate that the onset priming effect reflects the task goal of the need to generate a speech output for the target (Grainger and Ferrand, 1996; Kinoshita and Woollams, 2002; Kinoshita, 2003; Carreiras et al., 2005).

The onset priming effect has important methodological implications. The facilitation due to the mere overlap of onset in the naming task can be sizeable (up to about 30 ms in Grainger and Ferrand, 1996), and it can complicate the interpretation of priming effects in the read-aloud task. An example of this is seen in studies that investigated identity priming of abstract letter identities using the letter naming task (Arguin and Bub, 1995; Bowers et al., 1998). In these studies the size of identity priming effect for prime-target pairs that were visually similar across case (e.g., c-C, k/K) and dissimilar (e.g., a-A, b/B), which is consistent with the idea that there is an abstract letter identity which is invariant across shape and case. Bowers et al. however noted that this pattern was not found in other tasks such as the alphabet decision task and consonant-vowel decision task (as discussed above), and followed up the locus of the identity priming effect in the letter naming task by comparing it to priming produced by word homophone primes (e.g., sea-C, cue-Q), and phonologically similar letter primes (e.g., i-Y). Homophone primes produced facilitation that was as large as the identity primes, whereas the phonological primes produced little facilitation. These results are readily interpretable from the perspective that the priming effects in the naming task are mainly driven by the overlap in the phonemic onset between the prime and the target.

The task dependence of the onset priming effect also has important implications for theory development. Because reading aloud necessarily requires the generation of phonology, the naming task has been a task of choice for researchers interested in the role of phonology in word recognition (e.g., Plaut et al., 1996; Coltheart et al., 2001; Perry et al., 2007, 2010). To date, two computational models have provided accounts of the onset priming effect. According to the DRC model (Coltheart et al., 2001, for the most recent version with simulations of the onset priming effect, see Mousikou et al., 2010), the onset priming effect reflects a serial, left-to-right letter-to-phoneme mapping process implemented in the non-lexical route. The CDP+ (and CDP++) model (Perry et al., 2007, 2010), on the other hand, suggests the locus of the effect is the graphemic parsing process within the sublexical route. A sequence of letters is segmented into graphemes (which can include multi-letter graphemes such as SH, TCH) which are then placed into the Onset, Vowel, and Coda slots, and this parsing process is assumed to occur from left to right across the letter

sequence[7]. Both models account well for the serial, left-to-right nature of the effect (see Kinoshita, 2000; Montant and Ziegler, 2001; Malouf and Kinoshita, 2007): The phonemic overlap has little benefit if it is in the latter part of the stimuli but the onset differs (e.g., suf-SIB < muf-SIB but mub-SIB = muf-SIB; noon-MOON = need-MOON). However, both the DRC and CDP+ models lack an independent motivation (other than to account for the data) as to why the sublexical letter-to-phoneme mapping process or the sublexical graphemic parsing process operates in the left-to-right fashion. The models also have a difficulty accommodating the fact that in a task that does not require a speech output, the effect of sublexical phonology is not necessarily left-to-right. Recently, Kinoshita and Norris (2012) reported that pseudohomophone primes (e.g., cymptom-SYMPTOM, frajile-FRAGILE) produced greater priming effects than orthographic control primes (e.g., lymptom-SYMPTOM, franile-FRAGILE) and that the benefit did not differ for initial (e.g., cymptom) and medial (e.g., frajile) positions. Such a finding suggests that the phonology for the pseudohomophone was generated across the letter string in parallel, rather than serially, from left to right.

Recognizing the task-dependent nature of onset priming effect (that it occurs only when speech output needs to be generated) provides a rationale for proposing a different locus for a left-to-right process. In the speech production literature, there is much evidence for a serial, segment-to-frame association process (e.g., Levelt et al., 1999). In picture naming, the word-form phonology is retrieved as a whole, then (re)syllabified into metrical frames with stress markings. The phoneme segments are then slotted into the metrical frames, and this process is assumed to be left-to-right. The masked onset priming effect could have its origin in this frame-to-segment association process when preparing a speech response to the target. In line with the view, Roelofs (2004) showed that the serial, left-to-right process is observed both when naming objects and reading their printed names, and suggested that "the observed seriality is due to phonological encoding mechanisms shared by naming and reading rather than a grapheme-to-phoneme conversion in oral reading" (p. 221).

IS MASKED PRIMING STRATEGIC?
Before concluding this review, one comment is in order regarding the question of whether masked priming effect is "strategic." One of the major appeals of the masked priming procedure is that because subjects have no awareness of the prime, any priming effects should be immune to the influence of conscious strategies (e.g., Forster et al., 2003)[8]. This related to the belief that an "unconscious" prime will tap into automatic processes and can therefore be used to identify the obligatory representations and processes that support reading. We note that our claim that masked priming effects are driven by the goal of the task does not entail the

---

[7]The fact that unpronounceable consonant string primes (e.g., cdkm-CARO) do not produce onset priming effect (Dimitropoulou et al., 2010) is problematic for the DRC, but can be accounted for by the CDP+ model by assuming that such strings cannot be parsed into graphemes.

[8]An example can be seen in De Groot's (1983) investigation of semantic priming effects in a lexical decision task. By masking the prime, she successfully precluded subjects from using the "post-access coherence check," which creates a bias toward responding "non-word" when the prime and target are semantically unrelated.

view that masked priming is strategic, in the sense that subjects can choose how to make use of the prime-target relationship to facilitate responding to the target in a way that best suits the specific experimental context. Subjects "apply the task instruction to the prime" when the prime is masked because they are unaware that the prime and target are separate perceptual events. On this view, the relationship between the prime and target is necessarily veiled from awareness, and hence cannot be used strategically to facilitate responding to the target.

In apparent contradiction of this view, Bodner and Masson (2001, 2003, 2004) have reported finding in a number of studies that the proportion of related trials with masked primes modulated the size of priming effects: The size of masked priming effect is larger in a block containing a high proportion (0.8) of related trials and a low proportion (0.2) of unrelated trials relative to a block containing the opposite mix. This was found with the lexical decision task (Bodner and Masson, 2001; Bodner and Stalinski, 2008; note however that the effect of relatedness proportion in this task was variable across individual experiments), the read-aloud task (Bodner and Masson, 2004), and the parity judgment task (deciding whether a number is odd or even, Bodner and Dypvik, 2005). The interpretation offered by Bodner and Masson of these effects of relatedness proportion – which they referred to as the "prime validity effect" – is that "the processing operations applied to the prime to identity and interpret it form a new memory representation," and this memory episode can be recruited, without awareness, to assist with processing of a subsequent target. The degree of recruitment is modulated as a function of list context, with "a context containing a high proportion of task-useful primes" cuing the cognitive system to increase prime recruitment to facilitate target processing (Bodner et al., 2006, p. 1299). In other words, Bodner and Masson suggest that the recruitment of the prime, even when it is masked and not consciously available, is under strategic control, and is a function of the list-wide utility of the prime-target relationship.

While we agree broadly with the view that masked priming reflects the overlap of processing operations applied to the prime and the target, we do not believe that this implies that a new memory episode is established for masked primes. In fact, our view is that when the prime is masked, subjects are unaware that the prime and target are distinct perceptual events[9]. Consequently, we do not believe that the cognitive system is able to modulate the impact of the masked prime strategically as a function of the list-wide utility of the prime-target relationship – note that relatedness proportion is not necessarily the same as list-wide prime utility. For example, in a parity (odd-even) decision task, in a block containing a low proportion of category-congruent trials, the prime has high utility, as it predicts the opposite response to the prime (e.g., if the prime is odd, the target is likely to be even, and vice versa). Consistent with our view, the list-wide predictability of the target response from the prime modulates the size of priming when the prime is visible, but not when the prime is masked. In a parity decision task, when the relatedness proportion is 0.5 [i.e., when the

prime is an odd (even) number, the target is also odd (even) on half of the trials and even (odd) on the other half], the list-wide prime utility is zero, as the target parity cannot be predicted from the prime parity: Kinoshita et al. (2011) showed that in this condition there is no effect of parity congruence if the prime was visible, but there is a positive effect of congruence if the prime was masked. In a similar vein, in the same-different task, Kinoshita and Norris (2009) showed that the prime-target response contingency had a large impact on the size of priming when the prime was visible, but little impact when the prime was masked. With visible primes, the priming effect was large when the response to the target was the same as the prime on 75% of the trials (predictable contingency), and reduced to a negligible level when the they were the same on 50% of the trials (zero-contingency); in contrast, with masked primes, the priming effects was equally large and robust in the predictable- and zero-contingency conditions.

Given these dissociations between effects of relatedness proportion with the visible and masked primes, with only the former being a function of list-wide prime utility, how could the effects observed by Bodner and Masson with the masked primes be explained? Kinoshita et al. (2008, 2011) have presented an alternative account – termed the Adaptation to the statistics of the environment (ASE) – which explains the effect in terms of the adaptation of the response initiation process to the history of trial difficulty. The account is based on the assumption that in RT tasks, subjects attempt to meet the instruction to "respond as quickly as possible without making too many errors" by estimating the optimal point to initiate responding that minimizes the total cost of responding too early and risking an error, and delaying the response unnecessarily. The history of trial difficulty, in particular, that of immediately preceding trials, is used in conjunction with the evidence accumulated in the current trial to estimate the optimal point to respond. Consistent with this, RT of previous trial is positively correlated with the current trial RT (independent of post-error slowdown). When the condition is such that the easy trials show greater sensitivity to the previous trial RT than the hard trials, the relatedness proportion effect with masked primes falls out of the ASE model from the fact that related trials are "easy" trials (and unrelated trials are "hard" trials). Although the ASE model is silent with regards why easy trials should be more sensitive to the previous trial RT, this pattern is expected to hold in general from the assumption that when trials slow down by a fixed amount of time, the benefit in the reduction in error rate would be greater for the easy trials than for the hard trials due to its greater rate of evidence accumulation. Kinoshita et al. (2011) reanalyzed the masked priming experiment showing an apparent relatedness proportion effect (Experiment 3) as a function of previous trial RT, and found greater sensitivity of related trials than the unrelated trials to the previous RT, consistent with the ASE. Furthermore, they showed that the adaptation to the list-wide difficulty (as determined by the proportion of easy vs. hard trials) is a noisy process requiring many trials (over 300 trials) – many more than are standard in masked priming experiments.

In sum, the pattern of relatedness proportion effects are different with visible and masked primes, and the latter – referred to

---

[9]For evidence that an explicit episodic record of the prime capable of supporting long-term priming (priming spanning several intervening trials) is formed only with visible primes, see Humphreys et al. (1988).

as the prime validity effect and interpreted by Bodner and Masson as the main evidence for the view that masked priming is strategic – has an alternative explanation. We take this to argue that the concern that masked priming effects are strategic – in the sense that the impact of the prime on the target processing can be modulated as a function of the list-wide prime utility – is unwarranted.

CONCLUSION

From the perspective that different tasks should all tap into the output of a fixed lexical processing system, the diverse pattern of results found with different masked priming tasks makes little sense. Indeed, one might be tempted to conclude that because different tasks behave differently, none of them is particularly useful. After all, which task provides the true measure of lexical processing? In contrast, the view that we have presented is much more optimistic about the value of masked priming data. In fact, it implies that the different tasks all have something valuable to tell us. Different tasks tap into different facets of the word recognition process in ways that are quite systematic and lawful. We have argued that the variation in the pattern of priming is what should be expected if perception can be characterized as approximating optimal Bayesian decision making operating by sampling evidence accumulated from the perceptual input. As we explained earlier, we also need to assume that in masked priming the evidence from the prime and target is integrated in reaching a decision. That is, the prime and target are not treated as separate perceptual events. We illustrated this with the urn analogy. In order to make optimal decisions there is no need to know which samples or balls come from the "prime" and which from the target. Each sample, or ball, provides an independent piece of evidence that can be used in making the decision. This verbal description of the optimal decision making process is supported by simulations (see, e.g., Norris and Kinoshita, 2008; Norris et al., 2010).

Much of the data we have reviewed here focuses on the differences between lexical decision and the same-different task. This contrast is particularly illuminating because it allows us to investigate the source of these effects – whether it is in the orthographic representations, or in the lexicon. For example, the fact that TL priming effects are found with non-words (or even symbols) in the same-different task has told us that the effect reflects the perceptual uncertainty of positions of individual objects in a string. It is not surprising therefore that TL priming effects are found universally across languages in the same-different task. In contrast, in the lexical decision task, TL priming effects are modulated by language, and by morphological structure. From the perspective that priming effects reflect automatic activation of representations with a fixed property, that masked priming effects are task-dependent may be puzzling. However, from the view that masked priming reflects the accumulation of task-specific evidence contributed by the prime, that the pattern of task-dependent masked priming effects mirrors the task goal is exactly what is expected, given that the evidence needs to be accumulated is determined by the task. This framework provides a useful guide to interpreting what information is available to the readers in the first 250 ms of visual word recognition.

REFERENCES

Ansorge, U., Kiefer, M., Khalid, S., Grassl, S., and Konig, P. (2010). Testing the theory of embodied cognition with subliminal words. *Cognition* 116, 303–320.

Arguin, M., and Bub, D. (1995). Priming and response selection processes in letter classification and identification tasks. *J. Exp. Psychol. Hum. Mem. Percept. Perform.* 21, 1199–1219.

Bodner, G. E., and Dypvik, A. T. (2005). Masked priming of number judgments depends on prime validity and task. *Mem. Cognit.* 33, 29–47.

Bodner, G. E., and Masson, M. E. J. (1997). Masked repetition priming of words and nonwords: evidence for a nonlexical basis for priming. *J. Mem. Lang.* 37, 268–293.

Bodner, G. E., and Masson, M. E. J. (2001). Prime validity affects masked repetition priming: evidence for an episodic resource account of priming. *J. Mem. Lang.* 45, 616–647.

Bodner, G. E., and Masson, M. E. J. (2003). Beyond spreading activation: an influence of relatedness proportion on masked semantic priming. *Psychon. Bull. Rev.* 10, 645–652.

Bodner, G. E., and Masson, M. E. J. (2004). Beyond binary judgments: prime validity modulates masked repetition priming in the naming task. *Mem. Cognit.* 32, 1–11.

Bodner, G. E., Masson, M. E. J., and Richard, N. T. (2006). Repetition proportion biases masked priming of lexical decisions. *Mem. Cognit.* 34, 1298–1311.

Bodner, G. E., and Stalinski, S. M. (2008). Masked repetition priming and proportion effects under cognitive load. *Can. J. Exp. Psychol.* 62, 127–131.

Bowers, J. S. (2010). Does masked and unmasked priming reflect Bayesian inference as implemented in the Bayesian Reader? *Eur. J. Cogn. Psychol.* 22, 779–797.

Bowers, J. S., Vigliocco, G., and Haan, R. (1998). Orthographic, phonological, and articulatory contributions to masked letter and word priming. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 1705–1719.

Bueno, S., and Frenck-Mestre, C. (1999). Semantic features and semantic categories: differences in rapid activation of the lexicon. *Brain Lang.* 68, 199–204.

Bueno, S., and Frenck-Mestre, C. (2002). Rapid activation of the lexicon: further investigation with behavioral and computational results. *Brain Lang.* 81, 120–130.

Bueno, S., and Frenck-Mestre, C. (2008). The activation of semantic memory: effects of prime exposure, prime-target relationship and task demands. *Mem. Cognit.* 36, 882–898.

Carreiras, M., Ferrand, L., Grainger, J., and Perea, M. (2005). Sequential effects of phonological priming in visual word recognition. *Psychol. Sci.* 16, 585–589.

Carreiras, M., Perea, M., and Grainger, J. (1997). Effects of orthographic neighborhood in visual word recognition: cross-task comparisons. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 857–871.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. C. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychol. Rev.* 108, 204–256.

Damian, M. (2001). Congruity effects evoked by subliminally presented primes: automaticity rather than semantic processing. *J. Exp. Psychol. Hum. Mem. Percept. Perform.* 27, 154–165.

Davis, C., and Forster, K. I. (1994). Masked orthographic priming: the effect of prime-target legibility. *Q. J. Exp. Psychol. A* 47, 673–697.

De Groot, A. M. B. (1983). The range of automatic spreading activation in word priming. *J. Verbal Learn. Verbal Behav.* 22, 417–436.

De Jong, R., Liang, C.-C., and Lauber, E. (1994). Conditional and unconditional automaticity: a dual-process model of effects of spatial stimulus-response correspondence. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 731–750.

Dehaene, S., Naccache, L., Le Clec, H. G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van der Moortele, P. F., and Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature* 395, 597–600.

Dimitropoulou, M., Dunabeitia, J. A., and Carreiras, M. (2010). Influence of prime lexicality, frequency, and pronounceability on the masked onset priming effect. *Q. J. Exp. Psychol.* 63, 1813–1837.

Duñabeitia, J. A., Kinoshita, S., Carreiras, M., and Norris, D. (2010). Is morph-orthographic decomposition purely orthographic? Evidence from masked priming in the same-different task. *Lang. Cogn. Process.* 26, 509–529.

Duñabeitia, J. A., Perea, M., and Carreiras, M. (2007). Do transposed-letter similarity effects

occur at a morpheme level? Evidence for ortho-morphological decomposition. *Cognition* 105, 691–703.

Evett, L. J., and Humphreys, G. W. (1981). The use of abstract graphemic information in lexical access. *Q. J. Exp. Psychol.* 33, 325–350.

Finkbeiner, M., and Friedman, J. (2011). The flexibility of nonconsciously (deployed) cognitive processes: evidence from masked congruence priming. *PLoS ONE* 6, e17095. doi:10.1371/journal.pone.0017095

Forster, K. I. (1985). Lexical acquisition and the modular lexicon. *Lang. Cogn. Process.* 1, 87–108.

Forster, K. I. (1998). The pros and cons of masked priming. *J. Psycholinguist. Res.* 27, 203–233.

Forster, K. I. (2004). Category size effects revisited: frequency and masked priming effects in semantic categorization. *Brain Lang.* 90, 276–286.

Forster, K. I., and Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 680–698.

Forster, K. I., and Davis, C. (1991). The density constraint on form-priming in the naming task: interference effects from a masked prime. *J. Mem. Lang.* 30, 1–25.

Forster, K. I., Davis, C., Schoknecht, C., and Carter, R. (1987). Masked priming with graphemically related forms: repetition or partial activation? *Q. J. Exp. Psychol.* 39, 211–251.

Forster, K. I., and Hector, J. (2002). Cascaded and noncascaded models of lexical and semantic processing: the turple effect. *Mem. Cognit.* 30, 1106–1117.

Forster, K. I., Mohan, K., and Hector, J. (2003). "The mechanics of masked priming," in *Masked Priming: The State of the Art*, eds S. Kinoshita and S. J. Lupker (Hove: Psychology Press), 3–37.

Frost, R. (2009). "Reading in Hebrew vs. reading in English: is there a qualitative difference?" in *How Children Learn to Read: Current Issues and New Directions in the Integration of Cognition, Neurobiology, and Genetics of Reading and Dyslexia Research and Practice*, eds K. Pugh and P. McCradle (New York: Psychology Press), 235–254.

Frost, R., Forster, K. I., and Deutsch, A. (1997). What can we learn from the morphology of Hebrew? A masked priming investigation of morphological representation. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 829–856.

Frost, R., Kugler, T., Deutsch, A., and Forster, K. I. (2005). Orthographic structure versus morphological structure: principles of lexical organization in a given language. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 1293–1326.

García-Orza, J., Perea, M., and Muñoz, S. (2010). Are transposition effects specific to letters? *Q. J. Exp. Psychol.* 63, 1603–1618.

Gomez, P., Ratcliff, R., and Perea, M. (2008). The overlap model: a model of letter position coding. *Psychol. Rev.* 115, 577–600.

Grainger, J. (2008). Cracking the orthographic code: an introduction. *Lang. Cogn. Process.* 23, 1–35.

Grainger, J., and Ferrand, L. (1994). Phonology and orthography in visual word recognition: effects of masked homophone primes. *J. Mem. Lang.* 33, 218–233.

Grainger, J., and Ferrand, L. (1996). Masked orthographic and phonological priming in visual word recognition and naming: cross-task comparisons. *J. Mem. Lang.* 35, 623–647.

Grainger, J., and Holcombe, P. J. (2009). Watching the word go by: on the time course of component processes in visual word recognition. *Lang. Linguist. Compass* 3/1, 128–156.

Grainger, J., Rey, A., and Dufau, S. (2008). Letter perception: from pixels to pandemonium. *Trends Cogn. Sci. (Regul. Ed.)* 12, 381–387.

Grainger, J., and van Heuven, W. (2003). "Modeling letter position coding in printed word perception," in *The Mental Lexicon*, ed. P. Bonin (New York: Nova Science), 1–23.

Grainger, J., and Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Front. Psychol.* 2:54. doi:10.3389/fpsyg.2011.00054

Grondin, R., Lupker, S. J., and McRae, K. (2009). Shared features dominate semantic richness effects for concrete concepts. *J. Mem. Lang.* 60, 1–19.

Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: a survey and appraisal. *Behav. Brain Sci.* 9, 1–23.

Hommel, B. (1994). Spontaneous decay of response-code activation. *Psychol. Res.* 56, 261–268.

Horner, A. J., and Henson, R. N. (2009). Bindings between stimuli and multiple response codes dominate long-lag repetition priming in speeded classification tasks. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 757–779.

Humphreys, G. W., Besner, D., and Quinlan, P. T. (1988). Event perception and the word repetition effect. *J. Exp. Psychol. Gen.* 117, 51–67.

Kim, J., and Davis, C. (2002). Using Korean to investigate phonological priming effects without the influence of orthography. *Lang. Cogn. Process.* 17, 569–591.

Kinoshita, S. (2000). The left-to-right nature of the masked onset effect in naming. *Psychon. Bull. Rev.* 7, 133–141.

Kinoshita, S. (2003). "The nature of masked onset priming effects in naming," in *Masked Priming: The State of the Art*, eds S. Kinoshita and S. J. Lupker (Hove: Psychology Press), 223–238.

Kinoshita, S., and Hunt, L. (2008). RT distribution analysis of category congruence effects with masked primes. *Mem. Cognit.* 36, 1324–1334.

Kinoshita, S., Forster, K. I., and Mozer, M. C. (2008). Unconscious cognition isn't that smart: modulation of masked repetition effect in the naming task. *Cognition* 107, 623–649.

Kinoshita, S., Mozer, M. C., and Forster, K. I. (2011). Dynamic adaptation to history of trial difficulty explains the effect of congruency proportion on masked priming. *J. Exp. Psychol. Gen.* 140, 622–636.

Kinoshita, S., and Norris, D. (2009). Transposed-letter priming of pre-lexical orthographic representations. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 1–18.

Kinoshita, S., and Norris, D. (2011). Does the familiarity bias hypothesis explain why there is no masked priming for "NO" decisions? *Mem. Cognit.* 39, 319–334.

Kinoshita, S., and Norris, D. (2012). Pseudohomophone priming in lexical decision is not fragile in a sparse lexical neighborhood. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 764–775.

Kinoshita, S., Norris, D., and Siegelman, N. (2012). Transposed-letter priming effect in Hebrew in the same-different task. *Q. J. Exp. Psychol.* doi:10.1080/17470218.2012.655749

Kinoshita, S., and Woollams, A. M. (2002). The masked onset priming effect in naming: computation of phonology or speech-planning? *Mem. Cognit.* 30, 237–245.

Kinoshita, S., and Kaplan, L. (2008). Priming of abstract letter identities in the letter match task. *Q. J. Exp. Psychol.* 61, 1873–1885.

Knill, D. C., Kersten, D., and Yuille, A. (1996). "A Bayesian formulation of visual perception," in *Perception as Bayesian Inference*, eds D. C. Knill and W. Richards (Cambridge: Cambridge University Press), 1–21.

Kouider, S., and Dehaene, S. (2007). Levels of processing during non-conscious perception: a critical review of visual masking. *Philos. Trans. R. Soc. Lond. B.* 362, 857–875.

Kunde, W., Kiesel, A., and Hoffmann, J. (2003). Conscious control over the content of unconscious cognition. *Cognition* 88, 223–242.

Landauer, T. K., and Dumais, S. T. (1997). A solution to Plato's problem: the Latent Semantic Analaysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.* 104, 211–240.

Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–38.

Lupker, S. J., Perea, M., and Davis, C. J. (2008). Transposed-letter effects: consonants, vowels, and letter frequency. *Lang. Cogn. Process.* 23, 93–116.

Malouf, T., and Kinoshita, S. (2007). Masked onset priming effect for high-frequency words: further support for the speech-planning account. *Q. J. Exp. Psychol.* 60, 1155–1167.

Marcel, A. J. (1983). Conscious and unconscious perception: experiments on visual masking and word recognition. *Cogn. Psychol.* 15, 197–237.

McRae, K., and Boisvert, S. (1998). Automatic semantic similarity priming. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 558–572.

McRae, K., Cree, G. S., Seidenberg, M. S., and McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behav. Res. Methods Instrum. Comput.* 37, 547–559.

Montant, M., and Ziegler, J. (2001). Can orthographic rimes facilitate naming? *Psychol. Bull. Rev.* 8, 351–356.

Mousikou, P., Coltheart, M., Finkbeiner, M., and Saunders, S. (2010). Can the DRC computational model of reading offer a valid account of the masked onset priming effect? *Q. J. Exp. Psychol.* 63, 984–1003.

Naccache, L., and Dehaene, S. (2001). Unconscious semantic priming extends to novel unseen stimuli. *Cognition* 80, 223–237.

Norris, D. (2006). The Bayesian Reader: explaining word recognition as an

optimal Bayesian decision process. *Psychol. Rev.* 113, 327–357.

Norris, D. (2009). Putting it all together: a unified account of word recognition and reaction-time distributions. *Psychol. Rev.* 116, 207–219.

Norris, D., and Kinoshita, S. (2008). Perception as evidence accumulation and Bayesian inference: insights from masked priming. *J. Exp. Psychol. Gen.* 137, 433–455.

Norris, D., and Kinoshita, S. (2012). Orthographic processing is universal; it's what you do with it that's different: comment on Frost. *Behav. Brain Sci.* (in press).

Norris, D., Kinoshita, S., and van Casteren, M. (2010). A stimulus sampling theory of letter identity and order. *J. Mem. Lang.* 62, 254–271.

Perea, M., and Carreiras, M. (2008). Do orthotactics and phonology constrain the transposed-letter effect? *Lang. Cogn. Process.* 23, 69–92.

Perea, M., and Lupker, S. J. (2003). "Transposed-letter confusability effects in masked form priming," in *Masked Priming: The State of the Art*, eds S. Kinoshita and S. J. Lupker (New York: Psychology Press), 97–120.

Perea, M., and Lupker, S. J. (2004). Can CANISO activate CASINO? Transposed-letter similarity effects with nonadjacent letter positions. *J. Mem. Lang.* 51, 231–246.

Perea, M., and Perez, E. (2009). Beyond alphabetic orthographies: the role of form and phonology in transposition effects in Katakana. *Lang. Cogn. Process.* 24, 67–88.

Perry, C., Ziegler, J. C., and Zorzi, M. (2007). Nested modeling and strong inference testing in the development of computational theories: the CDP+ model of reading aloud. *Psychol. Rev.* 27, 301–333.

Perry, C., Ziegler, J. C., and Zorzi, M. (2010). Beyond single syllables: large-scale modeling of reading aloud with the connectionist dual process (CDP++) model. *Cogn. Psychol.* 61, 106–151.

Plaut, D., McClelland, J. L., Seidenberg, M. S., and Patterson, K. E. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychol. Rev.* 103, 56–115.

Quinn, W. M., and Kinoshita, S. (2008). Category congruence effect in semantic categorization with masked primes with narrow and broad categories. *J. Mem. Lang.* 58, 286–306.

Rastle, K., Davis, M. H., Marslen-Wilson, W. D., and Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition: a time-course study. *Lang. Cogn. Process.* 15, 507–537.

Rastle, K., Davis, M. H., and New, B. (2004). The broth in my brother's brothel: morpho-orthographic segmentation in visual word recognition. *Psychon. Bull. Rev.* 11, 1090–1098.

Reynvoet, B., Gevers, W., and Caessens, B. (2005). Unconscious primes activate motor codes through semantics. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 991–1000.

Roelofs, A. (2004). Seriality of phonological encoding in naming objects and reading their names. *Mem. Cognit.* 32, 212–222.

Rohde, D. L. T., Gonnerman, L. M., and Plaut, D. C. (2004). *An Improved Method for Deriving Word Meanings from Lexical Co-occurrence Statistics*. Available at: http://dlt4.mit.edu/%7Edr/COALS/Coals.pdf

Rosch, E., and Mervis, C. B. (1975). Family resemblances: studies in the internal structure of categories. *Cogn. Psychol.* 7, 573–605.

Rueckl, J. G., and Rimzhim, A. (2011). On the interaction of letter transpositions and morpheme boundaries. *Lang. Cogn. Process.* 26, 482–508.

Schiller, N. O. (2004). The onset effect in word naming. *J. Mem. Lang.* 50, 477–490.

Schiller, N. O. (2008). The masked onset priming effect in picture naming. *Cognition* 106, 952–962.

Schoonbaert, S., and Grainger, J. (2004). Letter position coding in printed word perception: effects of repeated and transposed letters. *Lang. Cogn. Process.* 19, 333–367.

Velan, H., and Frost, R. (2009). Letter-transposition effects are not universal: the impact of transposing letters in Hebrew. *J. Mem. Lang.* 61, 285–302.

Velan, H., and Frost, R. (2011). Words with and without internal structure: what determines the nature of orthographic and morphological processing? *Cognition* 118, 141–156.

Verdonschot, R. G., La Heij, W., Kiyama, S., Tamaoka, K., Kinoshita, S., and Schiller, N. O. (2011). The functional unit of Japanese word naming: evidence from masked priming. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 1458–1471.

Whitney, C., and Cornelissen, P. (2005). Letter-position coding and dyslexia. *J. Res. Read.* 28, 274–301.

Whitney, C., and Cornelissen, P. (2008). SERIOL reading. *Lang. Cogn. Process.* 23, 143–164.

Ziegler, J. C., Ferrand, L., Jacobs, A. M., Rey, A., and Grainger, J. (2000). Visual and phonological codes in letter and word recognition: evidence from incremental priming. *Q. J. Exp. Psychol.* 53A, 671–692.

Bilingual word recognition in a sentence context

Eva Van Assche*, Wouter Duyck and Robert J. Hartsuiker

Department of Experimental Psychology, Ghent University, Ghent, Belgium

This article provides an overview of bilingualism research on visual word recognition in isolation and in sentence context. Many studies investigating the processing of words out-of-context have shown that lexical representations from both languages are activated when reading in one language (language-non-selective lexical access). A newly developed research line asks whether language-non-selective access generalizes to word recognition in sentence contexts, providing a language cue and/or semantic constraint information for upcoming words. Recent studies suggest that the language of the preceding words is insufficient to restrict lexical access to words of the target language, even when reading in the native language. Eye tracking studies revealing the time course of word activation further showed that semantic constraint does not restrict language-non-selective access at early reading stages, but there is evidence that it has a relatively late effect. The theoretical implications for theories of bilingual word recognition are discussed in light of the Bilingual Interactive Activation+ model (Dijkstra and van Heuven, 2002).

Keywords: bilingualism, visual word recognition, sentence processing, eye tracking

INTRODUCTION

The task of reading is omnipresent in everyday life. People can read in their native language without apparent difficulty. It takes a skilled reader only a few hundred milliseconds to recognize a word. This is extremely fast given that the mental lexicon contains tens of thousands of words from among which the correct word has to be identified. Furthermore, many people have knowledge of more than one language. Recently, the process of reading by bilinguals has increasingly attracted the attention of the scientific community. Research on bilingualism includes issues such as: Are the words of one language activated when reading in the other? Are there any differences in cross-lingual activation between words presented in isolation and words presented in sentence context? What is the time course of cross-lingual activation and what factors may modulate this activation process? The most intuitively appealing idea would probably be that bilinguals have two separate lexicons that can be accessed selectively so that each language functions independently of the other. After all, most bilinguals can speak and read in each language without too many intrusions or errors (Poulisse and Bongaerts, 1994). However, in the last decade, more and more researchers have come to realize that "the bilingual does not equal the sum of two monolinguals" (Grosjean, 1989). Bilinguals do not recognize words in exactly the same way as monolinguals. It became clear that the two languages interact with each other when bilinguals are processing words in one language (e.g., Dijkstra et al., 1999; van Hell and Dijkstra, 2002; Duyck, 2005; Van Assche et al., 2009).

In this review, we focus on visual word recognition research in bilinguals and the lexical organization of the bilingual language system. First, we briefly summarize the main experimental findings in isolated word recognition. Then, we present the recently developed research line on bilingual word recognition in sentence contexts. Next, we discuss the most influential theoretical accounts on the lexical organization of the bilingual language system and

we present the theoretical implications of the research presented in this review for theories of bilingual word recognition, in particular the Bilingual Interactive Activation+ (BIA+) model (Dijkstra and van Heuven, 2002). Finally, we discuss future work directions for the study of the bilingual language system.

BILINGUAL VISUAL WORD RECOGNITION IN ISOLATION

An important issue in bilingualism research concerns the question of whether reading a word activates lexical representations in both languages, or in only the contextually relevant (target) language. Most of the research on this issue has focused on the cross-lingual interactions between orthographic representations. Evidence has accumulated that representations from both languages are activated in parallel (e.g., van Heuven et al., 1998; Dijkstra et al., 1999; van Hell and Dijkstra, 2002; Duyck, 2005; Van Assche et al., 2009). To our knowledge, Caramazza and Brones (1979) were the first to find evidence for the currently dominant theory that lexical representations in both languages are activated when reading in one language (i.e., language-non-selective access). In this study, Spanish-English bilinguals performed a lexical decision task (in which participants decide whether a string of letters is a word or a non-word) in their second language (L2). They found that bilinguals responded more quickly to cognates (i.e., translation equivalents with full or partial form overlap, e.g., Spanish-English *piano–piano*, *eco-echo*) than to matched non-cognates. This cognate facilitation effect is commonly attributed to the fact that a L2 cognate word also activates the L1 lexical representation of the cognate, mapped onto the same semantic representation, to a certain degree (see Dijkstra and van Heuven, 2002; Dijkstra et al., 2010, for more information on the representational structure of cognates). The cross-lingual activation of these representations speeds up the recognition of cognates compared to non-cognates.

Later, several studies have replicated this cognate facilitation effect in L2 for words presented out-of-context (e.g., Dijkstra

et al., 1999; Lemhöfer and Dijkstra, 2004; Duyck et al., 2007). In Lemhöfer et al. (2004), this effect is shown to even accumulate over languages. Lemhöfer et al. tested Dutch-English-German trilinguals performing a German (L3) lexical decision task and reported faster responses for L1-L2-L3 cognates than for L1-L3 cognates. Surprisingly, cognate facilitation even occurs when bilinguals perform a lexical decision task in their native and dominant language (L1; e.g., van Hell and Dijkstra, 2002; Van Assche et al., 2009). van Hell and Dijkstra (2002) investigated the influence of L2 and L3 on reading in the L1. Two groups of Dutch-English-French trilinguals with low and high proficiency in French performed a Dutch lexical decision task. The critical stimuli were L1-L2 cognates and L1-L3 cognates. For both groups of trilinguals, results yielded faster lexical decisions for L1-L2 cognates than for noncognates. However, only the trilinguals who were highly proficient in French showed cognate facilitation for L1-L3 cognates. These results provide strong evidence for language-non-selective access in the bilingual lexicon because the non-dominant languages exert an influence on the dominant L1. A minimal proficiency in the non-dominant language seems necessary however in order to obtain cross-lingual activation effects.

Other evidence for language-non-selective access comes from studies investigating the recognition of interlingual homographs (i.e., words that have the same orthographic form in both languages but that have a different meaning, e.g., Dutch-English *room*, meaning *cream* in Dutch; e.g., Dijkstra et al., 1998, 1999, 2000; Kerkhofs et al., 2006). In Dijkstra et al. (2000), Dutch-English bilinguals performed a go/no-go task in which they had to press a button only if the presented word was an English word. Reaction times for interlingual homographs were slower than for control words. Apparently, the Dutch reading of the homograph was activated and interfered with the recognition of the English word. The size and direction of this interlingual homograph effect can be modulated by task requirements, language intermixing and relative frequency of the homograph in the two languages. For instance, Dijkstra et al. (1998) observed facilitation for interlingual homographs when Dutch-English bilinguals performed a generalized lexical decision task (giving a yes-response when a word of either language was presented). It seems that participants responded as soon as one reading of the homograph was available, or even on the basis of the summed activity in the bilingual language system generated by the two readings of the homograph.

In addition to these cognate and homograph studies, there is further evidence for cross-lingual activation of lexical representations from neighborhood studies (e.g., van Heuven et al., 1998) and masked priming studies (e.g., Bijeljac-Babic et al., 1997). Monolingual studies have shown that word processing is influenced by the number (density) of orthographic neighbors (i.e., words differing by a single letter from the target, Coltheart et al., 1977; e.g., *house* is an intralingual neighbor of *mouse*) and their frequency (e.g., Grainger et al., 1989; Segui and Grainger, 1990). van Heuven et al. (1998) examined the claim of an integrated lexicon and language-non-selective lexical access by investigating whether word neighbors in both languages [e.g., *book* is a cross-lingual neighbor of the Dutch word *rook* (smoke)] affect word recognition. The results from Dutch-English bilinguals' performance

on two progressive demasking tasks showed that a higher number of Dutch word neighbors resulted in slower responses to English target words. This inhibitory effect of the number of neighbors was also present for word identification in the L1: Dutch-English bilinguals needed more time to identify a Dutch word with many English neighbors than a Dutch word with few English neighbors. van Heuven et al. also tested whether these results generalized to different task situations. As in the progressive demasking experiments, results of a generalized lexical decision task showed significant inhibition from Dutch neighbors on English word recognition. However, there was no effect of English neighbors on Dutch words. This suggests that the strength of neighborhood density effects is task dependent. An English lexical decision task with Dutch-English bilinguals showed an inhibitory effect from Dutch neighborhood on lexical decision times. This factor did not influence the responses of English monolinguals, ensuring that this effect was not due to any uncontrolled stimulus characteristics.

Other evidence for neighborhood density effects between languages comes from Bijeljac-Babic et al. (1997). They used the masked priming paradigm to test whether the inhibitory priming effect of orthographic neighbors on visual word recognition in monolinguals (e.g., Segui and Grainger, 1990) generalized to bilinguals. In Experiment 1, highly proficient French-English bilinguals made lexical decisions to L2 target words or non-words preceded by words from the same or a different language. Within each prime language condition, target words were preceded by either orthographically related primes (e.g., *less-LOSS*; *joie-JOIN*) or unrelated primes (*sore-LOSS*; *acte-JOIN*). When prime and target were from the same language, lexical decisions were slower after related primes than unrelated primes. More importantly, the same inhibition effect was found when prime and target were from different languages, providing evidence for language-non-selective access to the bilingual lexicon. In Experiment 2, the target language was changed and a different set of prime-target stimuli was tested in balanced and unbalanced bilinguals and in French monolinguals. The within-language effect was present in all three groups, while the between-language effect was larger for the balanced than for the unbalanced bilinguals. The French monolinguals showed no effect of English word primes. These cross-lingual activation effects from (masked) neighborhood studies strongly support the hypothesis of language-non-selective access to an integrated lexicon, even when subjects are performing a monolingual task. Note that converging evidence for language-non-selective access has also been obtained in other domains such as auditory word recognition (e.g., Spivey and Marian, 1999; Weber and Cutler, 2004; Lagrou et al., 2011) and word production (e.g., Costa et al., 1999).

We can conclude that there is a now a consensus in the bilingual literature about language-non-selective access of words in the two languages. However, in all of the studies discussed above, word recognition was always investigated for words presented out-of-context, using lab tasks (e.g., lexical decision) as operationalizations of reading. One of the key research questions for future bilingualism studies is whether these findings on lexical interactions between languages also generalize to word recognition in sentence contexts. The next section discusses the pioneer studies that have recently begun to assess this issue.

BILINGUAL VISUAL WORD RECOGNITION IN SENTENCES

Whereas most studies on lexical autonomy have investigated the recognition of isolated words, word recognition rarely occurs out-of-context. People usually read words embedded in meaningful sentences (e.g., in a newspaper article). The ecological validity of the studies on isolated word recognition can be put to the test by examining word recognition in sentences. The processing of words in isolation may differ in important ways from word processing during sentence reading. For instance, it is possible that the presentation of words in a sentence context restricts lexical activation to words of the target language only. This would actually be quite an efficient strategy to speed up word recognition, because it reduces the number of lexical candidates. And, indeed, in the monolingual domain, it has been shown that semantic and syntactic restrictions imposed by a sentence are used to speed up recognition of upcoming words (e.g., Schwanenflugel and LaCount, 1988). For instance, many studies have shown that context modulates lexical access for ambiguous words (e.g., *bank* as a riverside or a financial institution; e.g., Binder and Rayner, 1998). Also, previous research has shown that words embedded in a semantically constraining sentence context are processed faster than words embedded in a neutral sentence context (e.g., Stanovich and West, 1983; Rayner and Well, 1996). These monolingual studies indicate that sentence context can restrict semantic, syntactic, and lexical activation for word appearing later in the sentence.

The question now is whether such sentence context effects in monolinguals are also used by bilinguals to speed up lexical search through representations belonging to two different languages. Although there is one early study of Altarriba et al. (1996) that investigated word recognition in a sentence context for mixed-language sentences, all other studies examining bilingual sentence reading were carried out only very recently (e.g., Elston-Güttler et al., 2005; Schwartz and Kroll, 2006; Duyck et al., 2007; van Hell and de Groot, 2008; Libben and Titone, 2009; Van Assche et al., 2009, 2011; Titone et al., 2011).

L2 PROCESSING

In these studies investigating bilingual sentence reading, the cognate or interlingual homograph effect has often been used as a marker of non-selective activation. In a semantic priming study, Elston-Güttler et al. (2005) showed that cross-lingual activation is very sensitive to the influence of a sentence context and the previous activation state of the two languages. German-English bilinguals were presented with relatively low-constraint sentences in which a homograph (e.g., *The woman gave her friend a pretty GIFT*; *gift* meaning *poison* in German) or a control word was presented at the end (e.g., *The woman gave her friend a pretty SHELL*). The sentence was then replaced by a target word for lexical decision (*poison*). Targets were recognized faster after the related homograph sentence than after the unrelated control sentence, but only in the first block of the experiment and only for participants who saw a German film prior to the experiment, boosting L1 activation. This suggests that the L1 meaning of the homograph was activated while reading L2 sentences, but only after boosting L1 activation and for a limited amount of time because, as Elston-Güttler et al. put it, the bilingual language system quickly "zooms into" the L2 processing situation.

Furthermore, recordings of event-related potentials (ERPs), time-locked to the target word, showed this semantic priming effect in the modulations of the N200 and N400 components. The N200 component in the 150- to 250-ms time window has been linked to word access and/or orthographic processing (e.g., Bentin et al., 1999; but see Connolly and Phillips, 1994, where the N200 has also been linked to phonological processing). Elston-Güttler et al. (2005) suggested a translational word form link between *gift-poison* so that lexical access of the target *poison* is faster after the prime *gift*. The N400 component, present in the time window from 300 to 500 ms, has been linked to semantic integration processes (e.g., Brown and Hagoort, 1993). Target words (*poison*) are easier to integrate and therefore less negative in the N400 amplitude after a related prime (the L1 meaning of the homograph *gift*) than after an unrelated one (*shell*). This study showed that sentence context can prevent the activation of the homograph's non-target language representation and that this effect is very sensitive to task circumstances.

The study of Schwartz and Kroll (2006) tested cognate and homograph effects in Spanish-English bilinguals. They presented target words in low- and high-constraint sentences to investigate how the mere presentation of words in a sentence context, and the semantic constraint it provides, modulates language-non-selective activation in the bilingual lexicon. The words of the sentence were presented using rapid serial visual presentation and the target word (printed in red) had to be named. No homograph effects were found in either low- or high-constraint sentences, but less proficient bilinguals made more naming errors, especially in low-constraint sentences. These results for homographs were somewhat inconclusive and in this respect, it should be noted that results for interlingual homograph effects in isolation (e.g., Dijkstra et al., 2000) were also not always consistent and seem very sensitive to specific characteristics of the task. Therefore, cognate facilitation may be a more reliable marker of cross-lingual activation. Schwartz and Kroll observed cognate facilitation in low-constraint sentences, but not in high-constraint ones. This suggests that the semantic constraint of a sentence may restrict cross-lingual activation effects.

Similar results on cognate effects were obtained by van Hell and de Groot (2008) for Dutch-English bilinguals in an L2 lexical decision task and a translation task in forward (from L1 to L2) or in backward direction (from L2 to L1). Cognate facilitation was shown after the presentation of a low-constraint sentence, but cognate effects were no longer observed in high-constraint sentences in the lexical decision task and strongly diminished in the translation tasks.

In sum, data from studies using lexical decision, naming, or translation tasks suggest that the semantic constraint of a sentence modulates bilingual lexical access, reducing, or nullifying cross-lingual activation effects. However, this is possibly the result of processes occurring after lexical access had taken place. Lexical decision tasks may involve decision-making strategies or postlexical checking strategies. In the same way, naming requires a production component. As a result, these processes might disguise the actual effects reflecting lexical access in bilinguals. It is therefore important to explore this issue using more sensitive measurements such as eye tracking. This method has several important

advantages over lexical decision or naming. First, it allows reading as in everyday life and thereby provides the most natural experimental operationalization of reading. Second, there is no need for any overt response (e.g., as in lexical decision) that may be subject to strategic factors not directly related to word recognition. And finally, it allows to investigate the time course of lexical activation by dissociating several early (reflecting initial lexical access) and late reading time measures (reflecting higher-order processes; Rayner, 1998). Early measures typically include first fixations (i.e., the duration of the first fixation on the target word) and gaze durations (i.e., the sum of fixations from the moment the eyes land on the target for the first time until they move off again). Late reading time measures such as go-past times (i.e., the time elapsing from encountering a given target for the first time until a region to the right of the target is fixated) also include regressions originating from the target word.

The study of Duyck et al. (2007) used the eye tracking methodology to investigate the time course of cross-lingual activation effects in L2 sentence reading. Duyck et al. tested Dutch-English bilinguals while they read low-constraint sentences in which the cognate or its control were embedded (e.g., *Hilda bought a new RING-COAT and showed it to everyone*; *ring* is a cognate; *coat* is a control word). A pretest ensured that there were no differences in predictability between the cognate and control conditions. There was cognate facilitation from 249 ms onward after first encountering the target on early and late reading time measures, but only for identical cognates (i.e., cognates with identical orthographies across languages, e.g., *ring–ring*) and not for non-identical ones (e.g., *schip-ship*). The results indicate that when cross-lingual overlap was not complete, the cognate effect was not strong enough to be visible in a sentence. This shows that the amount of cross-lingual activation is a function of the similarity between the translation equivalents. Furthermore, the eye movement results indicate that the cross-lingual activations in the bilingual lexicon responsible for the cognate effect occur early in word recognition because cognate facilitation was already present on the first fixation of the target, and remained present in later eye tracking measures.

Van Assche et al. (2011) fine-tuned the distinction between identical and non-identical cognates of Duyck et al. (2007) by calculating the degree of orthographic overlap on van Orden's (1987) word similarity measure for each cognate and control word on a scale from 0 to 1 (e.g., the English-Dutch identical cognate *ring–ring*: 1.00; non-identical cognate *shoulder-schouder*: 0.81; control *witch-heks*: 0.06). Targets were presented in low- and high-constraint sentences. A cloze probability test ensured that cognates and controls were equally predictable in the sentences. In low-constraint sentences, discrete cognate facilitation (cognate vs. control) was again observed on first fixation durations, gaze durations and go-past times. Interestingly, this was shown to be a gradual and continuous effect: reading times were faster as the cross-lingual orthographic overlap between translation equivalents increased. In addition, cognate facilitation was already present on skipping rates (i.e., the probability that the word was not fixated): cognates were skipped more often than non-cognates, arguably reflecting the early origin of these cross-lingual activation effects in the time course of word processing. More

importantly, Van Assche et al. also examined how a strong semantic context modulates lexical activation spreading between languages in the bilingual lexicon by presenting cognates in high-constraint sentences. Cognate effects were observed in high-constraint sentences on both early and late measures and were present both when cognate status was taken as a discrete dichotomous variable and as a continuous variable. A control experiment with English monolinguals in which cognate effects disappeared ensured that the effects were genuinely due to the Dutch-English cross-lingual overlap. Thus, this study clearly finds evidence for cross-lingual interaction effects in the presence of a semantically constraining sentence at any stage of word recognition. This contrasts with the results of previous studies on this topic (e.g., Schwartz and Kroll, 2006; van Hell and de Groot, 2008). It seems that the use of the time-sensitive eye tracking measures uncovers the early interaction effects that were not observed in the naming task of Schwartz and Kroll (2006) or the lexical decision and translation tasks of van Hell and de Groot (2008).

The absence of an interaction between semantic constraint effects and the time course of cross-lingual lexical interactions (Van Assche et al., 2011) contrasts with the eye movement results of Libben and Titone (2009) who found cognate facilitation in semantically constraining sentences only on early comprehension measures. French-English bilinguals were presented with form-identical cognates and homographs in English sentences of low and high semantic constraint. Results showed cognate facilitation and homograph interference on all early and late measures in low-constraint sentences. However, in high-constraint sentences, these cross-lingual interaction effects were only observed on early stage reading time measures (i.e., first fixations, gaze durations, and skipping rates for cognates; gaze durations for homographs), but no effects were obtained on late stage measures. Libben and Titone suggested that lexical access in bilinguals is non-selective at early word processing stages, but that this dual-language activation is rapidly resolved by top-down factors (e.g., semantics) at later stages of comprehension.

Several factors may explain the inconsistent results across these studies. It is not the case that Van Assche et al. (2011) used a weaker semantic constraint manipulation. On the contrary, cloze probabilities in Van Assche et al. (0.86 for cognates and 0.89 for controls) were stronger than these in Libben and Titone (2009; 0.48 for cognates and 0.49 for controls). The specific bilingual population may be a key factor responsible for the different results. The bilinguals tested by Van Assche et al. were less balanced in their percentage of daily use of L1 and L2 and had acquired their L2 English later than Libben and Titone's. Therefore, Titone et al. (2011) argued that the L1 of the participants in Van Assche et al. may be more strongly activated, leading to greater L1-to-L2 cross-language activation so that semantic context may be insufficient to diminish cross-language activation.

In conclusion, these studies on L2 sentence processing indicate that the mere presentation of words in a sentence context and the language cue it provides does not nullify dual-language activation in the bilingual language system. Mixed results have been obtained for semantically constraining sentences, but recent studies using time-sensitive eye movement recordings suggest that even a strong semantic context does not necessarily eliminate

L1 PROCESSING

Although the vast majority of studies on bilingual word recognition have focused on L2 processing, there are a few studies that have investigated cross-language activation during native-language reading (e.g., Van Assche et al., 2009; Titone et al., 2011). van Hell and Dijkstra (2002) were the first to show that cognate facilitation for words out-of-context can be obtained in an exclusively native-language context. Van Assche et al. (2009) replicated this cognate effect in L1 for words out-of-context and they also investigated how a linguistic context provided by a sentence may restrict this cross-lingual activation. Dutch-English bilinguals were presented with low-constraint sentences that could include both the cognate and its control [e.g., *Ben heeft een oude OVEN/LADE gevonden tussen de rommel op zolder* (*Ben found an old OVEN/DRAWER among the rubbish in the attic*); *oven* is a Dutch-English cognate; *lade* is a control word]. Cognate facilitation was observed on early reading time measures, both as a discrete effect of cognates vs. controls and as a continuous facilitation effect of cross-lingual orthographic overlap. This implies that even when native-language processing is concerned, bilinguals are different from monolinguals: the mere knowledge of a second language affects a highly automated skill as sentence reading in the mother tongue. These findings provide strong evidence for language-non-selective access in the bilingual lexicon.

Titone et al. (2011) tested whether semantic constraint would modulate cross-language activation during L1 reading. Form-identical cognate facilitation and interlingual homograph interference was used as a marker of cross-lingual interactions. In a first experiment, English-French bilinguals read low- and high-constraint L1 sentences (e.g., *Because of the bitter custody battle over the kids, the expensive DIVORCE was a disaster*; *divorce* is an English-French cognate) while eye movements were recorded. Cognate facilitation was present on early reading time measures. This effect was independent of contextual constraint, but it was modulated by L2 age of acquisition: only bilinguals who acquired their L2 early in life showed cognate facilitation. The L2 age of acquisition did not affect the size of cognate facilitation on late reading time measures, but here, semantic constraint did: cognate effects were smaller in high- than low-constraint sentences.

In Experiment 2, Titone et al. (2011) intermixed French L2 sentences with the experimental English L1 sentences to assess whether making L2 more salient would increase cognate facilitation and interlingual homograph interference during L1 reading. And indeed, cognate effects on late reading time measures did not diminish in high-constraint sentences when L1 and L2 sentences were intermixed. Titone et al. suggested that the inclusion of the L2 sentences may have increased cross-language activation during L1 reading, which may have countered the effect of semantic constraint.

The homograph results showed no interference effects for first fixations, gaze durations, and go-past times in Experiments 1 and 2. There was, however, homograph interference for total reading times. It is striking how this pattern of results differs from the cognate results and the homograph results in an earlier study of L2 reading (Libben and Titone, 2009) because cognate and homograph effects are assumed to originate both from cross-lingual activation patterns in the bilingual language system. A possible explanation proposed by Titone et al. (2011) is that homographs and cognates are represented differently at the lexical level.

Summarizing, Van Assche et al. (2009) showed that a non-dominant language may affect native-language sentence reading, both at the earliest and at later reading stages. Titone et al. (2011) observed this cross-language activation at early reading stages only when the L2 was acquired early in life. They also showed that the semantic constraint provided by a sentence can attenuate cross-language activation at later reading stages.

THEORETICAL ACCOUNTS ON LEXICAL ORGANIZATION IN BILINGUALS

A theoretical explanation of the cross-lingual activation effects discussed in this review can be framed within bilingual language processing models such as the BIA+ model (Dijkstra and van Heuven, 2002). It is the successor of the original BIA model (Dijkstra and van Heuven, 1998), which was a bilingual extension of the Interactive Activation model (McClelland and Rumelhart, 1981). Two basic assumptions of the BIA+ model are that L1 and L2 words are represented in an integrated lexicon and that word recognition proceeds in a language-non-selective way. Upon the presentation of a word, orthographic, phonological, and semantic representations become activated (bottom-up) in both languages depending on the overlap with the input word. For homographs, orthographic representations in both languages will become strongly activated because of the identical orthography across languages, thereby activating two different semantic representations. Non-homographic control words on the other hand, will only activate lexical representations in the target language. This difference in activation level for homographs and control words gives rise to the homograph effect. For cognate words on the other hand, it is the high degree of cross-lingual orthographic, phonological, and semantic overlap that results in the cognate effect. The cross-lingual activation from these three codes speeds up the recognition of cognates compared to non-cognates.

Other theoretical accounts of the cognate effect attribute its origin to a morphological (e.g., Kirsner et al., 1993; Sánchez-Casas and García-Albea, 2005) or to a conceptual level (e.g., de Groot and Nas, 1991; van Hell and de Groot, 1998). For instance, Sánchez-Casas and García-Albea (2005) proposed that cognate translations share a morphological representation in bilingual memory whereas non-cognate translations have separate morphological representations in bilingual memory. Another account assumes that the conceptual representations of cognate translations are linked or shared across languages (e.g., van Hell and de Groot, 1998). The continuous effect of cognate status based on the degree of cross-lingual overlap in the two languages is more in line with the account that assumes cognate effects to arise from the convergent activation of orthographic, phonological, and semantic representations (e.g., van Hell and de Groot, 1998; Dijkstra and van Heuven, 2002), although a study of Lehtonen et al. (2006) also suggest a possibly different morphological representation for bilinguals and monolinguals.

In the BIA+ model (Dijkstra and van Heuven, 2002), there is a representational layer containing two language nodes, one for each language. These language nodes function as language tags, indicating to which language an item belongs, and they also reflect the global lexical activity of each language. In the earlier BIA model (Dijkstra and van Heuven, 1998), language nodes also served other functions such as language filters dependent on experimental variables or collectors of contextual activation coming from outside the lexicon. The language nodes could then facilitate activation of target language words through the inhibition of non-target language words. In this way, language nodes could account for top-down effects to the word level, although simulations have shown that language nodes cannot inhibit non-target language words sufficiently to obtain language selective access from the beginning of word recognition. Later, it became clear that combining both representational and functional aspects of language processing in one mechanism was not tenable and language nodes' function became purely representational. With respect to sentence context effects, Dijkstra and van Heuven (2002) suggested that language nodes can be pre-activated by the sentence, but as language nodes cannot inhibit non-target language words sufficiently, the mere presentation of words in a sentence does not constrain language-non-selective activation.

In order to account for differences between experiments and non-linguistic context effects (e.g., task features, instructions, participant's expectations), a distinction is made between the word identification system (containing orthographic, phonological, and semantic representation) and the task/decision system. Linguistic context, arising from lexical, syntactic, or semantic restrictions (e.g., a sentence context) is assumed to directly affect the word identification system. Non-linguistic context on the other hand, is assumed to affect the task/decision system. Dijkstra and van Heuven (2002) present the word identification system as part of a larger system in which sentence parsing and language production are also represented (e.g., Levelt et al., 1999). As the sentence parsing system may directly interact with the word identification system, syntactic and semantic context information may affect word recognition. Indeed, they explicitly state that such linguistic context information may restrict language-non-selective activation in bilinguals. However, they do not specify the exact mechanism that can give rise to these predicted top-down effects.

SUMMARY AND THEORETICAL IMPLICATIONS

The studies on bilingual sentence processing reviewed in the present paper showed that markers of language-non-selective access (such as cognate facilitation) were not nullified in the presence of a sentence context. It thus seems that the language of the preceding words is an insufficient cue to restrict lexical access to words of the target language (e.g., Schwartz and Kroll, 2006; Duyck et al., 2007; van Hell and de Groot, 2008), even when reading in the mother tongue (e.g., Van Assche et al., 2009). Furthermore, eye tracking studies revealing the time course of activation showed that semantic constraint does not necessarily restrict non-selective activation (Van Assche et al., 2011), although there is evidence that it has a relatively late effect (e.g., Libben and Titone, 2009; Titone et al., 2011), and that it affects cross-lingual activation in lexical decision, naming, and translation studies (e.g.,

Schwartz and Kroll, 2006; van Hell and de Groot, 2008). The difference in result patterns across studies suggests that the interaction between lexical activation and sentence processing is dependent on several experimental factors such as task demands (e.g., lexical decision vs. eye tracking; Duyck et al., 2007; van Hell and de Groot, 2008), type of bilinguals tested, lexical characteristics (e.g., identical vs. non-identical cognates; Duyck et al., 2007), and stimulus list composition (e.g., Titone et al., 2011).

These findings have important implications for the further development of models of bilingual word recognition. The BIA+ model (Dijkstra and van Heuven, 2002) for example, does not specify how linguistic context may exert effects in the bilingual language system. They did suggest that the language of the preceding words in the sentence does not restrict lexical activation. Indeed, the pre-activation of the language nodes by a sentence is not sufficient to restrict lexical access because language nodes cannot inhibit words to a considerable extent. Instead, lexical activation depends on the similarity of the input word with the representations in the lexicon and on the resting-level activation of the representations. The fact that cross-lingual activation was preserved in low-constraint sentences in L2 (e.g., Schwartz and Kroll, 2006; van Hell and de Groot, 2008; Libben and Titone, 2009; Van Assche et al., 2011) and in L1 (Van Assche et al., 2009; Titone et al., 2011) provides strong support for the assumption of limited influence of the sentence's language.

Furthermore, Dijkstra and van Heuven (2002) argued that the word identification system interacts with higher levels of linguistic processing (such as parsing), but they did not specify an exact mechanism for these top-down interactions from semantics to the orthographic and phonological levels. Given the data discussed in this review, how may these top-down interactions be interpreted within the BIA+ model? The reduction of homograph interference in high-constraint sentences (e.g., Libben and Titone, 2009) can easily be accounted for in the BIA+ model because it predicts that the semantic level feeds back activation to the orthographic level. As homographs have distinct semantic representations in each language, the semantic representation activated by the sentence context feeds back to the orthographic level so that the competition between the identical orthographic representations of homographs is resolved faster.

In order to explain the reduced cognate effects in semantically constraining sentences (e.g., Schwartz and Kroll, 2006; van Hell and de Groot, 2008), additional assumptions are needed regarding the role of semantic constraint on lexical activation. For instance, monolingual studies indicate that sentence context can restrict semantic, syntactic, and lexical activation for words appearing later in the sentence (e.g., Stanovich and West, 1983; Schwanenflugel and LaCount, 1988). Extrapolating this to bilinguals, we propose that, similar to the view of Altarriba et al. (1996), a semantically constraining sentence not only generates semantic and syntactic restrictions for upcoming words, but that these restrictions also result in the pre-activation of lexical representations. This may speed up lexical access for cognates so much that the convergent bottom-up activation from non-target lexical representations no longer exerts an effect.

Furthermore, recent eye tracking studies testing cognates (e.g., Libben and Titone, 2009; Van Assche et al., 2009; Titone et al., 2011)

showed clear cognate effects in early reading stages (reflected in measures such as first fixation duration and gaze duration), indicating that lexical restrictions only exert an influence during later stages of word recognition and after initial language-non-selective access had taken place. At present, it is not clear how the BIA+ model can explain the lexical restrictions generated by the sentence. The function of the language node may have an important role in this issue, but language nodes in the BIA+ model only have a representational function and cannot substantially inhibit words in the non-target language. In order to account for the lexical restrictions, it may be necessary to assume a feedback mechanism from the language nodes to the orthographic level, so that language nodes can have a direct effect on lexical selection. This way, we assure the possibility of selectivity, constrained by semantic and lexical restrictions provided by a sentence context, in the fundamentally language-non-selective bilingual language system.

It seems that the top-down modulation from semantics to the orthographic level only occurs during later stages of word recognition, but this conclusion is not fully supported by the empirical evidence. First, Van Assche et al. (2011) obtained no such modulation of the cognate effect on late reading time measures (e.g., go-past time), suggesting a very limited role of these top-down restrictions. A possible, tentative explanation for the fact that Van Assche et al. observed cognate facilitation on late reading time measures may be that if readers do not make many regressions from the target word, early reading time measures will be similar because they are completely included in late measures. Indeed, early and late reading time measures differed much more in the eye tracking studies of Libben and Titone (2009) than in Van Assche et al. Second, Titone et al. (2011) showed reduced cognate facilitation on late reading time measures in Experiment 1, but not in Experiment 2 when non-target language filler sentence were included. This indicates that the inclusion of fillers increased cross-lingual activation and may have countered the effect of sentential constraint. Here, global language processing context may have influenced bilingual word recognition, just as in Elston-Güttler et al. (2005), and this may also be linked to the language mode theory (Grosjean, 1997): lexical access may be more or less selective depending on the language context and/or the bilinguals' expectations.

FUTURE WORK DIRECTIONS

For the further development of the BIA+ model (Dijkstra and van Heuven, 2002) and other bilingual models, it is important

to note that the interactions between linguistic context and lexical variables in the BIA+ model (Dijkstra and van Heuven, 2002) may also interact with experimental/task factors (e.g., Duyck et al., 2007; van Hell and de Groot, 2008) or with participant characteristics such as age of acquisition of the L2 (e.g., Titone et al., 2011). For instance, it is important to examine whether the results generalize to other bilingual populations. For example, the bilinguals tested in Libben and Titone (2009) were more balanced and acquired their L2 earlier in life than Van Assche et al.'s (2011) bilinguals. A systematic test of the effects of proficiency and age of acquisition in future studies may help to explain whether these were the determining factors for the differences in results between these studies. Related to proficiency issues, it should be noted that many studies used self-ratings on reading, writing, speaking, and/or general proficiency. Although self-ratings provide an important indication of the proficiency level, in future studies, it is advisable to also use more direct measures to determine the L2 proficiency level such as measuring reaction times to words in both languages in lexical decision or naming tasks.

Future studies will also have to investigate how task effects influence the degree of language-non-selective access that is observed. There are important differences between results obtained with paradigms such as lexical decision, naming, and translation (e.g., Schwartz and Kroll, 2006; van Hell and de Groot, 2008) and those obtained with eye tracking (e.g., Libben and Titone, 2009; Van Assche et al., 2011). Only studies using eye tracking found evidence for cognate facilitation in semantically constraining sentences. It may well be that eye tracking constitutes a more sensitive paradigm. To examine this claim, Van Assche et al. (2011) ran an additional experiment in which the stimulus materials used in their eye tracking experiment were tested using the lexical decision paradigm of van Hell and de Groot (2008). They obtained cognate effects in low- and high-constraint sentences, but the latter effect was not very robust: cognate facilitation was weak and only emerged after testing many more bilinguals than van Hell and de Groot (2008) did. Another possibility, given in Libben and Titone (2009), is that lexical decision, naming, and/or translation tasks reflect comprehension processes subsequent to lexical access (during which cross-language activation is restricted by the semantic constraint for the target). Especially eye tracking may be sensitive enough to detect the earliest stages of word recognition and further studies are needed to clarify this issue.

REFERENCES

Altarriba, J., Kroll, J. F., Sholl, A., and Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: evidence from eye fixations and naming times. *Mem. Cognit.* 24, 477–492.

Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echalier, J. F., and Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels:

time course and scalp distribution. *J. Cogn. Neurosci.* 11, 235–260.

Bijeljac-Babic, R., Biardeau, A., and Grainger, J. (1997). Masked orthographic priming in bilingual word recognition. *Mem. Cognit.* 25, 447–457.

Binder, K., and Rayner, K. (1998). Contextual strength does not modulate the subordinate bias effect: evidence from eye-fixations and self-paced reading. *Psychon. Bull. Rev.* 5, 271–276.

Brown, C. M., and Hagoort, P. (1993). The processing nature of the N400: evidence from masked priming. *J. Cogn. Neurosci.* 5, 34–44.

Caramazza, A., and Brones, I. (1979). Lexical access in bilinguals. *Bull. Psychon. Soc.* 13, 212–214.

Coltheart, M., Davelaar, A., Jonasson, J. T., and Besner, D. (1977). "Access to the internal lexicon," in *Attention and Performance*, Vol. 6, ed. S. Dornic (New York: Academic Press), 535–555.

Connolly, J. F., and Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *J. Cogn. Neurosci.* 6, 256–266.

Costa, A., Miozzo, M., and Caramazza, A. (1999). Lexical selection in bilinguals: do words in the bilingual's two lexicons compete for selection? *J. Mem. Lang.* 41, 365–397.

de Groot, A. M. B., and Nas, G. L. J. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *J. Mem. Lang.* 30, 90–123.

Dijkstra, T., Grainger, J., and van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: the neglected role of phonology. *J. Mem. Lang.* 41, 496–518.

Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., and Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *J. Mem. Lang.* 62, 284–301.

Dijkstra, T., Timmermans, M., and Schriefers, H. (2000). On being blinded by your other language: effects of task demands on interlingual homograph recognition. *J. Mem. Lang.* 42, 445–464.

Dijkstra, T., and van Heuven, W. J. B. (1998). "The BIA model and bilingual word recognition," in *Localist Connectionist Approaches to Human Cognition*, eds J. Grainger and A. Jacobs (Mahwah, NJ: Lawrence Erlbaum Associates), 189–225.

Dijkstra, T., and van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: from identification to decision. *Biling. (Camb. Engl.)* 5, 175–197.

Dijkstra, T., Van Jaarsveld, H., and Ten Brinke, S. (1998). Interlingual homograph recognition: effects of task demands and language intermixing. *Biling. (Camb. Engl.)* 1, 51–66.

Duyck, W. (2005). Translation and associative priming with cross-lingual pseudohomophones: evidence for nonselective phonological activation in bilinguals. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 1340–1359.

Duyck, W., Van Assche, E., Drieghe, D., and Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: evidence for nonselective lexical access. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 663–679.

Elston-Güttler, K. E., Gunter, T. C., and Kotz, S. A. (2005). Zooming into L2: global language context and adjustment affect processing of interlingual homographs in sentences. *Cogn. Brain Res.* 25, 57–70.

Grainger, J., O'Regan, J. K., Jacobs, A. M., and Segui, J. (1989). On the role of competing word units in visual word recognition: the neighborhood frequency effect. *Percept. Psychophys.* 45, 189–195.

Grosjean, F. (1989). Neurolinguists, beware – the bilingual is not 2 monolinguals in one person. *Brain Lang.* 36, 3–15.

Grosjean, F. (1997). "Processing mixed language: issues, findings, and models," in *Tutorials in Bilingualism: Psycholinguistic Perspectives*, eds A. M. B. de Groot and J. F. Kroll (Mahwah, NJ: Erlbaum), 225–254.

Kerkhofs, R., Dijkstra, T., Chwilla, D. J., and de Bruijn, E. R. A. (2006). Testing a model for bilingual semantic priming with interlingual homographs: RT and N400 effects. *Brain Res.* 1068, 170–183.

Kirsner, K., Lalor, E., and Hird, K. (1993). "Exercise, meaning and morphology," in *The Bilingual Lexicon*, eds R. Schreuder and B. Weltens (Amsterdam: Benjamins), 215–248.

Lagrou, E., Hartsuiker, R. J., and Duyck, W. (2011). Knowledge of a second language influences auditory word recognition in the native language. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 952–965.

Lehtonen, M., Niska, H., Wande, E., Niemi, J., and Laine, M. (2006). Recognition of inflected words in a morphologically limited language: frequency effects in monolinguals and bilinguals. *J. Psycholinguist. Res.* 35, 121–146.

Lemhöfer, K., and Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: effects of code similarity in language-specific and generalized lexical decision. *Mem. Cogn.* 32, 533–550.

Lemhöfer, K., Dijkstra, T., and Michel, M. C. (2004). Three languages, one echo: cognate effects in trilingual word recognition. *Lang. Cogn. Process.* 19, 585–611.

Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 21, 1–38.

Libben, M. R., and Titone, D. A. (2009). Bilingual lexical access in context: evidence from eye movements during reading. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 381–390.

McClelland, J. L., and Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception. I. An account of basic findings. *Psychol. Rev.* 88, 375–407.

Poulisse, N., and Bongaerts, T. (1994). 1st Language use in 2nd-language production. *Appl. Linguist.* 15, 36–57.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 372–422.

Rayner, K., and Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: a further examination. *Psychon. Bull. Rev.* 3, 504–509.

Sánchez-Casas, R. M., and García-Albea, J. E. (2005). "The representation of cognate and noncognate words in bilingual memory: can cognate status be characterized as a special kind of morphological relation?" in *Handbook of Bilingualism: Psycholinguistic Approaches*, eds J. F. Kroll and A. M. B. de Groot (New York: Oxford University Press), 226–250.

Schwanenflugel, P. J., and LaCount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *J. Exp. Psychol. Learn. Mem. Cogn.* 14, 344–354.

Schwartz, A. I., and Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *J. Mem. Lang.* 55, 197–212.

Segui, J., and Grainger, J. (1990). Priming word recognition with orthographic neighbors: effects of relative prime-target frequency. *J. Exp. Psychol. Hum. Percept. Perform.* 16, 65–76.

Spivey, M. J., and Marian, V. (1999). Cross talk between native and second languages: partial activation of an irrelevant lexicon. *Psychol. Sci.* 10, 281–284.

Stanovich, K. E., and West, R. F. (1983). On priming by a sentence context. *J. Exp. Psychol. Gen.* 112, 1–36.

Titone, D., Libben, M., Mercier, J., Whitford, V., and Pivneva, I. (2011). Bilingual lexical access during L1 sentence reading: the effects of L2 knowledge, semantic constraint, and L1-L2 intermixing. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 1412–1431.

Van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., and Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *J. Mem. Lang.* 64, 88–107.

Van Assche, E., Duyck, W., Hartsuiker, R. J., and Diependaele, K. (2009). Does bilingualism change native-language reading? Cognate effects in a sentence context. *Psychol. Sci.* 20, 923–927.

van Hell, J. G., and de Groot, A. M. B. (1998). Conceptual representation in bilingual memory: effects of concreteness and cognate status in word association. *Biling. (Camb. Engl.)* 1, 193–211.

van Hell, J. G., and de Groot, A. M. B. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychol. (Amst.)* 128, 431–451.

van Hell, J. G., and Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychon. Bull. Rev.* 9, 780–789.

van Heuven, W. J. B., Dijkstra, T., and Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *J. Mem. Lang.* 39, 458–483.

van Orden, G. C. (1987). A rows is a rose: spelling, sound, and reading. *Mem. Cognit.* 15, 181–198.

Weber, A., and Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *J. Mem. Lang.* 50, 1–25.

The role of visual acuity and segmentation cues in compound word identification

Jukka Hyönä*

Department of Psychology, University of Turku, Turku, Finland

Studies are reviewed that demonstrate how the identification of compound words during reading is constrained by the foveal area of the eye. When compound words are short, their letters can be identified during a single fixation, leading to the whole-word route dominating word recognition from early on. Hence, marking morpheme boundaries visually by means of hyphens slows down the processing of short words by encouraging morphological decomposition when holistic processing is a feasible option. In contrast, the decomposition route dominates the early stages of identifying long compound words. Thus, visual marking of morpheme boundaries facilitates processing of long compound words, unless the initial fixation made on the word lands very close to the morpheme boundary. The reviewed pattern of results is explained by the visual acuity principle (Bertram and Hyönä, 2003) and the dual-route framework of morphological processing.

Keywords: word recognition, reading, compound words, eye movements, morphological processing, hyphenation, fovea

Research into the identification of compound words shows that word length is a central factor that should be taken into account when determining what happens in the first instances of the reading process. In the study of printed word recognition, it is tempting to restrict attention to words of a particular length. So far, the emphasis has been on the identification of short words (typically four to six letters). This choice may be motivated for many analytic languages, in which average word length tends to be quite short. However, in agglutinative languages where by default words comprise multiple morphemes, words tend to be significantly longer than in analytic languages. Finnish is a good example of a highly inflecting agglutinative language. For example, the multimorphemic word *autoissammekin* would be translated in an analytic language using multiple words: *also in our cars*. Thus, in order to gain a more complete picture of printed word recognition across structurally different languages, more emphasis should be placed on the investigation of recognizing long, multimorphemic words.

In the present review, the focus is on word compounding, which is a typical feature in agglutinative languages (e.g., Finnish), but it can also be found in more analytic languages (e.g., Dutch, German, Swedish). I review results regarding the identification of compound words during sentence reading. The reviewed studies have examined effects of two consequences of word compounding: (1) Compound words containing multiple morphemes tend to be rather long; (2) by not being marked by visually salient cues, such as spaces between morphemes, within-word morphemic units may become obscured, which may impede recognition. I particularly focus on studies examining effects of word length and salience of morpheme boundaries on the identification of compound words during reading. Thus, the present review does not cover all eye movement studies conducted on compound word reading (for a review of studies not fully covered here, see Pollatsek and Hyönä, 2006).

It is well documented that word length has a robust effect on word recognition. During reading, long words take more time to be recognized than short words (e.g., Just and Carpenter, 1980; Hyönä and Olson, 1995; Calvo and Meseguer, 2002; Kliegl et al., 2004; Juhasz et al., 2008). This is largely, but not entirely (McDonald, 2006; Hautala et al., 2011), due to long words being much more likely to require a second eye fixation on the word for successful recognition. A refixation is needed due to visual acuity limitations of the human eye.

Visual acuity drops dramatically as a function of the distance from the current fixation center. Vision is sharpest around the fovea, which spans about 2° of visual angle around the center of the fixation point. For adult readers the letter identity span (the region within which letter identities can be recognized) is no more than nine letters to the right of fixation (Häikiö et al., 2009). The span is also attentionally modulated so that it is greater toward right than left when reading text from left to right (Rayner et al., 1980, 1982); the leftward span is limited to the beginning of the currently fixated word. As the initial fixation tends to land somewhat left of the word center (Rayner, 1979), the letter (and word) identification span for adult readers is no more than 12 letters (asymmetric to the right). It should be noted, however, that the most typical fixation strategy for reading 12-letter compound words is nevertheless a two-fixation strategy (e.g., Hyönä and Pollatsek, 1998).

Long words do not only differ from short words in that they have more letters. As was briefly noted above, increased length also makes it more likely that words contain multiple morphemes. The fact that within-word morphemic units are not visually separable from each other results in the structure of multimorphemic words not being visually transparent, which in turn may lead to processing difficulties. Word length may exacerbate the impact of the structural opacity. With increased length, words are more likely to contain multiple morphemes. Moreover, decomposing

morphemes from each other may become more difficult the further away morphemes and morpheme boundaries are from the current fixation point. Thus, the probability of refixating a word as a function of morphological complexity is likely to increase even when all letters of the word are within the limits of the letter identity span.

In summary, longer word length increases chances of refixation in two ways: a higher number of letters reduces the visual acuity for words as a whole, and longer words are more likely to be made up of several morphemic units, which complicates interpretation. In the next two sections, I will discuss experiments investigating both issues. Finally, in the last section I will argue that the results of these studies strongly suggest that both number of letters and structural opacity affect early processing during reading by means of the visual acuity principle. I will also show how this explanation fits within the dual-route framework of morphological processing.

THE ROLE OF WORD LENGTH IN THE IDENTIFICATION OF COMPOUND WORDS

This section discusses the first main topic of this review: the effect of word length on the identification of morphologically complex words. As noted above, an increase in word length is accompanied by an increase in the probability of the reader not being able to read a word with a single fixation; instead, a refixation is programmed to the word region falling outside the letter identity span. In other words, when parallel processing of all letters of the word is rendered impossible due to visual acuity limitations, longer words are recognized sequentially. During the second stage (i.e., during the refixation), letters initially falling outside the letter identity span are subsequently identified. It should be noted, however, that during the initial stage the non-identified letters are pre-processed, leading to a subsequent processing benefit (called preview benefit) during their foveal processing (for recent reviews of parafoveal processing during reading, see Hyönä, 2011; Schotter et al., 2012).

But how does word length influence the identification of multimorphemic words? This question was investigated by Bertram and Hyönä (2003) with two-constituent Finnish compound words that were either relatively short (an average of 7.6 letters) or long (an average of 12.8 letters). The target words were embedded in sentences; participants' eye movements were tracked while they were reading these sentences for comprehension. In Experiment 1, the frequency of the first-constituent (as a separate word) was manipulated for both short and long compound words; in Experiment 2 the same was done for the whole-word frequency. According to the logic adopted from Taft and Forster's (1976) seminal work, an early effect of first-constituent frequency would suggest that the compound word is decomposed for its recognition (lexical access is initiated by the recognition of the first-constituent, followed by the recognition of the second constituent and the whole-word). On the other hand, parallel processing of the two components is implicated by an early effect of whole-word frequency and by the absence of an early effect of first-constituent frequency.

An attractive feature of the eye-tracking technique is that it can be used to tap into the time course of processing, particularly when word processing is distributed across multiple fixations. Hence, the duration of the first-fixation can be used to index early processes, while durations of subsequent fixations reflect processing done

at later stages. Despite being an aggregate measure, gaze duration (i.e., the summed duration of fixations landing on the word during its first-pass reading) is typically also used as an index of relatively late processing. This is due to gaze duration strongly reflecting the probability of refixating a word.

In Experiment 1, an early effect of first-constituent frequency, as indexed by first-fixation duration, was obtained for long compound words but not for short compound words; the latter only revealed a marginal effect in later processing indexed by gaze duration. In contrast, Experiment 2 revealed are liable early effect of whole-word frequency for short compound words but only a small and statistically marginal effect (4 ms) for long compound words; both types of words showed a whole-word frequency effect in later processing, as indexed by gaze duration. Thus, the pattern of data suggests that for short compound words the whole-word representation becomes active soon after the word is foveated. On the other hand, with long compound words the first-constituent is more strongly activated during the initial processing stage than the whole-word representation. It should be noted, however, that even though the whole-word representation receives early activation for short compound words, short compound words are not fully identified during the initial fixation, but often (roughly about half of the trials in the Bertram and Hyönä study) a refixation is needed to complete the lexical access.

To account for the observed pattern of results, Bertram and Hyönä (2003) put forth the visual acuity principle. According to this principle, word processing is initiated with whatever information is readily available in the foveal vision. When all (or a sufficient number of) letters of the word are within foveal reach, the whole-word representation becomes readily available early on. On the other hand, when only the initial morpheme is foveally available, as is the case with long compound words (longer than about 12 letters), word recognition is initiated by first accessing the initial constituent, followed by the second constituent and the whole-word. The claim that there is a strong sequential component in recognizing long compound words is further supported by the finding that the earliest point in time when the frequency of second constituent exerts an effect is when a second fixation is made on the word (Pollatsek et al., 2000). Note, however, that Inhoff et al. (2008) reported evidence indicating that the frequency of the second lexeme already exerts an effect on the first-fixation duration. This effect was obtained for the so-called tailed compound words, for which the second lexeme was the meaning-defining lexeme. However, the effect was not significant in the item analysis. Considering that their compound words varied in length between 8 and 11 letters (average length 9.1 letters), it is possible that the early second lexeme frequency effect was produced by the shorter compounds. If so, it would be evidence for compound word lexemes playing an active role early on during the identification of short compounds – a claim inconsistent with the visual acuity principle of Bertram and Hyönä (2003) and with the data of Pollatsek et al. (2000).

With regards to existing models of morphological processing, the data of Bertram and Hyönä (2003) are consistent with parallel dual-route morphological models (e.g., Schreuder and Baayen, 1995; Pollatsek et al., 2000). These models assume two routes to be in operation in tandem: the decomposition route and the full-form

route. Lexical access via the decomposition route takes place via the constituents, while the full-form route attempts access by finding a match between the visual input and a stored whole-word representation. To account for the above-summarized results, this framework needs to be complemented with the visual acuity principle. A head-start is won by the route favored by visual acuity. In other words, when only the first-constituent is fully available in the fovea, the decomposition route achieves a head-start; in contrast, the full-form route is either initially favored when the whole-word is within foveal reach or it quickly overtakes the initially favored decomposition route. The later effect of whole-word frequency for long compounds and the later effect of first-constituent frequency for short compounds observed by Bertram and Hyönä (2003) may in this framework be taken as reflecting the later activation of the slower route.

Does the full-form route provide access only to the lexical representation of the compound word or is access to meaning simultaneously also achieved? My present view is that with existing (lexicalized) compound words lexical access is very quickly followed by the activation of word meaning. My view is based on two eye-tracking experiments (Pollatsek and Hyönä, 2005; Frisson et al., 2008) that did not find evidence for disruption in processing when compound words were semantically opaque, in comparison to semantically transparent compounds (see, however, Juhasz, 2007). On the other hand, with novel compound words for which no mental representation exists, a meaning computation stage is quite naturally required (Pollatsek et al., 2011).

More recently, Fiorentino and Poeppel (2007) have studied the time course of compound word processing by registering brain activation via MEG (magnetoencephalography) when participants made lexical decisions to frequency- and length-matched compound and monomorphemic words in English. Their stimuli were comparable in length (an average of 7.8 letters) to the short compounds of Bertram and Hyönä (2003). All words were infrequent; however, the compound word constituents were all frequent (as separate words). There were two main findings: (a) lexical decision time was shorter for compound words than for frequency-matched monomorphemic words; (b) the MEG component presumed to index lexical access (M350) peaked earlier for compound than for monomorphemic words. These data were taken to suggest that compound words are always recognized via the decomposition route. This conclusion contrasts with the argument made by Bertram and Hyönä (2003). To recap, they posit that the full-form route has a head-start in processing compound words which are sufficiently short to fit in the foveal area of the eye.

There are two plausible explanations to account for the apparent discrepancy between the two set of results. First, Fiorentino and Poeppel (2007) presented their stimuli in a very large font; on average the words extended horizontally 6.4° of visual angle, which means that the words did not fit in the foveal area. Thus, the visual acuity principle predicts here that the decomposition route is initially favored – a claim consistent with their data. Second, the processing disadvantage for monomorphemic over compound words may be strengthened by the fact that a subset of the monomorphemic words was probably quite unknown to the participants. This possibility is hinted at by the fact that participants made 20% decision errors with the lowest-frequency

monomorphemic words. At any rate, further studies are needed to solve this discrepancy. An attractive possibility would be to combine eye-tracking with MEG recordings to examine whether the two methods provide converging evidence concerning the timing of different frequency effects. It would also serve as a methodological cross-validation.

Vergara-Martínez et al. (2009) conducted an ERP study in Basque, where they independently varied the frequency of the first and second constituent in two-constituent compound words. The target words were embedded in sentence context, and participants were asked to read the sentence for comprehension. Importantly for the present discussion, the length of the target words varied from 6 to 12 letters (average length 9.25 letters). Their spatial extent in terms of visual angle is not provided in the paper. However, Duñabeitia (personal communication) informed me that their standard procedure was to use Courier New font where one character subtended horizontally 0.41 cm. With their viewing distance of 80 cm, a 6-letter word subtended 1.76° and 12-letter word 3.52° of visual angle, respectively. Thus, their shorter words fitted in the foveal region, while their longest words did not.

Vergara-Martínez et al. (2009) obtained an earlier electrophysical response for the first-constituent frequency manipulation than for the second constituent frequency manipulation. Moreover, the nature and the scalp distribution were different. The early negativity effect in ERPs was greater for high- than low-frequency first-constituent compound words, whereas the amplitude of the later negativity effect was larger for low- than high-frequency second constituent compound words in the right hemisphere but not in the left hemisphere. This pattern of results was interpreted within the activation-verification framework that Duñabeitia et al. (2007) proposed for the processing of Basque and Spanish compound words. According to this framework, the early effect obtained for first-constituents reflects the activation of the morphological family triggered by first-constituents, with more activation produced by high-frequency than low-frequency first-constituents. The later effect associated with second constituents in turn reflects the selection of the final lexical candidate among those triggered by the first-constituent; the frequency of the second constituent affects the speed of verification.

The compatibility of the Vergara-Martínez et al. (2009) results with the visual acuity principle of Bertram and Hyönä (2003) cannot be readily assessed because it is unknown to what extent the observed effects should be attributed to the short and longer words. On the one hand, the average number of letters making up their compound words is closer to the average length of the short compounds used by Bertram and Hyönä. From that perspective, the early negativity effect obtained for the first-constituent manipulation may be tentatively interpreted to be inconsistent with the visual acuity principle. On the other hand, the negativity effect associated with the first-constituent frequency manipulation was also observed in the later time window – a finding compatible with those of Bertram and Hyönä (they found a suggestion for a later effect of first-constituent frequency for short compounds). It is also possible that a subset of longer compounds was responsible for the early effect of first-constituent frequency obtained by Vergara-Martínez et al. (2009), whereas a subset of short compounds would be responsible for the late effect.

Perhaps the most serious challenge to the visual acuity principle is provided by Juhasz (2008), who extended the work of Bertram and Hyönä (2003) by conducting studies in English rather than in Finnish. In Experiment 1, Juhasz employed two-constituent compounds that were either long (range: 10–13 letters; average: 10.9 letters) or short (range: six to seven letters; average: 6.6 letters), and manipulated first-constituent frequency. Contrary to Bertram and Hyönä (and several other studies), she found no early effect of first-constituent frequency for long compounds, as indexed by first-fixation duration. In contrast, for short compounds there was a nearly significant first-constituent frequency effect in first-fixation duration. The pattern of results was similar in gaze duration, with a nearly significant (not significant by items) first-constituent frequency effect for short compounds but not for long compounds (in fact, there was a marginal tendency for a reversed frequency effect).

In Experiment 2, Juhasz (2008) manipulated the rated whole-word frequency (i.e., familiarity) of short and long compounds. The frequency ratings were collected using a seven-point scale, yielding an average rating of 6.7 for the high-frequency compounds and an average rating of 3.1 for the low-frequency compounds. The short and long compounds were comparable in length to those in Experiment 1. There was a significant main effect of rated-frequency in first-fixation duration, but no interaction with word length; short and long compounds displayed an effect of similar size. However, when separate analyses were conducted for first-fixations when they were the single fixations on the word and when they were the first of multiple fixations, two opposing trends were observed. For single fixation duration, the rated-frequency effect was larger for long than short compounds, whereas the duration of first-fixation followed by at least one refixation displayed an opposite pattern. Given the fact that the latter type of trials was dominant in the Bertram and Hyönä (2003) data, these results are not completely inconsistent with their data. Finally, gaze duration revealed in the Juhasz (2008) study a larger rated-frequency effect for long than short compounds.

Taken together, the two experiments of Juhasz (2008) did not find evidence in English for the view advocated by the visual acuity principle that the decomposition route would be more powerful early on during long compound word processing, while the full-form route would quickly overtake the decomposition route when processing short compounds. At present, it is not clear how the differences in the results of Juhasz (2008) and those of Bertram and Hyönä (2003) could be explained. One possibility is that they may reflect inherent language differences: due to the morphological richness and complexity of Finnish, Finnish readers may be generally more prone to use the decomposition route than the English readers.

Before concluding the first section of the present review, I briefly discuss the possibility that the finding of early first-constituent frequency effects being somewhat more modest in English than Finnish studies may be explained by the visual acuity principle (note, however, that the study of Juhasz, 2008, speaks to the contrary). In the English studies, the length of the compound words tended to fall somewhere between the short and long compound words used by Bertram and Hyönä (2003). Juhasz et al. (2003) studied reading processes for two-constituent English

compound words that were all nine letters long. They obtained a marginal 11-ms effect of first-constituent frequency in the first-fixation duration indexing early effects. Similarly, Andrews et al. (2004) employed two-constituent compounds that were on average 8.5 letters (ranging from 6 to 11 letters) long, and found a marginal 7–8 ms effect of first-constituent frequency on first-fixation duration. On the other hand, Bertram and Hyönä (2003) observed a significant 16-ms early effect (i.e., in first-fixation duration) of first-constituent frequency for 12–14-letter Finnish compound words. These data are generally in line with the visual acuity principle, suggesting that the early involvement of first-constituents is attenuated for shorter compound words.

To sum up the first section, the data reviewed above provide relatively consistent support for the view that the identification of two-constituent compound words is constrained by word length. The results of most studies (but see Juhasz, 2008), support the hypothesis that the identification process for long compound words is initiated by first recognizing the initial constituent. In contrast, full-form access can be reached without going via the access of the constituents if compound words are short (provided that the full-form is sufficiently frequent in order to become readily available).

ROLE OF SEGMENTATION CUES IN IDENTIFYING COMPOUND WORDS

The second main topic of the present review concerns the effects of segmentation cues on the speed of identifying morphologically complex words. I have argued above that the lexical access of long compound words starts with the access of the initial constituent (i.e., via the decomposition route). If this claim is true, providing visual segmentation cues that make it easier to identify the morpheme boundary should speed up the processing of long compound words because it facilitates accessing the initial component. The same pattern should not be found for the short compound words because they are processed via the holistic route.

These claims were tested by Bertram and Hyönä (submitted) in an eye-tracking study in which participants read long (on average 12.1 letters) and short (on average 7.3 letters) compound words that were either hyphenated (e.g., *musiikki-ilta*) or concatenated (e.g., *yllätystulos* = surprise result; i.e., written without a hyphen at the constituent boundary). According to the Finnish spelling regulations (on the constraints of writing compound words in English, see Kuperman and Bertram, submitted), a hyphen has to be inserted at the constituent boundary when two identical vowels span the morpheme boundary (as in *musiikki-ilta* = music evening). Hyphens prevent possible misparses of the syllables at the boundary, and consequently that the word's morphological structure is misparsed. By explicitly marking the multimorphemic nature of words, hyphens are likely to benefit the decomposition route but should inhibit the whole-word route. If so, a hyphen at the constituent boundary would speed up the processing of long compound words but slow down the processing of short compound words. The hyphenated and non-hyphenated compounds were matched for word frequency as well as first- and second-constituent frequency. Moreover, the number of letters (not counting the hyphen) was equated separately for the two short and long compound conditions.

Bertram and Hyönä (submitted) obtained the predicted data pattern. The presence of hyphens in long compound words significantly affected subgaze duration (the gaze duration on the first-constituent) prior to making a saccade away from the first-constituent. Subgaze duration was 74 ms shorter in the hyphenated than in the concatenated condition. An effect of similar size (64 ms) was also observed in gaze duration of the whole-word. These data are in line with the view that the presence of hyphens supports the decomposition route that is presumed to prevail during the early stages of long compound word processing. In contrast, gaze duration on short compound words was significantly longer on hyphenated than concatenated words (a difference of 43 ms favoring concatenated short compounds). The gaze duration effect was largely due to hyphens attracting a second fixation on short compounds (typically landing on the second constituent). In other words, in short compound words a hyphen at the morpheme boundary seemed to have boosted the decomposition route in cases where holistic processing is a viable option, as claimed by the visual acuity principle. Interestingly, Häikiö et al. (2011) replicated the detrimental effect of hyphens on the processing of short compound words with elementary school children (Second, Fourth, and Sixth graders). All children, except the slowest Second grade readers, took longer to read short compounds when these were hyphenated than when they were concatenated. These findings suggest that even relatively young readers are capable of reading short compound words via the holistic route.

In addition to providing further support for the visual acuity principle, the study of Bertram and Hyönä (submitted) also demonstrated the usefulness of visually salient morpheme boundary cues (hyphens) in reading long compound words. The usefulness of hyphens was further examined by Bertram et al. (2011). In contrast to Bertram and Hyönä (submitted), they inserted hyphens at constituent boundaries despite them not being prescribed by spelling conventions. Thus, their study was a strong test of the usefulness of segmentation cues, as the hyphens were inserted illegally. Further differences with Bertram and Hyönä were that the words consisted of three constituents instead of two, and that not only Finnish stimuli were used (*lentokenttätaksi* = airport taxi), but also Dutch words (e.g., *voetbalbond* = football association). The average length of the Dutch stimuli was 14.5 letters (range 10–21 letters) and that of the Finnish stimuli 15.8 letters (range 13–24 letters). The target compound words were inserted in single sentences; native-language participants read these sentences while their eye movements were recorded. The processing of illegally hyphenated compounds was compared to that of concatenated compounds (i.e., written as required by the spelling conventions). Hyphens were inserted either at major or minor morpheme boundaries. Major boundaries demarcate the boundary between modifier and head, as in *voetbalbond* (football association) or *zaal-voetbal* (indoor football), while minor boundaries appear at morpheme boundaries of two-constituent modifiers (e.g., *voet-balbond* = foot-ball association) or head (e.g., *zaalvoet-ball* = indoor foot-ball). These two different word structures are called left-branching and right-branching, respectively. It was expected that hyphens would benefit processing when placed at major boundaries, whereas placing them at minor boundaries might lead to initially misparsing morphological structures.

For both Dutch and Finnish, Bertram et al. (2011) found a decrement in overall processing time (indexed by gaze duration on the whole-word) due to the insertion of hyphens at minor boundaries. The two sets of results differed from each other in that major-boundary hyphens speeded up gaze durations in Finnish, whereas in Dutch this condition did not differ from the concatenated words (i.e., legal spelling). More detailed analyses demonstrated early facilitation in processing hyphenated three-constituent Dutch compound words, as revealed by shorter subgaze durations on the left component (consisting of either one or two constituents, depending on branching) separated by a hyphen from the right component. In other words, subgaze duration on the modifier was shorter for illegally hyphenated compounds than for legally concatenated compounds. However, the early processing benefit was offset by a later processing cost associated with illegal hyphenation. Subgaze on the right component was significantly longer in the hyphenated than in the concatenated condition. The pattern was similar in Finnish for early processing. On the other hand, the later slowing down in processing the right component was not apparent in Finnish for the left-branching compounds (two-constituent modifier + one-constituent head) but was so for the right-branching compounds. In sum, both experiments of Bertram et al. (2011) demonstrate an early processing benefit due to hyphenation, presumably reflecting facilitation in morphological segmentation and in parsing the morphological structure (i.e., assigning the modifier-head relation) of three-constituent compound words. The later processing cost due to hyphenation is likely to reflect readers' response to illegal spelling. It is noteworthy, however, that in the course of the experiment Finnish readers became used to illegal hyphenation, to the extent that toward the end of the experiment gaze durations on the whole-word were significantly shorter for the hyphen-at-the-major-boundary compounds than for the legally concatenated ones. A similar type of learning was observed in the Dutch experiment; however, it did not result in faster processing of major-boundary hyphenation compounds over concatenated compounds.

The overall pattern of early facilitation offset by later slowing down in processing due to hyphenation is consistent with what Inhoff et al. (2000) found for processing illegally spaced German compound words. In other words, instead of inserting a hyphen at constituent boundaries they added spaces between the constituents in three-constituent compounds (e.g., *Daten-Schutz Experte*). They found shorter gaze durations on illegally spaced than legally unspaced compounds; on the other hand, the final fixation on the word tended to be longer in the spaced than unspaced condition. A similar pattern of results was obtained by Juhasz et al. (2005) for reading normally unspaced English compounds as spaced. First-fixation duration was shorter for spaced than unspaced compounds, but a disruption in processing due to spacing was observed in refixations. Yet, as detailed above, unlike spacing, hyphenation may lead to general processing benefits (see the Finnish results of Bertram et al., 2011; and those of Bertram and Hyönä submitted). This may be due to hyphens signaling that constituents belong to the same unit; on the other hand, spacing cannot accomplish this, which in turn may result in initially interpreting the compound word constituents as belonging to two separate phrases (Staub et al., 2007).

Bertram et al. (2004) were interested in whether orthographic-phonological cues that are more subtle than spaces or hyphens may signal the morphological boundary in long two-constituent Finnish compound words and hence aid in compound word identification. They studied how vowel harmony (vs. disharmony) at the constituent boundary affects the speed of processing two-constituent compound words in sentence contexts. Vowel harmony refers to a feature in Finnish[1], where back vowels (a, o, u) and front vowels (ä, ö, y) never appear together in word stems or case-inflected words. However, they can co-occur in compound word constituents; for example, the first-constituent may contain front vowels and the second constituent back vowels. Thus, it is also possible to have two vowels of different quality appear adjacent to each other at the morpheme boundary, as in *selkäongelma* (=back problem; the morpheme boundary is bolded). This is an unambiguous morpheme boundary cue, as the vowels ä and o have to belong to different lexemes. In contrast, the morpheme boundary appears more obscured when two vowels of the same quality stand next to each other at the boundary, as in *ryöstöyritys* (=robbery attempt; the morpheme boundary is bolded). In the latter case, it is possible to initially misparse the syllable structure of the word, as *töy* forms an existing syllable (note, however, that the target words never allowed two alternative morphological parses).

In Experiment 1, Bertram et al. (2004) embedded the two types of compound words described above (*selkäongelma* vs. *ryöstöyritys*) in sentences and recorded readers' eye movements on these words when silently reading these sentences for comprehension. Vowel quality at the constituent boundary had a significant effect on the speed of word recognition, as indexed by gaze duration on the word; gaze duration was 43 ms shorter in the vowel disharmony than in the vowel harmony condition. In a follow-up analysis, they compared the vowel harmony effect separately for short (four or five letters) and long (at least six letters) first-constituent compounds (word length was matched). This analysis showed that the effect was doubled in size for long than short first-constituent compounds (23 vs. 49 ms, respectively). The modulation of the effect size is interpreted to be due to visual acuity. The first-fixation on the word landed very close to the morpheme boundary for short first-constituent compounds, while for long first-constituent compounds it was some distance away from the initial fixation. In the former case the entire first-constituent is readily available in foveal vision, whereas in the latter case the morpheme boundary is not exactly at fixation, which then results in the boundary manipulation exerting a bigger effect. The modulation by first-constituent length was further confirmed in Experiment 2, where first-constituent length was systematically varied (three to five vs. seven to nine letters). There was a sizeable vowel harmony effect in gaze duration for long first-constituent compounds (114 ms), whereas it was non-existent (2 ms) for short first-constituent compounds. Thus, it seems that orthographic-phonological cues help to determine the constituent boundary with long first-constituent compounds, while these cues are ineffective with short first-constituent compounds,

presumably because the boundary is located in the center of the foveal vision when the word is initially fixated.

Two vowels of different quality (front vs. back) at the constituent boundary unavoidably create a bigram trough (Seidenberg, 1987; Rapp, 1992). Thus, the results of Experiment 1 and 2 may not necessarily reflect differences in vowel quality. However, *post hoc* analyses of Experiment 1 and 2 revealed that the vowel harmony effect was not merely due to differences in the frequency of the bigram spanning the morpheme boundary. Moreover, in Experiment 3 a 60-ms difference in gaze duration in favor of the disharmony condition over the harmony condition was observed when the critical vowels were not adjacent to each other (i.e., the first-constituent ended with a vowel but the second constituent started with a consonant) and the two vowel harmony conditions were matched for the frequency of the bigram spanning the morpheme boundary. Experiment 3 demonstrates that two vowels of different quality do not need to be adjacent to each other for the effect to emerge. Thus, these data suggest that vowel harmony appears to be a unique defining feature in Finnish for morpheme boundaries, perhaps operating at the phonological level.

In addition to vowel harmony, consonant type at the boundary was also manipulated. In Experiment 3, Bertram et al. (2004) compared two conditions: (a) the initial consonant of the second constituent was such that it cannot appear as the final letter in a lexeme (unambiguous condition), or (b) the consonant was one that can either appear at the end or the beginning of a lexeme (ambiguous condition). Consonant ambiguity produced an effect on gaze duration of similar size (52 ms) as vowel harmony. The consonant and vowel quality effects appeared independent of each other, as the two factors did not interact with each other. Finally, the analysis of the processing time course of the obtained effects suggested that boundary cue effects peaked at the third fixation made on the word; to a lesser extent they were also apparent during the second and fourth fixation. Thus, the relatively later appearance of the effect is generally inconsistent with the prelexical accounts of morphological decomposition (e.g., Taft, 1979, 1994; Rastle et al., 2004) predicting an early effect.

Interestingly, a recent lexical decision experiment conducted in Dutch (Lemhöfer et al., 2011) found a converging pattern of data to those reviewed above. Lemhöfer et al. observed that lexical decisions to compound words with extremely low-frequency bigrams at the morpheme boundary (e.g., sb in *fietsbel*) were 26 ms shorter than those to compounds with a frequent bigram at the boundary (e.g., sp in *fietspomp*). It should be noted that Inhoff et al. (2000) did not find an effect of uncommon bigrams at the constituent boundary on compound word reading in German. In a follow-up analysis Lemhöfer et al. found, similarly to Bertram et al. (2004), that the boundary cue exerted an effect on the identification of long (10–13 letters) but not of short (6–10 letters) compounds. Curiously, non-native Dutch speakers (German-Dutch bilinguals) did not show the modulation by length. This was taken to suggest that non-native speakers use the decomposition route to identify all compound words, irrespective of length.

In sum, the data summarized above suggest a dynamic interplay between lexical access and morphological parsing during the identification of long two-constituent compounds. Access to the first-constituent is readily achieved when it is short, as the whole

---

[1]Vowel harmony exists also in Hungarian, distantly related to Finnish, and in some Altaic languages (e.g., Turkish and Uighur).

constituent is within foveal reach during the initial fixation made on the word; thus, morphological parsing cues are of little value and can even be detrimental. In contrast, parsing cues become more valuable in facilitating access to the first-constituent when it is longer and the morpheme boundary resides some distance away from the center of the initial fixation.

WHAT HAPPENS DURING THE FIRST 250 MS OF COMPOUND WORD PROCESSING?

In this final section I present my view on the topic of the present special issue: what happens within the first 250 ms of compound word processing. My view is based on the data presented above, the visual acuity principle and the dual-route framework of morphological processing. For the identification of compound words, the dual-route model posits that the whole-word route and the decomposition route operate in parallel and possibly in interaction with each other.

When recognizing compound words that are sufficiently short to fit within the area of the foveal vision, all letters can be identified in parallel, which then enables the activation of the whole-word representation during the initial fixation of the word. Thus, the whole-word route is active early on during processing and dominates the identification of short compound words during those first 250 ms. As the whole-word representation becomes available early on, the initial fixation is often also the only fixation needed to recognize short compound words.

In contrast, simultaneous identification of all letters is impossible with longer compound words; only the letters of the first-constituent lie in the fovea and are thus recognizable. Consequently, the decomposition route dominates the first 250 ms of processing. During the initial processing stage, access to the first-constituent is achieved. A refixation is then needed to identify the remaining letters of the word. The holistic route also becomes fully active during this refixation; yet, the decomposition route is still in operation, as it takes care of the access to the second constituent. The decomposition route is aided by orthographic-phonological cues signaling the morpheme boundary, and with that, the morphological structure of the word. The facilitation in processing due to boundary cues is only achieved when the morpheme boundary is located some distance away from the location of the initial fixation. In other words, when the initial constituent is short, all its letters are clearly visible and boundary cues are not needed to separate its letters from those of the second constituent.

In conclusion, word length strongly affects word identification. Therefore, by widening their scope beyond short words, researchers cannot only generalize their findings to a larger pool of languages, but will also open a treasure trove of valuable new insights regarding early activities in the reading process. In addition, cross-linguistic and multi-language studies are also needed for building word recognition models capable of accounting for data derived from qualitatively different orthographies (see Frost, in press, for further arguments for the need of cross-linguistic studies of word recognition).

REFERENCES

Andrews, S., Miller, B., and Rayner, K. (2004). Eye movements and morphological segmentation of compound words: there is a mouse in the mousetrap. *Eur. J. Cogn. Psychol.* 16, 285–311.

Bertram, R., and Hyönä, J. (2003). The length of a complex word modifies the role of morphological structure: evidence from eye movements when reading short and long Finnish compounds. *J. Mem. Lang.* 48, 615–634.

Bertram, R., Kuperman, V., Baayen, R. H., and Hyönä, J. (2011). The hyphen as a segmentation cue in tri-constituent compound processing: it's getting better all the time. *Scand. J. Psychol.* 52, 530–544.

Bertram, R., Pollatsek, A., and Hyönä, J. (2004). Morphological parsing and the use of segmentation cues in reading Finnish compounds. *J. Mem. Lang.* 51, 325–345.

Calvo, M. G., and Meseguer, E. (2002). Eye movements and processing stages in reading: relative contribution of visual, lexical and contextual factors. *Span. J. Psychol.* 5, 66–77.

Duñabeitia, J. A., Perea, M., and Carreiras, M. (2007). The role of the frequency of constituents in compound words: evidence from Basque and Spanish. *Psychon. Bull. Rev.* 14, 1171–1176.

Fiorentino, R., and Poeppel, D. (2007). Compound words and structure in the lexicon. *Lang. Cogn. Process.* 22, 953–1000.

Frisson, S., Niswander-Klement, E., and Pollatsek, A. (2008). The role of semantic transparency in processing of English compound words. *Br. J. Psychol.* 99, 87–107.

Frost, R. (in press). Towards a universal model of reading. *Behav. Brain Sci.*

Häikiö, T., Bertram, R., and Hyönä, J. (2011). The development of whole-word representations in compound word processing: evidence from eye fixation patters of elementary school children. *Appl. Psycholinguist.* 32, 533–551.

Häikiö, T., Bertram, R., Hyönä, J., and Niemi, P. (2009). Development of the letter identity span in reading: evidence from the eye movement moving window paradigm. *J. Exp. Child Psychol.* 102, 167–181.

Hautala, J., Hyönä, J., and Aro, M. (2011). Dissociating spatial and letter-based word length effects observed in readers' eye movement patterns. *Vision Res.* 51, 1719–1727.

Hyönä, J. (2011). "Foveal and parafoveal processing during reading," in *The Oxford Handbook of Eye Movements*, eds S. P. Liversedge, I. Gilchrist, and S. Everling (Oxford: Oxford University Press), 819–838.

Hyönä, J., and Olson, R. K. (1995). Eye fixation patterns among dyslexic and normal readers: effects of word length and word frequency. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 1430–1440.

Hyönä, J., and Pollatsek, A. (1998). Reading Finnish compound words: eye fixations are affected by component morphemes. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 1612–1627.

Inhoff, A. W., Radach, R., and Heller, D. (2000). Complex compounds in German: interword spaces facilitate segmentation but hinder assignment of meaning. *J. Mem. Lang.* 42, 23–50.

Inhoff, A. W., Starr, M. S., Solomon, M., and Placke, L. (2008). Eye movements during reading of compound words and the influence of lexeme meanings. *Mem. Cognit.* 36, 675–687.

Juhasz, B. J. (2007). "The influence of semantic transparency on eye movements during English compound word recognition," in *Eye Movements: A Window on Mind, and Brain*, eds R. P. G. Van Gompel, M. H. Fischer, W. S. Murray, and R. L. Hill (Oxford: Elsevier), 373–389.

Juhasz, B. J. (2008). The processing compound words in English: effects of word length on eye movements during reading. *Lang. Cogn. Process.* 23, 1057–1088.

Juhasz, B. J., Inhoff, A. W., and Rayner, K. (2005). The role of interwordspaces in the processing of English compound words. *Lang. Cogn. Process.* 20, 291–316.

Juhasz, B. J., Starr, M. S., Inhoff, A. W., and Placke, L. (2003). The effects of morphology on the processing of compound words: evidence from naming, lexical decision, and eye fixations. *Br. J. Psychol.* 94, 223–244.

Juhasz, B. J., White, S. J., Liversedge, S. P., and Rayner, K. (2008). Eye movements and the use of parafoveal word length information in reading. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 1560–1579.

Just, M. A., and Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychol. Rev.* 87, 329–354.

Kliegl, R., Grabner, E., Rolfs, M., and Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *Eur. J. Cogn. Psychol.* 16, 262–284.

Lemhöfer, K., Koester, D., and Schreuder, R. (2011). When bicycle pump is harder to read than bicycle bell: effects of parsing cues in first and second language compound reading. *Psychon. Bull. Rev.* 18, 364–370.

McDonald, S. A. (2006). Effects of number-of-letters on eye movements during reading are independent from effects of spatial word length. *Vis. Cogn.* 13, 89–98.

Pollatsek, A., Bertram, R., and Hyönä, J. (2011). Processing novel and lexicalized Finnish compound words. *J. Cogn. Psychol. (Hove)* 23, 795–810.

Pollatsek, A., and Hyönä, J. (2005). The role of semantic transparency in the processing of Finnish compound words. *Lang. Cogn. Process.* 20, 261–290.

Pollatsek, A., and Hyönä, J. (2006). "Processing of morphologically complex words in context: what can be learned from eye movements," in *From Inkmarks to Ideas: Current Issues in Lexical Processing*, ed. S. Andrews (Hove: Psychology Press), 275–298.

Pollatsek, A., Hyönä, J., and Bertram, R. (2000). The role of morphological constituents in reading Finnish compound words. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 820–833.

Rapp, B. C. (1992). The nature of sublexical orthographic organization: the bigram trough hypothesis examined. *J. Mem. Lang.* 31, 33–53.

Rastle, K., Davis, M. H., and New, B. (2004). The broth in my brother's brothel: morpho-orthographic segmentation in visual word recognition. *Psychon. Bull. Rev.* 11, 1090–1098.

Rayner, K. (1979). Eye guidance in reading: fixation locations within words. *Perception* 8, 21–30.

Rayner, K., Well, A. D., and Pollatsek, A. (1980). Asymmetry of the effective visual field in reading. *Percept. Psychophys.* 27, 537–544.

Rayner, K., Well, A. D., Pollatsek, A., and Bertera, J. H. (1982). The availability of useful information to the right of fixation in reading. *Percept. Psychophys.* 31, 537–550.

Schotter, E., Angele, B., and Rayner, K. (2012). Parafoveal processing in reading. *Atten. Percept. Psychophys.* 74, 5–35.

Schreuder, R., and Baayen, R. H. (1995). "Modeling morphological processing," in *Morphological Aspects of Language Processing*, ed. L. B. Feldman (Hillsdale, NJ: Erlbaum), 131–154.

Seidenberg, M. S. (1987). "Sublexical structures in visual word recognition: access units or orthographic redundancy?" in *Attention and Performance XII: The psychology of Reading*, ed. M. Coltheart (Hove: Erlbaum), 245–263.

Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., and Majewski, H. (2007). The time course of plausibility effects of eye movements in reading: evidence from noun-noun compounds. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 1162–1169.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Mem. Cognit.* 7, 263–272.

Taft, M. (1994). Interactive-activation as a framework for understanding morphological processing. *Lang. Cogn. Process.* 9, 271–294.

Taft, M., and Forster, K. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *J. Verb. Learn. Verb. Behav.* 15, 607–620.

Vergara-Martínez, M., Duñabeitia, J. A., Laka, I., and Carreiras, M. (2009). ERP correlates of inhibitory and facilitative effects of constituent frequency in compound word reading. *Brain Res.* 1257, 53–64.

Morphological processing as we know it: an analytical review of morphological effects in visual word identification

Simona Amenta* and Davide Crepaldi

MoMo Lab, Department of Psychology, University of Milano-Bicocca, Milan, Italy

The last 40 years have witnessed a growing interest in the mechanisms underlying the visual identification of complex words. A large amount of experimental data has been amassed, but although a growing number of studies are proposing explicit theoretical models for their data, no comprehensive theory has gained substantial agreement among scholars in the field. We believe that this is due, at least in part, to the presence of several controversial pieces of evidence in the literature and, consequently, to the lack of a well-defined set of experimental facts that any theory should be able to explain. With this review, we aim to delineate the state of the art in the research on the visual identification of complex words. By reviewing major empirical evidences in a number of different paradigms such as lexical decision, word naming, and masked and unmasked priming, we were able to identify a series of effects that we judge as reliable or that were consistently replicated in different experiments, along with some more controversial data, which we have tried to resolve and explain. We concentrated on behavioral and electrophysiological studies on inflected, derived, and compound words, so as to span over all types of complex words. The outcome of this work is an analytical summary of well-established facts on the most relevant morphological issues, such as regularity, morpheme position coding, family size, semantic transparency, morpheme frequency, suffix allomorphy, and productivity, morphological entropy, and morpho-orthographic parsing. In discussing this set of benchmark effects, we have drawn some methodological considerations on why contrasting evidence might have emerged, and have tried to delineate a target list for the construction of a new all-inclusive model of the visual identification of morphologically complex words.

Keywords: morphological processing, visual identification, response times, ERPs, eye-tracking, benchmark effects, computational models

PAPER'S GOALS

Over the last 40 years, a growing number of studies have addressed the issue of morphological processing in the visual identification of complex words. While morphological effects have been consistently reported by a large number of studies, several issues are still matter of discussion, including whether processing unfolds along two different routes (e.g., Grainger and Ziegler, 2011) or just one (e.g., Crepaldi et al., 2010); whether semantics play a role since the very early processing stages (e.g., Feldman et al., 2009) or rather comes into play at a post-lexical level (e.g., Rastle et al., 2004); whether morphological analysis occurs automatically (e.g., Taft, 2004) or is context-dependent (e.g., Burani and Caramazza, 1987; Caramazza et al., 1988); and whether morphological effects need explicit morphemic representations to be accounted for (e.g., Baayen et al., 2006) or may simply emerge in the interaction between orthographic and semantic representation levels (e.g., Gonnerman et al., 2007; Baayen et al., 2011). General models of morphological processing conflict on how they deal with these issues, but the debate seems to have become somewhat inconclusive over the last decade: often new models are put forward

without previous models being clearly falsified, and without an explicit comparison that could clarify whether and how the new model extends the previous ones, both in its architecture and in its explanatory power. It is thus difficult to assign credit and blame to specific aspects of competing models, with the result that our knowledge in the field does not progress in a cumulative fashion (which means, someone might argue, that it does not progress at all). Several reasons lie behind this fact, but one fundamental issue, we believe, is that several pieces of evidence are still controversial: often scholars do not argue about the best interpretation of a given fact, but about whether that fact exists at all. Stated differently, we lack a list of uncontroversial experimental effects that any general theory should be able to explain. This is the issue that we have taken up in this paper, where we review morphological effects in visual word identification, trying to disentangle those that have received strong support from those that are still weak and require more experimental work. The aim of this paper is therefore to compile a list of reliable morphological effects in visual word identification that every model should be able to explain, in the hope that this will allow an easier adjudication process between existing theories

and, if necessary, the development of new theories in a cumulative, nested fashion (e.g., Grainger and Jacobs, 1996). Of course, this wish refers to general, all-inclusive models of the visual identification of complex words. In fact, the approach we are suggesting here does not exclude that specific models, more limited in scope, might be constructed to explain only a subset of the target list that we have illustrated above.

In achieving this goal, we will focus mainly on behavioral (i.e., response time based) effects for two reasons: first, in order to keep the discussion into manageable dimensions; and second, because all existing theories are defined in behavioral terms and thus can only license explicit and computationally defined predictions at this level. We also considered EEG and eye-tracking studies because their temporal resolution is fundamental in understanding the fine timing of behavioral effects, which is relevant for this special issue that is focused on the first 250 ms of visual word processing. Neuroimaging evidence will only be considered in support of behavioral data. We will also limit our review to those experimental paradigms that more directly tap onto visual word identification (such as masked priming and lexical decision), and in particular onto its early steps. Other tasks (such as, for example, word naming) or paradigms (long-SOA or cross-modal priming) will be considered only when the critical evidence can be reliably attributed to perceptual processes or to the purpose of contrasting early vs. late effects. Finally, in order to avoid any selection bias, we covered in this review any morphological effect in the visual identification of complex words that (i) we were aware of and (ii) could reliably be traced back to early processing steps. Any such effect that might be excluded from this review was only so because we failed to spot it in this vast literature.

MORPHOLOGICAL EFFECTS IN VISUAL WORD IDENTIFICATION

MORPHEME FREQUENCY EFFECTS

The morpheme frequency effect is generally interpreted as a diagnostic index of the use of morphemes as effective processing units in complex words recognition. Such effect has been repeatedly observed in psycholinguistic research, particularly in lexical decision experiments adopting a factorial approach (i.e., modeling frequency as a two-level variable – high vs. low). For example, Taft and colleagues (Taft, 1979; Taft and Ardasinski, 2006) described both surface and stem frequency effects in derived prefixed (e.g., reproach, dissuade) and inflected (e.g., sized, parents) words. These results for inflections were later confirmed in other languages (e.g., Italian: Burani et al., 1984; French: Colé et al., 1989; Dutch: Baayen et al., 1997; Finnish: Lehtonen et al., 2007). Morpheme frequency effects for both full form and constituents have also been observed with compound words using different methodologies (mainly eye-tracking and event-related potentials; see for example, Andrews, 1986; Juhasz et al., 2003; Pollatsek and Hyönä, 2005; Vergara-Martínez et al., 2009).

Obviously, stem frequency effects can only be appropriately studied when whole-word frequency is taken under control, which typically means that this latter variable was matched between the high and low-frequency stem words being compared. By adopting this approach, however, scholars were blind for years to the fact that stem frequency might be modulated by whole-word

frequency (Caramazza et al., 1988; Beauvillain, 1996; Baayen et al., 1997; Schreuder, 1997; Alegre and Gordon, 1999; Allen et al., 2003; Kuperman et al., 2008). This issue was explored by Baayen et al. (2007), who failed to find stem frequency effects in an experiment where only low-frequency words (derivations and inflections) were included. However, in a second experiment where target words spanned the entire whole-word and stem frequency range, stem frequency re-emerged as a significant factor, although modulated by whole-word frequency: stem frequency had in fact a facilitatory effect for the lowest frequency words, but an inhibitory effect for the highest frequency words. These findings emerged in an analysis of mean lexical decision times for around 8,000 words across 816 subjects as reported in the English Lexicon Project database (Balota et al., 2004), and are thus to be considered as the most reliable estimate of the stem frequency effect available to date.

Other studies have investigated whether frequency effects emerge independently of the context, or are rather contingent to, e.g., the presence of some specific type of filler items. Andrews (1986) showed that a stem frequency effect was present in the recognition of suffixed words only when compounds were also included in the experiment. A more recent study by Taft (2004) investigated word frequency effects in a lexical decision task where non-words had real vs. non-existent stems ("mirths" vs. "milphs"). This study showed that, when lexical decision is performed against nonsense stem non-words, high base-frequency words are easier to recognize than low base-frequency words as one would normally expect; but the reverse happens when lexical decision involves real-stem non-words. It does seem, then, that the overall characteristics of the entire experimental list presented to the subjects have an effect on stem frequency effects. (We point out, however, that this might not be relevant in simulation studies, where, typically, word response times are estimated as theoretical identification times with no reference to specific experimental contexts).

Some studies have gone more in depth and have tried to analyze the relationship that holds between stem and affix frequency effects. Burani and Thornton (2003), for example, demonstrated that lexical decision latencies depend on the interaction between root and suffix frequency in Italian derived words and pseudo-words. In a series of lexical decision experiments, they showed that suffixed pseudo-words (e.g., galmy, tudness) with higher frequency affixes present increased decision latencies and higher error rates, in comparison to pseudo-words with lower frequency affixes. They also showed an asymmetrical pattern for high-frequency and low-frequency roots whereby the former showed quicker and more accurate responses, while the latter did not differ from non-derived words, irrespectively of affix frequency. Results were interpreted to indicate that the main factor responsible for lexical decision performance is root frequency, with only a marginal role for affix frequency.

Finally a few studies addressed the role of affix productivity in modulating frequency effects. Bradley's (1979) study showed a stem frequency effect only for derived words with productive endings like "-ness" or "-ment," while derived words with less productive affixes showed only a surface frequency effect. These results were partially replicated by Vannest and Boland (1999; Experiment 1): however, the authors also report a lack of impact for root frequency when enlarging the item list to include 10

suffixes (productive: "-ship," "-ness," "-less," "-hood," "-er"; non-productive: "-ous," "-ory," "-ity," "-ian," "-ation") instead of the three used originally in Experiment 1 ("-less," "-ity," and "-ation"), therefore weakening the original claim that affix productivity is a crucial factor in the modulation of frequency effects.

In sum, there is strong evidence that stem frequency influences the identification times of complex words independently of affix characteristics (e.g., frequency and productivity). Substantial evidence (although without replication as yet) is also available that stem frequency effect interacts with whole-word frequency, namely, that it is facilitatory for low-frequency words, but inhibitory for high-frequency words. Finally, evidence shows that stem frequency effects might depend on testing condition, in particular on the composition of the stimulus list.

MORPHOLOGICAL PRIMING EFFECT

Morphological priming has been so extensively observed (e.g., Forster et al., 1987; Grainger et al., 1991; Marslen-Wilson et al., 1994; Frost et al., 1997; Rastle et al., 2000; Gonnerman et al., 2007; Crepaldi et al., 2010) that it does not make any sense to ask ourselves whether it exists or not: it is an established fact that prior exposure to a morphological relative – whether briefly or for relatively long time, in the same modality or in a different one – makes the visual identification of any given word faster and more accurate. It is interesting, however, to ask which variables affect morphological priming; this is much less obvious and likely to provide constraints on morphological theories of visual word identification.

Frequency

When the prime is consciously visible to participants, there is evidence showing that low-frequency primes yield larger time savings than high-frequency primes, at least for derived words (Raveh, 2002). This is confirmed by data in cross-modal priming experiments, which tap on central levels of processing similarly to what long-SOA paradigms do. For example, Meunier and Segui (1999) compared high- and low-frequency spoken primes (suffixed derived words) in a visual lexical decision task, and found reliable morphological effect only for the latter. Effects of target frequency on morphological priming appear to be weaker: to the best of our knowledge, they were only reported once and with derived targets (Meunier and Segui, 1999), which is not the standard condition under which morphological priming is evaluated.

However, data from masked priming paradigms are unclear as to whether prime frequency actually matters in early stages of the word identification process. For example, McCormick et al. (2009) are clear-cut in showing no sign of interaction between prime frequency and morphological facilitation in a study on derived words. These data seem to suggest that morphological decomposition is applied to all complex words regardless of their frequency. However, Giraudo and Grainger (2000) report larger effects with high-frequency derived primes than with low-frequency derived primes. One possibility is that the different results obtained in the two studies depend on the fact that Giraudo and Grainger (2000) used a longer SOA (57 ms vs. 42 ms), but this is clearly a speculation that calls for more direct experimental support.

Affix and stem priming

Morphological priming is typically investigated in experiments where primes and targets share their stem (e.g., dealer-DEAL). However, most of the recent morphological models do not attribute different roles to stems and affixes in visual identification (e.g., Crepaldi et al., 2010; Baayen et al., 2011; Grainger and Ziegler, 2011) and thus we should also be able to observe affix priming.

Giraudo and Grainger (2003) did report such an effect (both with prefixes and suffixes, at least when these latter coincided with a syllable), but only in comparison with an unrelated baseline (e.g., *enjeu-ENVOL* – in English: stake-FLIGHT – vs. *biche-ENVOL* – in English: deer-FLIGHT); affixed primes never yielded significant time savings as compared to pseudo-affixed primes (e.g., *engin-ENVOL* – in English: device-FLIGHT) where the initial (or final) letter sequences did not contribute any piece of meaning to the whole-word. Giraudo and Grainger (2003) do not specify whether words in their pseudo-affixed condition were entirely decomposable into existing morphemes (similar to the English example "corner"), which might justify why they did not differ from truly affixed words. In fact, given Longtin et al.'s (2003); Rastle et al.'s (2004) and several others' data on morpho-orthographic priming (see Rastle and Davis, 2008 for a review), a proper control condition for affix priming should be orthographically matched with the critical one, but should also contain undecomposable primes (similar to the form condition tested in those experiments, e.g., brothel-BROTH). Curiously, three affix priming studies include such a control condition, but their results are inconsistent. Chateau et al. (2002) tested prefix priming in English against an orthographically matched, monomorphemic condition (e.g., dislike-DISPROVE vs. violin-VIOLATE) and reported no significant effect. On the contrary, Dominguez et al. (2010) – working on prefixes – and Duñabeitia et al. (2008) – working on suffixes – obtained significant affix priming over and above orthographic effects. Although this might just be cross-linguistic variability, there is no obvious reason why affix priming should emerge in Spanish, but not in English. One obvious difference between these languages is that English is morphologically impoverished as compared to Spanish (perhaps in a reflection of a more general distinction between Germanic and Roman languages), but this does not seem to be related to affix saliency. More work is clearly required on this issue.

Semantic transparency

A series of studies, adopting a wide range of paradigms, have shown that semantics play a crucial role in modulating morphological priming in derived words (Sandra, 1990; Marslen-Wilson et al., 1994; Zwitserlood, 1994; Drews and Zwitserlood, 1995; Schreuder, 1997; Rastle et al., 2000; Longtin et al., 2003; Zwitserlood et al., 2005; Gonnerman et al., 2007; Meunier and Longtin, 2007; Rueckl and Aicher, 2008; Paterson et al., 2011). There seems to be universal agreement now that when primes are presented overtly (for at least 70 ms) or in the auditory modality, facilitation only emerges for semantically related prime-target pairs (Marslen-Wilson et al., 1994), or at least that facilitation is significantly larger with transparent than opaque pairs (Frost et al., 2000).

It has been hotly debated, however, whether this is also the case in masked priming experiments (i.e., when the prime is presented

for less than 60 ms, anticipated – and sometimes followed – by a visual mask). A substantial number of studies have reported that: (i) pseudo-related pairs of words (e.g., corner-CORN) give more facilitation than what would be expected on the basis of orthographic overlap; and (ii) that this facilitation is equivalent to that yielded by truly related words (e.g., dealer-DEAL; see Rastle et al., 2000; Longtin et al., 2003; Devlin et al., 2004; Feldman et al., 2004; Rastle et al., 2004; Gold and Rastle, 2007; Lavric et al., 2007; Kazanina et al., 2008; Marslen-Wilson et al., 2008; Kazanina, 2011). However, some studies do report different results (Diependaele et al., 2005, 2009; Morris et al., 2007; Feldman et al., 2009). Some of this apparently inconsistent evidence can be reconciled on methodological grounds (see Davis and Rastle, 2010). Diependaele et al. (2005), for example, used a backward mask, mixed written and spoken targets in the same experiment, and showed three repetitions of each prime-target pairs to their participants, one of which might have been visible to some of them (SOA = 67 ms). Morris et al. (2007) also made use of a backward mask. Feldman et al. (2009) had instead several prime-target pairs in their opaque set characterized by non-systematic changes in the stem (e.g., bliss-BLISTERY, coin-COYNESS, relay-RELATION, sack-SACCADE), which was much less frequently the case in their transparent set. It seems, then, that the only genuine failure to replicate the pattern described above is reported in Diependaele et al.'s (2009) Experiment 4. A first thing to note is that, in fact, this experiment confirmed that morpho-orthographic priming is larger than form priming; where Diependaele et al.'s (2009) results depart from the streamline is in showing that transparent pairs yield larger time savings than opaque pairs. One possibility to account for this result is quite unrelated to any specific feature of Diependaele et al.'s (2009) experiment. It would just be that transparent priming is indeed numerically larger than opaque priming, but by a margin that is too small to overcome consistently the standard RT variability in priming experiments, and is thus typically not able to reach significance in the vast majority of the cases. This state of affairs could explain Diependaele et al.'s (2009) result on the basis of normal cross-experiment variability, which might determine occasional significant results. Related to that, Morris et al. (2007) propose that there is a significant linear trend in the effect size across transparent, opaque, and orthographic condition. It is suggested that semantic transparency effects might be graded, with semantic pairs holding the greatest effects and orthographic pairs the smallest. Clearly, this is just speculation at present; more direct experimental work is needed before one can take into question the general result that morpho-orthographic priming is (i) larger than form priming and (ii) statistically indistinguishable from transparent priming, at least in the standard masked priming paradigm.

In fact, in a recent study by Duñabeitia et al. (2011) equal facilitatory effects were reported for morpho-semantic (walker-WALK), morpho-orthographic (corner-CORN), and form-related pairs (brothel-BROTH). This experiment involved a cross-case same-different task, a variant of the Forster and Davis (1984) paradigm that was originally designed by Norris and Kinoshita (2008) to tap onto very peripheral orthographic processing. These data clearly show that morpho-orthographic effects do not depend entirely on a fixed relationship between primes and targets, but are

sensitive to the task required to participants (see also Deutsch et al., 2003; Duñabeitia et al., 2007; Paterson et al., 2011); any complete model of the visual identification of complex words should be able to account for this fact.

Regularity

Irregularly inflected words such as "bought" are an issue for standard morphological theories. In fact, these latter consider morphemes as the smallest meaning-bearing orthographic/phonological units, thus implying a one-to-one mapping between orthography/phonology and semantics that is clearly absent in irregular words (e.g., there is no way of breaking down "bought" so that one orthographic element tells the reader what the word is about – i.e., buying something – and one orthographic element tells the reader that the word is a past tense form). This consideration has driven some scholars to propose a dual-route theory of morphology, whereby regular complex words are analyzed morphologically, whereas irregular words are stored as undivided wholes (and processed as such) in the mental lexicon (e.g., Pinker, 1991; Marslen-Wilson and Tyler, 1998; see Lavric et al., 2001 for discussion). Such proposals have implications for priming effects: because irregular words are not decomposed into their constituent morphemes, the visual identification system should fail to appreciate the morphological relationship with their stems, and so morphological priming should be absent between, e.g., "bought" and "buy," or "drove" and "drive" (once orthography and semantics are properly controlled).

It is not clear whether this prediction is met in response time, long-lag priming experiments. Stanners et al. (1979) found that irregular past tense forms prime their base form to a lesser extent than the base form itself (Experiment 2), but because no unrelated baseline was employed, we do not know whether irregular priming was present overall. Interestingly, somewhat different results emerged with irregular derivations (e.g., "descriptive," from "describe"), which appear to prime their base form to the same extent as regular derivations do (Stanners et al., 1979; Fowler et al., 1985). But this is a quite different issue, because, contrary to what happens in irregular inflected words, irregular derivations are still decomposable into separate and well-identified morphemes (e.g., "descriptive" into "descript-" and "-ive"), even if the stem does change in form.

In contrast, as far as masked priming is concerned, data seem to be clear-cut in showing that irregular inflected forms do facilitate the visual identification of their stems. In addition to the seminal work by Forster et al. (1987), the existence of morphological priming between irregular inflections and their base forms was documented by Kielar et al. (2008), Meunier and Marslen-Wilson (2004), and Pastizzo and Feldman (2002). Although these experiments all suffered from some methodological problems with control primes, their result were recently replicated in a study by Crepaldi et al. (2010), who provided new evidence that indeed masked irregular inflections prime their base forms, also showing that this does not depend on the system capturing morpho-orthographic sub-regularity in "lexical islands" (such as "meet," "bleed," "feed" and "breed," whose past tense forms are "met," "bled," "fed" and "bred"; or "spend," "send," "bend" and "lend," whose past tense forms are "spent," "sent," "bent" and "lent"): in

fact, there was no significant facilitation with pseudo-irregular past tense forms (e.g., red-REED, tent-TEND).

In the ERP literature, several studies using long-lag priming report dissociation in the ways regular and irregular inflected verbs are processed (Weyerts et al., 1996; Münte et al., 1999; Rodriguez-Fornell et al., 2002). For example, Weyerts et al. (1996) showed that regular infinitives prime their inflected forms (present participle or simple present), while priming effect for irregular verbs does not reach statistical significance. Moreover ERPs patterns for regular and irregular forms diverged in waveform, peak latencies, and amplitudes. For example, regular past participle forms primed by their infinitive forms showed a P200 effect as opposed to irregular past participle forms (Weyerts et al., 1996). Interestingly, this same component was reported for repetition priming trials within the same experiment, suggesting that (i) similar mechanisms, at least in terms of their time-course, underlie repetition and regular-form priming; and (ii) regular and irregular forms processing is, at least in terms of timing, qualitatively different (Weyerts et al., 1996). In an ERP repetition priming paradigm, Münte et al. (1999) found a reduced N400 effect for regular verb pairs (stretched-STRETCH) as compared to irregular verb pairs (fought-FIGHT), which could not be linked to phonological and orthographic factors. N400 is a well-known – although highly discussed – component in the psycholinguistic literature (Kutas and Federmeier, 2011). As far as morphological processing is concerned, it has been suggested to reflect facilitated access to word stems (Morris et al., 2007). Therefore, the decreased N400 observed for regular-forms priming may indicate that regular primes are able to activate their word stems more effectively than irregular primes.

More recently however, contrasting evidence emerged in a series of studies employing ERPs (Kielar and Joanisse, 2009) and event-related magnetic fields (Stockall and Marantz, 2006). In a visual lexical decision task (SOA = 200 ms), Kielar and Joanisse (2009), compared neural responses to regular (baked-BAKE), vowel-change irregular (sang-SING), and suffixed irregulars (slept-SLEEP) prime-target pairs. The authors reported a strong N400 effect only for regular verbs seemingly indicating that regular and irregular verbs are processed differently. However, subsequent analyses differentiating early vs. late components of the N400 revealed temporal changes in the ERP pattern: while the early time interval (324–400 ms) showed the influence of formal relationship between prime and target (N400 effect for regular and ortho-phonologically overlapping pairs), the late time interval (400–476 ms) showed an effect for morphologically related pairs (regular and irregular). It appears that the difference between regular and irregular pairs might be graded and affected by the interaction of formal, semantic and phonological factors.

These results seem to confirm what was previously reported by Stockall and Marantz (2006) in a long-term priming, lexical decision, MEG study. These authors compared magnetic responses to regular and irregular prime (past participle)-target (base form) pairs, where orthographic overlap and priming direction were manipulated so as to build eight conditions tested in two separated experiments: irregular low overlap (taught/TEACH) vs. irregular high overlap (gave/GIVE) vs. identity (boil/BOIL) vs. orthographic overlap (curt/CART; Experiment 1); and irregular low overlap (teach-TAUGHT) vs. irregular high overlap (give-GAVE)

vs. regular (date-DATED) vs. orthographic and semantic relation (boil-BROIL; Experiment 2). In both experiments, regular and irregular participle primed their base forms to a similar extent, with similar latencies of the M350 component – an index of root activation – in all morphologically related conditions. However it was shown that the M350 effect depended crucially on orthographic overlap and on priming direction. High orthographic overlap pairs (gave-GIVE) showed priming effects in both directions (gave-GIVE and give-GAVE); on the contrary, low orthographic overlap pairs showed an effect only when the inflected form was used as a prime (teach-TAUGHT). More interestingly, pairs that shared orthographic and semantic elements, like "boil-BROIL," failed to show any priming effect. This data was interpreted as evidence that morphological effects cannot be explained solely on the bases of orthographic, phonological or semantic relatedness.

Taken altogether, the pattern shown in electrophysiological studies seem to suggest that regularity effects emerge only at later stages of lexical processing and that they are sensitive to pattern of sub-regularities which could be represented as the probabilistic combination of orthographic, phonological, and semantic elements (Justus et al., 2008). In conclusion, then, both behavioral and electrophysiological evidence suggests that regular and irregular inflections are processed in a similar fashion early after stimulus presentation, thus providing support for the existence of a single mechanisms operating at least during the initial stages of lexical access.

Free and bound stems

Morphological theories differ substantially as to whether free stems (stems that are existing words themselves; e.g., "form") and bound stems (stems that cannot be used as words in isolation; e.g., "-mit," as in "submit," "permit," and "commit") have the same mental representation (e.g., Taft and Kougious, 2004; Crepaldi et al., 2010). It is thus not obvious whether these two types of morphemes should give rise to equivalent priming effects.

Forster and Azuma (2000) investigated this issue and discovered that bound and free stems produce equivalent facilitation, which in both cases could not be attributed solely to orthographic factors. Moreover, they found that priming with bound stems depends on affix and stem productivity (roughly, the number of different complex words where they appear). Forster and Azuma's (2000) data were closely replicated by Pastizzo and Feldman (2004; see also Järvikivi and Niemi, 2002), using both orthographic and unrelated pairs as a baseline. In particular, these authors reported that bound stem priming correlates with the number of morphological relatives (in line with Forster and Azuma, 2000), whereas free stem priming does not.

In conclusion, there is consistent evidence that free and bound stem give rise to equivalent priming effects, even though bound stem priming seems to depend on affix and stem distributional properties.

TRANSPOSED-LETTER EFFECTS AND MORPHEME BOUNDARIES

After the seminal report by Forster et al. (1987) showed that transposed-letter (TL) primes ("anwser" for "answer") are as effective as identity primes in facilitating visual word identification, a

number of experiments have documented the so-called "jumbled word effect" (Grainger and Whitney, 2004), namely, that the word identification system tolerates imprecisions in letter position so that it tends to identify some kind of transposed-letter non-words as their corresponding words (e.g., "jugde" as "judge"; e.g., Perea and Lupker, 2003, 2004; Schoonbaert and Grainger, 2004; Lupker et al., 2008; Duñabeitia et al., 2009b). This phenomenon has crossed the morphology literature when it was shown that primes containing letter transpositions within morphemes (e.g., sunhsine) facilitate naming as much as correctly spelled primes, but primes with letter transpositions across morpheme boundaries (e.g., susnhine) do not yield any time saving as compared to substituted-letter primes (e.g., sumzhine; Christianson et al., 2005). This effect also held for pseudo-compounds (e.g., mayhem) and derived words (e.g., grinder), and was replicated by Duñabeitia et al. (2007) (i) in two more languages (Basque and Spanish), (ii) in a more standard lexical decision paradigm, and (iii) with stronger statistical support. These results were taken to show that morphological decomposition operates early, most likely before lexical identification has taken place. In line with this suggestion, Lemhöfer et al. (2011) showed that Dutch readers are quicker at recognizing compounds when their morpheme boundary is flagged by a low-frequency letter bigram (at least when the compound word was a long one). Because bigram frequency is sub-lexical information, these results strengthen the idea that morphological segmentation kicks off well before lexical identification has taken place.

However, the difference between cross-morpheme and within-morpheme TL effects does not prove to be very solid. In fact, neither Rueckl and Rimzhim (2011) in English nor Perea and Carreiras (2006) in Spanish provide converging evidence that TL effects decrease over morphemic boundaries. There are differences between these contrasting experiments that might explain inconsistencies; for example, Perea and Carreiras (2006) used compound words, whereas Duñabeitia et al. (2007) used affixed words. However, taking this into consideration does not help to reconcile the existing evidence into a coherent and clear frame. For example, on the basis of the Spanish data one might suggest that morphological modulation of TL effects emerges in affixed, but not in compound words. This proposal is contradicted by the English data, where compound words generate interaction between morphemic boundaries and TL effects (Christianson et al., 2005), but mixed results were obtained on affixed words (Christianson et al., 2005, and Rueckl and Rimzhim, 2011). Clearly, more work is necessary before it will be possible to take a stand on this issue.

MORPHOLOGICAL EFFECTS IN NON-WORD PROCESSING
It has long been debated whether the visual word identification system gets access to morphological information *before* lexical identification (readers would identify morphemes first, and then words; e.g., Taft, 1994), or rather *upon* lexical identification (readers would identify words first, and then become aware of their morphological structure; e.g., Giraudo and Grainger, 2001). Crucial for this debate is what happens to non-words that are morphologically structured (e.g., shootment), for which, clearly, lexical identification never occurs; observing morphological effects on

this type of stimuli would thus be strong evidence for pre-lexical morphological processing.

In a seminal study, Taft and Forster (1975) reported that non-words composed of an existing prefix and an existing stem (dejuvenate) are slower to be rejected than non-words composed of an existing prefix and a non-stem (depertoire). In a similar way, compound non-words where the first constituent is a word (footmilge) take longer to be rejected as non-word in comparison to compound non-words where the second constituent is a word (thernlow; Taft and Forster, 1976). This pattern was more recently confirmed by an Italian ERP study using a lexical decision task to compare neural responses to compound and simple words and non-words (El Yagoubi et al., 2008). This study provided clear evidence that non-words composed by an existing word and a non-word (*drillococco* – in English: drilecoconut) elicited a more negative N400 than non-words composed by two existing words (*spadapesce* – in English: fishsword), thus suggesting that existing stems embedded in non-words might trigger lexical access, mitigating the difference between words and non-words (see also, Fiorentino and Poeppel, 2007).

This morpheme interference effect was then generalized to the inflectional domain and to derived, pseudo-suffixed words (although with more controversial data). Caramazza et al. (1988) showed that pseudo-inflected Italian non-words ("*cantevi,*" similar to the English "buyed") were rejected more slowly than non-words made up of a real-stem and a non-suffix ("*cantovi,*" similar to "buyel"), a non-stem and an existing suffix ("*canzevi,*" similar to "beyed"), and a non-stem and a non-suffix ("*canzovi,*" similar to "beyel"; see also Leinonen et al., 2009, Experiment 1, for convergent ERP data in Finnish). Again testing Italian readers, Burani et al. (1997) reported that suffixed non-words (e.g., "*vetrezza,*" lit. "glassness") are more difficult to reject in a lexical decision task than non-words composed of an existing stem and a non-suffix (e.g., "*vetralle,*" similar to "glassmilp" in English), but only when the final part of the word is a frequent word-ending. In apparent contrast with these data, Burani et al. (2002) obtained no difference between rejection times on suffixed non-words (e.g., "*donnista,*" lit. "womanist") and rejection times on orthographically controlled non-words that did not contain any morpheme (e.g., "*dennosto,*" similar to "wemanost" in English); a difference between the two conditions, however, emerged in the analysis of the error rates. More recently, Crepaldi et al. (2010) investigated the same issue with English material, and confirmed the pattern of results obtained by Burani et al. (1997), i.e., that suffixed non-words (e.g., gasful) take longer to be rejected than orthographic controls with non-morphological endings (e.g., gasfil). In consideration of the fact that similar morpheme interference effects have also been reported for pseudo-compounds (e.g., "pipemeal"; Taft, 1985), we would conclude that, even if some inconsistent result does appear in the literature, there is sufficient evidence to hold that morphologically structured non-words are more difficult to reject than appropriately matched orthographic controls. Incidentally, this pattern of results fits well with the ERP evidence provided by McKinnon et al. (2003), who showed similar brain responses for real words and morphologically structured non-words, thus indicating similar processes for the two types of stimuli.

Interestingly, the importance of these data on the role of morphemes in non-word processing was further strengthened by the report of masked morphological priming with non-word primes. For example, Meunier and Longtin (2007) found that response times on stem words such as "sport" are made faster by morphologically related non-word primes, such as "sportation." This was shown to be independent from whether non-words were semantically interpretable (e.g., quickify vs. sportation), or designed to be synonymous with existing words (e.g., "brightment," which most people would consider to mean the same thing as "brightness"). These data were confirmed in English by McCormick et al. (2009).

On the whole, then, it is clear that non-words with a morphological structure are analyzed in terms of their morphemes, thus questioning seriously any theory that suggests morphological processing to kick off upon lexical identification.

MORPHEME POSITION EFFECTS

Capitalizing on the morpheme interference effects described in the previous paragraph, scholars have recently started to investigate how morpheme position is coded in the visual identification system. This is an important issue from a theoretical point of view, because no morphological model proposed so far has taken a stand in this respect.

Crepaldi et al. (2010) have reported evidence that suffix position coding is locked to word-final positions (or at least to post-stem positions). These authors showed that, while "shootment" is slower to be rejected than its orthographic control "shootmant" (see Burani et al., 1997), "mentshoot" and "mantshoot" are equally difficult; this was taken as a proof that "ment" is not identified as a suffix in "mentshoot" (i.e., in word-initial position), which is evidence that its representation in the visual identification system is position-specific.

More work was carried out on free stem position coding, i.e., on constituent coding in compounds (and pseudo-compounds; e.g., Taft, 1985; Taft et al., 1999; Shoolman and Andrews, 2003; Duñabeitia et al., 2009a). The evidence accumulated so far is suggestive of two facts, namely, (i) that free stems are coded in a position-independent fashion (i.e., they are identified even when they lie in unusual positions, as for "honey" and "moon" in "moonhoney"), and (ii) that their position is coded flexibly, so that, e.g., "moon" in "moonhoney" drives some activation to the word "honeymoon," even if the position of the stem in the stimulus (word-initial) and in the target word (word-final) do not match. These conclusions are based on the observation that reversed compounds (e.g., "doorback") seem to take longer to be rejected than control pseudo-compounds (e.g., pipemeal; Taft, 1985; Taft et al., 1999; Shoolman and Andrews, 2003), and that constituent priming occurs in a cross-position fashion (e.g., "hang*over*" primes "*over*come"; Duñabeitia et al., 2009a). A word of caution is necessary here however, because this evidence comes either from experiments where morpheme position was not the main issue, and thus some methodological details were not clear of problems (Taft, 1985; Shoolman and Andrews, 2003). More direct evidence on this issue would be desirable.

STEM HOMOGRAPHS EFFECT

Stem homographs are complex words with stems that are orthographically identical, but semantically and – theoretical linguists might say – morphologically unrelated. Examples of these words abound in Neo-Latin languages such as Italian ("*colp-a*," "fault," and "*colp-o*," "stroke") and Spanish ("*mor-os*," "moors," and "*mor-ir*," "to die"), and have been quite extensively studied in the nineties (Laudanna et al., 1989, 1992; Allen and Badecker, 1999, 2002; Badecker and Allen, 2002). This type of words is interesting because of its close relationship with morpho-orthographic effects: stem homographs share in fact an orthographically defined stem (just as "corner" and "corn" do) and are entirely decomposable into existing morphemes.

In two very early studies, Laudanna et al. (1989, 1992) reported an inhibitory effect by stem homographs in Italian, which was later confirmed by Allen and Badecker (1999) in Spanish (see also Barber et al., 2002; Carreiras et al., 2005; and Domínguez et al., 2004 for converging eye-tracking and ERP evidence). These were all long-SOA priming studies that allowed participants to fully process primes; it is not surprisingly, then, that stem allomorphs inhibit each other (most likely because of competition at the semantic level). In line with this consideration, and with the more recent literature on morpho-orthographic segmentation, stem homographs were found to facilitate each other in a masked priming experiment (Badecker and Allen, 2002), where instead participants were prevented from processing primes up to the semantic level.

Interestingly, Domínguez et al. (2004), using event-related potentials, were able to trace the time-course of the stem-inhibition effect reported in long-SOA priming studies, and to disentangle the effect from orthographic confounds. In a lexical decision, long-lag priming experiment (SOA = 200 ms), the authors reported an early (250–350 ms time window) overlap of stem homographic (*foco-FOCA* – in English: floodlight-SEAL) and morphological (*hijo-HIJA* – in English: son-DAUGHTER) priming waves. However, starting from 350 ms, the two wave patterns started to differ, with stem homographs producing a delayed N400 effect. Interestingly, orthographic pairs (*rasa-RANA*- in English: flat-FROG) did not produce any facilitative effect in the 250–350 ms time window, but later showed a N400 effect comparable to the one elicited by unrelated pairs.

The evidence available thus indicates that at early steps in lexical access, stem homographs have access to a common representation; however, at a later stage of semantic processing, they seem to activate two different and competing mental representations, thus resulting in the inhibitory effect commonly observed in long-SOA priming studies.

PARADIGMATIC EFFECTS: FAMILY SIZE AND ENTROPY

Two morphological effects were described over the last 15 years in the lexical decision task that do not refer to the morphological structure of the word-to-be-processed itself, but rather to the morphological family where that word belongs. This refers to the family size effect (e.g., Schreuder, 1997; Bertram et al., 2000a; Pylkkänen et al., 2004; Juhasz and Berkowitz, 2011), whereby words with more morphological relatives are processed faster than words with a few morphological relatives, and to entropy effects (e.g., Moscoso del Prado Martín et al., 2004), whereby words with equally frequent morphological relatives are processed faster than words whose morphological family is characterized by a few very

dominant members. These effects were observed in the processing of both simple (e.g., Baayen et al., 2006) and complex words (e.g., Bertram et al., 2000a; Kuperman et al., 2010; Baayen et al., 2011), and were also shown to hold independently of other, more established, lexical variables, such as cumulative family frequency, surface frequency, and neighborhood density (Schreuder, 1997). Interestingly, Schreuder (1997) also showed how family size effect progressively decreases with priming demasking, thus indicating that the effect is most likely semantic in nature, and emerges at a later, post-identification stage of lexical processing (see also De Jong et al., 2000). This effect is also one of the very few which have been shown to hold across different language families (Indo-European vs. Semitic; Moscoso del Prado Martín et al., 2005), which strengthens its reliability.

AFFIX DISTRIBUTIONAL PROPERTIES: ALLOMORPHY AND PRODUCTIVITY

Other factors that might affect how a morphologically complex word is processed are connected to the distributional properties of its constituent morphemes, in particular, allomorphy and productivity. These features have been suggested to concur to determine affix salience (Schreuder and Baayen, 1994; Laudanna and Burani, 1995; Burani et al., 1997; Järvikivi et al., 2006), and, in turn, to affect the probability of an affix to be activated as a specific processing unit during word recognition (Allen and Badecker, 1999; Bertram et al., 1999, 2000b), thus balancing storage and parsing processes for what concerns both inflected and derived words (Bertram et al., 1999, 2000b).

In lexical decision studies, words including affixes with several allomorphs resulted in longer latencies (Laudanna and Burani, 1995; Järvikivi et al., 2006). Moreover, Allen and Badecker (1999) showed an inhibitory effect for Spanish targets that were preceded by primes allomorphically related to their homographs (e.g., "*cierra*," (he) closes, whose stem, "*cierr-*," is an allomorph of the main stem of the verb "to close", "*cerr-*", inhibited "*cerro*", hill) (see Linares et al., 2006, Experiment 2, for convergent ERP results).

Affix productivity has been defined in several different ways, which makes quite difficult to establish its role in the visual identification of complex words. Laudanna et al. (1994) used as an index of productivity the proportion between the number of words in which a given affix appeared as such (e.g., "driver" for "-er") and the number of words in which the same affix did not play any morphological role (e.g., "corner" for "er"). Adopting this definition, they found that non-words including productive affixes were harder to reject than non-words including non-productive affixes. Investigating Finnish and Dutch, Bertram et al. (1999, 2000b) came to somewhat different conclusions. Without giving any exact definition of productivity, but using affixes supposedly lying at the opposite extremes of its distribution, Bertram and colleagues conclude that productivity does not have a well-identifiable effect on processing times, but interacts with word formation type (derivation vs. inflection) and affixal homonymy (an interaction that has received no independent confirmation). Finally, Plag and Baayen (2009) report effects of the number of words including any given affix on word naming times, but not on lexical decision times, again in apparent contrast with what found by Laudanna et al. (1994). All in all, there does not seem to be clear evidence to hold that productivity, however defined, influences word identification times.

INFLECTION, DERIVATION, AND COMPOSITION

In closing this review, we turn our attention to an issue that is cause of pain to many scholars in the field, namely, that the literature on inflection, derivation, and (in particular) compounding appears to be somewhat disconnected, perhaps under the assumption that these morphological processes are too different from each other to be reciprocally informative.

Indeed inflection, derivation and composition are very different morphological processes. Inflectional processes do not result in a new lexical entity, while derivation and composition always do (Kurylowicz, 1964). Inflection never involves a change in grammatical class, which is instead most frequently the case in derivational processes (e.g., deal-dealer). Inflection generally preserves the meaning of the stem, whereas this is not always the case in derivation (e.g., angel-angelic; Aronoff, 1976). Again, whereas inflection implies a consistent and predictable semantic change ("table" and "tables" entertain the same semantic relationship that holds between "idea" and "ideas" or "cat" and "cats"), this is much less the case in the derivational domain (e.g., while a "gardener" is a professional who takes care of gardens, a "juicer" is a kitchen appliance) and in compounding ("honey" has very different meanings in "honeycomb" and "honeymoon").

Most of these differences are based on syntactic and semantic processes, which are unlikely to be in action very early after stimulus presentation. In fact, we would claim that, at least for what concerns the more peripheral stages of visual word identification, there is not much psycholinguistic evidence suggesting different processing of inflected, derived and compound words.

In support of this statement, Leinonen et al. (2008) and Álvareza et al. (2011) reported that ERPs patterns for inflected and derived words start to diverge around the 300–450 ms time window, with effects spilling over to the 450–550 ms time window for inflected words, thus suggesting that differences between inflection and derivation is apparent only at a later stage of lexical processing, when semantics is more likely to come into play.

Support in this direction also comes from a paper by Raveh (2002), where – in a rare direct comparison between derivational and inflectional priming – inflected and derived words yielded equivalent time savings in the identification of their stems at a brief SOA (50 ms), whereas a difference emerged later on (inflected words gave more priming at SOAs of 150 ms and 250 ms).

Substantial similarity between morphological effects with derived and compound words also emerges when considering morpho-orthographic segmentation. The vast majority of this literature has investigated derived and pseudo-derived words (see above), but in a recent paper Fiorentino and Fund-Reznicek (2009) reported significant and equivalent masked priming effects for transparent (teacup-TEA) and opaque compounds (honeymoon-HONEY, carpet-CAR), as compared to orthographic, non-morphological controls (penguin-PEN). The effect held for both initial and final constituent word priming (flagpole-FLAG vs. classroom-ROOM), and clearly mirrors what has been reported for derived words, thus suggesting that the early morpho-orthographic segmentation proposed by Rastle et al. (2004) generalizes to all types of morphologically complex words.

Perhaps even more strikingly, data gathered on inflected and compound words are closely similar for what concerns the rejection time of morphologically structured non-words in

lexical decision tasks. In fact, it has been documented, for both pseudo-inflected and pseudo-compound non-words, that non-words made up entirely by non-existing morphemes (e.g., "iblish" and "thrimnade") or by a non-existing first element and an existing second morpheme (e.g., "ibvive" and "flurbpair") are easier to reject than non-words made up of a real morpheme as a first element followed by a non-existing second element (e.g., "inlish" and "spellcung"). In turns, these latter non-words are easier to judge than non-words entirely made up of real morphemes (e.g., "invive" and "toastpull"; see Taft and Forster, 1975; Lima and Pollatsek, 1983; Taft et al., 1986; and Monsell, 1985).

Clearly, this evidence is far from suggesting that the visual identification system processes inflected, derived and compound words in exactly the same way. However, it does suggest that at least some (peripheral) processing steps are common to all types of complex words and, more generally, that there should be a more tight integration between the literature on inflected, derived and compound words.

THE TARGET LIST

In this paper we reviewed the behavioral literature on the visual identification of complex words with the aim of building a list of established facts that might help in adjudicating between existing theories, and eventually in developing a comprehensive computational model of how complex words get identified by the visual system.

The list should include these effects:

- Stem frequency has a facilitatory effect on low-frequency words, and an inhibitory effect on high-frequency words;
- Non-words that are morphologically structured are more difficult to reject in lexical decision, no matter whether they are pseudo-prefixed, pseudo-suffixed, or pseudo-inflected;
- Non-words that are morphologically structured, but that contain a suffix at their onset are as easy to reject in lexical decision than orthographic control non-words;
- Words with larger family size are identified more quickly;
- Words with higher entropy are identified more quickly;
- Words including affixes with several allomorphs yield longer lexical decision times;

In unmasked priming:

- Low-frequency complex words yield time savings on the identification of their stems more than high-frequency complex words do;
- Morphological effects emerge only for semantically related prime-target pairs;
- Stem homographs (and their allomorphs) inhibit each other;
- Inflectional priming is larger than derivational priming;

In masked priming:

- Morphological effects emerge to the same extent for transparent and opaque prime-target pairs (but when masked priming is employed in tasks other than lexical decision, facilitation might not emerge at all for both transparent and opaque pairs);

- Morphologically structured non-words facilitate the identification of words sharing their stem;
- Irregularly inflected words prime their stems;
- Both free and bound stems determine time savings when they are shared between primes and targets;
- Bound stem priming is proportional to stem productivity (i.e., the number of different complex words where they appear);
- Stem homographs facilitate each other;
- Inflectional and derivational priming are equivalent;

From a theoretical point of view, it is not easy to see in a glimpse whether these effects speak clearly against or in favor of any existing theory. Surely, morphological effects in non-words exclude the possibility that morphological information only comes into play after lexical identification. For what concerns the other big dichotomies illustrated at the beginning of the paper (e.g., one vs. dual-route models; PDP vs. localist models), there is no clear indication popping out. This is exactly where computational modeling comes as a useful tool; in fact, by implementing theories in a computer program it becomes easier to understand unequivocally which model survives confrontation with the data (in particular for what concerns the simulation of several effects with the same system settings), and which does not.

Obviously, this list is by no means definitive (new evidence is continuously arising on what seems to be a hotly debated topic), nor necessarily complete. We made all our efforts to ensure that we covered all the relevant data, but with such a huge amount of evidence amassed over the last 40 years, it is possible that we have missed some important results. We encourage anyone to flag possible gaps, also taking advantage of the brilliant "Comment" tool made available upon the open-access policy adopted by this Journal.

The main point that we want to make with this paper, however, is not about the list *per se*; rather, we hope that having a list of benchmark effects will help the field to move forward in a more cumulative and cooperative fashion. In the spirit of the nested modeling principle put forward more than a decade ago in the related field of reading aloud (Grainger and Jacobs, 1996), we hope that in the near future (i) existing models will confront on the basis of their ability to account for these (or other) benchmark effects; (ii) credit and blame will be assigned to specific parts of each theory for their successes and failures in this attempt; (iii) in proposing any new theory, substantial effort will be spent in explaining how the new theory relates with its predecessors, how it extends them, why it does that in the way that it does, which new effects it is able to explain that its predecessors were not able to explain, and which effects it is still not able to explain that were also outside the grasp of its predecessors.

ACKNOWLEDGMENTS

REFERENCES

Alegre, M., and Gordon, P. (1999). Frequency effects and the representational status of regular inflections. *J. Mem. Lang.* 40, 41–61.

Allen, M., and Badecker, W. (1999). Stem homograph inhibition and stem allomorphy: representing and processing inflected forms in a multilevel lexical system. *J. Mem. Lang.* 41, 105–123.

Allen, M., and Badecker, W. (2002). Stem homographs and lemma level representations. *Brain Lang.* 81, 79–88.

Allen, M., Badecker, W., and Osterhout, L. (2003). Morphological analysis in sentence processing: an ERP study. *Lang. Cogn. Processes* 18, 405–430.

Álvareza, C. J., Urrutiaa, M., Domínguez, A., and Sánchez-Casas, R. (2011). Processing inflectional and derivational morphology: electrophysiological evidence from Spanish. *Neurosci. Lett.* 490, 6–10.

Andrews, S. (1986). Morphological influences on lexical access: lexical or nonlexical effects? *J. Mem. Lang.* 25, 726–740.

Aronoff, M. (1976). *Word Formation in Generative Grammar*. Massachusetts: MIT Press.

Baayen, R. H., Dijkstra, T., and Schreuder, R. (1997). Singulars and plurals in Dutch: evidence for a parallel dual-route model. *J. Mem. Lang.* 37, 94–117.

Baayen, R. H., Feldman, L. B., and Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *J. Mem. Lang.* 55, 290–313.

Baayen, R. H., Milin, P., Durdevic, D. F., Hendrix, P., and Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychol. Rev.* 118, 438–481.

Baayen R. H., Wurm, L. H., and Aycock, J. (2007). Lexical dynamics for low-frequency complex words: a regression study across tasks and modalities. *The Mental Lexicon* 2, 419–463.

Badecker, W., and Allen, M. (2002). Morphological parsing and the perception of lexical identity: a masked priming study of stem homographs. *J. Mem. Lang.* 47, 125–144.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., and Yap, M. (2004). Visual word recognition of single-syllable words. *J. Exp. Psychol. Gen.* 132, 283–316.

Barber, H., Dominguez, A., and de Vega, M. (2002). Human brain potentials indicate morphological decomposition in visual word recognition. *Neurosci. Lett.* 318, 149–152.

Beauvillain, C. (1996). The integration of morphological and whole-word form information during eye fixations on prefixed and suffixed words. *J. Mem. Lang.* 35, 801–820.

Bertram, R., Baayen, R. H., and Schreuder, R. (2000a). Effects of family size for complex words. *J. Mem. Lang.* 42, 390–405.

Bertram, R., Schreuder, R., and Baayen, R. H. (2000b). The balance of storage and computation in morphological processing: the role of word formation type, affixal homonymy, and productivity. *Cognition* 26, 489–511.

Bertram, R., Laine, M., and Karvinen, K. (1999). The interplay of word formation type, affixal homonymy, and productivity in lexical processing: evidence from a morphologically rich language. *J. Psycholinguist. Res.* 28, 213–226.

Bradley, D. C. (1979). "Lexical representation of derivational relation," in *Juncture*, eds M. Aronoff and M. L. Kean (Cambridge: MIT Press), 37–55.

Burani, C., and Caramazza, A. (1987). Representation and processing of derived words. *Lang. Cogn. Process* 2, 217–227.

Burani, C., Dovetto, F. M., Thornton, A. M., and Laudanna, A. (1997). "Accessing and naming affixed pseudo-words," in *Yearbook of Morphology 1996*, eds G. Booij, and J. von Marle (Dordrecht: Kluer Academic Publishers), 55–72.

Burani, C., Marcolini, S., and Stella, G. (2002). How early does morpho-lexical reading develop in readers of a shallow orthography? *Brain Lang.* 81, 568–586.

Burani, C., Salmaso, D., and Caramazza, A. (1984). Morphological structure and lexical access. *Visible Lang.* 18, 342–352.

Burani, C., and Thornton, A. M. (2003). "The interplay of root, suffix and whole-word frequency in processing derived words," in *Morphological Structure in Language Processing*, eds R. H. Baayen, and R. Schreuder (Berlin: Mouton de Gruyter), 157–208.

Caramazza, A., Laudanna, A., and Romani, C. (1988). Lexical access and inflectional morphology. *Cognition* 28, 297–332.

Carreiras, M., Perdomo, A., and Meseguer, E. (2005). Are stem homographs and orthographic neighbours processed differently during silent reading? *Lang. Cogn. Process* 20, 317–339.

Chateau, D., Knudsen, V. E., and Jared, D. (2002). Masked priming of prefixes and the influence of spelling–meaning consistency. *Brain Lang.* 81, 587–600.

Christianson, K., Johnson, R. L., and Rayner, K. (2005). Letter transpositions within and across morphemes. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 1327–1339.

Colé, P., Beauvillain, C., and Segui, J. (1989). On the representation and processing of prefixed and suffixed derived words: a differential frequency effect. *J. Mem. Lang.* 28, 1–13.

Crepaldi, D., Rastle, K., Coltheart, M., and Nickels, L. (2010). "Fell" primes "fall," but does "bell" prime "ball"? Masked priming with irregularly-inflected primes. *J. Mem. Lang.* 63, 83–99.

Davis, M. H., and Rastle, K. (2010). Form and meaning in early morphological processing: comment on Feldman, O'Connor, and Moscoso del Prado Martin (2009). *Psychon. Bull. Rev.* 17, 749–755.

De Jong, N., Schreuder, R., and Baayen, R. H. (2000). The morphological family size effect and morphology. *Lang. Cogn. Process* 15, 329–365.

Deutsch, A., Frost, R., Pelleg, S., Pollatsek, A., and Rayner, K. (2003). Early morphological effects in reading: evidence from parafoveal preview benefit in Hebrew. *Psychon. Bull. Rev.* 10, 415–422.

Devlin, J. T., Jamison, H. L., Matthews, P. M., and Gonnerman, L. M. (2004). Morphology and the internal structure of words. *Proc. Natl. Acad. Sci. U.S.A.* 101, 14984–14988.

Diependaele, K., Sandra, D., and Grainger, J. (2005). Masked cross-modal morphological priming: unravelling morpho-orthographic and morpho-semantic influences in early word recognition. *Lang. Cogn. Process* 20, 75–114.

Diependaele, K., Sandra, D., and Grainger, J. (2009). Semantic transparency and masked morphological priming: the case of prefixed words. *Mem. Cogn.* 37, 895–908.

Dominguez, A., Alija, M., Rodriguez-Ferreiro, J., and Cuetos, F. (2010). The contribution of prefixes to morphological processing of Spanish words. *Eur. J. Cogn. Psychol.* 22, 569–595.

Domínguez, A., de Vega, M., and Barber, H. (2004). Event-related brain potentials elicited by morphological, homographic, orthographic, and semantic priming. *J. Cogn. Neurosci.* 16, 598–608.

Drews, E., and Zwitserlood, P. (1995). Morphological and orthographic similarity in visual word recognition. *J. Exp. Psychol. Hum. Percept. Perform.* 21, 1098–1116.

Duñabeitia, J. A., Kinoshita, S., Carreiras, M., and Norris, D. (2011). Is morpho-orthographic decomposition purely orthographic? Evidence from masked priming in the same – different task. *Lang. Cogn. Process* 26, 509–529.

Duñabeitia, J. A., Laka, I., Perea, M., and Carreiras, M. (2009a). Is Milkman a superhero like Batman? Constituent morphological priming in compound words. *Eur. J. Cogn. Psychol.* 21, 615–640.

Duñabeitia, J. A., Molinaro, N., Laka, I., Estévez, A., and Carreiras, M. (2009b). N250 effects for letter transpositions depend on lexicality: "casual" or "causal"? *Neuroreport* 4, 381–387.

Duñabeitia, J. A., Perea, M., and Carreiras, M. (2007). Do transposed-letter similarity effects occur at a morpheme level? Evidence for morpho-orthographic decomposition. *Cognition* 105, 691–703.

Duñabeitia, J. A., Perea, M., and Carreiras, M. (2008). Does darkness lead to happiness? Masked suffix priming effects. *Lang. Cogn. Process* 23, 1002–1020.

El Yagoubi, R., Chiarelli, V., Mondini, S., Perrone, G., Danieli, M., and Semenza, C. (2008). Neural correlates of Italian nominal compounds and potential impact of headedness effect: an ERP study. *Cogn. Neuropsychol.* 25, 559–581.

Feldman, L. B., O'Connor, P. A., and Del Prado Martín, F. M. (2009). Early morphological processing is morphosemantic and not simply morpho-orthographic: a violation of form-then-meaning accounts of word recognition. *Psychon. Bull. Rev.* 16, 684–691.

Feldman, L. B., Soltano, E. G., Pastizzo, M. J., and Francis, S. E. (2004). What do graded effects of semantic transparency reveal about morphological processing? *Brain Lang.* 90, 17–30.

Fiorentino, R., and Fund-Reznicek, E. (2009). Masked morphological priming of compound constituents. *Ment. Lex.* 4, 159–193.

Fiorentino, R., and Poeppel, D. (2007). Compound words and structure in the lexicon. *Lang. Cogn. Process* 22, 953–1000.

Forster, K. I., and Azuma, T. (2000). Masked priming for prefixed words with bound stems: does submit prime permit? *Lang. Cogn. Process* 15, 539–561.

Forster, K. I., and Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 680–698.

Forster, K. I., Davis, C., Schoknecht, C., and Carter, R. (1987). Masked priming with graphemically related forms: repetition or partial activation? *Q. J. Exp. Psychol.* 39, 211–251.

Fowler, C. A., Napps, S. E., and Feldman, L. (1985). Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Mem. Cogn.* 13, 241–255.

Frost, R., Deutsch, A., Gilboa, O., Tannenbaum, M., and Marslen-Wilson, W. (2000). Morphological priming: dissociation of phonological, semantic, and morphological factors. *Mem. Cogn.* 28, 1277–1288.

Frost, R., Forster, K. I., and Deutsch, A. (1997). What can we learn from the morphology of Hebrew? A masked-priming investigation of morphological representation. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 829-56. Retrieved from

Giraudo, H., and Grainger, J. (2000). Effects of prime word frequency and cumulative root frequency in masked morphological priming. *Lang. Cogn. Process* 15, 421–444.

Giraudo, H., and Grainger, J. (2001). Priming complex words: evidence for supralexical representation of morphology. *Psychon. Bull. Rev.* 8, 127–131.

Giraudo, H., and Grainger, J. (2003). "On the role of derivational affixes in recognizing complex words: evidence from masked priming," in *Morphological Structure in Language Processing*, eds R. H. Baayen, R. Shreuder (Berlin: Mouton de Gruyter), 209–232.

Gold, B. T., and Rastle, K. (2007). Neural correlates of morphological decomposition during visual word recognition. *J. Cogn. Neurosci.* 19, 1983–1993.

Gonnerman, L. M., Seidenberg, M. S., and Andersen, E. S. (2007). Graded semantic and phonological similarity effects in priming: evidence for a distributed connectionist approach to morphology. *J. Exp. Psychol. Gen.* 136, 323–345.

Grainger, J., Colé, P., and Segui, J. (1991). Masked morphological priming in visual word recognition. *J. Mem. Lang.* 30, 370–384.

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

Grainger, J., and Whitney, C. (2004). Does the huamn mnid raed wrods as a wlohe? *Trends Cogn. Sci.* 8, 58–59.

Grainger, J., and Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Front. Psychol.* 2, 54.

Järvikivi, J., Bertram, R., and Niemi, J. (2006). Affixal salience and the processing of derivational morphology: the role of suffix allomorphy. *Lang. Cogn. Process* 21, 394–431.

Järvikivi, J., and Niemi, J. (2002). Form-based representation in the mental lexicon: priming (with) bound stem allomorphs in Finnish. *Brain Lang.* 81, 412–423.

Juhasz, B. J., and Berkowitz, R. N. (2011). Effects of morphological families on English compound word recognition: a multitask investigation. *Lang. Cogn. Process* 26, 653–682.

Juhasz, B. J., Starr, M. S., Inhoff, A. W., and Placke, L. (2003). The effects of morphology on the processing of compound words: evidence from naming, lexical decisions and eye fixations. *Br. J. Psychol.* 94, 223–244.

Justus, T., Larsen, J., de Mornay Davies, P., and Swick, D. (2008). Interpreting dissociations between regular and irregular past-tense morphology: evidence from event-related potentials. *Cogn. Affect. Behav. Neurosci.* 8, 178–194.

Kazanina, N. (2011). Decomposition of prefixed words in Russian. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 1371–1390.

Kazanina, N., Dukova-Zheleva, G., Geber, D., Kharlamov, V., and Tonciulescu, K. (2008). Decomposition into multiple morphemes during lexical access: a masked priming study of Russian nouns. *Lang. Cogn. Process* 23, 800–823.

Kielar, A., and Joanisse, M. F. (2009). Graded effects of regularity in language revealed by N400 indices of morphological priming. *J. Cogn. Neurosci.* 22, 1373–1398.

Kielar, A., Joanisse, M. F., and Hare, M. L. (2008). Priming English past tense verbs: rules or statistics? *J. Mem. Lang.* 58, 327–346.

Kuperman, V., Bertram, R., and Baayen, R. H. (2008). Morphological dynamics in compound processing. *Lang. Cogn. Process* 23, 1089–1132.

Kuperman, V., Bertram, R., and Baayen, R. H. (2010). Processing trade-offs in the reading of Dutch derived words. *J. Mem. Lang.* 62, 83–97.

Kurylowicz, J. (1964). *The Inflectional Categories of Indo-European.* Heidelberg: Winter.

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.*, 62, 621–647.

Laudanna, A., Badecker, W., and Caramazza, A. (1989). Priming homographic stems. *J. Mem. Lang.* 28, 531–546.

Laudanna, A., Badecker, W., and Caramazza, A. (1992). Processing inflectional and derivational morphology. *J. Mem. Lang.* 31, 333–348.

Laudanna, A., and Burani, C. (1995). "Distributional properties of derivational affixes: implications for processing," in *Morphological Aspects of Language Processing*, ed. L. B. Feldman (Hillsdale, NJ: Lawrence Erlbaum Associates), 345–364.

Laudanna, A., Burani, C., and Cermele, A. (1994). Prefixes as processing units. *Lang. Cogn. Process* 9, 295–316.

Lavric, A., Clapp, A., and Rastle, K. (2007). ERP evidence of morphological analysis from orthography: a masked priming study. *J. Cogn. Neurosci.* 19, 866–877.

Lavric, A., Pizzagalli, D., Forstmeier, S., and Rippon, G. (2001). Mapping dissociations in verb morphology. *Trends Cogn. Sci.* 5, 301–308.

Lehtonen, M., Cunillera, T., Rodríguez-Fornells, A., Hultén, A., Tuomainen, J., and Laine, M. (2007). Recognition of morphologically complex words in Finnish: evidence from event-related potentials. *Brain Res.* 1148, 123–137.

Leinonen, A., Bratticoa, P., Järvenpääc, M., and Krausea, C. M. (2008). Event-related potential (ERP) responses to violations of inflectional and derivational rules of Finnish. *Brain Res.* 1218, 181–193.

Leinonen, A., Grönholm-Nyman, P., Järvenpää, M., Söderholm, C., Lappia, O., Laine, M., and Krause, C. M. (2009). Neurocognitive processing of auditorily and visually presented inflected words and pseudowords: evidence from a morphologically rich language. *Brain Res.* 1275, 54–66.

Lemhöfer, K., Koester, D., and Schreuder, R. (2011). When bicycle pump is harder to read than bicycle bell: effects of parsing cues in first and second language compound reading. *Psychon. Bull. Rev.* 18, 364–370.

Lima, S., and Pollatsek, A. (1983). Lexical access via an orthographic code? The basic orthographic syllabic structure (BOSS) reconsidered.

J. Verb. Learn. Verb. Behav. 22, 310–332.

Linares, R. E., Rodriguez-Fornell, A., and Clahsen, H. (2006). Stem allomorphy in the Spanish mental lexicon: evidence from behavioral and ERP experiments. *Brain Lang.* 97, 110–120.

Longtin, C. -M., Segui, J., and Halle´, P. A. (2003). Morphological priming without morphological relationship. *Lang. Cogn. Process* 18, 313–334.

Lupker, S. J., Perea, M., and Davis, C. J. (2008). Transposed-letter effects: consonants, vowels and letter frequency. *Lang. Cogn. Process* 23, 93–116.

Marslen-Wilson, W., and Tyler, L. K. (1998). Rules, representations, and the English past tense. *Trends Cogn. Sci.* 2, 428–435.

Marslen-Wilson, W., Tyler, L. K., Waksler, R., and Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychol. Rev.* 101, 3–33.

Marslen-Wilson, W. D., Bozic, M., and Randall, B. (2008). Early decomposition in visual word recognition: dissociating morphology, form, and meaning. *Lang. Cogn. Process* 23, 394–421.

McCormick, S. F., Brysbaert, M., and Rastle, K. (2009). Is morphological decomposition limited to low-frequency words? *Q. J. Exp. Psychol.* 62, 1706–1715.

McKinnon, R., Allen, M., and Osterhout, L. (2003). Morphological decomposition involving nonproductive morphemes: ERP evidence. *Neuroreport* 6, 883–886.

Meunier, F., and Longtin, C. -M. (2007). Morphological decomposition and semantic integration in word processing. *J. Mem. Lang.* 56, 457–471.

Meunier, F., and Marslen-Wilson, W. (2004). Regularity and irregularity in French verbal inflection. *Lang. Cogn. Process* 19, 561–580.

Meunier, F., and Segui, J. (1999). Morphological priming effect: the role of surface frequency. *Brain Lang.* 68, 54–60.

Monsell, S. (1985). "Repetition and the lexicon," in *Progress in the Psychology of Language*, ed. A. W. Ellis (Hillsdale, NJ: Lawrence Erlbaum Associated), 147–195.

Morris, J., Frank, T., Grainger, J., and Holcomb, P. J. (2007). Semantic transparency and masked morphological priming: an ERP investigation. *Psychophysiology* 44, 506–521.

Moscoso del Prado Martín, F., Deutsch, A., Frost, R., Schreuder, R., De Jong, N. H., and Baayen, R. H. (2005). Changing places: a cross-language

perspective on frequency and family size in Dutch and Hebrew. *J. Mem. Lang.* 53, 496–512.

Moscoso del Prado Martín, F., Kostic, A., and Baayen, R. H. (2004). Putting the bits together: an information theoretical perspective on morphological processing. *Cognition* 94, 1–18.

Münte, T. F., Say, T., Clahsen, H., Schiltz, K., and Kutas, M. (1999). Decomposition of morphologically complex words in English: evidence from event-related brain potentials. *Cogn. Brain Res.* 7, 241–253.

Norris, D., and Kinoshita, S. (2008). Perception as evidence accumulation and Bayesian inference: insights from masked priming. *J. Exp. Psychol. Gen.* 137, 434–455.

Pastizzo, M. J., and Feldman, L. B. (2002). Discrepancies between orthographic and unrelated baselines in masked priming undermine a decompositional account of morphological facilitation. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 244–249.

Pastizzo, M. J., and Feldman, L. B. (2004). Morphological processing: a comparison between free and bound stem facilitation. *Brain Lang.* 90, 31–39.

Paterson, K. B., Alcock, A., and Liversedge, S. P. (2011). Morphological priming during reading: evidence from eye movements. *Lang. Cogn. Process* 26, 600–623.

Perea, M., and Carreiras, M. (2006). Do transposed-letter effects occur across lexeme boundaries? *Psychon. Bull. Rev.* 13, 418–422.

Perea, M., and Lupker, S. J. (2003). Does judge activate COURT? Transposed-letter similarity effects in masked associative priming. *Mem. Cognit.* 31, 829–841.

Perea, M., and Lupker, S. J. (2004). Can CANISO activate CASINO? Transposed-letter similarity effects with nonadjacent letter positions. *J. Mem. Lang.* 51, 231–246.

Pinker, S. (1991). Rules of language. *Science* 253, 530–254.

Plag, I., and Baayen, R. H. (2009). Suffix ordering and morphological processing. *Language* 85, 109–152.

Pollatsek, A., and Hyönä, J. (2005). The role of semantic transparency

in the processing of Finnish compound words. *Lang. Cogn. Process* 20, 261–290.

Pylkkänën, L., Feintuch, S., Hopkins, E., and Marantz, A. (2004). Neural correlates of the effects of morphological family frequency and family size: an MEG study. *Cognition* 91, B35–B45.

Rastle, K., Davis, M., Marslen-Wilson, W., and Tyler, L. (2000). Morphological and semantic effects in visual word recognition: a time-course study. *Lang. Cogn. Process* 15, 507–537.

Rastle, K., and Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Lang. Cogn. Process* 23, 942–971.

Rastle, K., Davis, M. H., and New, B. (2004). The broth in my brother's brothel: morpho-orthographic segmentation in visual word recognition. *Psychon. Bull. Rev.* 11, 1090–1098.

Raveh, M. (2002). The contribution of frequency and semantic similarity to morphological processing. *Brain Lang.* 81, 312–325.

Rodriguez-Fornell, A., Münte, T. F., and Clahsen, H. (2002). Morphological priming in Spanish verb forms: an ERP repetition priming study. *J. Cogn. Neurosci.* 14, 443–454.

Rueckl, J. G., and Aicher, K. A. (2008). Are CORNER and BROTHER morphologically complex? Not in the long term. *Lang. Cogn. Process* 23, 972–1001.

Rueckl, J. G., and Rimzhim, A. (2011). On the interaction of letter transpositions and morphemic boundaries. *Lang. Cogn. Process* 26, 482–508.

Sandra, D. (1990). On the representation and processing of compound words: automatic access to constituent morphemes does not occur. *Q. J. Exp. Psychol.* 42, 529–567.

Schoonbaert, S., and Grainger, J. (2004). Letter position coding in printed word perception: effects of repeated and transposed letters. *Lang. Cogn. Process* 19, 333–367.

Schreuder, R. (1997). How complex simplex words can be. *J. Mem. Lang.* 37, 118–139.

Schreuder, R., and Baayen, H. (1994). Prefix stripping re-visited. *J. Mem. Lang.* 33, 357–375.

Shoolman, N., and Andrews, S. (2003). "Racehorses, reindeer, and sparrows: using masked priming to investigate morphological influences on compound word identification," in *Masked Priming: The State of the Art*, eds S. Kinoshita and S. J. Lupker (New York, NY: Psychology Press), 241–278.

Stanners, R. F., Neiser, J. J., Hernon, W. P., and Hall, R. (1979). Memory representation for morphologically related words. *J. Verb. Learn. Verb. Behav.* 18, 399–412.

Stockall, L., and Marantz, A. (2006). A single route, full decomposition model of morphological complexity. *Ment. Lex.* 1, 1871–1340.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Mem. Cogn.* 7, 263–272.

Taft, M. (1985). The lexical account of word naming considered further. *Behav. Brain Sci.* 8, 727.

Taft, M. (1994). Interactive-activation as a framework for understanding morphological processing. *Lang. Cogn. Process* 9, 271–294.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Q. J. Exp. Psychol.* 57, 745–765.

Taft, M., and Ardasinski, S. (2006). Obligatory decomposition in reading prefixed words. *Ment. Lex.* 1, 183–199.

Taft, M., and Forster, K. (1975). Lexical storage and retrieval of prefixed words. *J. Verb. Learn. Verb. Behav.* 14, 638–647.

Taft, M., and Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *J. Verb. Learn. Verb. Behav.* 15, 607–620.

Taft, M., Hambly, G., and Kinoshita, S. (1986). Visual and auditory recognition of prefixed words. *Q. J. Exp. Psychol.* 38, 351–365.

Taft, M., and Kougious, P. (2004). The processing of morpheme-like units in monomorphemic words. *Brain Lang.* 90, 9–16.

Taft, M., Zhu, X., and Peng, D. (1999). Positional specificity of radicals in

Chinese characters recognition. *J. Mem. Lang.* 40, 498–519.

Vannest, J., and Boland, J. E. (1999). Lexical morphology and lexical access. *Brain Lang.* 68, 324–332.

Vergara-Martínez, M., Duñabeitia, J. A., Laka, I., and Carreiras, M. (2009). ERP correlates of inhibitory and facilitative effects of constituent frequency in compound word reading. *Brain Res.* 1257, 53–64.

Weyerts, H., Münte, T. F., Smid, H. G. O. M., and Heinze, H. (1996). Mental representations of morphologically complex words: an event-related potential study with adult humans. *Neurosci. Lett.* 206, 125–128.

Zwitserlood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Neuroreport* 9, 341–368.

Zwitserlood, P., Bolwieder, A., and Drews, E. (2005). Priming morphologically complex verbs by sentence contexts: effects of semantic transparency and ambiguity. *Lang. Cogn. Process* 20, 295–415.

Future morphology? Summary of visual word identification effects draws attention to necessary efforts in understanding morphological processing

Dirk Koester[1,2]*

[1] Neurocognition and Action – Biomechanics Research Group, Faculty of Psychology and Sport Science, Bielefeld University, Bielefeld, Germany
[2] Center of Excellence – Cognitive Interaction Technology, Bielefeld University, Bielefeld, Germany
*Correspondence: dkoester@cit-ec.uni-bielefeld.de

Edited by:
Carlo Semenza, Università degli Studi di Padova, Italy

Reviewed by:
Carlo Semenza, Università degli Studi di Padova, Italy
Marco Marelli, University of Trento, Italy

A commentary on

Morphological processing as we know it: an analytical review of morphological effects in visual word identification
by Amenta, S., and Crepaldi, D. (2012). Front. Psychol. 3:232. doi: 10.3389/fpsyg.2012.00232

This commentary discusses the insights and suggestions of a review by Amenta and Crepaldi (2012) in *Frontiers in Psychology*. The authors have diagnosed a controversial accumulation of findings in the field of visual word identification and, hence, have provided an overview of the field with the aim to separate substantial effects from findings that need further confirmation. The authors aim to provide a broad basis for theory development. Amenta and Crepaldi (2012) are the first to attempt a comprehensive psycholinguistic review of the major forms of word formation, namely inflection, derivation, and compounding. The authors summarize 17 robust experimental effects and suggest that "any theory should be able to explain" this set of experimental effects (cf. abstract). Thus, the listed effects are supposed to help to decide which among competing theories have more explanatory power and might thus be considered scientifically superior.

Without repeating all effects, Amenta and Crepaldi propose that stem frequency, family size, word entropy, and the number of affix allomorphs are main determinants in visual word identification. Furthermore, non-word processing is suggested to be relevant for morphological theories, if the non-words are made from morphemes. Other relevant properties are proposed for

methodologically specified situations. For example, if stimuli are fully visible, morphological priming effects are only proposed for semantically related words, and inflectional priming yields greater effects than derivational priming. In contrast, in masked priming, morphological effects are comparable in magnitude for semantically transparent and opaque words. Also, inflectional and derivational priming is suggested to yield comparable effect sizes.

Amenta and Crepaldi's review point toward relevant linguistic (e.g., productivity) and psycholinguistic variables (e.g., frequency measures) and their relations regarding visual word identification. Such knowledge will guide future investigations and, hence, impact also models of language performance. The authors suggest that these findings provide a basis for the evaluation of competing theories and, in doing so, to contribute to future theory development; in their own words, to construct an "all-inclusive model of visual identification of morphologically complex words." In light of the specificity of the insights, these broad suggestions leave the reader with the impression of a gap between insights and suggestions. The authors deal with a specific functional step (visual word identification) of the more complex human ability of (single word) reading and it is not necessary that the relevant variables for identification can be extrapolated to other functional steps (e.g., morphosyntactic and/or semantic combination of morpho-orthographic segments).

As the authors mentioned themselves, their list of effects is not exhaustive. However, one would like to know whether, and if so, what role further variables such as surface frequency, word length, word class, abstractness, or cues to morpheme boundaries are supposed to play in word identification (e.g., Caramazza and Hillis, 1991; Inhoff et al., 2000; Taft, 2004; Baayen et al., 2007; Juhasz, 2008; Kuperman et al., 2009; Juhasz and Pollatsek, 2011; Hyönä, 2012). The impact of Amenta and Crepaldi's (2012) target list of relevant effects on future experimentation and theory development will depend on the relative contribution of all these variables. Consequently, one needs to discuss whether the additional variables mentioned here affect only later reading stages or what their role could be during word identification.

More generally, the authors seem to aim for a psycholinguistic, i.e., cognitive model of language behavior rather than a linguistic theory. (This is not the same; morphological effects can, for example, be simulated without an implementation of morphology; cf. Baayen et al., 2011.) They also refer to some eye-tracking and electrophysiological studies which provide neural evidence. It remains unclear how the neural evidence is to be incorporated into a strictly cognitive model. Alternatively, one may aim for a neuro-cognitive model of language behavior and visual word identification in particular. If one is to construct a complete model of such a phenomenon, the effects (behavior) but also the causes (neural activity) appear to be relevant and should be considered. While Amemta and Crepaldi's (2012) work to consider the large body of behavioral evidence is certainly ambitious, future work should take neural evidence also into account because cognitive and

neural evidence can be mutually informative and helpful in understanding language performance (Grimaldi, 2012).

Another methodological issue arises from the suggestion that models of visual word identification should explain the effects listed by Amenta and Crepaldi (2012) because the list comprises aspects of experimental techniques (masking). Masking does not pertain to the phenomenon in question but experimental paradigms can be modeled. For example, Norris and Kinoshita (2008) proposed that in masked priming, prime, and target are perceptually fused into a single percept or object. As a consequence, masked priming effects may depend on the task requirements rather than on the relation between prime and target representations. Norris and Kinoshita (2008) show that priming effects can be shifted from word stimuli to non-word stimuli by using a same-different task rather than a lexical decision task. For the present discussion, one can doubt whether psycholinguistic models of word identification have to combine methodological aspects such as masked priming with the processes of interest, i.e., visual word identification.

Although some questions remain, Amenta and Crepaldi's (2012) review is bound to stimulate scientific discussions regarding visual word identification and provoke further research efforts to better understand morphological processing. Next steps of enquiry may focus on the different domains of morphology, inflection, derivation, or compounding drawing on as many sources of evidence as possible (e.g., different populations, methodologies, or language families) to comprehensively describe each domain (cf. Niemi et al., 1994; Marslen-Wilson and Tyler, 2007). Another major challenge is the theoretical unification of different sensory modalities. Finally, future reviews would be highly informative, if they use quantitative evaluations of research findings. One might perform meta-analyses or quantify the frequency of replications of particular effects. This way, our understanding of the connection between morphology and how it is represented and controlled by the human brain may be fostered (cf. Grimaldi, 2012).

REFERENCES

Amenta, S., and Crepaldi, D. (2012). Morphological processing as we know it: an analytical review of morphological effects in visual word identification. *Front. Psychol.* 3: 232. doi: 10.3389/fpsyg.2012.00232

Baayen, R. H., Milin, P., Ðurđević, D. F., Hendrix, P., and Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychol. Rev.* 118, 438–482.

Baayen, R. H., Wurm, L. H., and Aycock, J. (2007). Lexical dynamics for low frequency complex words: a regression study across tasks and modalities. *Ment. Lex.* 2, 419–463.

Caramazza, A., and Hillis, A. E. (1991). Lexical organization of nouns and verbs in the brain. *Nature* 349, 788–790.

Grimaldi, M. (2012). Toward a neural theory of language: old issues and new perspectives. *J. Neurolinguistics* 25, 304–327.

Hyönä, J. (2012). The role of visual acuity and segmentation cues in compound word identification. *Front. Psychol.* 3:188. doi: 10.3389/fpsyg.2012.00188

Inhoff, A. W., Radach, R., and Heller, D. (2000). Complex compounds in German: interword spaces facilitate segmentation but hinder assignment of meaning. *J. Mem. Lang.* 42, 23–50.

Juhasz, B. (2008). The processing of compound words in English: effects of word length on eye movements during reading. *Lang. Cogn. Process.* 23, 1057–1088.

Juhasz, B., and Pollatsek, A. (2011). "Lexical influences on eye movements in reading," in *The Oxford Handbook of Eye Movements*, eds S. Liversedge, I. Gilchrist, and S. Everling (Oxford: Oxford University Press), 873–893.

Kuperman, V., Schreuder, R., Bertram, R., and Baayen, R. H. (2009). Reading polymorphemic Dutch compounds: toward a multiple route model of lexical processing. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 876–895.

Marslen-Wilson, W., and Tyler, K. K. (2007). Morphology, language and the brain: the decompositional substrate for language comprehension. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 823–836.

Niemi, J., Laine, M., and Tuominen, J. (1994). Cognitive morphology in Finnish: foundations of a new model. *Lang. Cogn. Process.* 9, 423–446.

Norris, D., and Kinoshita, S. (2008). Perception as evidence accumulation and Bayesian inference. Insights from masked priming. *J. Exp. Psychol. Gen.* 137, 434–455.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Q. J. Exp. Psychol. A* 57, 745–765.