UNIVERSITE D'AIX-MARSEILLE

ECOLE DOCTORALE 356 : Cognition Langage Education

LABORATOIRE DE PSYCHOLOGIE COGNITIVE, UMR 7290

Thèse présentée pour obtenir le grade universitaire de docteur

Discipline : Psychologie

Joshua Snell

Titre de la thèse : Readers are parallel processors

Soutenue le 07/09/2018 devant le jury :

Richard Shillcock	University of Edinburgh	Rapporteur
Walter van Heuven	University of Nottingham	Rapporteur
Sarah Risse	Universität Potsdam	Examinateur
Ludovic Ferrand	Université Clermont Auvergne & CNRS	Examinateur
Martijn Meeter	Vrije Universiteit Amsterdam	Examinateur
Johannes Ziegler	Aix-Marseille Université & CNRS	Examinateur
Jonathan Grainger	Aix-Marseille Université & CNRS	Directeur de thèse

Index

Preface

Three years prior to writing this thesis, Jonathan Grainger and I presented plans for carrying out investigations into how orthographic processing occurs across multiple words. In adhering to the proposed three-year plan we were doomed to fail, for our mutual research interests were a bit broader than orthographic processing—and thankfully so: given the many interactions among the 'sub-components' of the reading system (at least, that is what I will argue in this thesis), it may be futile to explore any sub-component in isolation without considering the reading system as a whole. Hence, in three years, we theorized about visuo-spatial attention, letter processing, lexical processing, syntactic processing, semantic processing and sentence processing. In this endeavor, Martijn Meeter—who has never shied away from investigating anything—has been the perfect partner-in-crime in his role as co-supervisor.

I am incredibly grateful to many friends and colleagues with whom I have shared publications and/or fruitful interactions: Mathieu Declerck, Sebastiaan Mathôt, Jonathan Mirault, Stéphane Dufau, Gabriela Meade, Elizabeth Beyersmann, Daisy Bertrand, Yun Wen, Françoise Vitu, Jean-Baptiste Bernard, Jean-Baptiste Melmi, Cécile Mahnich, Eva Vives, Johannes Ziegler, Philip Holcomb, Benchi Wang.

Summary

This thesis addresses one of the most hotly debated issues in reading research: *Are words processed serially or in parallel during reading*? One could argue that this is primarily a question of visuo-spatial attention: is attention distributed across multiple words during reading? Naturally, if one were to establish that attention is strictly directed to one word at a time, then one would necessarily conclude that readers are serial processors. However, if one were to establish that attention is allocated to multiple words at once, then matters become more complicated: it is possible, for instance, that a small portion of attention dedicated to surrounding words allows the reader to pick up low-level information (e.g., information about word lengths or letters) from those words, but that actual lexical access would still occur in a serial fashion. Thus, in case of a widespread attentional distribution, one would need means to determine whether higher-order processing occurs for multiple words at once, in order to truly evidence parallel processing.

The research presented here suggests that attention can indeed be allocated to multiple words at once. It is further established that attention is a key factor driving (sub-lexical) orthographic processing. The next question, then, is whether multiple lexical representations can be activated in parallel. This thesis comprises a wealth of evidence for parallel lexical activation: firstly we have found that readers activate embedded words (e.g., '*use*' in '*houses*') alongside the word that is to be recognized, indicating that parallel lexical processing would occur even if readers could effectively focus their attention on single words. Moreover, we have found that semantic and syntactic categorization decisions about foveal target words are influenced by the semantic and syntactic aspects of surrounding words, even when all these words are presented for a duration shorter than the average time needed to recognize a single word. Hence, given that readers' attention is spread across multiple words and that multiple lexical representations can be activated in parallel, it seems reasonable to claim that the reading system is in principle a parallel processing system.

This generates another important question: *How does the parallel processing reading system organize incoming information*? Indeed, proponents of serial processing have claimed that a parallel processing system would have difficulty processing text, as it might mix up linguistic information across words and lose track of word order. In this thesis I present a theoretical framework that addresses these issues. The crux is that the process of determining the identity and order of words is heavily influenced by sentence-level representations. This conception is supported by the finding that the recognition of individual words is much better when surrounding words form a coherent sentence, compared to when the same surrounding words are in a different, syntactically incorrect order. Such effects again emerge with very short stimulus presentation times, and moreover, are equally strong for words in the left and right visual hemifields respectively, hence arguing against rapid serial left-to-right processing. Strikingly, our theoretical framework also correctly predicted that readers can in fact recognize words out of order—and that this is not so problematic as proponents of serial processing have thought.

Finally, this thesis also presents a computational model that integrates most of these insights. The OB1-Reader model accounts for more phenomena than its predecessors, and additionally offers new means to explore how and when reading may go awry.

Résumé

Le traitement en parallèle des mots pendant la lecture

Une question centrale des recherches sur la lecture concerne la nature séquentielle ou parallèle de l'identification des mots pendant la lecture de phrases. L'hypothèse dominante postule que l'attention spatiale est allouée à un seul mot à la fois, et qu'avec cette contrainte, l'identification des mots doit forcément s'opérer de manière séquentielle. Cependant, un certain nombre de résultats suggèrent, au contraire, que l'attention spatiale peut être allouée à plusieurs mots à la fois, de manière distribuée. Cette attention disbribuée pourrait permettre l'identification en parallèle de plusieurs mots de manière simultanée, et les travaux présentés dans cette thèse cherchent à déterminer la viabilité de cette hypothèse. Notamment, nos travaux visent à préciser le niveau de traitement (visuel, orthographique, lexical, sémantique ou syntaxique) permis par cette attention distribuée.

En premier lieu, les travaux de la thèse confirment que l'attention spatiale peut être allouée à plusieurs mots à la fois, et montrent que cette attention distribuée est un facteur important pour permettre le traitement en parallèle des informations orthographiques issues de plusieurs mots. Concernant l'identification en parallèle des mots, les travaux de la thèse présentent plusieurs preuves que non seulement cela est possible, démontré à l'aide des paradigmes de lecture simplifiée, mais que cela se passe également en situation de lecture normale. D'abord, il a été montré que les mots enchassés dans des mots plus longs (ex : « mat » dans « matière ») sont activés pendant la lecture, influençant ainsi l'activité oculomotrice du lecteur. Ensuite, nous avons montré que les décisions sémantiques, syntaxiques ou lexicales prises sur un mot cible central sont influencées par les caractéristiques sémantiques, syntaxiques ou lexicales des mots adjacents, présentés à gauche et à droite du mot cible pour une durée très brève de 170 ms. Ces résultats montrent que les lecteurs experts sont en principe capables de traiter les informations sémantiques, syntaxiques et lexicales issues de plusieurs mots en parallèle.

Ces résultats posent donc une question fondamentale : comment un système de traitement en parallèle des mots pendant la lecture organise-t-il ces informations pour ne pas créer de confusions ? En effet, les défenseurs d'une approche strictement séquentielle de la lecture ont notamment déclaré qu'un système de traitement en parallèle créerait une confusion au niveau de l'ordre des mots dans les phrases. La thèse présentée ici décrit un cadre théorique pour la lecture experte qui explique comment un système de traitement en parallèle peut gérer correctement les informations qu'il traite et fournir l'ordre correct des mots. Une caractéristique centrale de cette proposition concerne l'importance des influences du traitement des informations à l'échelle de la phrase sur le traitement des informations aux niveaux inférieurs, et notamment au niveau lexical. Ces influences ont été révélées de manière expérimentale en montrant que l'identification d'un mot est facilitée lorsque le mot est présenté dans une phrase correcte par rapport à une séquence agrammaticale. De plus, nous avons pu vérifier une prédiction centrale de ce cadre theorique, à savoir que dans certaines conditions, les mots peuvents être identifiés comme étant dans un ordre différent de l'ordre réel dans lequel ils étaient présentés.

Enfin, la thèse est conclue par la présentation d'un modèle computationnel de la lecture qui intègre la plupart des caractéristiques du modèle conceptuel. Ce modèle, OB1-Reader, possède un pouvoir explicatif plus grand que ses prédécesseurs, et pour la première fois offre un moyen d'explorer, au-délà de la lecture des mots isolés, les conditions susceptibles de provoquer un dysfonctionnement de la lecture.

Introduction

The ability to read is arguably one of the most delicate, complex and societally important tricks the human brain has up its sleeve. It may therefore come as no surprise that the roots of reading research can be traced back to the very beginnings of cognitive psychology, initiated primarily at the Leibniz laboratory of Wilhelm Wundt in 1879. Important works include that of James McKeen Cattell (a doctoral student of Wundt), who in 1886 investigated *the time it takes to see and name objects* (which was also the title of his seminal paper); Benno Erdmann and Raymond Dodge, who in 1898 published 356 pages worth of theory and experimental investigations into reading[1]; and Edmund Burke Huey, who marked the apex of early reading research with his 1908 publication *The psychology and pedagogy of reading*, a book that comprised a review of all that had been known thus far about the reading system, as well as descriptions of new methodologies for studying reading behavior.[2]

My mentioning of these founding fathers of reading research is not merely to honor the people whose work I intend to extend. Rather, I feel the need to highlight a key characteristic of their research: that is, what they did at the time was, by today's standards and definitions, *interdisciplinary*. Indeed, these pioneers endeavored to capture the reading process in full, and this was not driven by mere ignorance about the reading system's complexity; as noted by Huey (1908), "[…] to completely analyze what we do when we read would almost be the acme of a psychologist's achievements, for it would be to describe very many of the most intricate workings of the human mind." (p. 6). Not shying away from this challenge, Cattell (1886), for example, investigated letter processing, word processing, sentence processing, as well as interactions between these cognitive stages. As such, he established the well-known *word superiority effect*, whereby letter recognition is better when the letter is in a word than when it is presented in a nonword string; and interestingly, he also provided a first glance at the lesser-known *sentence superiority effect*, whereby recall of individual words is better when the words are presented in a coherent sentence, compared to when the words are presented in a random sequence.

In light of these ambitious beginnings, it is striking that after reading research's 40-year hiatus—which started in the 1920's and whose existence is attributed to the behaviorist revolution (Rayner, Pollatsek, Ashby & Clifton Jr., 2012)—the endeavor, or at least its partakers, had largely given up on interdisciplinarity. Cognitive psychology had by then set its course toward sophistication via reductionism, whereby in order to understand cognitive processes, those processes were conceptually divided into sub-processes. By the early 1990's, reading research was no longer a single discipline, but rather an amalgamation of numerous domains, including e.g. *orthographic processing*, *syntax, semantics, eye movements in reading, reading development* and *dyslexia*. Some of these domains have been quite isolated. A stunning point in case is provided by the fact that the most influential computational models of text reading, E-Z Reader (Reichle, Pollatsek & Rayner, 2006) and SWIFT (Engbert, Nuthmann, Richter & Kliegl, 2005), do not

---

[1] Their work was titled *Psychologische untersuchen über das lessen auf experimenteller grundlage* (English: Psychological investigations into reading based on experiments).

[2] Most notably, Huey invented the first eye-tracker. He tracked the reader's eye position with a piece of Paris cup that was placed on the eye like a contact lens, with a tiny hole in the center to allow light onto the pupil. The lens was attached to a pointer via a light type of iron wire, so that the reader's eye movements caused displacements of the pointer. To help overcome the discomfort of this intrusive method, subjects were administered drug agents such as cocaine.

comprise mechanisms of letter and word recognition. Researchers investigating letter and word recognition, in turn, have mostly tested recognition of isolated words, as such disregarding potential influences of surrounding information and context.

Possibly, the disregard of neighboring domains has allowed researchers to formulate assumptions somewhat more freely. One such assumption, which has held reading research in its grip for a few decades now, concerns the belief that words are processed in a strict serial, one-by-one fashion (e.g., Reichle, Pollatsek, Fischer & Rayner, 1998; Reichle et al., 2006; Reichle, Liversedge, Pollatsek & Rayner, 2009; Schotter, Angele & Rayner, 2012)—an assumption not driven by data, but rather by the general notion that a serial processing system would have an easier time dealing with text than a parallel processing system. This notion is most likely true. Context comprehension is acquired by determining the identities of words, as well as the order in which they appear (for '*dog eats*' alludes to a different context than '*eats dog*'). In a serial processing system, words would simply be appended to a sentence-level representation in the order in which they are recognized, and the ability to dedicate all processing resources to one word at a time would prevent confusion of the identity or location of a word with that of surrounding words. In contrast, as noted by Reichle et al. (2009), how a parallel processing system keeps track of word order is not clear. Moreover, it has been assumed that parallel processing of multiple words would inevitably lead to integration of (orthographic, lexical, semantic) information across those words, which presumably makes the identification of individual words more difficult. For these reasons it seems fair to state that readers would ideally be serial processors. The question, then, is whether the brain is indeed able to effectively focus its processing resources on single words.

While serial processing has been the dominant view, primarily driven by the popularity of the E-Z Reader model (Reichle et al., 1998; 2006), some theoretical frameworks have assumed parallel processing (e.g., the SWIFT model of Engbert et al., 2005, and the Glenmore model of Reilly & Radach, 2006). Now that two decades have passed since E-Z Reader's first conception, it is safe to say that these modeling efforts alone were not able to provide conclusive evidence in favor of serial or parallel processing. A big issue here is that all these models have treated the word recognition process as a black box; and it turns out that such simplified models can simulate normal reading behavior (i.e., eye movements) equally well.

Therefore, in this thesis, I posit that two ingredients are essential for truly evidencing serial or parallel processing: firstly, we need to tread outside the realms of normal (sentence) reading and test behavior in artificial settings that push the reading system to its limit. Secondly, we need to develop (and implement) more sophisticated theoretical frameworks that can predict and account for such artificially induced reading behavior. Returning to the roots of reading research, successful theoretical frameworks must be *interdisciplinary* in the sense that they have to integrate evidence from multiple domains: visuo-spatial attention, orthographic processing, lexico-semantic processing, as well as syntactic (sentence-level) processing.

Below, I will report a set of studies that, in my sense, meet these requirements. The wealth of evidence obtained with these studies allows me to claim that *readers are parallel processors.*

This work comprises five chapters: (**i**) *Parallel word activation in single word processing*; (**ii**) *Orthographic integration and the attentional gradient*; (**iii**) *Higher-order integration: eye-movements versus decisions*; (**iv**) *Top-down expectations*; and (**v**) *OB1-Reader: A successful parallel processing system*.

Chapter **i** addresses the claim by proponents of serial processing that lexical processing of multiple words is not consistent with any model of word identification (Reichle et al., 2009). This is a misconception: all word identification models assume parallel activation of multiple lexical representations (e.g. McClelland & Rumelhart, 1981; Grainger & van Heuven, 2003; Davis, 2010). I will report two studies that support the notion that readers process multiple words in parallel, even when they only view single words.

Chapter **ii** shows that orthographic parafoveal-on-foveal effects (whereby words are recognized faster when they share many letters with surrounding words) are driven by a widespread distribution of visuo-spatial attention.

Chapter **iii** raises the possibility that higher-order (e.g., syntactic, semantic) processing of multiple words truly occurs in parallel; that is, without integrating these types of information across words.

Chapter **iv** tests a few claims of the theoretical framework that is outlined in the studies of the preceding chapter. Specifically, it is hypothesized that sentence-level representations can constrain ongoing identification of individual words. Further, it is hypothesized that top-down expectations influence which recognized word is associated with which location in the sentence-level representation.

Finally, Chapter **v** presents a computational model that integrates the domains of word recognition and eye movements in text reading. This model, OB1-Reader, is able to account for more phenomena than all its predecessors, and may therefore mark an important step toward the goal of accounting for the reading process in full.

Chapter 1: Parallel word activation in single word processing

A word on words in words: How do embedded words affect reading?

Joshua Snell, Jonathan Grainger & Mathieu Declerck

Citation: Snell, J., Grainger, J. & Declerck, M. (2018). A word on words in words: How do embedded words affect reading? *Journal of Cognition*, in press.

Abstract

A surprisingly small portion of reading research has been dedicated to investigating how the visual word recognition process is influenced by embedded words (e.g. '*arm*' in '*charm*'), and no research has yet investigated embedded words in a natural reading setting. Covering this issue, the present work reports analyses of eye-tracking data from the GECO bilingual book reading corpus. Word viewing times were analyzed as a function of the number, frequency and proportional length of embedded words. We anticipated two scenarios: embedded words would either facilitate processing due to increased word-letter feedback, or inhibit processing due to increased lexical competition. A main facilitatory effect of embedded words on the recognition process was established, with an increasing number of embedded words resulting in shorter word viewing times and fewer fixations. This pattern was depicted by readers of Dutch as well as readers of English. Long, high-frequency embedded words formed an exception however, as these led to inhibition (Dutch participants) or a null-effect (English participants). The present results indicate that both scenarios outlined above are at play, but do so whilst putting a theoretical constraint on the role of word-to-word inhibitory connections. Specifically, such connections may predominantly exist among words of similar length. Hence, embedded words generally facilitate processing through word-letter feedback, but this facilitatory effect is countered by word-to-word inhibition if the embedded word's length approximates that of its superset.

1. Introduction

A primary goal of reading research has been to determine how recognition of letters leads to activation and recognition of word representations in the brain. Much at the heart of the endeavor of 'cracking the orthographic code' (Grainger, 2008) lies the principle that the recognition of words is influenced by other words that are either in temporal or spatial proximity. This principle has been pivotal in two ways. Firstly, observing how characteristics of one word influence the other has allowed researchers to deduce key aspects of the word recognition process—especially with respect to the encoding of letter identity and position. Secondly, word-to-word influences per se can be regarded as an intrinsic part of the word recognition process; that is, a complete understanding of how words are recognized obligates understanding of why and how words are influenced by other words. This will be the central theme of the present paper. Specifically, we cover how the recognition of words is impacted by words that are in the uttermost spatial and temporal proximity: namely, embedded words (e.g. '*arm*' in '*charm*').

Having been addressed in but a handful of studies,[3] many issues concerning reading embedded words remain open to inquiry (Bowers, Davis & Hanley, 2005; Davis & Taft, 2005; Davis, Perea & Acha, 2009). The most obvious question is whether embedded words are lexically accessed at all (as noted in the title of Bowers, Davis and Hanley's (2005) study: "*Is there a 'hat' in 'that'?*"). In their study, Bowers et al. found that semantic categorization decisions about target words (e.g., '*Does* hatch *refer to a piece of clothing?*') were slower and less accurate if the embedded word ('*hat*') was associated with a different response than the target ('*hatch*'), hence suggesting that representations associated with embedded words may indeed be activated—and semantically processed—in parallel with representations associated with the target word. It is further worth noting that within the closely related research domain investigating morphological processing, evidence has been found for parallel processing of affixes, stems and suffixes (e.g., '*de*', '*brief*' and '*ing*' in '*debriefing*'), and even so in the case of pseudo-derived words (e.g., '*corner*', which is not a derivation of '*corn*'), thus again arguing for embedded word activation (see Grainger & Beyersmann, 2017, for a review).

The next question is what effect embedded words may have on target word processing. Note that the work of Bowers et al. (2005) does not evidence an influence of embedded words on semantic processing of the target: after all, semantic processing of the target and embedded word may have proceeded independently, with the embedded word's semantic category only impacting at the stage of decision-making.

Concerning the embedded word's influence on target processing, four scenarios are conceivable: either the embedded word (i) facilitates processing of the target, (ii) inhibits processing of the target, (iii) does both, potentially resulting in a null-effect, or (iv) does neither, hence certainly resulting in a null-effect. Below, we will explain why embedded words may have a facilitatory and/or inhibitory influence on processing of the word by which they are contained (their so-called *supersets*). In doing so, we will firstly draw a critical analogy to *orthographic neighbors*. Secondly, we will address the potential role of morphological relationships between embedded words and their supersets.

---

[3] Note that this pertains to visual word recognition specifically. Quite a few studies have investigated embedded words in spoken word recognition: see Bowers, Davis and Hanley (2005) for a discussion of these studies.

1.1 *Orthographic neighbors versus embedded words*

By the classic definition, orthographic neighbors are words that are identical to one another in all but one letter (e.g., '*rock*' and '*rack*'). They play a key role in neighborhood size effects, whereby words that have a larger amount of orthographic neighbors (e.g. '*ball*', '*bill*', '*hall*', '*hill*') are recognized faster than words with a smaller neighborhood size (Andrews, 1989; Forster & Shen, 1996; Johnson & Pugh, 1994). Accounting for this effect is the conception that letter nodes are connected to the nodes of all words in which the letter occurs (such that the '*r*' in '*rock*' would not only activate '*rock*' but also '*rocket*', '*rack*', '*race*', et cetera), and that word nodes would in turn provide feedback activation of letter nodes (e.g. McClelland & Rumelhart 1981). Viewing a word with a large neighborhood size results in letter detectors receiving feedback activation from a higher amount of strongly activated word nodes, in turn leading to faster recognition (but see Grainger & Jacobs, 1996, for an alternative account expressed in terms of decision-level processes).

Orthographic neighbors have also been shown to facilitate processing when being presented next to one another in the visual field. In sentence reading, Snell, Vitu and Grainger (2017a) found that a target word at position *n* was recognized faster (evidenced by shorter word viewing times) if an orthographic neighbor was presented at position *n*+1, compared to if an unrelated word of the same length and frequency was presented at that location. In a flanker paradigm wherein participants made lexical decisions about foveal target words that were flanked by words on the left and right, Snell, Bertrand, Meeter and Grainger (2018a) found that responses were faster and more accurate with neighbor flankers, compared to unrelated flankers.

It should be noted however, that orthographic neighbors have also been found to slow down, rather than speed up, word recognition. Using a masked-priming lexical decision task, Segui and Grainger (1990) found that processing of target words (e.g. '*rack*') was inhibited if those targets were preceded by briefly presented higher-frequency neighbor prime words ('*rock*'), compared to unrelated prime words of the same frequency ('*step'*) (see also de Moor & Brysbaert, 2000; Davis & Lupker, 2006). This has led researchers to argue for the presence of inhibitory connections among word nodes—the idea being that co-active words would have to compete for recognition (e.g. Segui & Grainger, 1990; Grainger, 2008). It stands to question, then, why neighbors facilitate processing when they are spatially adjacent to the target, as has been shown in the studies of Snell et al. (2017a; 2018a). The answer offered by Snell et al. (2018a) is that multiple lexical representations may be processed in parallel, and that mutual inhibition would only occur between words that are associated with the same spatial location. Indeed, whereas Snell et al. established a facilitatory effect of neighbor flankers, they also observed an inhibitory effect when the same stimuli were used as masked primes at the target location.

Clearly, how an embedded word relates to its superset is to some extent similar to how a word relates to its orthographic neighbor. Like orthographic neighbors, embedded words and their supersets are orthographically related, but are represented by different word nodes. One might consequently expect that the embedded word should influence processing of its superset similarly to how an orthographic neighbor would.

The central problem here, then, is that it is unclear to what extent influences of embedded words should mimic neighborhood size- and neighbor flanker effects (facilitation) or masked neighbor priming effects (inhibition). On the one hand one may argue that words should be recognized faster upon the presence of an embedded word, given that the embedded word contributes to the orthographic neighborhood size of its superset, hence leading to stronger feedback activation of letters. On the other hand, it is possible that the word-to-word inhibitory connections that are postulated to exist among orthographic neighbors, also exist between embedded words and their supersets.

The few studies that have directly investigated how embedded words may influence visual word recognition, seem to argue for the latter. Investigating English and Spanish reading respectively, Davis and Taft (2005) and Davis et al. (2009) found that high-frequency embedded words slowed down responses in a lexical decision task—a finding reminiscent of the inhibitory effects of orthographic neighbors in lexical decision-making (e.g. Segui & Grainger, 1990; de Moor & Brysbaert, 2000; Davis & Lupker, 2006).

However, several uncertainties remain. For one, the inhibitory effects reported by Davis and colleagues (2005; 2009) were found with the exclusive use of low-frequency targets that contained high-frequency embedded words (e.g. '*come*' in '*comet*'). It is possible that different behavioral patterns emerge when the respective frequencies of targets and embedded words change, given that the strength of lateral inhibition depends on the word's initial activation level, which in turn is influenced by its frequency. A low-frequency neighbor is not likely to be strongly activated, and might therefore not yield an observable inhibitory effect. Additionally, the embedded words in aforementioned studies were always equal to the target minus one letter (coined 'deletion neighbors' by Davis et al.). Snell, Van Leipsig, Grainger and Meeter (2018b) have proposed that lateral inhibitory connections may exclusively exist among words of a sufficiently similar length[4], suggesting that while embedded words covering almost the entire superset would exert inhibition, shorter embedded words could instead facilitate due to the presence of word-to-letter feedback but absence of an inhibitory connection between the embedded word and its superset. The justification of this assumption is that if an embedded word is considerably shorter than its superset, such word length information should dispel any ambiguity that would otherwise invoke the need for a lexical competition mechanism (see Snell et al., 2018b, for the implementation of this principle in a computational model).

1.2 *Morphemes versus embedded words*

Yet another factor that may influence the embedded word's impact on recognition of its superset, is the morphological relationship between these words. While orthographic neighbors presumably exert mutual inhibition, morphologically related words (e.g., '*farming*' and '*farmer*') have been theorized to activate one another through feedback from morpho-semantic nodes ('*farm*') that are connected to all members of the morphological family. The principle, shared by contemporary frameworks of morphological processing in reading, is that activation of the orthographic representation '*farming*' would lead to activation of the morpho-semantic node '*farm*', which would in turn provide feedback activation of the orthographic representation '*farmer*' (Beyersmann, Castles & Coltheart, 2012; Crepaldi, Rastle, Coltheart & Nickels, 2010; Diependaele, Sandra & Grainger, 2009; Giraudo & Grainger, 2001; Grainger & Beyersmann, 2017).

Evidence for such a mechanism, mostly obtained with the masked priming lexical decision task, is mixed. On the one hand, lexical decisions about target words were made faster and more accurately when targets were preceded by morphologically related primes (e.g., '*teacher*' — '*teach*') than unrelated primes ('*finally*' — '*teach*'), while no such effects were observed when comparing non-morphological primes (e.g., '*dialog*' — '*dial*') to unrelated primes (Rastle, Davis & New, 2004; Longtin, Segui & Hallé, 2003). On the other hand, similar priming effects were

[4] Concretely, in our model of reading, OB1 reader (Snell et al., 2018b), we have proposed that word lengths are estimated with a 0.20 error margin. As such, a 6-letter word would be sufficiently similar to a 7-letter word.

reported for pseudo-derived primes that clearly do not belong to the same morphological family as the target (e.g., '*number*' — '*numb*'), hence indicating that there must be additional mechanisms at play.

The masked priming paradigm taken aside, the literature to date is scarce as well as equivocal about facilitation from morphemes. Testing single word reading, Carlisle and Stone (2005) found that elementary school students read bi-morphemic words (e.g., '*shady*') faster than mono-morphemic words (e.g., '*lady*'). Burani, Marcolini, De Luca and Zoccolotti (2008) obtained similar results with dyslexic children, but established no such effect with skilled children and adults. On the other hand, in a single-word lexical decision task (without primes), Hasenäcker, Schröter and Schroeder (2017) observed facilitation from compound, prefixed and suffixed words when compared to mono-morphemic words in both children and adults. Finally, in sentence reading, Niswander, Pollatsek and Rayner (2000) found that words comprising a high-frequency stem were read faster than words comprising a low-frequency stem (whole-word frequency being equal); however, to our knowledge, no sentence reading research to date has investigated the effects of the presence of stems per se.

1.3 *The present study*

Summing up the previous sections, embedded words might either have a facilitatory or inhibitory impact, depending on multiple factors. In particular, the scarce literature on this issue has suggested that long, high-frequency embedded words may inhibit processing of low-frequency supersets (Davis & Taft, 2005; Davis et al., 2009). It is unknown whether such effects hold for embedded words that are shorter and/or of higher frequency. Further, morphemes might facilitate processing, although evidence for this is largely restricted to the realm of masked priming lexical decision-making, with facilitation from pseudo-derived morphemes ('*corner*' — '*corn*') raising doubts about whether these effects are truly morphological in nature (Rastle et al., 2004; Longtin et al., 2003).

It is clear that the field lacks thorough investigations into the impact of embedded words. Of particular concern here, is the fact that most of the aforementioned studies employed highly artificial reading settings. Prior research has shown that discrepancies between such settings and more natural reading may lead to quite different behavioral patterns and, as a potential consequence, contrasting theories (e.g. Snell, Meeter & Grainger, 2017b; Snell, Declerck & Grainger, 2017c). To our knowledge, no prior research has provided a direct test of the impact of embedded words, morphologically related to the superset or not, in sentence reading.[5] While Davis et al. (2009) have reported a sentence reading experiment comparing targets with deletion neighbors versus targets without deletion neighbors, this experiment used targets for which the neighbor required deletion of an inner-letter (e.g., '*house*', which contains '*hose*'), meaning that they did not test true intact embedded words (comprised of contiguous letter combinations).

Hence, attempting to draw the bigger picture concerning embedded words, the present paper assesses embedded words in a natural reading setting. This investigation invokes analyses of the GECO book reading corpus (Cop, Dirix, Drieghe & Duyck, 2017), which contains eye-tracking data of 18 Dutch-English bilingual subjects and 14 English monolingual subjects reading

[5] Weingartner, Juhasz and Rayner (2012) have also tested embedded words in sentence reading; however, their study did not include a baseline condition using targets without embedded words (rather, they compared targets with a high-frequency embedded word to targets with a low-frequency embedded word). This study is therefore not informative about the impact of embedded words.

an entire novel.[6] The eye-tracking data comprise a multitude of variables, including, for instance, word viewing times, which are used as a measure of word recognition speed (e.g. Rayner, 1998).

Importantly, the vast size of the GECO corpus provides us abundant statistical power to determine whether and how effects of embedded words are modulated by factors such as length and frequency. Following the proposition of Snell et al. (2018b), we hypothesize that embedded words will only inhibit processing of their superset if they are of a sufficiently similar length (alongside the criterion of having a higher frequency than the superset, in line with aforementioned studies). Embedded words that are considerably shorter than their superset should thus lead to facilitation, due to feedback activation of letters but absence of inhibitory word-to-word connections. Lastly, as outlined in Section 2.2.3, we will carry out a preliminary investigation into the role of morphology, the exploratory nature of which prevents us to formulate concrete hypotheses.

2. Analysis of the GECO corpus

In their study, Cop et al. (2017) let 18 Dutch-English bilingual subjects[7] (F = 16, age range = 18–24) read the novel *The mysterious affair at Styles*, by Agatha Christie. This novel was chosen for being freely accessible on the internet in many different languages (allowing for replication in other languages) as well as having a below-average reading difficulty (Cop et al., 2017). The novel was read both in Dutch and in English, with 9 subjects reading the first half of the book in Dutch and the second half in English, while the 9 other subjects read the first half in English and the second half in Dutch. Language proficiency tests pointed out that the subjects' L2 proficiency was of an upper-intermediate (60–80%) level. Additionally, the authors tested 14 English monolinguals (F = 8, age range = 18–36) whom read the whole book in English.

Eye movement data was collected for 59,716 words (per two subjects) in the Dutch version of the novel, and for 54,364 words in the English version. The text was presented in paragraphs (with a maximum of 145 words per display) with black 14-point Courier New font on a light grey background and with triple line spacing. Subjects could press a button on a control pad to move from one paragraph to the next. After each of the 18 book chapters, subjects were presented multiple-choice questions to ensure that they had paid attention throughout the chapter. Reading of the entire novel took approximately four hours and was carried out in four 1-hour sessions. Eye movements were tracked with the EyeLink 1000 system (SR Research, Canada) with a sampling rate of 1 kHz.

2.1 *Our datasets*

From the GECO corpus we abstracted two datasets, representing Dutch reading (by 18 Dutch bilingual subjects) and English reading (by 14 monolingual English subjects). We retrieved each subject's first 2000 words with a length between 4 and 15 characters, amounting to a total of 36,000 datapoints for Dutch reading and 28,000 datapoints for English reading. Names (e.g. '*John*') and contractions (e.g. '*don't*') were avoided.

For each datapoint we determined the set of embedded words contained by the target word. This was done by checking for each of the target's contiguous letter combinations (e.g., '*ro*', '*oc*', '*ck*', '*roc*' and '*ock*' in '*rock*') whether it occurred in the Dutch Lexicon Project database of

[6] The corpus data is publicly available at http://expsy.ugent.be/downloads/geco/.
[7] Data from a 19th bilingual subject was also reported in Cop et al. (2017); however, this subject only read half of the book, and is thus not included in our analyses.

Keuleers, Diependaele and Brysbaert (2010) for Dutch words, or the British equivalent by Keuleers, Lacey, Rastle and Brysbaert (2012) for English words. The resulting number of embedded words was one of the independent variables used in our analyses. We further determined the length of embedded words relative to the target word, by dividing the average embedded word length by the target word length. Embedded word frequency was calculated by averaging the embedded words' log-transformed book frequencies (as reported in Keuleers et al., 2010; 2012).

We also marked items that exclusively comprised edge-aligned embedded words, as well as items that exclusively comprised non-edge-aligned embedded words (e.g., '*or*' in '*word*'). This was done to inspect the potential role of morphological relationships between embedded words and their supersets, with the rationale that edge-aligned embedded words are considerably more often morphemes (e.g., '*farm*' in '*farmer*') than non-edge-aligned embedded words (Grainger & Beyersmann, 2017).

Finally, for each target word we retrieved the gaze duration (GD) and the total viewing time (TVT) from the GECO corpus. GD reflects the sum of all first-pass fixation durations on a word, and arguably provides the most direct measure of word recognition speed (e.g., Rayner, 1998). TVT represents the sum of all fixation durations on a word (that is, including fixations following a regression), and reflects contextual comprehension. We further determined how many times each target was fixated (with zero times indicating a word skip).

Some descriptive statistics about our datasets are shown in Figures 1 to 3. Unsurprisingly, longer targets contain more embedded words on average (which is why target length was added as a factor in our models; Section 2.2.1). The number of embedded words does not go up with target frequency. In Dutch targets, the position of embedded words is on average slightly shifted to the right (by about 2–3%) compared to English targets (Figure 3).

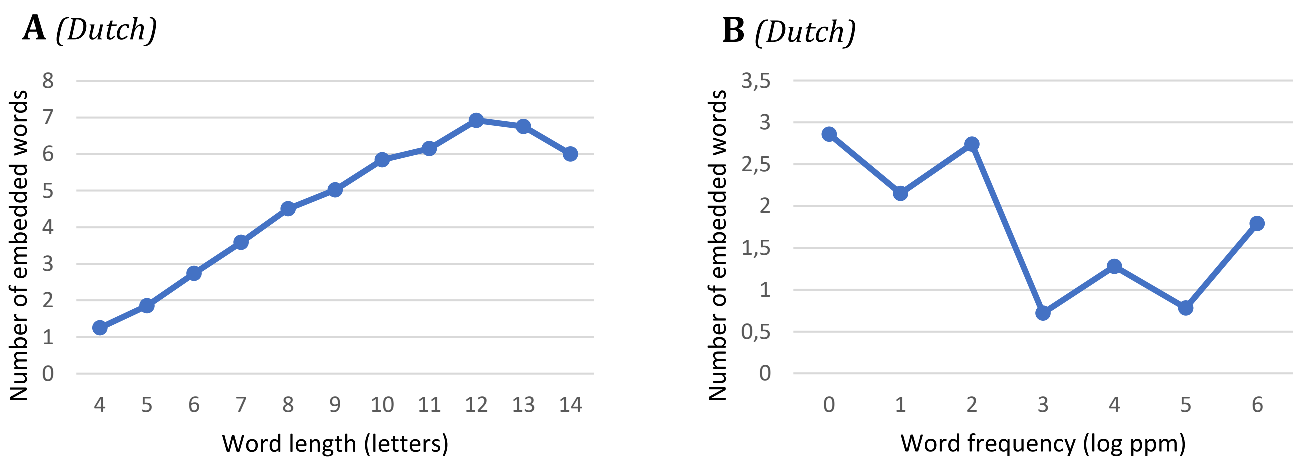

Figure 1. Descriptive statistics of the average number of embedded words as a function of word length (A) and word frequency (B) in Dutch reading.

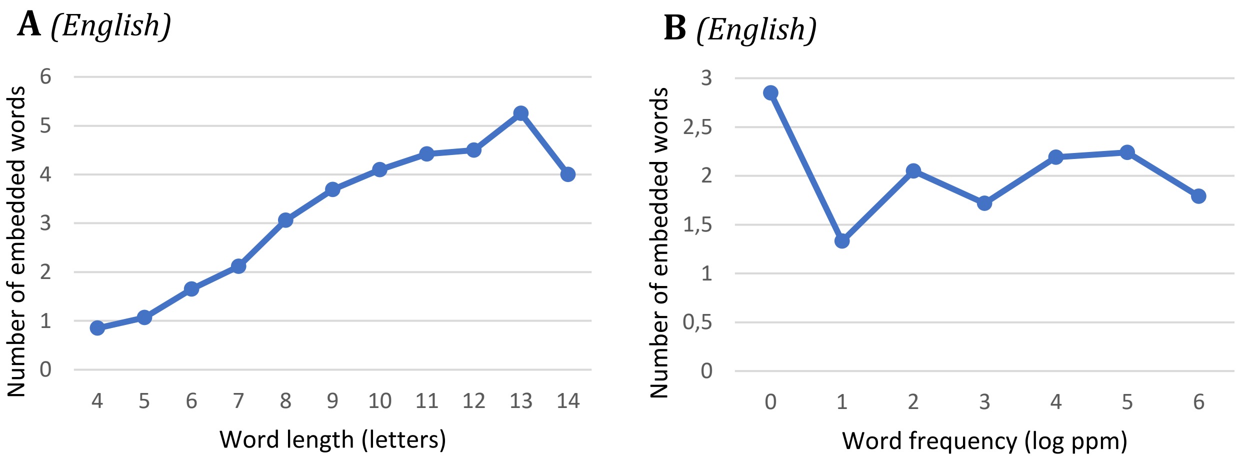

Figure 2. Descriptive statistics of the average number of embedded words as a function of word length (A) and word frequency (B) in English reading.

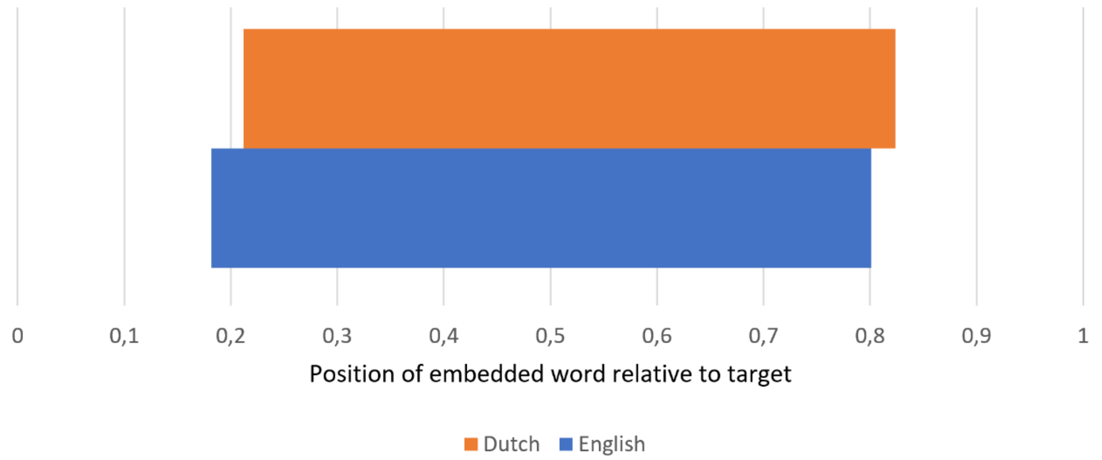

Figure 3. Average position of embedded words relative to the target, with the target's left and right boundary corresponding to 0 and 1 respectively.

2.2 *Results*

Below, our analyses are presented separately for Dutch and English reading. Skipped words (33.2%) were excluded from the word viewing time analyses. Words with a word viewing time (as reflected in GD) beyond 2.5 SD from the grand mean (1.81% of words) were excluded from the analyses of word viewing times as well as number of fixations.

We employed linear mixed-effect models (LMMs) with subjects entered as random factor (e.g. Baayen, 2008). The models were fitted with the lmer function from the lme4 package (Bates, Maechle, Bolker & Walker, 2015) in the R statistical computing environment. Following Barr, Levy, Scheepers and Tily (2013) we determined the maximal random effect structure permitted by the data. This led us to include by-subject random slopes alongside the random intercept. We report *b*-values, standard errors (SEs) and *t*-values for all factors, with *t*-values of |1.96| and beyond deemed significant.

2.2.1 *Dutch reading*

We firstly tested for a main effect of the number of embedded words on target word recognition speed. The LMM for this analysis included the number of L1 embedded words and target word length as factors. The latter variable was included to account for its cofound with the number of embedded words: i.e., longer words, which are usually recognized slower, [8] are likely to contain more embedded words (Figure 1A).

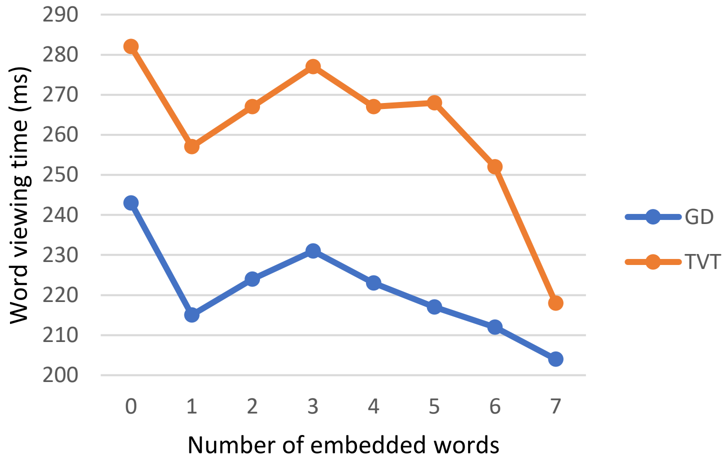

Figure 4. Word viewing times (gaze duration (GD) and total viewing time (TVT)) as a function of the number of embedded words in 6-letter Dutch targets.

As it turned out, word viewing times decreased as the number of embedded words increased: in GD, b = -2.93, SE = 0.79, t = -3.71; in TVT, b = -4.91, SE = 1.48, t = -3.31 (see word viewing times on 6-letter words in Figure 4). Significantly fewer fixations were made upon an increasing number of embedded words: b = -0.03, SE = 0.01, t = -4.46. We established an interaction of the number of embedded words and target word length, with the effect of embedded words on GD and number of fixations being more strongly pronounced in longer target words: in GD, b = 0.37, SE = 0.07, t = 5.31; in TVT, b = 0.45, SE = 0.12, t = 3.80; in number of fixations, b = 2.49×10^{-3}, SE = 5.30×10^{-4}, t = 4.70.

Next, we assessed whether the effect of embedded words on target word recognition was modulated by embedded word length. This was done by including embedded word length as a variable in the model alongside number of embedded words and target word length. A significant interaction of embedded word length and the number of embedded words on word viewing times was established, with b = 12.02, SE = 5.19, t = 2.32 in GD, and b = 27.54, SE = 9.92, t = 2.78 in TVT. This interaction was also expressed in the number of fixations: b= 0.09, SE = 0.04, t = 2.16. Aiming to explore the nature of this interaction, we split the dataset in two, based on embedded word length being above or below the median of 0.46 (this is a normalized value with 1 corresponding to the target word length). Crucially, when running the original model on these two datasets, we found that relatively short embedded words facilitated target processing (GD: b = -4.70, SE = 0.95, t = -4.94; TVT: b = -6.28, SE = 1.89, t = -3.32; number of fixations: b = -0.03, SE = 0.01, t = -4.84) whereas no effect was found for the relatively longer embedded words (GD: b = 1.70, SE = 1.81, t = 0.94; TVT: b = 2.09, SE = 3.15, t = 0.67; number of fixations: b = -0.01, SE = 0.01, t = -0.69).

Our assessment of embedded word frequency followed the same procedure. Here, a marginally significant interaction was established between embedded word frequency and the

[8] We indeed found that the fixation duration increased with word length: b = 5.75, SE = 0.50, t = 11.54.

number of embedded words, (GD: b = 1.18, SE = 0.66, t = 1.78; TVT: b = 2.01, SE = 1.17, t = 1.71; number of fixations: b = 1.18, SE = 0.66, t = 1.78), such that the facilitatory main effect of embedded words was slightly stronger when they were low-frequency (i.e., below the frequency median of 4.95: b = -3.60, SE = 0.94, t = -3.84) compared to when they were high-frequency (i.e., above the frequency median: b = -2.09, SE = 1.11, t = -1.88). Considering that prior single word reading research has only established inhibitory effects using high-frequency embedded words (Davis & Taft, 2005; Davis et al., 2009), we further scrutinized the subset of data with long embedded words for which we found no significant effect. Aligning with the observations of Davis and colleagues, an inhibitory effect emerged with long embedded words above the frequency median (GD: b = 9.74, SE = 3.16, t = 3.08; TVT: b = 13.75, SE = 4.69, t = 2.93; number of fixations: b = 0.10, SE = 0.02, t = 3.92) while no effect was observed for long embedded words below the frequency median (b = -0.79, SE = 2.37, t = -0.33; TVT: b = -5.61, SE = 4.26, t = -1.32; number of fixations: b = 0.04, SE = 0.02, t = 1.44).

2.2.2 *English reading*

Like the Dutch-English bilingual readers, the English monolingual readers showed a facilitatory main effect of the number of embedded words on the word viewing time (GD: b = -2.00, SE = 1.16, t = -1.72; TVT: b = -6.07, SE = 1.95, t = -3.12) and number of fixations (b = -0.02, SE = 0.01, t = -2.05) (Figure 5). The number of embedded words again interacted with target word length (GD: b = 0.27, SE = 0.14, t = 2.02; TVT: b = 0.78, SE = 0.23, t = 3.43; number of fixations: b = $3.70*10^{-3}$, SE = $1.25*10^{-3}$, t = 2.95).

However, the English monolinguals did not show an interaction between number of embedded words and embedded word length (GD: b = -3.69, SE = 55.50, t = -0.07; TVT: b = -22.63; SE = 13.57; t = -1.67; number of fixations: b = 0.08, SE = 0.07, t = 1.13) or embedded word frequency (GD: b = 2.69, SE = 3.22, t = 0.83; TVT: b = 1.58, SE = 1.09, t = 1.45; number of fixations: $3.81*10^{-3}$, SE = 0.01, t = 0.40). Indeed, this time around, no inhibition of embedded words was observed for cases above the length- and frequency medians (0.50 and 4.48 respectively), with b = -2.46, SE = 5.52, t = -0.45 for GD; b = 5.39, SE = 6.03, t = 0.90 for TVT; and b = -0.02, SE = 0.03, t = -0.61 for the number of fixations.

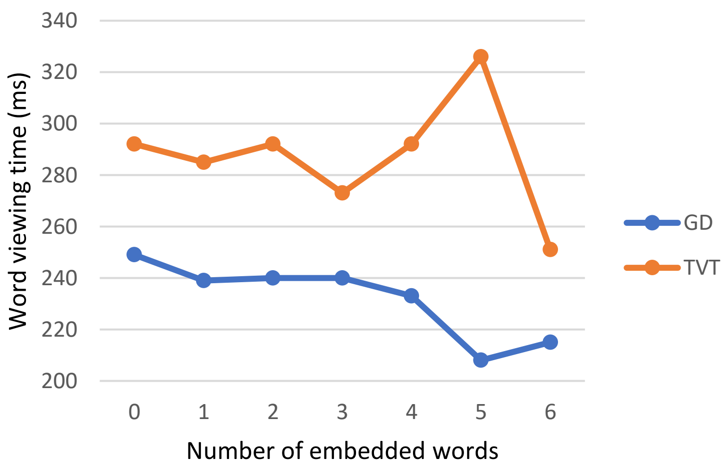

Figure 5. Word viewing times as a function of the number of embedded words in 6-letter English targets, (one might note that the facilitatory effect of embedded words on TVT, as reported in the text, is not apparent in this figure; this is likely because the facilitatory effect is driven by targets of other lengths).

2.2.3 *Morphology*
Lastly, we assessed whether the main facilitatory effect of number of embedded words may have been driven by morphological relationships between the embedded words and their supersets. Given the obvious impracticality of manually determining morphological relationships between all items and each of their respective embedded words (necessitating one to evaluate more than 175,000 word pairs within the present datasets), we opted for an alternative strategy whereby we made use of the fact that stems (e.g., '*farm*' in '*farmer*') are most often edge-aligned (Grainger & Beyersman, 2017). We thus reasoned that, if the facilitatory effects were driven by morphemes, these effects should be stronger when isolating items that exclusively comprise edge-aligned embedded words.[9]

As it turned out, in Dutch reading, no influence of number of embedded words was observed in the subset of items with exclusively edge-aligned embedded words (56.49% of all items): in GD, b = -0.85, SE = 3.15, t = -0.27; in TVT, b = -1.16, SE = 4.66, t = -0.25; in number of fixations, b = 0.03, SE = 0.02, t = 1.35. On the other hand, inhibition was found in items that exclusively contained non-edge-aligned embedded words (17.25% of all items), albeit exclusively in the GD measure: in GD, b = 12.26, SE = 5.77, t = 2.13; in TVT, b = -3.87, SE = 8.13, t = -0.48; in number of fixations, b = -0.02, SE = 0.05, t = -0.53.

In the English dataset, edge-aligned embedded words (53.20% of all items) were found to facilitate: in GD, b = -18.52, SE = 5.26, t = -3.52; in TVT, b = -10.60, SE = 6.00, t = -1.77; in number of fixations, b = -0.05, SE = 0.03, t = -1.82. No effects were observed in items exclusively containing non-edge-aligned embedded words (26.00% of all items): in GD, b = 4.02, SE = 5.70, t = 0.71; in TVT, b = -5.46, SE = 8.35, t = -0.65; in number of fixations, b = -0.03, SE = 0.05, t = -0.68.

3. Discussion

Through decades of reading research, the question of whether and how embedded words might influence the visual word recognition process has received relatively little attention. Yet, the investigation of embedded words may provide a valuable contribution to our understanding of the reading process, given that observations of facilitation or inhibition from embedded words would be informative about the existence of word-to-letter feedback connections and word-to-word inhibitory connections in the brain. Concretely, if embedded words were found to speed up the recognition process, this would provide further evidence for word-to-letter feedback, with the rationale that a stimulus comprising more embedded words would lead to the activation of more word nodes. Those word nodes in turn provide more feedback activity to letter nodes, leading to faster word recognition (cf. the orthographic neighborhood size effect; e.g. Andrews, 1989; Forster & Shen, 1996; Johnson & Pugh, 1994). Alternatively, if embedded words were found to inhibit the recognition process, this would support the idea that word-to-word inhibitory connections exist among lexical competitors (e.g. Segui & Grainger, 1990; DeMoor & Brysbaert, 2000; Davis & Lupker, 2006).

Our analyses show that, in Dutch reading, embedded words can have both a facilitatory and inhibitory impact on target word recognition: that is, when embedded words are short (less than half the target's length), they tend to speed up recognition of the target. In contrast, longer embedded words can slow down the recognition process, on the premise that they are of high

[9] It must be acknowledged that this approach only allows for tentative conclusions concerning the role of morphology. Complete insight into the role of morphology warrants a direct comparison of morphologically related versus unrelated embedded words.

frequency. These results reaffirm earlier claims that embedded words may inhibit the recognition process (Davis & Taft, 2005; Davis et al., 2009), although it should be noted that, overall, embedded words rather tend to facilitate the recognition process (as illustrated by a facilitatory main effect of the number of embedded words).

Strikingly, in English reading, no inhibitory effect was observed at all, even when embedded words were long and of high frequency. This null-result contrasts with the pattern of effects in lexical decision times observed by Davis and Taft (2005) as well as with the pattern of effects depicted by Dutch subjects of the GECO corpus. These discrepancies may have been caused by cross-lingual differences (Dutch versus English reading) and the different natures of the respective tasks (lexical decision versus natural reading). With respect to task differences, for example, it is worth noting that while lexical decisions are facilitated by an increased orthographic neighborhood size, sentence reading is slowed (Pollatsek, Perea & Binder, 1999). Cross-lingual discrepancies are not uncommon either: as outlined by Andrews (1997), neighborhood frequency effects have produced more reliable inhibition in Spanish and French than in English.

Morphology is one potentially relevant factor. Prior research has shown facilitated processing of morphologically complex words (i.e., words consisting of more than one morpheme) comprising a high-frequency root or lexeme, compared to morphologically complex words comprising a low-frequency root or lexeme, in English sentence reading (Niswander et al., 2000). One might in this light wonder whether a higher proportion of target words comprised a root or lexeme in the English data compared to the Dutch data in the present analyses—which, in a more general sense, would be the result of key differences between the Dutch and English morphology. However, as shown in Section 2.2.3, the proportion of items with exclusively edge-aligned embedded words was approximately equal in Dutch and English (56.94% and 53.20% of all items, respectively). Hence, although practical constraints prevented us from obtaining an absolute count of the number of morphemes in both languages, we estimate that the number of morphemes must have been fairly equal. Here, then, it is noteworthy that edge-aligned embedded words, which more often consist of stems (e.g., '*farm*' in '*farmer*') facilitated word processing in English, while no effects of edge-aligned embedded words were observed in Dutch. This means that if morphology was a critical factor driving differences between Dutch and English reading, then such differences would be reflected in the way morphemes affect processing in each respective language, rather than in the number of morpheme occurrences per se. Naturally, such accounts of aforementioned discrepancies are at this point mere speculation, but would be worthy of further investigation in future research.

Overall, the main facilitatory effect of embedded words (observed both in Dutch and English reading) puts a theoretical constraint on the role of word-to-word inhibitory connections. In line with the proposition of Snell et al. (2018b), it is possible that such connections do not simply exist among all words that share letters. Rather, a prerequisite for such connections to exist—or at least to have a considerable impact—seems to be that words must have a sufficiently similar length. Indeed, from a theoretical perspective this makes sense: word-to-word connections were theorized to function as a means to prevent that multiple words are recognized upon viewing a single word (e.g., without these connections, viewing '*rack*' may lead to simultaneous recognition of '*rack*' and '*rock*'). If an embedded word is considerably shorter than the word by which it is contained, however, such word length information should dispel any ambiguity that would otherwise drive the need for a lexical competition mechanism.

Our exploratory investigation into the role of morphology has not provided a clear answer about whether the observed facilitatory effects are driven by morphologically related embedded words. On the one hand, the inhibitory effect of non-edge-aligned embedded words (i.e., words that are presumed to be largely morphologically unrelated to the superset) in Dutch reading might be taken as evidence that the facilitatory effects must have been driven by morphemes. On the other hand, the absence of this effect in English, alongside the absence of a facilitatory effect of edge-aligned embedded words in Dutch, casts doubt on the idea that morphological relationships were the main driving force here. As it is more generally unclear how morphemes influence the word recognition process (e.g., reports of facilitation from pseudo-derived morphemes) it seems that the field of morphology research is in dire need of means to overcome these knowledge gaps before we can reasonably start disentangling effects of morphemes on the one hand, and general effects of embedded words on the other.

In sum, in the present paper we have aimed to shed some light on the impact of embedded words in natural reading. The finding that words are generally recognized faster when they contain an increased amount of embedded words supports the idea that word nodes provide feedback activation to letter nodes. However, when embedded words are long and highly frequent, the recognition process may be slowed, suggesting that word-to-word inhibitory connections may only exist among words of similar length.

Acknowledgements

This research was supported by the European Research Council (grant number ERC742141 awarded to Jonathan Grainger) and the Brain & Language Research Institute (ANR-11-LABX-0036).

References

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

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

Baayen R. (2008). *Analyzing Linguistic Data*: *A pratical introduction to statis*tics. Cambridge University Press.

Barr D., Levy R., Scheepers C. & Tily H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255-278.

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

Beyersmann, E., Castles, A. & Coltheart, M. (2012). Morphological processing during visual word recognition in developing readers: Evidence from masked priming. *The Quarterly Journal of Experimental Psychology, 65,* 1306-1326.

Bowers, J., Davis, C. & Hanley, D. (2005). Automatic semantic activation of embedded words: Is there a 'hat' in 'that'?. *Journal of Memory and Language, 52,* 131-143.

Carlisle, J. & Stone, C. (2005). Exploring the role of morphemes in word reading. *Reading Research Quarterly, 40,* 428-449.

Cop, U., Dirix, N., Drieghe, D. & Duyck, W. (2017). Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. *Behavior Research Methods, 49,* 602-615.

Crepaldi, D., Rastle, K., Coltheart, M. & Nickels, L. (2010). 'Fell' primes 'fall', but does 'bell' prime 'ball'? Masked priming with irregularly-inflected primes. *Journal of Memory & Language, 63,* 83-99.

Davis, C. & Lupker, S. (2006). Masked inhibitory priming in English: Evidence for lexical inhibition. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 668-687.

Davis, C., Perea, M. & Acha, J. (2009). Re(de)fining the orthographic neighborhood: The role of addition and deletion neighbors in lexical decision and reading. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 1550-1570.

Davis, C. & Taft, M. (2005). More words in the neighborhood: Interference in lexical decision due to deletion neighbors. *Psychonomic Bulletin & Review, 12,* 904-910.

de Moor, W. & Brysbaert, M. (2000). Neighborhood-frequency effects when primes and targets are of different lengths. *Psychological Research, 63,* 159-162.

Diependaele, K., Sandra, D. & Grainger, J. (2009). Semantic transparency and masked morphological priming: The case of prefixed words. *Memory & Cognition, 37,* 895-908.

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

Giraudo, H. & Grainger, J. (2001). Priming complex words: Evidence for supralexical representation of morphology. *Psychonomic Bulletin & Review, 8,* 127-131.

Grainger, J. (2008). Cracking the orthographic code: An introduction. *Language and Cognitive Processes, 23*, 1-35.

Grainger, J. & Beyersmann, E. (2017). Edge-aligned embedded word activation initiates morpho-orthographic segmentation. In B. Ross (Ed.), *Psychology of Learning and Motivation, 67* (pp. 285–317). Elsevier academic press.

Grainger, J. & Jacobs, A. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103,* 518-565.

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

Johnson, N. & Pugh, K. (1994). A cohort model of visual word recognition. *Cognitive Psychology, 26,* 240-346.

Keuleers E., Diependaele K. & Brysbaert M. (2010). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology*, *1,* 1-15.

Keuleers, E., Lacey, P., Rastle, K. & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods, 44,* 287-304.

Longtin, C., Segui, J. & Hallé, P. (2003). Morphological priming without morphological relationship. *Language and Cognitive Processes, 18,* 313-334.

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

Niswander, E., Pollatsek, A. & Rayner, K. (2000). The processing of derived and inflected suffixed words during reading. *Language and Cognitive Processes, 15,* 389-420.

Pollatsek, A., Perea, M. & Binder, K. (1999). The effects of neighborhood size in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 1142-1158.

Rastle, K., Davis, M. & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition*. Psychonomic Bulletin & Review, 11*, 1090-1098.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372-422.

Snell, J., Bertrand, D., Meeter, M. & Grainger, J. (2018a). Integrating orthographic information across time and space: Masked priming and flanker effects with orthographic neighbors. *Experimental Psychology* (in press).

Snell, J., Declerck, M. & Grainger, J. (2017c). Parallel semantic processing during reading revisited: Effects of translation equivalents in bilingual readers. *Language, Cognition & Neuroscience*, doi: 10.1080/23273798.2017.1392583.

Snell, J., Meeter, M. & Grainger, J. (2017b). Evidence for simultaneous syntactic processing of multiple words during reading. *PLoS ONE, 12*, e0173720.

Snell, J., Van Leipsig, S., Grainger, J. & Meeter, M. (2018b). OB1-Reader: A model of word recognition and eye movements in text reading. *Psychological Review,* (in press).

Snell, J., Vitu, F. & Grainger, J. (2017a). Spatial integration of parafoveal orthographic information: Beyond the sub-lexical level? *Quarterly Journal of Experimental Psychology, 70,* 1984–1994.

Weingartner, K., Juhasz, B. & Rayner, K. (2012). Lexical embeddings produce interference when they are morphologically unrelated to the words in which they are contained: Evidence from eye movements. *Journal of Cognitive Psychology, 24,* 179-188.

Integrating orthographic information across time and space: Masked priming and flanker effects with orthographic neighbors

Joshua Snell, Daisy Bertrand, Martijn Meeter & Jonathan Grainger

Citation: Snell, J., Bertrand, B., Meeter, M. & Grainger, J. (2018). Integrating orthographic information across time and space: Masked priming and flanker effects with orthographic neighbors. *Experimental Psychology, 65,* 32-39.

Abstract

Research has suggested that the word recognition process is influenced by the integration of orthographic information across words. The precise nature of this integration process may vary, however, depending on whether words are in temporal or spatial proximity. Here we present a lexical decision experiment, designed to compare temporal and spatial integration processes more directly. Masked priming was used to reveal effects of temporal integration, while the flanker paradigm was used to reveal effects of spatial integration. Primes/flankers were high-frequency orthographic neighbors of the target (*blue-blur*) or unrelated control words (*head-blur*). We replicated prior observations of inhibition in trials where the neighbor was used as a masked prime, while facilitation was observed in trials where the neighbor was presented as flanker. We conclude that sub-lexical orthographic information is integrated both temporally and spatially, but that spatial information is used to segregate lexical representations activated by spatially distinct sources.

1. Introduction

The processing of orthographic information during reading involves both the temporal and spatial integration of information. Temporal integration of orthographic information concerns the accumulation over time of information extracted from the same spatial location, and is typically evaluated by presenting successive orthographic stimuli (words and nonwords) at the same location (Grainger & Jacobs, 1999). Spatial integration of orthographic information concerns the combination of information extracted from different word locations, at the same point in time (e.g., Dare & Shillcock, 2013). In the present study we investigate the mechanisms that may underlie these integration processes, and in particular, to what extent they may differ.

The masked priming paradigm (Forster & Davis, 1984) has been the paradigm of choice for investigating the temporal integration of information during single word reading. Brief presentation of the prime stimulus is thought to prevent it from being perceived as a distinct perceptual event (Humphreys, Evett, & Quinlan, 1990) hence facilitating integration of information across prime and target (Grainger & Jacobs, 1999). Temporal integration of orthographic information can then be investigated by manipulating the orthographic overlap across prime and target stimuli (e.g., Ferrand & Grainger, 1992; 1994; Forster & Davis, 1984; Humphreys et al., 1990).

More recently, spatial integration of orthographic information has been revealed in a paradigm introduced by Dare and Shillcock (2013), the flanking letters lexical decision (FLLD) task, whereby a central target stimulus is flanked by two letters on each side, separated from the target by a space (e.g. '*ro rock ck*'). Here, spatial integration is investigated by means of manipulating the orthographic overlap between the target word and the two flanking stimuli.

In the present study, we focus on one manipulation that has produced contrasting effects in the masked priming and flanker paradigms. The manipulation in question is one where primes / flankers can be orthographic neighbors of target words (e.g., *blue-blur*) or unrelated words (e.g., *head-blur*). Prior research has revealed inhibitory effects of orthographic neighbor primes in masked priming (e.g., De Moor & Brysbaert, 2000; Davis & Lupker, 2006; Segui & Grainger, 1990). On the contrary, orthographic neighbor flanking stimuli have been found to facilitate target word processing (Snell, Vitu, & Grainger, 2017a).

The inhibitory effects of word neighbor primes found with masked priming have been taken as evidence for competitive processes operating between lexical representations (lexical competition) during visual word recognition (Segui & Grainger, 1990). In support of this interpretation, Jacobs and Grainger (1992) demonstrated that lateral inhibition across co-activated lexical representations in the interactive-activation model (McClelland & Rumelhart, 1981) accurately simulated inhibitory priming effects from orthographic neighbors. It is further known that these effects are affected by word frequency and -lexicality: the strongest inhibition is obtained with a combination of high-frequency prime words and low-frequency target words (Segui & Grainger, 1990), while non-word neighbor primes either generate facilitatory priming or null effects (Forster & Davis, 1991; Forster, Davis, Schoknecht, & Carter, 1987; Van Heuven, Dijkstra, Grainger & Schriefers, 2001).

The inhibitory effects of neighbor primes but concurrent facilitatory effects of non-word neighbor primes suggest that the *temporal* integration of orthographic information takes place both at the sub-lexical level as well as the lexical level. Following this reasoning, considering that orthographic neighbors facilitated target processing in the flanker paradigm (Snell et al., 2017a), Snell et al. concluded that the *spatial* integration of orthographic information operates at the sub-lexical level but not beyond. We further elaborate on this reasoning in the Discussion.

The facilitatory parafoveal-on-foveal effect reported by Snell et al. (2017a) speaks against a single channel "one-word-at-a-time" approach to word identification and reading (e.g., Andrews & Reichle, 2017; Grainger, Dufau, & Ziegler, 2016). According to Grainger et al. (2016), orthographic information provided by flanking stimuli is integrated into a single channel that outputs a unique word identity. Given a flanker condition *'bl blur ue'*, the flanking letters '*bl*' and '*ue*' should combine with orthographic information extracted from the target '*blur*' and provide evidence for the competing word *'blue'*, leading to inhibition and not to the facilitation observed by Snell et al. (2017a). Instead, their results suggest that despite the spatial integration of sub-lexical orthographic information, the lexical representations that are consequently activated continue to be processed independent from one another – as long as these are associated with spatially distinct sources.

On the other hand, one could argue that this pattern was obtained because orthographic information concerning the competing word was split across the left and right flankers in the Snell et al. experiment, whereas in masked priming the competing word was intact. This caused individual flankers to bear no lexical status (e.g., neither '*ro*' nor '*ck*' in '*ro rock ck*' is a word), as such possibly activating sub-lexical integration processes but not lexical integration processes. It is therefore important to examine effects of word neighbor flankers when these are intact, such as in the example '*blue blur blue'* – while ensuring, crucially, that no facilitation is obtained with the same stimuli and participants in the masked priming paradigm. This was the primary goal of the present study.

2. Method

2.1 Participants

32 students from Aix-Marseille University gave informed consent to participate in this experiment for €4. All participants reported to be native to the French language, non-dyslexic, and had normal or corrected-to-normal vision. All participants were naïve to the purpose of the experiment.

2.2 Materials

Using the same procedure as in Snell et al. (2017a), we retrieved a list of 74 triplets (target word (e.g. '*brut'*), orthographic neighbor (e.g. '*bout'*) and orthographically unrelated control word (e.g. '*noix'*)) from the French Lexicon Project database (Ferrand et al., 2010). All words consisted of four letters, were non-conjugated and contained no diacritics. Word pairings were chosen such that orthographic neighbors and control words had a lower lexical decision time (LDT) than their respective target word (for targets, neighbors and controls, the mean LDT was 671 ms, 618 ms, and 615 ms respectively).[10] Targets and neighbors only differed in an inner-positioned letter. In a similar fashion we retrieved a list of 74 pseudoword triplets from the French Lexicon Project pseudoword database (Ferrand et al., 2010). These were filler stimuli, not to be included in data analyses. We present the complete stimulus list in the Appendix.

2.3 Design

[10] Following Snell et al. (2017a) we selected stimuli based on the LDT measure because it more directly reflects the speed with which words become active and reach recognition. Words with a low LDT value are activated faster, and as such exert stronger inhibition on lexical competitors; hence the choice for low-LDT primes and high-LDT targets.

We used a 2 × 2 × 2 factorial design, with word lexicality (*word / pseudoword*), trial type (*masked priming / flanking*) and relatedness of the prime/flanker (*neighbor / control*) as factors. Participants were Latin-squared into two groups, such that every stimulus was presented twice to each participant (once in the neighbor condition and once in the control condition) and in both trial types per two participants. The experiment thus consisted of 296 trials per participant (148 of which were included in the analyses), and these were presented in randomized order.

2.4 Apparatus

The experiment was implemented with OpenSesame (Mathôt, Schreij & Theeuwes, 2012) and presented on a 1024x768 px, 150 Hz computer monitor. Participants were seated in a comfortable office chair at a distance of 50 cm from the display, so that each character space subtended 0.40 degrees of visual angle. Responses were collected with a keyboard.

2.5 Procedure

Before commencing the experiment, participants received instructions both verbally by the experimenter and visually onscreen. Participants were instructed to fixate in between two centrally located vertical fixation bars that were presented throughout the experiment. Figure 1 shows the procedure for each trial type. Both trial types would start with a 500 ms mask consisting of four hashmarks. In masked priming trials, the mask would be replaced by the neighbor/control for 70 ms, followed by the target word. In the flanker trials, the mask would be replaced by the target, with the neighbor/control being presented left and right of the target (separated by a single character space). Following Snell et al. (2017a), all words were presented in 18-point monospaced font (droid sans mono; the default monospaced font in OpenSesame) and in lower case. The target would stay onscreen until participants pressed a left- or right-handed key for *pseudoword* or *word* respectively. Participants were instructed to respond as quick and accurate as possible, and the maximum allowed response time was 1800 ms after the target onset. Participants received feedback in the form of a briefly presented centrally located green or red dot, for correct and incorrect responses respectively. The next trial began immediately after the 600 ms feedback signal.

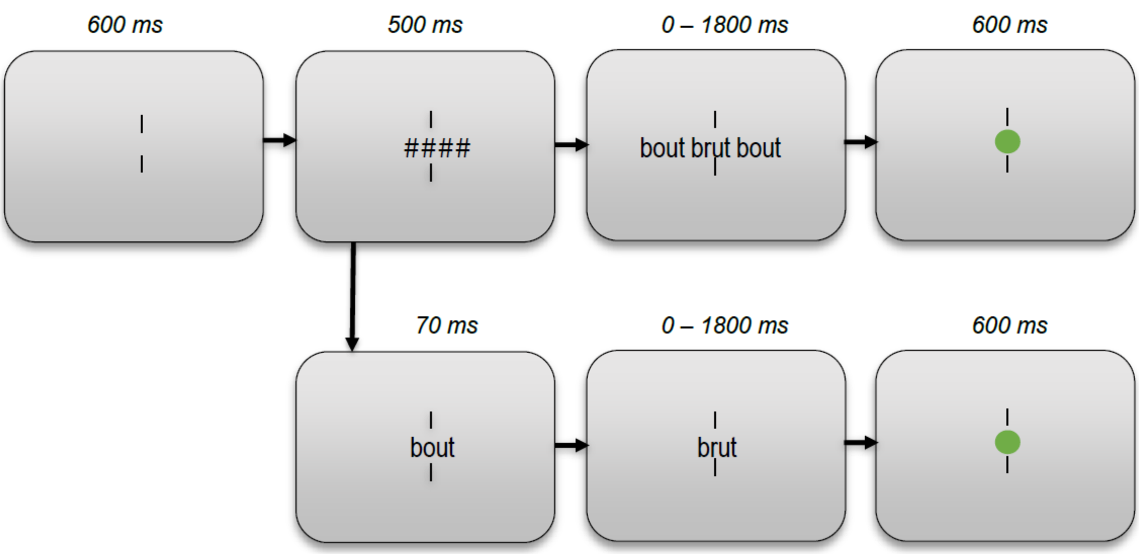

Figure 1. Overview of the trial procedure in the flanker setting (top) and masked priming setting (bottom). The size of stimuli relative to the screen is exaggerated in these examples.

3. Results

Only correctly answered trials (93.14%) were included in the analysis of response times (RT).[11] For our analyses of RTs and error rates we used linear mixed-effect models (LMMs) with items and participants as crossed random effects (Baayen, 2008). To meet the models' assumption that the data were normally distributed, RTs were inverse-transformed (-1000/RT) prior to the analyses. The models were fitted with the lmer function from the lme4 package (Bates, Maechle, Bolker & Walker, 2015) in the R statistical computing environment. Following Barr, Levy, Scheepers and Tily (2013) we determined the maximal random effect structure permitted by the data. This led us to include by-item and by-participant random intercepts, as well as by-item and by-participant random slopes. We report regression coefficients (b), standard errors (SE) and t-values. Fixed effects were deemed reliable if $|t| > 1.96$ (Baayen, 2008). Logistic LMMs (fitted with the glmer function) were used to analyze the error rates. Below we present separate analyses for flanker trials and masked priming trials, followed by a direct comparison of the two trial types.

3.1 Flanker trials

We replicated the finding of Snell et al. (2017a) that target processing is facilitated by orthographic neighbor flankers, as RTs were significantly lower in the neighbor condition as compared to the control condition (b = 0.03, SE = 0.01, t = 2.54; condition means in Table 1). The error rate did not differ significantly between conditions (b = 0.25, SE = 0.19, z = 1.36).

3.2 Masked priming trials

Whereas our neighbor stimuli were found to facilitate target processing in the flanker condition, the opposite pattern was found in the masked priming trials. An inhibitory effect was found in the error rates, with significantly more errors following neighbor primes than control primes (b = 0.43, SE = 0.22, z = 1.98). The pattern of RTs followed the same direction numerically (see Table 1), but did not reach significance (b = -0.02, SE = 0.01, t = -1.27).[12]

Table 1. Response times (ms) and error rates (probability)

	Response times		*Error rates*	
Condition	*Neighbor*	*Control*	*Neighbor*	*Control*
Flanker trials	731 (171)	742 (164)	.072 (.045)	.059 (.050)
Masked priming	707 (161)	697 (149)	.058 (.044)	.039 (.048)

Note: values in between parentheses indicate standard deviations.

[11] The raw datafiles can be accessed online at https://osf.io/tq38d/.

[12] It should be noted that inhibitory prime effects are not always established in RT data. Zimmerman and Gomez (2012) have argued that the amount of attentional resources spent on processing of the prime directly affects the chance of finding an inhibitory effect, and that longer prime durations might as such lead to stronger inhibitory effects.

3.3 Comparison of trial types

To compare the two trial types directly,[13] we entered the interaction of relatedness × trial type in a separate model. The effect in RTs of prime/flanker on target processing turned out to interact significantly with trial type (b = 0.05, SE = 0.02, t = 3.15), thus confirming the significance of the opposite pattern of effects found in the two trial types (Table 1). No significant interaction was established in the error rates (b = 0.19, SE = 0.28, z = 0.69; Table 1).

Lastly, there was a noteworthy main effect of trial type on RTs, with increased RTs in the flanker setting compared to the masked priming setting (b = 0.06, SE = 0.01, t = 5.67). This suggests that flankers generally perturbed target processing more than primes.

4. Discussion

A lexical decision experiment examined the effects of orthographic neighbors on target word recognition when the neighbors were either presented as masked primes immediately before the target word at the same location, or presented as flanking words simultaneously with, and to the left and to the right of the target word. The general aim was to compare the temporal integration of orthographic information as revealed by masked priming, with the spatial integration of orthographic information as revealed by the flanker task. Insight into the respective natures of these different types of integration further provides a means to test two opposing accounts of word identification and reading: a single-channel, one-word-at-a-time account (e.g., Grainger et al., 2016; Reichle, Pollatsek & Rayner, 2006) and a multi-channel, parallel word identification account (e.g., Snell, Meeter, & Grainger, 2017b).

According to the single channel model proposed by Grainger et al. (2016), orthographic processing operates in parallel across multiple words during sentence reading (cf., Engbert, Nuthmann, Richter & Kliegl, 2006; Reilly & Radach, 2006), and the orthographic information extracted from different words is integrated into a single processing channel that outputs a unique word identity. We reasoned that if this were the case, then the presence of an orthographic neighbor as flanking stimulus should lead to inhibition of target word processing, mimicking the effects seen when orthographic neighbors are presented as masked prime stimuli (De Moor & Brysbaert, 2000; Lupker & Davis, 2006; Segui & Grainger, 1990). Snell et al. (2017a) put this reasoning to test and, on the contrary, found that parafoveal orthographic neighbors facilitated target word processing both in a sentence reading setting as well as in the flanker task (e.g., '*bl blur ue*'). When used as flanker, however, the orthographic neighbor was divided either side of the target, contrary to the use of complete prime stimuli in masked priming experiments. The results of the present experiment show that presenting whole-word flankers on either side of the target (e.g., '*blue blur blue*') produces a similar facilitatory effect, the size of the effect being 11 ms compared with the 14 ms effect in the Snell et al. (2017a) study.[14] Crucially, in the present study, the same participants showed an inhibitory priming effect with the same stimuli when these were presented as primes and targets in a masked priming procedure.

[13] We acknowledge that such a comparison is complicated by the many differences between the two paradigms (we further elaborate on these differences in the Discussion), but believe that a direct comparison is nonetheless relevant in the context of the present study.

[14] We further note the presence of a small speed-accuracy trade-off in the present study, with a non-significant increase in errors arising in the presence of related flankers.

Why then do orthographic neighbor flanking stimuli facilitate target word identification in the flanker task? The answer offered by Snell et al. (2017b) is that orthographic information extracted from distinct spatial locations is integrated sub-lexically (see also Angele, Tran & Rayner, 2013; Grainger, Mathôt, & Vitu, 2014; Snell et al., 2017a), hence facilitating target word recognition when there is orthographic overlap. What is novel in Snell et al.'s (2017b) account is that spatial information is used to keep track of which activated word representation belongs to which spatial location, hence enabling parallel higher-level processing of multiple stimuli. The fact that this parallel processing is geared to output several distinct word identities means that flanker and target stimuli do not interfere at the level of lexical processing and beyond. Thus, whereas sub-lexical orthographic information is integrated across spatially *and* temporally distinct stimuli, lexical integration takes place within- rather than across spatial locations.

On a methodological note, it is important to consider the various differences between the masked priming and flanker trials – in particular with respect to the availability of the prime / flanker stimulus during target processing – and whether or not such differences may have contributed to the opposing (facilitatory vs. inhibitory) effects obtained in each respective setting. Concretely, one might argue that the neighbor could have been processed to a further extent in flanker trials than in masked priming trials, given that the neighbor was only available for 70 ms in the latter trial type whereas it was available during the whole stimulus-response interval in the former trial type. On the other hand one might argue that the constraints imposed by visual acuity cause foveal processing of the prime stimulus to be of higher quality than parafoveal processing of the flanker stimulus, as such compensating for their different presentation time. Importantly, we opted to keep flanking stimuli onscreen rather than to have them disappear after 70 ms because the offset of these stimuli would have directed attention away from the fovea (similar to a stimulus onset). Crucially, even if flankers were processed to a further degree than primes, this should have then only increased the effects that were established here. Indeed, it is clear that the potential increased processing of flankers compared to primes did not lead to inhibition, as might otherwise be expected following deeper integration of information between orthographic neighbors.

In sum, the present results underline the idea that the integration of orthographic information from multiple words can impact the recognition process in various ways. The outcome of this integration process seems to depend strongly on the words' spatial locations, in line with the idea that readers keep track of which word belongs to which position: when word representations are tied to the same spatial location, the integration of information is carried on to the lexical level, where lexical competition perturbs the recognition process. In contrast, when word representations are tied to different spatial locations, this segregation allows for parallel independent lexical processing, resulting in stronger activation and faster word recognition.

Acknowledgements

This research was supported by grants ANR-15-CE33-0002-01 and ANR-11-LABX-0036 (BLRI) from the French National Research Agency (ANR).

References

Andrews, S. & Reichle, E.D. (2017). The cognitive architecture of reading: The organization of an acquired skill. In P. Hagoort (Ed.), *Human Language: From Genes and Brains to Behavior*, MIT press.

Angele, B., Tran, R. & Rayner, K. (2013). Parafoveal–foveal overlap can facilitate ongoing word identification during reading: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 526–538.

Baayen, R. (2008). *Analyzing Linguistic Data: A pratical introduction to statistics*. Cambridge, UK: Cambridge University Press.

Barr D., Levy R., Scheepers C. & Tily H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015), Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, *67,* 1–48.

Dare, N. & Shillcock, R. (2013). Serial and parallel processing in reading: Investigating the effects of parafoveal orthographic information on nonisolated word recognition. *The Quarterly Journal of Experimental Psychology, 66*, 417–428.

Davis, C., & Lupker, S. (2006). Masked inhibitory priming in English: Evidence for lexical inhibition. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 668–687.

De Moor, W. & Brysbaert, M. (2000). Neighborhood-frequency effects when primes and targets are of different lengths. *Psychological Research, 63,* 159–162.

Engbert, R., Nuthmann, A., Richter, E. & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112,* 777–813.

Ferrand, L. & Grainger, J. (1992). Phonology and orthography in visual word recognition: Evidence from masked nonword priming*. Quarterly Journal of Experimental Psychology, 45,* 353–372.

Ferrand, L., et al. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods, 42,* 488–496.

Forster, K., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*, 680–698.

Forster, K., & Davis, C. (1991). The density constraint on form-priming in the naming task: Interference effects from a masked prime*. Journal of Memory and Language*, *30,* 1–25.

Forster, K., Davis, C., Schoknecht, C., & Carter, R. (1987). Masked priming with graphemically related forms: Repetition or partial activation*? Quarterly Journal of Experimental Psychology*, *39,* 211–251.

Grainger, J., Dufau, S., & Ziegler, J. (2016). A vision of reading. *Trends in Cognitive Sciences, 20,* 171–179.

Grainger, J. & Jacobs, A. (1999). Temporal integration of information in orthographic priming. *Visual Cognition*, *6,* 461–492.

Grainger, J., Mathôt, S. & Vitu, F. (2014). Test of a model of multi-word reading: Effects of parafoveal flanking letters on foveal word recognition. *Acta Psychologica, 146,* 35–40.

Humphreys, G., Evett, L., & Quinlan, P. (1990). Orthographic processing in visual word identification. *Cognitive Psychology*, *22,* 517–560.

Jacobs, A. & Grainger, J. (1992). Testing a semistochastic variant of the interactive activation model in different word recognition experiments. *Journal of Experimental Psychology, Human Perception and Performance, 20,* 1311–1334.

Mathôt, S., Schreij, D. & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*, 314–324.

McClelland, J. & Rumelhart, D. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*, *88,* 375–407.

Reichle, E., Pollatsek, A. & Rayner, K. (2006). E-Z Reader: A cognitive-control, serial-attention model of eye movement behavior during reading. *Cognitive Systems Research, 7,* 4–22.

Reilly, R. & Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research*, 7, 34–55.

Snell, J., Vitu, F. & Grainger, J. (2017a). Integration of parafoveal orthographic information during foveal word reading: Beyond the sub-lexical level? *Quarterly Journal of Experimental Psychology*, *70,* 1984–1994.

Snell, J., Meeter, M. & Grainger, J. (2017b). Evidence for simultaneous syntactic processing of multiple words during reading. *PLoS ONE, 12*, e0173720.

Segui, J., & Grainger, J. (1990). Priming word recognition with orthographic neighbors: Effects of relative prime-target frequency. *Journal of Experimental Psychology: Human Perception and Performance*, *16,* 65–76.

Van Heuven, W., Dijkstra, T., Grainger, J. & Schriefers, H. (2001). Shared neighborhood effects in masked orthographic priming. *Psychonomic Bulletin & Review*, *8,* 96–101.

Zimmerman, R. & Gomez, P. (2012). Drawing attention to primes increases inhibitory word priming effects. *The Mental Lexicon, 7,* 119–146.

Chapter 2: Orthographic integration and the attentional gradient

Parafoveal letter-position coding in reading

Joshua Snell, Daisy Bertrand & Jonathan Grainger

Citation: Snell, J., Bertrand, D. & Grainger, J. (2018). Parafoveal letter-position coding in reading. *Memory & Cognition, 46,* 589-599.

Abstract

The masked-priming lexical decision task has been the paradigm of choice for investigating how readers code for letter identity and position. Insight into the temporal integration of information between prime and target words has pointed out, among other things, that readers do not code for the absolute position of letters. This conception has spurred various accounts of the word recognition process, but the results at present do not favor one account in particular. Thus, employing a new strategy, the present study moves out of the arena of temporal information integration, and into the arena of *spatial* information integration. We present two lexical decision experiments that tested how the processing of 6-letter target words is influenced by simultaneously presented flanking stimuli (stimulus on-time 150 ms). We manipulated the orthographic relatedness between targets and flankers, both in terms of letter identity (same/different letters and using the target's outer-/inner- letters) and letter position (intact/ reversed order of letters and flankers, and contiguous/non-contiguous flankers). Target processing was strongly facilitated by same-letter flankers, and this facilitatory effect was modulated by letter/flanker order and contiguity. However, when flankers consisted of the target's inner-positioned letters alone, letter order no longer mattered. These findings suggest that readers may code for the relative position of letters using words' edges as spatial points of reference. We conclude that the flanker paradigm provides a fruitful means to investigate letter-position coding in the fovea and parafovea.

Introduction

There is general consensus that the process of visual word recognition is largely driven by orthographic processing, that is, the coding of letter identity and letter position and the consequent activation of lexical representations. However, decades of reading research have made it clear that the cognitive mechanisms underlying letter-position coding are not easy to pinpoint, and today's researchers are still faced with the challenge of "cracking the orthographic code" (Grainger, 2008). Nonetheless, the development of good, testable theoretical models and experimental paradigms fit to put those models to the test, has advanced our understanding of the recognition process considerably.

The masked-priming lexical decision task, pioneered by Forster and Davis (1984), has arguably made the strongest impact. In this task, participants make lexical decisions (*word, non-word*) about target words that are preceded by briefly presented (sometimes unnoticed) prime words. Manipulating the way in which prime words relate to target words, for example in terms of orthography, phonology or semantics, has allowed researchers to gain insight into whether and how these types of information are processed and integrated over time.

Concerning orthographic processing, masked priming has revealed at least three things. Firstly, orthographic information can be integrated across multiple stimuli, such that same-letter primes lead to faster target recognition compared to different-letter primes (e.g. Forster & Davis, 1984; Forster, Davis, Schoknecht & Carter, 1987). Secondly, this integration process may be carried on to the lexical level, such that orthographically related word pairs (i.e., orthographic neighbors: *rock* – *rack*) compete for recognition by exerting mutual inhibition, leading to slower recognition as compared to unrelated word pairs (e.g. *stop* - *rack*) (e.g. Segui & Grainger, 1990; Davis & Lupker, 2006; De Moor & Brysbaert, 2000; De Moor, Van der Herten, Verguts, 2007). Thirdly, evidence suggests that readers do not code for the absolute position of letters. For instance, it has been shown that target words are recognized faster after a transposed-letter prime (e.g. *mohter* – *mother*) as compared to a prime with different letters at the same positions (e.g. *monder* – *mother*) (Perea & Lupker, 2004; Schoonbaert & Grainger, 2004). Additionally, Peressotti and Grainger (1999) found that the processing of 6-letter target words was facilitated by 4-letter relative-position primes (e.g. *mthr* – *mother*) as compared to unrelated primes (e.g. *lndn* – *mother*) (see also Bowers, Davis & Hanley, 2005; Van Assche & Grainger, 2006; Grainger, Granier, Farioli, Van Assche & Van Heuven, 2006).

The conception that readers do not code for the absolute position of letters has led researchers to argue against the seminal Interactive Activation model (IAM) of McClelland & Rumelhart (1981; 1982) wherein letters are processed in a rigid slot-based fashion (e.g. *rock* would strictly activate a detector for *'r'* at position 1, a detector for '*o*' at position 2, et cetera), and the search for theoretical alternatives has since spurred at least three modeling approaches: *noisy slot-based coding, spatial coding* and *relative-position coding*. Noisy slot-based coding refers to the addition of Gaussian noise to the slot-based scheme of the IAM (Gomez, Perea & Ratcliff, 2007; Norris & Kinoshita, 2012), meaning that each letter of a stimulus would not only activate the node representing that letter at its specific slot (*n*), but also in slots *n-2, n-1, n+1, n+2* etc., with increasing eccentricity from the letter's true position leading to weaker activation. This Gaussian noise renders the system less efficient, but more flexible, and allows it to account for the effects of transposed-letter primes and relative-position primes (although some extreme cases of relative-position priming might be problematic for noisy slot-coding).

Spatial coding (Davis, 1999; 2010) implements flexibility in a fairly similar way, by adding letter position uncertainty to a spatial code of letter representations. The SOLAR model additionally buys a certain amount of length-independent flexibility, such that the word '*backstop*' would activate representations for '*back*' and '*stop*' as well (Davis, 2010).

The third modeling approach, relative-position coding, assumes that orthographic input activates location-invariant nodes that represent the relative position of within-word letter pairs (e.g. the stimulus *'rock'* would activate nodes for *'ro', 'rc', 'rk', 'oc', 'ok'* and *'ck'*; see Whitney, 2001; Grainger & van Heuven, 2003). These so-called *open-bigram* units in turn activate all lexical representations that they belong to. The unit *'ro'*, for example, would activate *'rock'*, but also *'rose'* and *'ribbon'*. Accounting for the transposed-letter priming effect (e.g. *'rock'* is primed more strongly by *'rcok'* than by *'rduk'*), the lexical representation of *'rock'* would be activated by a larger subset of open-bigram units with the former prime (*'rc', 'ro', 'rk', 'ck', 'ok'*) as compared to the latter prime (*'rk'*) (Grainger & Whitney, 2004).

Beyond the isolated word

The three modeling approaches seem to do equally well in explaining results obtained with the masked-priming paradigm (e.g. Grainger, 2008, for a review), suggesting that the question of which account of orthographic processing is the most plausible cannot be answered by the masked-priming paradigm alone. In the present paper we therefore shift focus away from temporal information integration as revealed by masked-priming, and instead build on evidence suggesting that readers can also integrate information across spatially distinct stimuli.

Using a paradigm where subjects had to identify two words that were briefly presented together (e.g. *sand lane*), McClelland and Mozer (1986) showed that letter migration errors can occur (e.g. *land sane*). Davis and Bowers (2004) showed that such illusory identifications do not have to respect position: given a word pair like *step soap,* participants could also respond '*stop*', indicating a migration of the letter 'o' from position two to position three.

In sentence reading, fixation durations on word *n* are found to be shorter when *n* is orthographically related to word *n+1* (Inhoff, Radach, Starr & Greenberg, 2000; Angele, Tran & Rayner, 2013; Dare & Shillcock, 2013; Kennedy & Pynte, 2005; Snell, Vitu & Grainger, 2017), indicating that orthographic information from multiple words is simultaneously processed and integrated. Further, using the novel Flanking Letters Lexical Decision (FLLD) task, Dare and Shillcock (2013) found that lexical decisions about foveal target words were made faster when the target letters were repeated as flankers on each side (e.g. *'ro rock ck'*) as compared to when flanker letters were unrelated ('*st rock ep'*). Crucially, the same facilitatory effect was obtained when the order of flanker bigrams was reversed ('*ck rock ro*'). As pointed out by Grainger, Mathôt and Vitu (2014), this particular finding is in line with the Open-Bigram model of Grainger and Van Heuven (2003) which operates on the principle that open-bigram nodes (representing the relative position of within-word letter pairs) are location-invariant. In a similar experiment, Grainger et al. went on to replicate the absence of an effect of bigram order, but at the same time found that letter order did matter (i.e. *'or rock kc'* let to longer RT's than '*ro rock ck*'), confirming the importance of relative- rather than absolute letter position.

Indeed, it is difficult to conceive how noisy slot-based coding could account for these findings, as the greater distance between letter occurrences in the switched bigram condition should have led to less activation and therefore longer RTs (e.g. the parafoveal '*k*' is only three slots away from the target's '*k*' in '*ro rock ck*', while it is five slots away in '*ck rock ro*'). The only solution to this problem would be to assume that the target and flankers were processed as three separate units, such that the '*k*' in the flanker '*ck*' was coded for position 2, irrespective of the position of the flanker relative to the target. The '*k*' in the target '*rock*' would as such receive additional activation from a '*k*' that is only two slots away, in both the original and the switched bigram condition.

The letter order effect reported by Grainger et al. (2014) suggests that both the coding of parafoveal letters as well as the spatial integration of this information is quite fine-grained. However, more research needs to be done in order to gain full understanding of the mechanisms underlying (parafoveal) letter-position coding. For instance, according to the Open-Bigram model of Grainger and Van Heuven (2003), similar facilitatory parafoveal-on-foveal effects should be obtained with contiguous flankers (e.g. *'ro rock ck*') and non-contiguous flankers (e.g. '*rc rock ok'*). Further, it can be argued that the stimuli used in aforementioned FLLD studies were within the limits of what readers can process in a single glance, and it would be worthwhile to see whether the presence of a letter order effect and the absence of a bigram order effect (i.e., location-invariance) persist in a setting using longer targets and flankers. Finally, previous research has suggested that outer-positioned letters may have a special role in orthographic processing due to reduced crowding effects for these letters as compared to inner-positioned letters, and it has been argued that the edges of a word may be used as anchoring points during letter-position coding (e.g. Fischer-Baum, McCloskey & Rapp, 2010; Fischer-Baum, Charny & McCloskey, 2011; Jacobs, Rey, Ziegler, & Grainger, 1998). The FLLD paradigm would be suited to put this conception to the test, for instance by comparing intact vs. switched letter flankers using only inner-positioned letters (e.g. '*ar barrel re'* vs. '*ra barrel er*').

Below we report two experiments that address these issues. Experiment 1 was designed as a test of the importance of letter order and contiguity, in a setting using 6-letter targets and 3-letter flankers. Experiment 2 provides a further test of the Open-Bigram scheme of Grainger and Van Heuven (2003) by comparing intact vs. switched flankers in a setting using 6-letter targets and 3-letter flankers. Additionally, Experiment 2 compares intact vs. reversed letter flankers using inner-positioned letters, to test an account of letter-position coding according to which the word's edges are used as spatial points of reference.[15]

Experiment 1

Method

Participants

Twenty students volunteered to participate in this study, carried out at the Aix-Marseille Université in Aix-en-Provence, France. All students reported to be non-dyslexic and native to the French language, and had normal or corrected-to-normal vision. All students were naïve to the purpose of the experiment.

Stimuli and design

From the French Lexicon Project lexicon and its pseudo-word lexicon counterpart (Ferrand et al., 2010) we retrieved 150 6-letter word targets and 150 6-letter pseudo-word targets (the latter of which were used as filler stimuli). These targets were selected from a lexical decision time range (LDT) of 500–700 ms (mean frequency of the word targets was 4.79 Zipf; see Van Heuven, Mandera, Keuleers and Brysbaert, 2014, for more on the Zipf scale). Word targets were noun, adjective or non-conjugated verbs. Targets contained no diacritics (e.g. *á, ï, û, ç*), which are quite

[15] All data of the present work are available at https://osf.io/y2vj8/.

typical in the French orthography, but whose effects are not well-documented in word recognition research.

In every trial, the target was flanked by three letters on the left and three letters on the right. We manipulated these flanking letters across five conditions, as can be seen in Table 1. Four of the five conditions had flankers that were orthographically related to the target, whereas the fifth condition served as a control with unrelated flankers. For this control condition we retrieved a word from the lexicon for each target such that the LDT values were equal (difference <50 ms) and that there was no orthographic overlap (mean frequency of the control words was 4.83 Zipf). Pseudo-word targets were coupled to an unrelated pseudo-word in a similar way. The unrelated words were split into two halves that were used as left- and right-sided flankers respectively (e.g., target '*laptop*' and unrelated word *'sprint'* would become *'spr laptop int'*). The four related conditions followed a 2 × 2 factorial design, with letter order (intact / reversed) and contiguity (contiguous / non-contiguous) as factors. The exact same conditions were also implemented for pseudo-word trials. We used a Latin-square design to ensure that all targets were seen across all conditions, but only once per participant. The experiment thus contained 300 experimental trials, and these were presented in randomized order.

Table 1. Experiment 1 conditions.

	Intact letter order	*Reversed letter order*
Contiguous letters	123 123456 456	321 123456 654
Non-contiguous letters	135 123456 246	531 123456 642
Control (unrelated letters)	ddd 123456 ddd	

Note: The target's letters are indicated by the digits 123456. The flanker digits indicate letter identities based on their respective location in the target. Unrelated letters, used in the control condition, are indicated with '*d'.*

Apparatus and software

The stimuli and experimental design were implemented with OpenSesame (Mathôt, Schreij, & Theeuwes, 2012) and presented on a 17-inch, 1024x768 pixel, 150 Hz display. Participants were seated at a 60cm distance from the display, so that each character space subtended 0.30º of visual angle.

Procedure

Participants were seated in a comfortable office chair in a dimly-lit room. Centrally positioned vertical fixation bars, separated from one another by 0.60º of visual angle, were shown throughout each trial. 500 ms after the start of every trial, a target stimulus with flankers (separated from the target by one character space) was shown for 150 ms in between the fixation bars, after which participants had a maximum of 2000 ms to respond. Responses were given by means of a right- or left-handed button press, to indicate *word* or *non-word* respectively. In case of a correct button response, a green fixation dot was shown at the center of the screen for 700 ms. In case of an incorrect- or no response, a red fixation dot was shown for 700 ms. The display then returned to the beginning state, 500 ms after which a new trial would commence. Participants were offered a short break after every block of 100 trials. Participants would start every block with three practice trials for which we did not collect data. The experiment lasted approximately 20 minutes.

Results

Only correctly answered trials with a word target[16] were included in the analysis of response times (RTs). Trials with an RT beyond 2.5 SD from the grand mean (2.8%) were discarded. We used linear mixed-effect models (LMMs) with items and participants as crossed random effects (Baayen, 2008). We followed Barr, Levy, Scheepers and Tily (2013) in determining the maximal random effects structure permitted by the data. This led us to include by-item and by-participant random slopes alongside random intercepts[17]. The models were fitted with the lmer function from the lme4 package (Bates, Maechler, Bolker & Walker, 2015) in the R statistical computing environment. We report regression coefficients (b), standard errors (SE) and t-values. Fixed effects were deemed reliable if $| t | > 1.96$. Logistic LMMs were used to analyze the error rates.

Prior to the analyses, we determined for each of the 3000 trials whether the flankers were in themselves words, as well as whether the flankers were pronounceable. The rationale here is that the lexicality and pronounceablity of flankers may have biased participants toward a certain response (e.g., non-word flankers and/or unpronounceable flankers may have induced a *non-word* response). In a similar vein we determined the flanker bigram frequency, considering that flankers with a low bigram frequency could have been experienced as more unusual, again biasing participants toward the *non-word* response. Flanker bigram frequencies represent the average frequency of all bigrams (as indicated by the Lexique bigram database of New and Pallier[18]) occurring in the flanker (e.g. '*ye*', '*ys*' and '*es*' in '*yes*'). These variables were not taken into consideration during stimulus selection, but their potential relevance was realized at a later stage. Hence, to exclude the possibility that our results would be (partially) driven by these variables, flanker lexicality, pronounceability and bigram frequency (log ppm.) were included in the LMMs for all analyses. The variables are also listed in Table 2.

As will be seen below, none of our effects of interest were influenced by flanker lexicality, pronounceability and bigram frequency. Note that one might nonetheless also attempt to control these variables in the stimulus selection phase[19], rather than in the analyses.

[16] In pseudo-word trials, no significant effects in RT were observed (compared to the control condition, b = 10.69, SE = 12.96, t = 0.83 for contiguous intact letter order flankers; b = 0.26, SE = 13.06, t = 0.02 for contiguous reversed letter order flankers; b = 0.73, SE = 13.09, t = 0.06 for non-contiguous intact letter order flankers; and b = -7.76, SE = 13.05, t = -0.59 for non-contiguous reversed letter order flankers). No main effects of letter order (b = 10.25, SE = 9.41, t = 1.09) and contiguity (b = 9.91, SE = 9.31, t = 1.06) were observed. Similar non-significant values were observed in the error rates.

[17] It should be noted that there is as of yet no consensus on the most appropriate random effects structure of LMMs, with some peers being in favor of the maximal random effects structure (e.g., Barr et al., 2013) while others prefer models with fewer parameters (e.g. Baayen, Davidson & Bates, 2008). Indeed, one may argue that the models used in the present study were over-parameterized as the correlations between random slope and intercept were 1.0 for subjects and 0.8 for items. Importantly, the patterns of effects reported in this work persist with a simpler model.

[18] http://www.lexique.org/listes/liste_bigrammes.php

[19] To control for the lexicality and pronounceability of flankers, we checked whether flankers occurred in the word and pseudoword lexicon of Ferrand et al. (2010). All flankers occurring in either lexicon were marked as *pronounceable*; all flankers specifically occurring in the word lexicon were marked as *word.* All other flankers were marked as *nonword* and/or *unpronounceable*. One may apply the same procedure to filter out stimuli during the stimulus selection phase. As for flanker bigram frequency, one will have to delete stimuli until the average flanker bigram frequency is equal across conditions.

Table 2. Number of word target trials with word flankers or pronounceable flankers in Experiment 1.

	Trials with word flankers	*Trials with pronounceable flankers*	*Average bigram frequency (log ppm.)*
Contiguous intact	0	180	3.78
Contiguous reversed	132	213	3.72
Non-contiguous intact	0	70	3.72
Non-contiguous reversed	0	33	3.64
Control (unrelated letters)	0	186	3.87

Response times

Table 3 shows the mean RTs and error rates across conditions. There were significantly lower RTs in all conditions with target-related flankers as compared to the control condition (Table 4). Meanwhile, no significant influences of flanker lexicality (b = -10.78, SE = 10.67, t = -1.01), pronounceability (b = 4.20, SE = 4.94, t = 0.85) and bigram frequency (b = 0.70, SE = 3.18, t = 0.22) were established.

There was a significant main effect of flanker letter-order, such that RTs in trials with intact letter-order flankers were lower than RTs in trials with reversed letter-order flankers, with b = 14.00, SE = 3.90, t = 3.59. This effect of letter-order was not modulated by flanker lexicality (b = -4.42, SE = 8.76, t = -0.51), pronounceability (b = 3.75, SE = 9.86, t = 0.38) or bigram frequency (b = 3.31, SE = 5.47, t = 0.61) (these variables did not produce main effects in this analysis and subsequent analyses either).

We also found a significant main effect of contiguity, with contiguous letter flankers yielding shorter RTs than non-contiguous letter flankers: b = 7.39, SE = 3.31, t = 2.23. Like letter-order, the effect of contiguity was not modulated by flanker lexicality (b = -4.42, SE = 8.76, t = -0.51), pronounceability (b = 8.14, SE = 11.76, t = 0.69) or bigram frequency (b = -3.16, SE = 6.71, t = 0.47). There was also no significant interaction between letter-order and contiguity: b = 2.46, SE = 6.61, t = 0.37.

Table 3. Condition means

Condition	*Flanker letters*		*RT (ms)*	*Error rate*
Contiguous intact	*123*	*456*	562 (110)	0.019 (0.025)
Contiguous reversed	*321*	*654*	575 (120)	0.024 (0.031)
Non-contiguous intact	*135*	*246*	572 (121)	0.032 (0.033)
Non-contiguous reversed	*531*	*642*	579 (113)	0.031 (0.009)
Control (unrelated letters)	*ddd*	*ddd*	605 (124)	0.040 (0.010)

Note: values in between parentheses indicate standard deviations.

Table 4. Analysis of response times and error rates (ref.: control)

	Response time			Error rate		
Condition	*b*	*SE*	*t*	*b*	*SE*	*z*
(intercept):	594.80	30.72	**14.48**	-3.80	0.30	**12.60**
Contiguous intact	-42.03	5.31	**-7.92**	-0.81	0.39	**-2.11**
Contiguous reversed	-28.18	5.83	**-4.84**	-0.77	0.39	**-1.99**
Non-contiguous intact	-34.58	5.61	**-6.16**	-0.17	0.33	-0.52
Non-contiguous reversed	-20.96	5.78	**-3.62**	-0.43	0.35	-1.23

Note: Significant values are shown in bold.

Error rates

Compared to the control condition, the error rate was significantly lower in the contiguous bigram conditions (Table 4). Again, flanker lexicality (*b* = -0.46, SE = 0.68, *z* = -0.67), pronounceability (*b* = -035, SE = 0.32, *z* = -1.10) and bigram frequency (*b* = -0.02, SE = 0.21, *z* = -0.08) had no influence here. There was no main effect of letter-order on the error rate (*b* = 0.13, SE = 0.28, *z* = 0.46), and no interaction of letter-order and flanker lexicality (*b* = -0.24, SE = 0.56, *z* = -0.42), pronounceability (*b* = 0.20, SE = 0.33, *z* = 0.60) or bigram frequency (*b* = 0.49, SE = 0.29, *z* = 1.67).

There was a marginally significant effect of contiguity, with contiguous flankers leading to fewer errors than non-contiguous flankers (*b* = 0.50, SE = 0.29, *z* = 1.76). This effect was again not modulated by flanker lexicality (*b* = -0.69, SE = 0.59, *z* = -1.17), pronounceability (*b* = 0.28, SE = 0.31, *z* = 0.88) or bigram frequency (*b* = 0.53, SE = 0.34, *z* = 1.57). As in the RTs, no significant interaction between letter-order and contiguity was found in the error rates: *b* = 0.30, SE = 0.57, *z* = 0.52.

Discussion

Experiment 1 replicated the letter order effect reported by Grainger et al. (2014), this time with 6-letter targets and 3-letter flankers, suggesting that even at the greater eccentricities implicated by the use of longer stimuli in Experiment 1, parafoveal orthographic processing is sensitive to letter position information. The fact that the reversed letter order flankers nonetheless yielded considerably shorter RTs than the control flankers, suggests that readers also code for separate letter identities, completely irrespective of position, and that the coding of letter identity and position may thus be represented by two distinct cognitive components (e.g., Grainger et al. (2014) argued for the presence of a "bag-of-letters" alongside a "bag-of-bigrams"; see also Peressotti & Grainger, 1995, for an earlier proposal). We also found an effect of contiguity, with contiguous flankers (e.g. 123 123456 456) yielding shorter RTs than non-contiguous flankers (e.g. 135 123456 246).

Within the Open-Bigram scheme of Grainger and Van Heuven (2003), the effect of contiguity may be explained by the idea that bigrams are only formed between letters that are not too far apart from each other in the stimulus, meaning that the target 123456 would activate bigrams for 12, 13 and 14, but not 15 or 16 (see also Hannagan & Grainger, 2012; Whitney, 2001). Similarly, the target would activate bigrams for 56, 46 and 36, but not 26. Now, the contiguous flankers 123 and 456 would activate the bigrams 12, 13, 23, 45, 46 and 56, all of which belong to the target's set of bigrams. In contrast, the non-contiguous flankers 135 and 246 activate the bigrams 15 and 26, which are not part of the target's set of bigrams, thus leading to increased RTs (however, see also the Discussion section of Experiment 2 below, for an alternative explanation). Importantly, the letter-order effect persisted in the non-contiguous flanker trials, in line with a relative-position coding scheme.

Our aim for Experiment 2 was twofold. Firstly, since we replicated the letter-order effect reported by Grainger et al. (2014) using 6-letter targets, we now wanted to provide a further test of the second claim of their study, i.e., that bigram order does *not* matter. We thus compared a condition with intact flanker order (123 123456 456) against one with switched flanker order (456 123456 123). Secondly, according to the relative-position coding approach, a letter-order effect should be obtained when using inner-positioned letters in the flankers (e.g., comparing 23 123456 45 vs. 32 123456 54). However, it has been argued that outer-positioned letters may play an important role in letter-position coding (e.g. Fischer-Baum et al., 2010; 2011; Jacobs et al., 1998), and this alternative account would predict that there should be no letter-order effect when using inner-positioned letters alone. We thus included these conditions in Experiment 2 as well. It should therefore be noted that Experiment 2 did not follow a 2 × 2 factorial design, but rather tested two pairs of conditions in isolation. We also included a control condition with unrelated flanker letters to evaluate overall effects of flanker relatedness.

Experiment 2

Method

Twenty-six students volunteered to participate in this study, carried out at the Aix-Marseille Université in Aix-en-Provence, France. The experimental conditions for Experiment 2 are shown in Table 5. Other than that, the entire methodology for Experiment 2 was left unchanged from Experiment 1.

Table 5. Experiment 2 conditions.

Intact flanker order	123 123456 456
Switched flanker order	456 123456 123
Intact inner letters	23 123456 45
Reversed inner letters	32 123456 54
Control (unrelated letters)	ddd 123456 ddd

Note: Targets' constituent letters are indicated by the digits 123456. The flanker digits indicate letter identities based on their respective location in the target. Unrelated letters, used in the control condition, are indicated with '*d'.*

Results

We again included only correctly answered trials in the analysis of RTs. The 2.5 SD cut-off led to the exclusion of 2.94% of trials. Data were analyzed using LMMs in the same way as in Experiment 1. Flanker lexicality, pronouncability and bigram frequency (Table 6) were again included as factors in all analyses.

Table 6. Number of word target trials with word flankers or pronounceable flankers in Experiment 2.

	Trials with word flankers	*Trials with pronounceable flankers*	*Average bigram frequency (log ppm.)*
Intact flanker order	0	180	3.78
Switched flanker order	0	180	3.78
Intact inner-letter flankers	343	426	4.18
Reversed inner-letter flankers	238	293	4.08
Control (unrelated letters)	0	186	3.87

Table 5. Condition means

Condition	*Flanker letters*		*RT (ms)*	*Error rate*
Intact flanker order	*123*	*456*	600 (111)	0.019 (0.025)
Switched flanker order	*456*	*123*	612 (115)	0.024 (0.031)
Intact inner-letter order	*23*	*45*	605 (111)	0.032 (0.033)
Reversed inner-letter order	*32*	*54*	609 (112)	0.031 (0.009)
Control (unrelated letters)	*ddd*	*ddd*	645 (129)	0.040 (0.010)

Note: values in between parentheses indicate standard deviations.

Table 6. Analysis of response times and error rates (ref.: control)

	Response time			*Error rate*		
Condition	*b*	*SE*	*t*	*b*	*SE*	*z*
(intercept):	638.56	29.55	**16.54**	-3.30	1.43	**2.32**
Intact flanker order	-44.10	4.74	**-9.30**	-0.80	0.29	**-2.75**
Switched flanker order	-31.60	4.76	**-6.64**	-0.60	0.28	**-2.18**
Intact inner-letter order	-36.18	5.72	**-6.32**	-0.34	0.30	-1.13
Reversed inner-letter order	-31.86	5.72	**-5.57**	-0.35	0.30	-1.17

Note: Significant values are shown in bold.

We again found decreased RTs in all conditions with orthographically related flankers as compared to the control condition (Tables 5 and 6), with no significant contribution of flanker lexicality (b = -2.97, SE = 7.35, t = -0.40), pronounceability (b = 0.81, SE = 5.01, t = 0.16) and bigram frequency (b = 0.48, SE = 3.09, t = 0.15).[20]

Flanker order

Unlike Grainger et al. (2014) and Dare & Shillcock (2013) who found that bigram order (flanker order in the present study) does not matter, here RTs were found to be significantly decreased in

[20] As in Experiment 1, no significant effects were observed for pseudo-word trials.

the intact flanker order condition compared to the switched flanker condition: b = 10.04, SE = 4.66, t = 2.16. Flanker order did not interact with flanker lexicality (neither condition contained trials with word flankers), pronounceability (b = 15.01, SE = 10.22, t = 1.47) or bigram frequency (b = -0.36, SE = 11.46, t = -0.03).

It could have been the case that the bigram frequency of flankers consisting of letters 123 differed from flankers consisting of letters 456, and that asymmetrical processing of the flankers drove our flanker order effect. Investigating this scenario, we found that flankers 123 had a higher bigram frequency than flankers 456 (respectively 3.93 and 3.69 log ppm.). Interestingly however, we also found that higher bigram frequencies of *rightward* flankers significantly reduced RT (b = -11.87, SE = 5.90, t = -2.01), while the bigram frequency of leftward flankers had no influence (b = 2.66, SE = 5.90, t = 0.45). If the bigram frequency of rightward flankers played a crucial role, then we would expect shorter RTs in the switched flanker order condition. Given that we found the opposite, it is safe to assume that our flanker order effect was not cofounded with differences in flanker bigram frequency combined with asymmetrical processing (note also that there was no interaction of flanker order and left- or rightward flanker bigram frequency: b = 11.07, SE = 12.13, t = 0.91 and b = 4.93, SE = 12.01, t = 0.41, respectively).

Flanker order did not influence the error rate: b = 0.02, SE = 0.35, z = 0.06. Nevertheless, the switched flanker condition did lead to faster RTs and lower error rates than the unrelated flanker condition, in line with the findings of Dare and Shillcock (2013) and Grainger et al. (2014) for bigram flankers.

Inner-positioned letter order

Response times and error rates did not differ significantly between the intact vs. reversed letter order flankers, now that these flankers consisted of target's inner-positioned letters alone: for RTs, b = 4.25, SE = 4.68, t = 0.91 (no modulation of flanker lexicality (b = 13.33, SE = 10.91, t = 1.22), pronounceability (b = 13.33, SE = 10.91, t = 1.22) and bigram frequency (b = 11.62, SE = 6.92, t = 1.68)); for errors, b = 0.04, SE = 0.31, z = 0.14 (again no modulation of flanker lexicality (b = -0.58, SE = 0.70, z = -0.82), pronounceability (b = -0.58, SE = 0.70, z = -0.82) and bigram frequency (b = -0.11, SE = 0.43, z = -0.26)).

Discussion

The results from Experiment 2 are quite clear-cut. The key finding of Dare and Shillcock (2013) and Grainger et al. (2014), that flanker order does not modulate the effect of related flankers, could not be confirmed in a setting using 6-letter targets and 3-letter flankers, as the switched flanker condition yielded significantly longer RTs than the intact flanker condition. Thus, these results argue against the idea that the absolute position of letters does not matter at all, at least concerning the integration of orthographic information across spatially distinct stimuli. Instead, taking the present results and the work of Grainger et al. together, it seems that location-invariance only persists within certain spatial limits (see General Discussion). Nonetheless, the fact that the switched bigram flankers yielded lower RTs than the control flankers indicates that the integration of orthographic information is, at least to a considerable degree, location-invariant.

Experiment 2 further showed that when flankers do not involve the target's outer-positioned letters, the effect of letter order disappears. This points to the importance of outer-positioned letters with respect to letter-position coding, in line with the proposals of Fischer-Baum et al. (2011) and Jacobs et al. (1998). According to their *both-edges* account of letter-position coding (Fischer-Baum et al., 2011), the location of an inner-positioned letter would be represented by its distance to the first (ß) as well as the last letter (Ɛ) of the word. Thus, the '*r*' in '*target*' would have position representations ß+2 and Ɛ-3, whereas the '*e*' would have position representations ß+4 and Ɛ-1. Combining this approach with a relative-position coding account, it may be that words only activate bigrams that consist of an outer-positioned letter: '*target*', for instance, would activate '*ta*', '*tr*', '*tg*', '*rt*', '*gt*' and '*et*', but not '*ar*', '*rg*' or '*ge*'. Such a type of letter position coding could take on more importance in peripheral vision, where position information becomes more noisy (Chung & Legge, 2009) while outer-positioned letters remain highly visible (e.g. Chanceaux & Grainger, 2012).

This approach to relative-position coding, using the word's edges as anchoring points, raises the question of how readers would be able to distinguish words (e.g., '*target*') from their jumbled counterparts ('*tgerat*'), considering that these stimuli would activate the same set of bigrams. Here, the answer would be that an increasing distance between a bigram's constituent letters should lead to increased activation of the bigram. The rationale behind this assumption is that there should be more certainty about the relative position of two objects when those objects are further apart from each other. Lexical representations would be activated by specific bigram activity patterns. The representation for '*target*', for example, would correspond to a highly active bigram node representing '*te*', and a less active node representing '*ta*'. This mechanism would then also account for the flanking letter contiguity effect obtained in Experiment 1. Comparing the flankers 123 and 135, the distances between letters in the former flanker match that of the target's, thus leading to a similar pattern of bigram activities. The latter flanker, in contrast, causes a different pattern of bigram activities, leading to slower target word recognition.

The concept of lexical activation through bigram pattern activation bears quite some resemblance to spatial coding, as employed in Davis's SOLAR model (1999; 2010). In the SOLAR model, location-specific letter detectors respond to the orthographic input in a quick, sequential left-to-right fashion, meaning, for instance, that the visual input '*target*' will lead to activation of a node coding for '*t*', followed by activation of a node coding for '*a*', et cetera; (the level of activation over time for each node follows a Gaussian function, meaning that multiple letters can be active at once, accounting for letter position uncertainty). This dynamic activation pattern is coined the spatial code. Word nodes, in turn, are tuned to specific spatial codes, and the degree of similarity between the 'learned' spatial code (i.e., the lexical representation) and the incoming signal pattern determines whether the word is recognized. In a both-edges bigram coding scheme, word nodes would be tuned to bigram node activities rather than letter node activities. The theoretical advantage of this approach is that multiple letters – or indeed, multiple stimuli – would be processed in parallel, rather than sequentially, in line with the kind of parallel processing that is revealed by the flanker paradigm.

It should not be forgotten that the flankers without outer-positioned letters nonetheless led to lower RTs than the control flankers. In this light, it must be stressed that a both-edges bigram coding scheme would serve to account for letter position coding more so than letter identity coding. As argued in the Discussion of Experiment 1, letter identity and letter position may be encoded by two distinct cognitive mechanisms. As a consequence, letters would always to some degree activate lexical representations, irrespective of the absolute position of those letters.

General Discussion

Previous work aiming to "crack the orthographic code" has predominantly employed the masked-priming paradigm (Forster & Davis, 1984), and results obtained with this paradigm have led to the development of various cognitive accounts of the word recognition process. However, the bulk of masked-priming data, which mostly pertains to the temporal integration of orthographic information, does not favor any model in particular. In light of this, the present paper builds on recent evidence that information about letter identity and -position is integrated not only in the temporal dimension, but also in the spatial dimension (e.g. Dare & Shillcock, 2013; Grainger et al., 2014; Snell et al., 2017). Exploiting this principle, The FLLD paradigm pioneered by Dare and Shillcock (2013) provides an important novel approach to pinpointing the mechanisms underlying letter-position coding in reading.

The two experiments reported in this work tested the processing of 6-letter target words, when those targets were flanked by two or three letters on each side. In both experiments we found that readers processed target words considerably faster when these were flanked by related letters, as compared to unrelated letters. This effect continued to be quite strong (b = 31.52 ms) when the target's right- and left-sided letters were used as left- and right-sided flankers respectively (e.g. '*get target tar*'; Experiment 2). Harmonizing this finding with a noisy slot-based account is not so straightforward, considering that each of the target's letters is seven positions away from its repetition in the flankers in this condition. Allowing letters to influence one another at such a distance would impair a noisy slot-based model's performance greatly (e.g. Davis, 2010). On the other hand, as has been argued in this paper, it is likely that letter identities are to some extent activated irrespective of the position of these letters in the visual input (see also Grainger et al., 2014). Yet, while this assumption would allow a noisy slot-based model to account for general effects of flanker relatedness, it would not allow such a model to account for the letter order effect reported here and in Grainger et al. (2014).

While relative-position coding accounts for these findings quite effectively, we did find a difference between the aforementioned switched flanker condition ('*get target tar*') and the intact flanker condition ('*tar target get*'), contrasting with Dare and Shillcock (2013) and Grainger et al. (2014) who found that flanker order did not matter when using 4-letter targets and 2-letter flankers (e.g. '*ck rock ro*'). Taking these results together, there may be factors at play that cause orthographic information to be tied to specific spatial locations more strongly under certain conditions. For instance, the 3-letter flankers in the present experiments may have borne more processing weight than the 2-letter flankers in the studies of Dare and Shillcock (2013) and Grainger et al. (2014), causing increased lateral activation at the early visual processing stages, consequently allowing higher processing levels to make a stronger distinction between information stemming from the left- and right visual hemifield. Alternatively, it might be that the flanker order effect is driven by differences in processing of initial vs. final letters in the target word, combined with a difference in the impact of left- vs. rightward flankers. It must be acknowledged that such explanations are at this point speculative, but worthy of investigation in future research.

Experiment 1 replicated the flanker letter-order effect reported by Grainger et al. (2014), suggesting that both the processing of parafoveal information, as well as the parafoveal-foveal integration of this information, is sensitive to letter position information – even at the greater eccentricities implicated by the use of 6-letter targets and 3-letter flankers in the present work. Although numerically reduced in size, the letter-order effect persisted in conditions using non-contiguous letter flankers (e.g. 531 123456 642), underlining the importance of relative- rather than absolute position.

Results from Experiment 2 suggest that the word's edges (i.e., outer-positioned letters) play an important role with respect to letter-position coding, as there was no effect of flanker letter-order when using the target's inner-positioned letters alone (23 123456 45 vs. 32 123456 54). It is conceivable that outer-positioned letters act as spatial points of reference, in relation to which the location of inner-positioned letters is determined. Such a scheme is much in line with the both-edges account of letter-position coding proposed by Fischer-Baum et al. (2010; 2011) – with, however, the important difference that readers would code for the relative position, rather than the absolute distance, between outer- and inner-positioned letters. At the same time this both-edges bigram coding scheme bears resemblance to spatial coding (Davis, 1999; 2010), in that word nodes would be tuned to specific activation patterns. In Davis's SOLAR model, these patterns consist of letter node signals that are ordered in the temporal dimension. In both-edges bigram coding, the activation pattern consists of combined bigram activities (e.g. the word '*target*' corresponds to a highly active bigram node for '*tg*' and a less active node for '*ta*'). The theoretical difference between these two approaches is that letter processing would take place in a serial fashion in the SOLAR model, whereas it would take place in a parallel fashion in the scheme here proposed, in line with evidence for parallel processing during reading (e.g. Dare & Shillcock, 2013; Grainger et al., 2014; Snell et al., 2017). It should be noted that this type of letter position coding could take on more importance in peripheral vision, where position information becomes more noisy (Chung & Legge, 2009) while outer-positioned letters remain highly visible (e.g. Chanceaux & Grainger, 2012).

Finally, it must be stressed that a both-edges bigram coding scheme would serve to account for letter-position coding more so than letter-identity coding. Indeed, the switched flanker conditions and the conditions using flankers without outer-positioned letters yielded considerably shorter RTs than the control condition, suggesting that a portion of the word recognition process is driven by the coding of letters irrespective of their absolute position (e.g., similarly, Grainger et al. (2014) have argued for the presence of a "bag-of-letters" alongside a "bag-of-bigrams"; and see also Peresssotti and Grainger's proposal (1995) for position-independent letter detectors).

A question that remains is whether the present findings speak to the reading system in general, and thus, whether the cognitive mechanisms underlying letter-position coding and spatial integration would operate similarly in more natural reading settings such as sentence reading. In regard to this, it is noteworthy that Snell et al. (2017) employed both the FLLD paradigm as well as sentence reading to investigate parafoveal-on-foveal effects of orthographic neighbors, and found similar patterns of effects in the two settings. Further, as was mentioned in the Introduction, various other studies have shown that orthographic information is integrated across spatially distinct words in sentence reading (e.g. Inhoff et al., 2000; Angele et al., 2013; Dare & Shillcock, 2013). Hence, given that mechanisms underlying spatial integration seem to operate similarly in natural reading settings, we at this point see no reason why mechanisms underlying letter-position coding would operate differently. Additionally, one may argue that the FLLD paradigm may be unveiling task-specific processes rather than word recognition processes, and that the locus of the flanker effects established here would be at the decision-level, rather than at the level of orthographic processing. Our argument against this is that if our flanker effects concerned task-specific processes, then we would expect an effect of flanker lexicality, given that the task at hand is a lexical decision task. This was not the case however. Considering that we established flanker letter-order and –identity effects but no effects of flanker lexicality, pronounceability or bigram frequency, the most straightforward conclusion is that our flankers affected the cognitive stages of letter-position and -identity coding, i.e., orthographic processing.

In conclusion, the novelty of the FLLD paradigm is attested by the fact that many flanker configurations are still to be tested. The claims raised in this paper provide a tentative explanation for results obtained so far, but will need to be further consolidated – for instance, by testing flanker order in a setting using 4-letter targets and 3-letter flankers (e.g. '*roc rock ock'* vs. '*ock rock roc*') as well as a setting using 6-letter targets and 2-letter flankers (e.g. '*sp sprint nt*' vs. '*nt sprint sp*'). A further test of the both-edges relative-position scheme, as discussed above, may be provided by a setting comparing 'edge flankers' (e.g. '*rk rock rk*') to 'non-edge flankers' (e.g. '*oc rock oc*'). With respect to mechanisms underlying spatial integration rather than letter-position coding, it would further be worth investigating possible asymmetries between influences from leftward and rightward flankers in the FLLD paradigm. Future implementations of the FLLD paradigm should reveal how much merit the investigation of spatial information integration has in helping to crack the orthographic code.

Acknowledgements

This research was supported by grants ANR-11-LABX-0036 and ANR-15-CE33-0002-01 from the French National Research Agency (ANR).

References

Angele, B., Tran, R. & Rayner, K. (2013). Parafoveal–foveal overlap can facilitate ongoing word identification during reading: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 526–538.

Baayen, R. (2008). *Analyzing Linguistic Data: A pratical introduction to statistics*. Cambridge, UK: Cambridge University Press.

Baayen, R., Davidson, D. & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390-412.

Barr, D., Levy, R., Scheepers, C. & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015), Fitting Linear Mixed-Effects Models using lme4*. Journal of Statistical Software*, 67, 1–48.

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Automatic semantic activation of embedded words: Is there a 'hat' in 'that'? *Journal of Memory and Language, 52*, 131–143.

Chanceaux, M. & Grainger, J. (2012). Serial position effects in the identification of letters, digits, symbols, and shapes in peripheral vision, *Acta Psychologica, 141,* 149–158.

Dare, N. & Shillcock, R. (2013). Serial and parallel processing in reading: Investigating the effects of parafoveal orthographic information on nonisolated word recognition. *The Quarterly Journal of Experimental Psychology, 66*, 417–428.

Davis, C. (1999). The self-organizing lexical acquisition and recognition (SOLAR) model of visual word recognition. Unpublished doctoral dissertation. University of New South Wales, Australia.

Davis, C. (2010). SOLAR versus SERIOL revisited. *European Journal of Cognitive Psychology, 22,* 695–724.

Davis, C., & Bowers, J. (2004). What do letter migration errors reveal about letter position coding in visual word recognition? *Journal of Experimental Psychology: Human Perception and Performance, 30*, 923–941.

Davis, C., & Lupker, S. (2006). Masked inhibitory priming in English: Evidence for lexical inhibition. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 668–687.

De Moor, W. & Brysbaert, M. (2000). Neighborhood-frequency effects when primes and targets are of different lengths. *Psychological Research,* 63, 159–162.

De Moor, W., Van der Herten, L., & Verguts, T. (2007). Is masked neighbor priming Inhibitory? Evidence using the incremental priming technique. *Experimental Psychology, 54,* 113–119.

Chung, S. & Legge, G. (2009). Precision of position signals for letters. *Vision Research, 49*, 1948-1960.

Ferrand, L., et al. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods,* 42, 488–496.

Fischer-Baum, S., Charny, J. & McCloskey, M. (2011). Both-edges representation of letter position in reading. *Psychonomic Bulletin & Review, 18,* 1083–1089.

Fischer-Baum, S., McCloskey, M. & Rapp, B. (2010). The representation of grapheme position: Evidence from acquired dysgraphia. *Cognition, 115,* 466–490.

Forster, K., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*, 680–698.

Forster, K., Davis, C., Schoknecht, C., & Carter, R. (1987). Masked priming with graphemically related forms: Repetition or partial activation*? Quarterly Journal of Experimental Psychology*, 39, 211–251.

Gomez, P., Ratcliff, R. & Perea, M. (2008). The Overlap Model: A model of letter position coding. *Psychological Review, 115,* 577–600.

Grainger, J. (2008). Cracking the orthographic code: An introduction. *Language and Cognitive Processes, 23*, 1–35.

Grainger, J., Granier, J., Farioli, F., Van Assche, E. & van Heuven, W. (2006). Letter position information and printed word perception: The relative-position priming constraint. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 865–884.

Grainger, J., Mathôt, S. & Vitu, F. (2014). Test of a model of multi-word reading: Effects of parafoveal flanking letters on foveal word recognition. *Acta Psychologica,* 146, 35–40.

Grainger, J., & van Heuven, W. (2003). Modeling letter position coding in printed word perception. In P. Bonin (Ed.), *The mental lexicon* (pp. 1-23). New York: Nova Science.

Grainger, J., & Whitney, C. (2004). Does the huamn mind raed wrods as a wlohe? *Trends in Cognitive Sciences, 8,* 58–59.

Hannagan, T. & Grainger, J. (2012). Protein analysis meets visual word recognition: A case for string kernels in the brain. *Cognitive Science, 36,* 575–606

Inhoff A., Radach R., Starr M. & Greenberg S. (2000). Allocation of visuospatial attention and saccade programming during reading. In Kennedy A., Radach R., Heller D. & Pynte J. (Eds.). *Reading as a perceptual process*. Oxford, UK: Elsevier.

Jacobs, A., Rey, A., Ziegler, J. & Grainger, J. (1998). MROM-P: An interactive activation, multiple read-out model of orthographic and phonological processes in visual word recognition. In J. Grainger & A. Jacobs (Eds.), *Localist connectionist approaches to human cognition*. Mahwah, NJ.: Erlbaum (pp. 147-188).

Mathôt, S., Schreij, D. & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*, 314–324.

McClelland, J., & Mozer, M. (1986). Perceptual interactions in two-word displays: Familiarity and similarity effects. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 18–35.

McClelland, J. & Rumelhart, D. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*, 88, 375–407.

Norris, D., & Kinoshita, S. (2012). Orthographic processing is universal; it's what you do with it that's different. *Behavioral and Brain Sciences, 35,* 296–297.

Perea, M., & Lupker, S. (2004). Can CANISO activate CASINO? Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language, 51,* 231–246.

Peressotti, F. & Grainger, J. (1995). Letter position coding in random consonant arrays. *Perception & Psychophysics*, 57, 875–890.

Peressotti, F., & Grainger, J. (1999). The role of letter identity and letter position in orthographic priming. *Perception and Psychophysics, 61,* 691–706.

Schoonbaert, S., & Grainger, J. (2004). Letter position coding in printed word perception: Effects of repeated and transposed letters. Language and Cognitive Processes, 19, 333–367.

Segui, J., & Grainger, J. (1990). Priming word recognition with orthographic neighbors: Effects of relative prime-target frequency. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 65–76.

Snell, J., Vitu, F. & Grainger, J. (2017). Integration of parafoveal orthographic information during foveal word reading: Beyond the sub-lexical level? *Quarterly Journal of Experimental Psychology*, *70,* 1984–1994.

Van Assche, E., & Grainger, J. (2006). A study of relative-position priming with superset primes. *Journal of Experimental Psychology: Learning, Memory and Cognition, 32,* 399–415.

Van Heuven, W., Mandera, P., Keuleers, E. & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology, 67*, 1176–1190.

Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin and Review, 8,* 221–243.

An electrophysiological investigation of orthographic spatial integration in reading

Joshua Snell, Gabriela Meade, Martijn Meeter, Phillip Holcomb & Jonathan Grainger

Citation: Snell, J., Meade, G., Meeter, M., Holcomb, P. & Grainger, J. (2018). *An electrophysiological investigation of orthographic spatial integration in reading.* Manuscript submitted for publication.

Abstract

The recognition of words is influenced by surrounding information. Particularly intriguing is the finding that words are recognized faster when they share many letters with surrounding words. The nature of this phenomenon is not understood: some theories attribute it to low-level visual operations (parafoveal feature detectors influencing foveal letter detectors), while other theories assume that location-invariant sub-lexical nodes (e.g., bigrams) are activated by information from parallel processed words. To arbitrate between the two, the present study reveals the time course of orthographic spatial integration with electroencephalography in a lexical decision task. Foveal target words were flanked on each side by parafoveal words, manipulated across three conditions: repetition flankers (e.g. *rock rock rock*), unrelated flankers (e.g. *step rock step*) and a no-flanker condition. Linear mixed-effect models were constructed to analyze electro-encephalographic data on a trial-to-trial basis and for each 10 ms timestep. Word recognition was worse in the unrelated flanker condition than in the repetition and no-flanker conditions, and this behavioral pattern was accompanied by increased negativity in the N250 and N400 windows, associated with the activation of sub-lexical and lexico-semantic representations, respectively. Crucially, the absence of earlier effects ruled out accounts involving low-level visual processes. We conclude that orthographic spatial integration is driven by parallel processing of multiple words, which leads to the activation of a larger set of sub-lexical nodes in the presence of unrelated flankers, and consequently more difficult processing at the lexical level.

1. Introduction

One of the main endeavors of reading research has been to determine how processing of letter strings leads to word recognition. This has led to sophisticated accounts of single word recognition (e.g., Grainger, 2008, for a review). However, having focused predominantly on the recognition of isolated words, relatively little attention has been paid to the fact that the recognition of words is influenced by surrounding information. The present study addresses this characteristic of the reading process. Specifically, electroencephalography (EEG) is employed with the aim to understand how parafoveal orthographic information impacts on foveal word processing.

Recent lines of research have pointed out that information about the identity and relative position of letters is to some extent integrated across words. For instance, lexical decisions about target words (e.g., '*rock*') are made faster and more accurately when those targets are flanked by orthographically related stimuli (e.g., '*ro rock ck*') compared to unrelated stimuli ('*st rock ep*') (Dare & Shillcock, 2013; Grainger, Mathôt & Vitu, 2014; Snell, Vitu & Grainger, 2017; Snell, Bertrand, Meeter & Grainger, 2018a; Snell, Bertrand & Grainger, 2018b). Such so-called orthographic parafoveal-on-foveal effects have not only been established in target word recognition tasks, but also in natural (sentence) reading, with readers spending less time viewing words that are orthographically related to upcoming words (Inhoff, Starr & Greenberg, 2000; Dare & Shillcock, 2013; Angele, Tran & Rayner, 2013; Snell et al., 2017).

There is as of yet no consensus about the cognitive mechanisms underlying the spatial integration of orthographic information. Some researchers have argued that the locus of orthographic parafoveal-on-foveal effects is at an early stage of visual processing. For example, Angele et al. (2013) have proposed that foveal letter detectors may be connected to parafoveal feature detectors, as such driving parafoveal-foveal integrative effects.

Others have argued that orthographic information may be integrated at later stages of (sub-)lexical processing. As outlined by Grainger et al. (2014) and Snell et al. (2017; 2018a; 2018b), the Open-Bigram model of word recognition (Grainger & van Heuven, 2003) would account for orthographic parafoveal-on-foveal effects by assuming parallel processing of letters from multiple words. This would lead to stronger activation of bigrams (sub-lexical nodes that encode the relative position of letter pairs; e.g., '*ro*' and '*ok*' in '*rock*') if adjacent words are orthographically related, which in turn leads to stronger lexical activation and faster word recognition (see Snell, van Leipsig, Grainger & Meeter, 2018c, for the implementation of this process in a computational model).

Going further still, one could argue that, in the case of lexical decision tasks, parafoveal information might impact on the post-lexical stage of decision-making, such that decisions are influenced by the experienced convergence of independently processed target and flanking stimuli. A set of unrelated stimuli may be evaluated as being 'more complex', thereby biasing the reader to make a nonword response. However, contending this conception is the fact that orthographic parafoveal-on-foveal effects are typically not observed in pseudoword target trials (e.g., Snell et al., 2018b). Additionally, this account would not cover parafoveal-on-foveal effects observed in the realm of sentence reading.[21]

As the scenarios outlined above point to different temporal loci for orthographic parafoveal-on-foveal effects, EEG may well provide the perfect tool for discerning the nature of

[21] One could argue, however, that the decision of whether and when to move the eyes during sentence reading may be influenced by experienced stimulus complexity in a similar manner.

these effects. Indeed, the temporal resolution with which event-related potentials (ERPs) are recorded from the scalp has already provided ample insight into the time course of the recognition of isolated words (Holcomb, O'Rourke & Grainger, 2002; Grainger, Kiyonaga & Holcomb, 2006; Holcomb & Grainger, 2006; Petit, Grainger, Midgley & Holcomb, 2006; Dufau, Grainger & Holcomb, 2008; Grainger & Holcomb, 2009; Massol, Grainger, Dufau & Holcomb, 2010; Dufau, Grainger, Midgley & Holcomb, 2015; Meade, Grainger, Midgley, Emmorey & Holcomb, 2018). Specifically, this line of work has led to the establishment of a cascade of ERP components that can be linked to the onset and interplay of various cognitive stages involved in reading; from the earliest stages of visual processing up to the levels of lexical and semantic access. Below, we briefly summarize these ERP components. Subsequently, we will declare our hypotheses concerning the time course of the spatial integration of orthographic information, as reflected in ERPs.

1.1 ERP components of visual word recognition

As reviewed by Grainger and Holcomb (2009), ERP investigations of single word reading have revealed at least three key components: N/P150, N250 and N400.[22] The first of these, the N/P150, is a domain-unspecific bipolar component, consisting of a more positive deflection over anterior sites, and a concurrent negative deflection over posterior sites, that is associated with processing at the level of visual features. More specifically, this component is thought to reflect mapping of visual features onto abstract representations (e.g., letters), a process that takes place in the time frame of approximately 90 to 200 ms after stimulus onset (Grainger & Holcomb, 2009). Evidence for the relevance of this component to the domain of reading was obtained with the masked priming paradigm (e.g., Forster & Davis, 1984), which tests lexical decisions about target words as a function of prime words that are briefly presented (and likely go unnoticed) prior to the onset of the target. Using ERPs, increased visual similarity between the target and prime stimulus led to smaller amplitude N/P150s, indicative of reduced processing difficulty (Petit et al. 2006).

Arguably, by the rationale of Angele et al. (2013)'s account of orthographic parafoveal-on-foveal effects—i.e., that letter detectors are influenced by parafoveal feature detectors—one might predict that the N/P150 component elicited by a target word is affected by its orthographic relatedness to flanking stimuli. However, Dufau et al. (2008) reported that N/P150 effects of prime-target featural overlap disappear when the prime and target are not presented at the exact same location—a result that does not hold true for all later components. It is therefore uncertain that the N/P150 elicited by a target word can be influenced by simultaneously presented flankers. What is certain, nonetheless, is that the account of Angele et al. (2013) would predict some effect to take place around this timepoint, for instance in the time window of 130—190 ms associated with single letter processing (Bann & Herdman, 2016).

The N/P150 is followed by a negative-going peak around 250 ms. This N250 component has a widespread scalp distribution, but is usually strongest over anterior sites. Indicative of its involvement in visual word recognition, the component is more negative if a target word is preceded by an orthographically unrelated prime, compared to a repetition prime (Grainger & Holcomb, 2009). Importantly, unlike the N/P150, the N250 is not affected by manipulations of similarity in prime-target pairs consisting of pictures or single letters (Eddy, Schmid & Holcomb, 2006; Petit et al., 2006). On the other hand, N250 effects do occur when manipulating the

[22] Note that Grainger and Holcomb (2009) have described a fourth component, the P325; however, this component has been rather elusive in subsequent research and is thus not discussed here.

orthographic relatedness of pseudoword prime-target pairs, suggesting that this component reflects mapping of activated sublexical representations (e.g., letters, bigrams) onto whole-word form representations (Grainger & Holcomb, 2009).

Finally, the negative-going posterior-oriented N400 is influenced by prime-target relatedness when these are words but not when these are pseudowords (Kiyonaga, Midgley, Holcomb & Grainger, 2007). When preceded by a semantically related prime, target words elicit a weaker N400 deflection (Bentin, McCarthy & Wood, 1985; Holcomb, 1988). As argued by Grainger and Holcomb (2009), these findings suggest that the N400 represents the ease of mapping activated word identities onto higher-level semantic representations.

1.2 The present study

In the present study, the approach to discerning the nature of orthographic parafoveal-on-foveal effects consists of drawing an analogy from temporal information integration, as seen in the masked priming paradigm, to spatial information integration, as observed in the flanker paradigm (e.g., Snell et al., 2018a). We report an experiment that tested, by means of a lexical decision task, the processing of briefly presented target words that were flanked either by the same words ('*rock rock rock*'; the related flanker condition), unrelated control words of the same frequency ('*step rock step*'; the unrelated flanker condition) or by nothing ('rock'; the no-flanker condition).

The methodological focus of this paper is on time course analyses of the EEG signal elicited by target stimuli across the three flanker conditions. Voltage differences across conditions were analyzed at every 10 ms timestep, which allowed us to determine precisely at what point in time processing of the target is impacted by its orthographic relatedness to surrounding stimuli, and, based on this, to infer the mechanisms involved. This work further exhibits the relatively novel use of linear mixed-effect models (LMEs) in analyzing EEG data (e.g. Emmorey, Midgley, Kohen, Sehyr & Holcomb, 2017; Payne, Lee & Federmeier, 2015; Winsler, Midgley, Grainger & Holcomb, 2018). LMEs, which are rapidly becoming the golden standard in the field of language research, enabled us to analyze ERP variance on a single-trial basis, taking into account not only inter-individual differences, but also item-to-item variance. Moreover, LMEs alleviate the need to exclude participants with a high proportion of contaminated trials (as is necessary when using classic ANOVAs), hence preserving more data and more statistical power.

We expected to obtain evidence in favor of one out of two scenarios: either the electrophysiological signal in the related versus unrelated flanker conditions would start to differ at an early timepoint (<200 ms), in line with Angele et al. (2013)'s account stating that letter detectors are influenced by parafoveal feature detectors; or alternatively, the signal would start to differ in the N250 window, in line with our own account (e.g., Grainger et al., 2014; Snell et al., 2017; 2018a; 2018b) stating that sub-lexical (bigram) nodes are activated by the target as well as parallel-processed flankers. Analogous to masked priming evidence, we expected that the larger set of activated sub-lexical nodes would be reflected in a greater N250 amplitude in the unrelated flanker condition compared to the related flanker condition. A subsequent greater N400 amplitude in the unrelated flanker condition would reflect increased lexical processing difficulty (e.g., Meade et al., 2018, for arguments linking increased N400 negativity to increased processing difficulty).

Comparisons against the no-flanker condition were carried out as an exploratory investigation of the effects of flanker presence. Compared to the no-flanker condition, we expected the unrelated flanker condition to again elicit more negative deflections in the N250 window as the result of the activation of a greater set of sublexical nodes. Additionally, increased N250 and N400 amplitudes were expected in the unrelated flanker condition given the increase in processing difficulty when resources are shared between the flanker and the target compared to when the target is presented alone.

2. Methods

2.1 Participants

We recruited 24 participants (F = 13) in the age range of 18—28 years old from the Aix-Marseille University. All participants reported to be native French speakers, non-dyslexic, and to have normal or corrected-to-normal vision. Participants received monetary compensation, and gave informed consent to their participation in accordance with the Institutional Review Board of Aix-Marseille University. In the analyses, three participants were excluded due to lost EEG data as the result of experimenter error.

2.2 Stimuli and design

We retrieved 80 4-letter word targets 80 4-letter pseudoword targets from the French Lexicon Project database (Ferrand et al., 2010). Stimuli contained no diacritics. For each target, we retrieved an orthographically unrelated stimulus with the same length, frequency, and lexical status from the database. The unrelated stimuli were to act as unrelated flankers, whereas repetitions of the targets were used as related flankers.

The experiment followed a 2×3 design with target lexicality (*word, pseudoword*) and flanker (*repetition, unrelated, no-flanker*) as factors. Note that pseudoword trials were merely used to induce the task and were thus not analyzed. All participants saw all targets in all flanker conditions, meaning there were 480 trials per participant in total. These were presented in random order.

2.3 Procedure

Participants were seated in a comfortable office chair in a dimly lit room. Prior to the experiment, participants received task instructions both by the experimenter as well as visually on the screen.

The trial procedure is shown in Figure 1. Stimuli were shown in black on a luminance-neutral gray background. Each trial started with a 500 ms fixation display consisting of two centrally positioned vertical fixation bars. This display was followed by a 150 ms stimulus display with the target being positioned in between the two fixation bars. Flankers were separated from the target by a single character space. The stimulus display was followed by the fixation display allowing participants a maximum of 2000 ms to respond with a right- or left-handed trigger button to indicate '*word*' or '*pseudoword*', respectively. Upon the participant's response (or the 2000 ms time-out), a centrally positioned green or red fixation dot was presented for 500 ms to indicate that the response was correct or incorrect, respectively.

The experiment was implemented with OpenSesame (Mathôt, Schreij & Theeuwes, 2012). Stimuli were presented on a 1024×78 pixel 0.48 kHz CRT monitor, the display of which was positioned at a 100 cm distance from the participant's eyes, so that each letter in the display subtended 0.30 degrees of visual angle. Responses were collected with left- and right-handed trigger buttons on a gamepad at a polling rate of approximately 0.10 kHz.

Three breaks were offered during the experiment. The 480 experimental trials were preceded by 12 practice trials. The experiment lasted approximately 25 minutes.

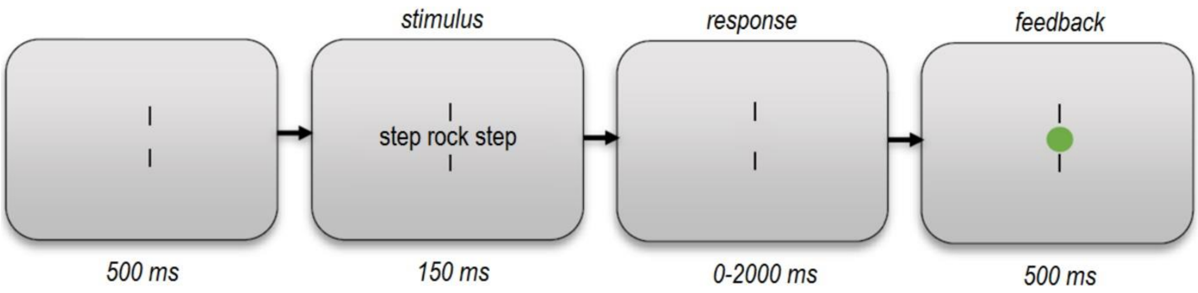

Figure 1. Trial procedure. The size of stimuli relative to the screen is exaggerated in this example.

2.4 EEG recording and analysis

Participants were fitted with an elastic cap with 64 electrodes. A reference signal was recorded from an electrode on the left mastoid, while recordings from electrodes below the left eye and on the outer canthi of both eyes were used to detect blinks and horizontal eye movements, respectively. EEG was sampled with a Biosemi ActiveTwo system at a rate of 2048 Hz, but subsequently downsampled to 512 Hz. Each epoch began 100 ms before stimulus onset (used for baseline correction) and ended 600 ms after stimulus onset and was filtered between 0.1 and 15 Hz. Trials contaminated by artifacts within this window (13% on average) were excluded from all analyses, as were trials with incorrect responses. Channels with consistent artifacts across trials were replaced using spherical spline interpolation.

Two types of analyses were performed: first, to differentiate between the theories outlined in the Introduction, we performed time course analyses that consisted of contrasting conditions with one another in every 10ms time step of the interval between the baseline and 600 ms post stimulus onset. This allowed for a fine-grained inspection of the onset of processes involved in the spatial integration of orthographic information. Within each trial and electrode, potential differences were averaged across every 10ms time step. We thus acquired a three-dimensional data matrix with potential differences per trial, electrode and timepoint. Next, an LME including items and participants as random effects, and flanker condition as factor, was run for each timepoint and for each electrode, meaning the model performed 70 * 64 contrasts (10 * 64 of which were for the baseline interval and 60 * 64 of which were for the post-stimulus interval), each on a number of data points equal to the total number of trials. The model included by-item and by-participant random intercepts but no random slopes, as the model did not converge under this maximal random structure for a portion of the analyses.

Second, we analyzed distinct time-windows corresponding to the N150, N250 and N400 components. Here, data was averaged within the intervals of interest (90—180 ms, 175—250 ms and 250—450 ms, for the N150, N250 and N400 respectively) within each trial and for each individual electrode. These intervals were based on visual inspection of the grand average waveforms (see Figure 4) and are consistent with previous masked priming studies. Each trial thus yielded 64 data points. Each component was analyzed with a single model encompassing data from all electrodes. These models tested for distributional effects by assigning a spatial coordinate to each electrode according to a 3×3 grid (Figure 2), such that the deflection measured by each electrode was tested as a function of three independent variables and their interactions: flanker condition, laterality (*left, center, right*) and anteriority (*anterior, center, posterior*). We refer to these models as the Distributional Effects Models (DEMs). Each DEM was run twice: once after isolating the unrelated and related flanker condition (to test the effect of flanker relatedness) and once after isolating the unrelated and no-flanker condition (to test the effect of flanker presence).

To visualize effects observed in each window, we obtained separate LME solutions for each individual electrode (hence without the spatial factors) and plotted the resulting *t*-values across the scalp using interpolated topographic maps.

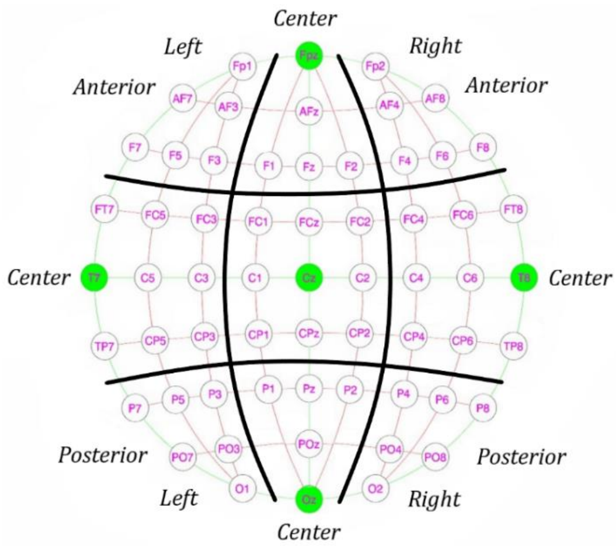

Figure 2. 64 Scalp electrode sites divided into a 3 × 3 grid.

3. Results

3.1 Response times and errors

Trials with an RT beyond 2.5 SD from the grand mean (2.22% of trials) were excluded from the analyses of RTs and errors. Trials with an incorrect response (7.83%) were excluded from the RT analyses.

The behavioral data were analyzed with LME models that included by-participant and by-item random intercepts as well as random slopes. We report *b*-values, standard errors (SEs) and *t*-values (RTs) or *z*-values (errors), with $|t|$ and $|z|$ beyond 1.96 deemed significant (Baayen, 2008).

The average RT per condition is plotted in Figure 3. RTs to words were significantly longer in the presence of unrelated flankers compared to repetition flankers (b = 39.21, SE = 4.45, t = 8.82). No difference was observed between the repetition flanker condition and no-flanker condition (b = 3.64, SE = 3.61, t = 1.01). A similar pattern was observed in the error rate, with more errors in the unrelated flanker condition than in the repetition flanker condition (b = 0.64, SE = 0.27, z = 2.33) while the repetition flanker condition and no-flanker condition were again similar (b = 0.14, SE = 0.30, z = 0.45).

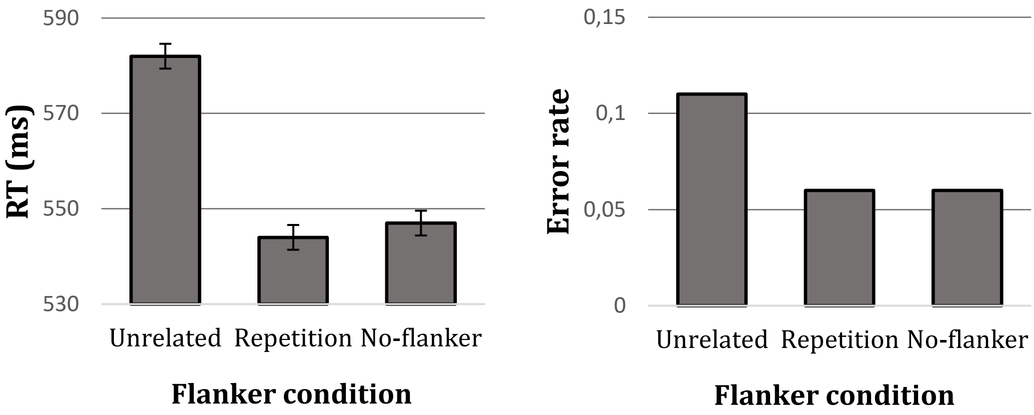

Figure 3. Average RTs and error rate per flanker condition for word trials. Error bars indicate standard errors.

3.2 Electrophysiological data

ERPs for each condition at electrode sites Fz, Cz and POz are plotted in Figure 4.

3.2.1 Time course analyses

Prior to all analyses, trials in which the potential difference of the electrode at the timepoint of interest was beyond 2.5 SD from the grand mean (2.60% of trials on average) were discarded. Analyses are plotted in Figures 5 and 6. For readability's sake we did not plot analyses of all 64 electrodes, but rather of one electrode per cell as marked in Figure 2.

As can be seen in Figure 5, significant differences between the unrelated and repetition flanker conditions began around 200 ms and continued beyond 500 ms.[23] Deflections were more negative in the unrelated flanker condition. Importantly, and in line with previous descriptions of the N250 and N400 (Section 1.1), effects appeared to be larger at anterior sites around 250 ms and larger at posterior sites around 450 ms.

When contrasting the unrelated flanker condition against the no-flanker condition, an even clearer difference can be observed around the N250 window (Figure 6). Words flanked by unrelated words elicited larger negativities than those flanked by the same words. The effect was significant throughout the N400 window, where it was strongest across posterior sites.

[23] We are well aware of the multiple-comparisons problem, which entails that the chance of encountering false positives increases as the number of contrasts being carried out increases. Crucially, therefore, it must be observed in Figure 5 that the intervals of significance are sustained for considerable amounts of time (>100 ms) and, moreover, are preceded and followed by trends towards significance. These aspects of the data indicate that the observed effects are not type I errors (e.g., Guthrie & Buchwald, 1991). The same reasoning applies for the analyses of flanker presence (see Figure 6).

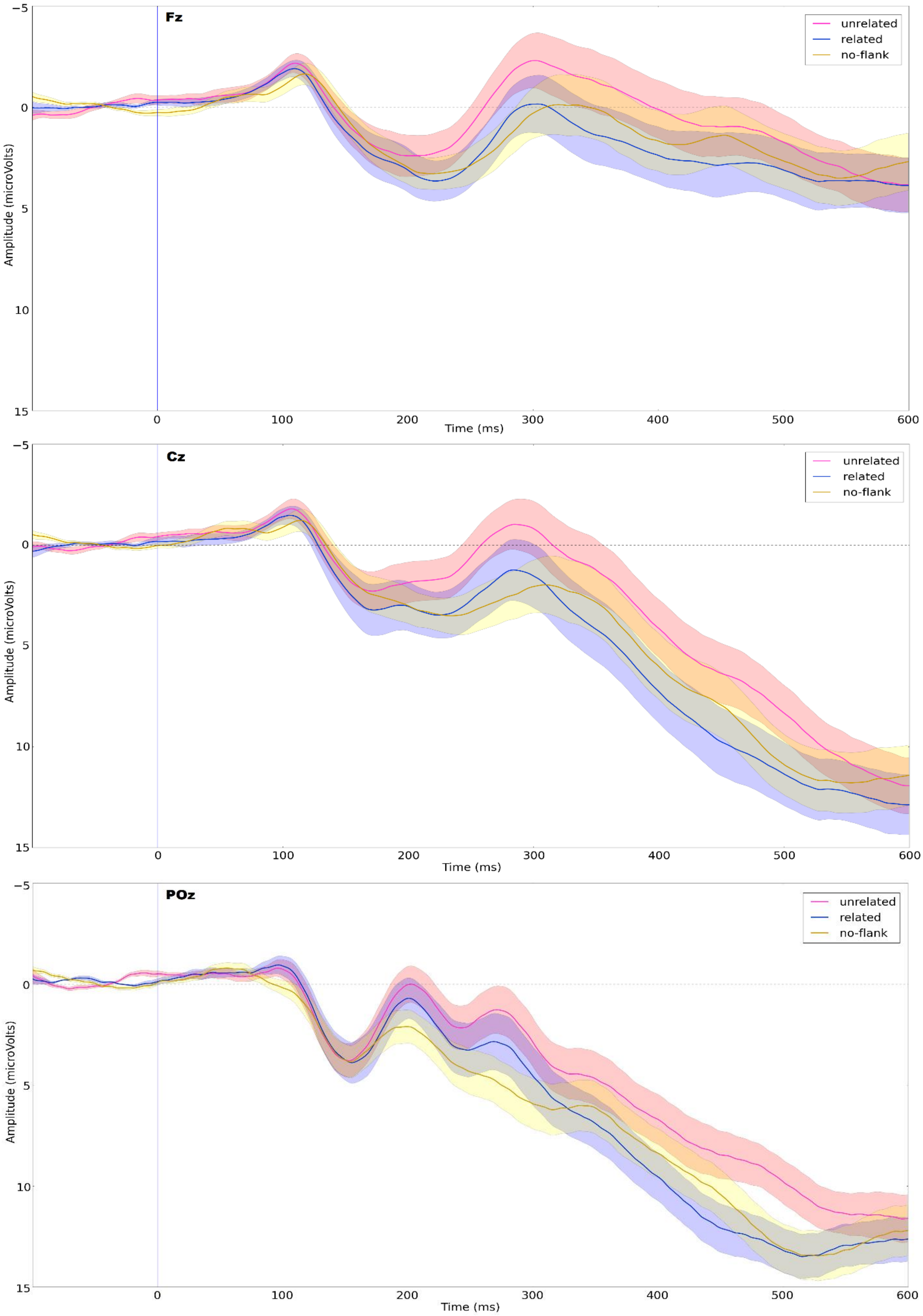

Figure 4. ERPs per condition, as registered by the Fz electrode (top panel), Cz electrode (middle panel), and POz electrode (bottom panel). Shaded areas around the curves represent standard errors.

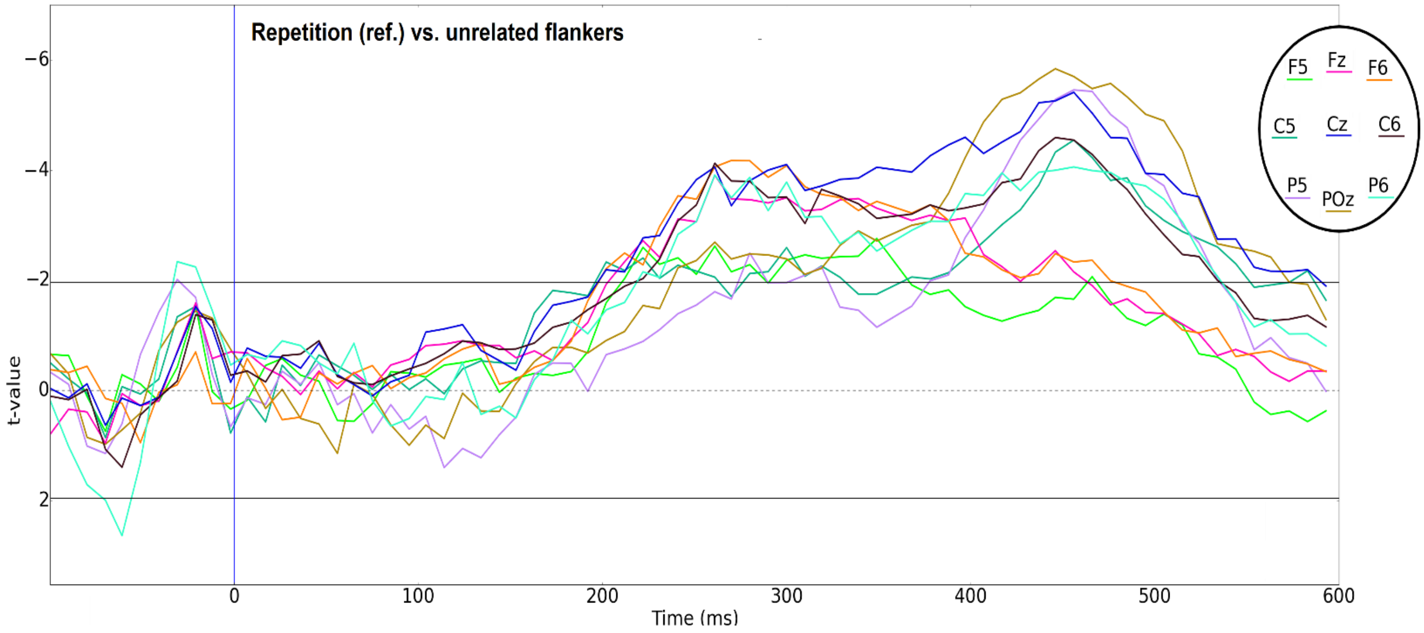

Figure 5. Contrast of the Repetition (ref.) vs. unrelated flanker condition, across the baseline to 600 ms post target onset interval. The significance thresholds are indicated by the black horizontal lines at $t = |1.96|$.

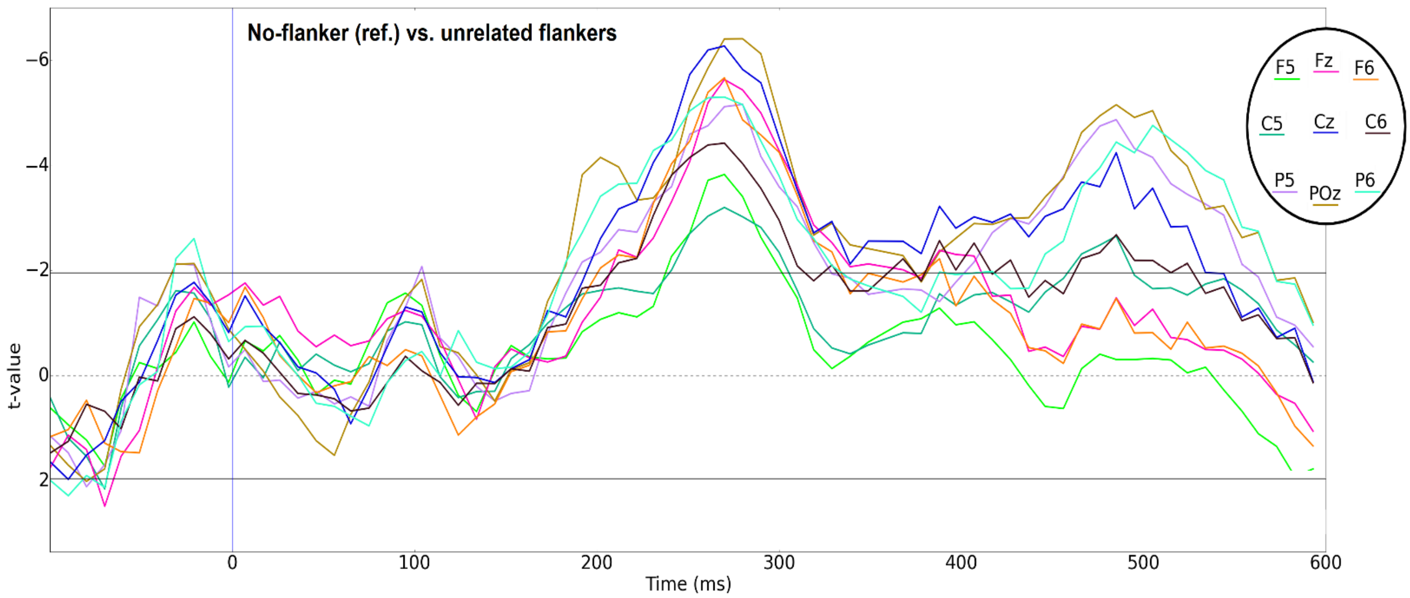

Figure 6. Contrast of the No-flanker (ref.) vs. unrelated flanker condition, across the baseline to 600 ms post target onset interval. The significance thresholds are indicated by the black horizontal lines at $t = |1.96|$.

3.2.2 N150, N250 and N400

No significant effects of flanker relatedness (Figure 7A; Table 1) or flanker presence (Figure 7B; Table 2) occurred in the N150 window.

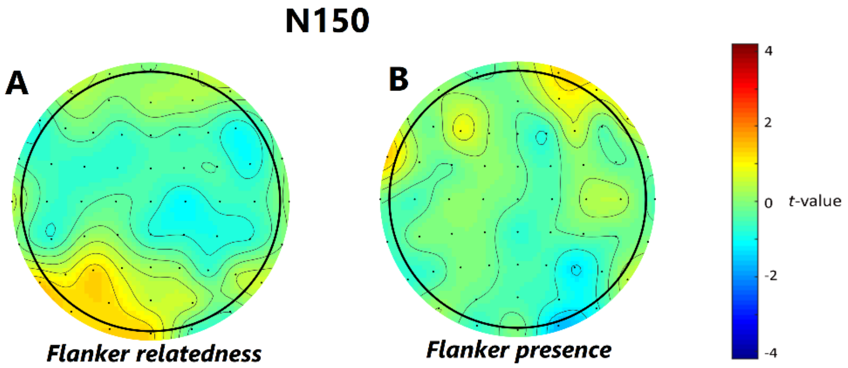

Figure 7. *t*-values across the scalp, illustrating N150 effects of flanker relatedness (ref.: repetition flankers, panel A) and flanker presence (ref.: no-flankers, panel B) respectively.

Table 1. Contrast of unrelated vs. repetition flankers in the N150 window (ref.: repetition flankers). Significant values are shown in bold.

	b	*SE*	*t*
Intercept	0.06	0.44	0.14
Flanker relatedness	-0.16	0.13	-1.23
Relatedness × Anteriority	0.18	0.10	1.91
Relatedness × Laterality	0.01	0.10	0.12
Three-way interaction	-0.07	0.07	-0.96

Table 2. Contrast of unrelated vs. no-flankers in the N150 window (ref.: no-flankers). Significant values are shown in bold.

	b	*SE*	*t*
Intercept	0.10	0.41	0.26
Flanker presence	0.06	0.13	0.50
Presence × Anteriority	-0.10	0.10	-1.01
Presence × Laterality	0.03	0.10	0.31
Three-way interaction	-0.04	0.07	-0.50

As can be seen in Figure 8A, the negativity within the N250 window was significantly larger in the presence of unrelated flankers compared to repetition flankers. In line with the research outlined in Section 1.1, this effect was strongest at anterior sites. Indeed, when isolating the two flanker conditions, the DEM for this time window showed a significant interaction between flanker relatedness and anteriority (Table 3). Neither an interaction of flanker relatedness and laterality, nor a three-way interaction of flanker relatedness, laterality and anteriority was observed.

Compared to the no-flanker condition, N250 amplitudes were significantly more negative across a large portion of the scalp (Figure 8B). Whereas the effect of flanker relatedness was more strongly expressed over anterior sites, the effect of flanker presence was stronger over posterior sites. No interaction of flanker presence and laterality nor a three-way interaction of these factors was observed (Table 4).

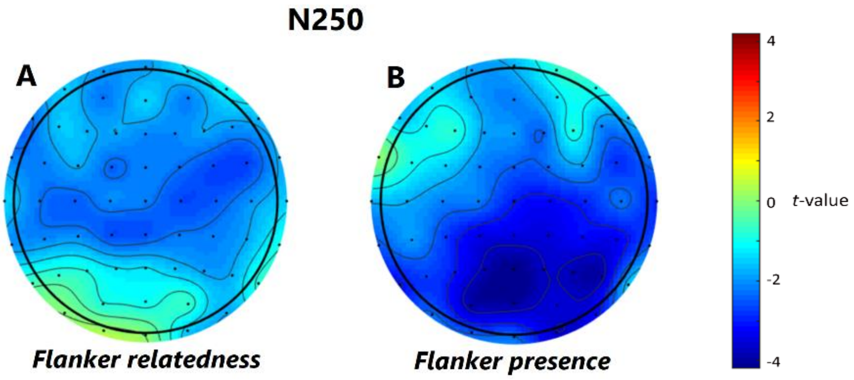

Figure 8. *t*-values across the scalp, illustrating N250 effects of flanker relatedness (ref.: repetition flankers, panel A) and flanker presence (ref.: no-flankers, panel B) respectively.

Table 3. Contrast of unrelated vs. repetition flankers in the N250 window (ref.: repetition flankers). Significant values are shown in bold.

	b	*SE*	*t*
Intercept	2.85	0.70	**4.07**
Flanker relatedness	-0.99	0.15	**-6.55**
Relatedness × Anteriority	0.33	0.11	**2.89**
Relatedness × Laterality	0.06	0.12	0.51
Three-way interaction	-0.10	0.09	-1.09

Table 4. Contrast of unrelated vs. no-flankers in the N250 window (ref.: no-flankers). Significant values are shown in bold.

	b	*SE*	*t*
Intercept	2.81	0.62	**4.53**
Flanker presence	-0.57	0.15	**-3.78**
Presence × Anteriority	-0.26	0.11	**-2.27**
Presence × Laterality	-0.11	0.12	-0.98
Three-way interaction	-0.03	0.09	-0.36

Within the N400 window, a strong main effect of flanker relatedness was observed (Figure 9A; Table 5). No interactions among flanker relatedness, anteriority, or laterality were significant within this window.

Unrelated flanker trials also elicited significantly more negative amplitudes than no-flanker trials within the N400 window (Figure 9B). This effect had a posterior and rightward distribution, as reflected in the interactions between flanker presence and anteriority, and flanker presence and laterality (Table 6).

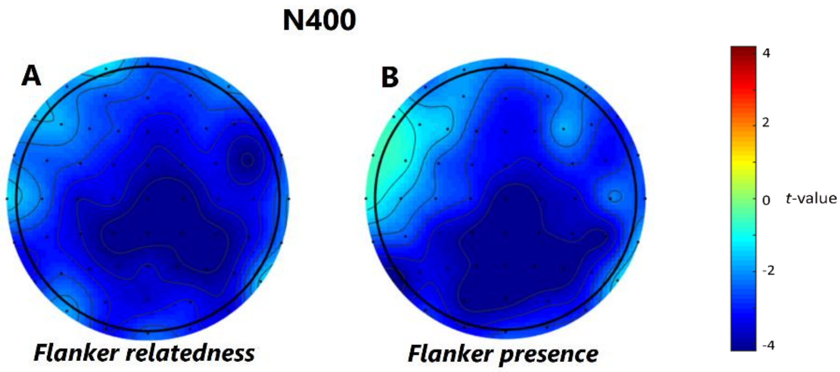

Figure 9. t-values across the scalp, illustrating N400 effects of flanker relatedness (ref.: repetition flankers, panel A) and flanker presence (ref.: no-flankers, panel B) respectively.

Table 5. Contrast of unrelated vs. repetition flankers in the N400 window (ref.: unrelated flankers). Significant values are shown in bold.

	b	*SE*	*t*
Intercept	0.13	0.88	0.15
Flanker relatedness	-1.27	0.16	**-7.89**
Relatedness × Anteriority	0.01	0.12	0.10
Relatedness × Laterality	-0.10	0.12	-0.80
Three-way interaction	0.02	0.09	0.19

Table 6. Contrast of unrelated vs. no-flankers in the N400 window (ref.: no-flankers). Significant values are shown in bold.

	b	*SE*	*t*
Intercept	0.04	0.88	0.05
Flanker presence	-0.80	0.16	**-4.97**
Fl. presence × Anteriority	-0.34	0.12	**-2.82**
Fl. presence × Laterality	-0.29	0.12	**-2.36**
Three-way interaction	0.15	0.09	1.60

Discussion

The present study provides the first electrophysiological examination of orthographic parafoveal-on-foveal effects in a flanker paradigm. We show that such effects onset in the time window of the N250, a component associated with the mapping of activated sub-lexical representations (e.g., letters, bigrams) onto whole-word form representations (Grainger & Holcomb, 2009). We argue that the greater negativity in the presence of unrelated flankers is caused by the activation of a larger set of sub-lexical nodes. This in turn translates into the increased processing difficulty at the lexical level, characterized by more negative deflections in the N400 window.

The absence of flanker relatedness effects prior to 200 ms (Figure 5) speaks against accounts of orthographic integration effects positing that such effects take place at levels of low-level visual processing (e.g., feature detectors; Angele et al., 2013). Instead, the present findings converge with recent observations that processing at the level of orthographic representations occurs for multiple words in parallel (e.g., Snell, Mathôt, Mirault & Grainger, 2018c).

Interestingly, while the repetition flanker condition and no-flanker condition yielded equal RTs (Figure 3), the EEG data make clear that these two conditions differ from the unrelated flanker condition for different reasons. While the effect of flanker relatedness in the N250 was stronger over anterior sites, the effect of flanker presence had a more posterior locus. This dichotomy coincides quite well with the recent claim of Snell and Grainger (2018) that parafoveal words impact on foveal word processing in at least two distinct ways: firstly, through the spatial integration of orthographic information, and secondly, through shifting some attentional resources away from the fovea. This suggests that the effects observed in Figures 8B and 9B are not only orthographic, but also attentional in nature. This can be tested, for instance, in a flanker paradigm that implements spatial cues (e.g., Posner, 1980), with the prediction that such cues should modulate the effects of flanker presence depicted in Figures 8B and 9B to a greater degree than the effects of flanker relatedness depicted in Figures 8A and 9A.

Methodologically, the present study joins other recent endeavors (e.g. Emmorey et al., 2017; Payne et al., 2015; Winsler et al., 2018) in the employment of linear mixed-effects models (LMEs) to analyze EEG data. The justification for using LMEs to analyze EEG data follows the exact same line of reasoning that has driven LMEs to become the golden standard in behavioral data analyses—at least within the realm of language research. That is, LMEs allow one to take into account by-participant and by-item variance, hence reducing noise; moreover, the trial-by-trial approach of LMEs omits the ambiguity of within-subject variance inherent to ANOVAs, thus eliminating the need to exclude participants with a high proportion of contaminated trials and preserving more statistical power.

It is our sense that the use of LMEs is particularly fruitful when performing time course analyses (Section 3.2.1). Indeed, given that temporal resolution is EEG's strong suit, it is arguably somewhat ironic that the classic approach for analyzing ERPs consists of averaging data across large intervals of interest, which inevitably masks the more intricate activation dynamics that unfold within those intervals. A case in point is provided in Section 3.2.2, where we report the absence of an interaction between flanker relatedness and anteriority within the conventional 250—450 ms window of the N400. Yet, the time course analyses plotted in Figure 5 hint at an interaction with anteriority, as the frontal electrodes in specific regressed towards non-significance beyond the 400 ms time point while effects continued to increase for central and posterior electrodes. The reason for the absence of an interaction in the DEM analysis, then, is the relatively large interval across which deflections were averaged. At the conventional start of this interval, 250 ms, the effect of relatedness is still strong over frontal sites (in line with the frontal orientation of the N250). This compensates for the weaker effect over frontal sites at later time points, hence eliminating an interaction effect. Hence, as LMEs boast enough power and sensitivity to the data to analyze the data at a higher temporal resolution, LME-based time course analyses arguably maximize what can be learned from a given EEG dataset.

In sum, in the present paper we have shed light on the nature of orthographic parafoveal-on-foveal effects. Our analyses have made it clear that such effects unfold approximately 200 ms post-stimulus onset. We conclude that parafoveal and foveal words jointly activate sub-lexical nodes, which in turn leads to faster word activation if the parafoveal and foveal word are identical.

References

Angele, B., Tran, R. & Rayner, K. (2013). Parafoveal–foveal overlap can facilitate ongoing word identification during reading: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 526-538.

Baayen, R. (2008). *Analyzing Linguistic Data: A practical introduction to statistics*. Cambridge: Cambridge University Press (2008).

Bann, S. & Herdman, A. (2016). Event related potentials reveal early phonological and orthographic processing of single letters in letter-detection and letter-rhyme paradigms. *Frontiers in Human Neuroscience, 10,* 176.

Bentin, S., McCarthy, G., Wood, C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology, 60, 343*-55.

Dare, N. & Shillcock, R. (2013). Serial and parallel processing in reading: Investigating the effects of parafoveal orthographic information on nonisolated word recognition. *The Quarterly Journal of Experimental Psychology, 66*, 417-428.

Dufau, S., Grainger, J. & Holcomb, P. (2008). An ERP investigation of location invariance in masked repetition priming. *Cognitive, Affective and Behavioral Neuroscience, 8,* 222-228.

Dufau, S., Grainger, J., Midgley, K., & Holcomb, P. (2015). A thousand words are worth a picture: Snapshots of printed word processing in an ERP megastudy. *Psychological Science*, *26,* 1887-1897.

Eddy, M., Schmid, A. & Holcomb, P. (2006). A new approach to tracking the time-course of object perception: Masked repetition priming and event-related brain potentials. *Psychophysiology, 43,* 564-568.

Emmorey, K., Midgley, K., Kohen, C., Sehyr, Z. & Holcomb, P. (2017). The N170 ERP component differs in laterality, distribution, and association with continuous reading measures for deaf and hearing readers. *Neuropsychologia, 106,* 298-309.

Ferrand, L., New, B., Brysbaert,M., Keuleers, E., Bonin, P.,Méot, A., Augustinova, M. & Pallier, C. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods, 42,* 488-496.

Forster, K. & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10, 680*-698.

Grainger, J. (2008). Cracking the orthographic code: An introduction. *Language and Cognitive Processes, 23*, 1-35.

Grainger, J. & Holcomb, P. (2009). Watching the word go by: On the time course of component processes in visual word recognition. *Language and Linguistics Compass, 3,* 128-156.

Grainger, J., Kiyonaga, K. & Holcomb, P. (2006). The time-course of orthographic and phonological code activation. *Psychological Science, 17,* 1021-1026.

Grainger, J., Mathôt, S. & Vitu, F. (2014). Test of a model of multi-word reading: Effects of parafoveal flanking letters on foveal word recognition. *Acta Psychologica,* 146, 35-40.

Grainger, J., & van Heuven, W. (2003). Modeling letter position coding in printed word perception. In P. Bonin (Ed.), *The mental lexicon* (pp. 1-23). New York: Nova Science.

Guthrie, D. & Buchwald, J. (1991). Significance testing of difference potentials. *Psychophysiology, 28,* 240-244.

Holcomb, P. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and Language, 35,* 66-85.

Holcomb, P. & Grainger, J. (2006). The time-course of masked repetition priming: An event-related brain potential investigation. *Journal of Cognitive Neuroscience, 18,* 1631-1643.

Holcomb, P., O'Rourke, T. & Grainger, J. (2002). An event-related brain potential study of orthographic similarity. *Journal of Cognitive Neuroscience, 14,* 938-950.

Inhoff A., Radach R., Starr M. & Greenberg S. (2000). *Allocation of visuospatial attention and saccade programming during reading*. In Kennedy A., Radach R., Heller D. & Pynte J. (Eds.). Reading as a perceptual process. Oxford, UK: Elsevier.

Kiyonaga, K., Midgley, K., Holcomb, P. & Grainger, J. (2007). Masked cross-modal repetition priming: An ERP investigation. *Language and Cognitive Processes, 22,* 337-376.

Massol, S., Grainger, J., Dufau, S. & Holcomb, P. (2010). Masked priming from orthographic neighbors: an ERP investigation. *Journal of Experimental Psychology: Human Perception & Performance, 36*, 162-174.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An opensource, graphical experiment builder for the social sciences. *Behavior Research Methods, 44,* 314-324.

Meade, G., Grainger, J., Midgley, K., Emmorey, K. & Holcomb, P. (2018). From sublexical facilitation to lexical competition: ERP effects of masked neighbor priming. *Brain Research, 1685,* 29-41.

Payne, B., Lee, C. & Federmeier, K. (2015). Revisiting the incremental effects of context on word processing: Evidence from single-word event-related brain potentials. *Psychophysiology, 52,* 1456-1469

Petit, J., Grainger, J., Midgley, K. & Holcomb, P. (2006). On the time-course of processing in letter perception: A masked priming ERP investigation. *Psychonomic Bulletin and Review, 13,* 674-681.

Posner, M. (1980). Orienting of attention. Quarterly Journal of Experimental Psychology, 32, 3-25.

Snell, J., Bertrand, D. & Grainger, J. (2018b). Parafoveal letter-position coding in reading. *Memory & Cognition.* doi:10.3758/s13421-017-0786-0

Snell, J., Bertrand, D., Meeter, M. & Grainger (2018a). Integrating orthographic information across time and space: Masked priming and flanker effects with orthographic neighbors. *Experimental Psychology, 65,* 32-39.

Snell, J. & Grainger, J. (2018). Parallel word processing in the flanker paradigm has a rightward bias. *Attention, Perception & Psychophysics,* doi: 10.3758/s13414-018-1547-2.

Snell, van Leipsig, Grainger, J. & Meeter, M. (2018c). OB1-reader: A model of word recognition and eye movements in text reading. *Psychological Review,* (in press).

Snell, J., Vitu, F. & Grainger, J. (2017a). Integration of parafoveal orthographic information during foveal word reading: Beyond the sub-lexical level? *Quarterly Journal of Experimental Psychology*, *70*, 1984-1996.

Winsler, K., Midgley, K., Grainger, J. & Holcomb, P. (2018). An electrophysiological megastudy of word recognition. *Language, Cognition & Neuroscience*, doi: 10.1080/23273798.2018.1455985.

Parallel graded attention in reading: A pupillometric study

Joshua Snell, Sebastiaan Mathôt, Jonathan Mirault & Jonathan Grainger

Citation: Snell, J., Mathôt, S., Mirault, J. & Grainger, J. (2018). Parallel graded attention in reading: A pupillometric study. *Scientific Reports, 8*, 3743. doi: 10.1038/s41598-018-22138-7

Abstract

There are roughly two lines of theory to account for recent evidence that word processing is influenced by adjacent orthographic information. One line assumes that multiple words can be processed simultaneously through a parallel graded distribution of visuo-spatial attention. The other line assumes that attention is strictly directed to single words, but that letter detectors are connected to both foveal and parafoveal feature detectors, as such driving parafoveal-foveal integrative effects. Putting these two accounts to the test, we build on recent research showing that the pupil responds to the brightness of *covertly attended* (i.e., without looking) locations in the visual field. Experiment 1 showed that foveal target word processing was facilitated by related parafoveal flanking words when these were positioned to the left and right of the target, but not when these were positioned above and below the target. Perfectly in line with this asymmetry, in Experiment 2 we found that the pupil size was contingent with the brightness of the locations of horizontally but not vertically aligned flankers, indicating that attentional resources were allocated to those words involved in the parafoveal-on-foveal effect. We conclude that orthographic parafoveal-on-foveal effects are driven by parallel graded attention.

Introduction

One of the most hotly debated issues in reading research concerns the question of whether words are processed serially or in parallel[1,2,3,4,5]. Of particular relevance in this debate are the recent lines of research pointing out that the recognition of words in the fovea is influenced by parafoveal orthographic information. The principal finding is that words are recognized faster when they are orthographically related to surrounding words or letters. For instance, several studies have found the fixation duration on word *n* to be decreased when it was orthographically related to *n+1* during sentence reading[1,2,3,4,5]. Furthermore, in the flanker paradigm, lexical decisions about isolated target words were found to be faster and more accurate when those targets were flanked by related letters on each side (e.g. '*ro rock ck*') as compared to unrelated letters (e.g. '*st rock ep*')[4,5,6].

The conception that foveal word processing is influenced by surrounding words might lead one to conclude that attentional resources must have been allocated to the foveal word and surrounding words in parallel[8,9,10,11,12,13]. According to the parallel processing approach, the attentional distribution would follow a Gaussian shape centered on the attentional focus[10], such that while processing of the fixated word would normally be the strongest, surrounding words and letters may nonetheless exert some influence, hence explaining the so-called parafoveal-on-foveal effects listed above.

Yet, these effects are not considered by all to provide conclusive evidence that attention is distributed across multiple words. Indeed, one alternative theory proposes that visuo-spatial attention is strictly directed to one word at a time (i.e. serial processing accounts of reading[7]), but that foveal letter detectors would be connected to both foveal and parafoveal feature detectors. Parafoveal orthographic information would as such influence foveal letter processing without having received any attentional resources[3]. Importantly however, syntactic and semantic variants of the flanker paradigm have shown that syntactic decisions (e.g., *noun/verb*) and semantic decisions (e.g., *natural/artifactual object*) about foveal target words are made faster if those targets were flanked by congruent words as compared to incongruent words (stimulus on-time 170 ms)[8,9]. Crucially, there was no orthographic overlap between targets and flankers, implying that the idea of parafoveal feature detectors influencing foveal letter detectors cannot account for these particular findings. It must be noted however, that while these higher-order (e.g. syntactic, semantic) parafoveal-on-foveal effects show up in artificial reading tasks such as the flanker paradigm, they are not expressed in more natural measures of reading speed, such as fixation durations in sentence reading[3,8,9] – possibly because higher-order information is simply not integrated across words during normal reading.

The debate of serial versus parallel processing still ongoing, it is apparent that the field is in dire need of a more direct measure of the allocation of attention during reading. This is precisely what we aim to provide with the present work. Specifically, we report a novel methodology that builds on recently obtained evidence that the pupillary light response can reflect visual attention.

Using the pupil to track attention

The pupillary light response consists of a constriction of the pupil in brightness (to reduce the amount of light entering the eye) and a dilation of the pupil in darkness (to maximize the amount of light entering the eye). It was historically considered to be a low-level reflex without any cognitive component[14]. However, in recent years it has become clear that pupil size changes may reflect a multitude of higher-order phenomena, such as awareness[15], interpretation[16] and mental imagery[17,18,19].

Concerning attention, in a paradigm where participants responded to target stimuli appearing in the left- or right visual hemifield, with the screen background being vertically split into a white and a black half with a luminance-neutral (gray) band in the middle, it was found that presenting an auditory cue prior to target onset ('left' or 'right', indicating the probable location of the target) caused the pupil to dilate more if the cued side was black, as compared to if the cued side was white, indicating that the shift of covert (i.e., without looking) visual attention triggered a pupillary light response[20]. Crucially, the eye position was carefully tracked, so as to ensure that the overall amount of light entering the pupil was equal across conditions.

A more recently implemented paradigm showed that the pupil size could reveal which one out of several parafoveal orthographic stimuli participants covertly attended[21]. In this paradigm, packaged as a 'mind-writing' interface, eight letters were presented in a circle around a central fixation point. Each letter was presented on a background that oscillated between white and black, with four letters being on a black background during the time that the other four letters were on a white background, and vice versa. Attending one of the letters (while continuing to focus on the screen center) caused the pupil size to oscillate in cadence with the background of that letter, thus allowing the computer to bring the amount of possibly attended letters from eight to four. These four letters were then again divided into two groups with opposing background brightness, until the pupil size led the computer to select two candidate letters. A final repetition of this procedure led the computer to deduce the truly attended letter.

Involving pupillometry in the flanker paradigm

It is clear then, that pupil size is contingent with the brightness of covertly attended locations in the visual field. This sparks a straightforward prediction concerning the role of attention in parafoveal-on-foveal effects: if attentional resources are indeed allocated to parafoveal stimuli, then the pupil size should be influenced by the brightness of the locations of these stimuli (keeping the overall screen brightness equal). To this end we devised a series of experiments wherein we could obtain varying degrees of parafoveal-foveal integration, while manipulating the brightness of parafoveal locations. The flanker paradigm was particularly suited for this, as it simulates the conditions of reading in a controlled manner that does not necessitate the reader to make saccades, while nevertheless allowing us to determine to which extent readers are engaged in additional processing by surrounding words. The aim was to see, firstly, whether pupil size would be affected by the brightness of flanker locations, and secondly, whether this pupil size effect would covary with the degree of parafoveal-foveal integration.

In order to obtain varying degrees of parafoveal-foveal integration, we hypothesized that flankers positioned above and below the target word should have a smaller impact on target processing than flankers positioned left and right of the target, given that attention is mainly distributed along the horizontal axis during reading (in scripts that are aligned horizontally)[22]. Experiment 1 tested this hypothesis by letting participants make lexical decisions about foveal target words, while orthographically related or unrelated words were presented either left and right of the target, or above and below the target (stimulus on-time 150 ms).

All experiments were carried out with approval of the Scientific and Ethical committee of the Aix-Marseille University, and were in accordance with the declaration of Helsinki. All data gathered for this work are publicly available at https://osf.io/jm938/.

Experiment 1: A first test of horizontal vs. vertical integration

Method

Participants

Twenty students from Aix-Marseille University gave informed consent to participate in this study, carried out at the Laboratoire de Psychologie Cognitive (Marseille, France) for €4,- or its equivalent in course credit. All participants were native to the French language, non-dyslexic, and naïve to the purpose of the study.

Materials

We retrieved a list of 80 4-letter target words from the French Lexicon Project database (Ferrand et al., 2010). These targets were noun or non-conjugated verb and contained no diacritics (e.g. *é, ô, ç*). Each target was paired up with a 4-letter control word that met the same criteria as the target and that shared no letters with the target. The average frequency of targets and control words was equal, at 5.41 Zipf (a log10-based frequency measure[23]).

In a similar fashion we retrieved 80 4-letter non-word targets from the French Lexicon Project pseudoword database[24], and paired each of these with a non-word control. The non-word targets were used as filler stimuli and were not included in the data analyses.

Design

Experiment 1 followed a 2 × 2 factorial design with flanker relatedness (related / unrelated to the target) and flanker position (above and below the target / left and right of the target) as factors (note that the target lexicality factor is disregarded here). In the related flanker conditions, the target word was repeated at the flanker locations (Figure 1). In the unrelated conditions, the control word was shown at the flanker locations. Participants were Latin-squared into two groups to ensure that all stimuli were shown in all four conditions, but only twice per participant (with the same flanker shown in both flanker positions). The experiment thus consisted of 320 trials (including non-word trials), and these were presented in random order.

Apparatus and software

The experiment was implemented with OpenSesame[25] and presented on a 1024x768 px, 150 Hz computer monitor. Participants were seated in a comfortable office chair at a distance of 50 cm from the display, so that each character space subtended 0.40 degrees of visual angle. Responses were collected with a gamepad (using two trigger buttons for the left and right index finger) at a polling rate of approximately 125 Hz.

Procedure

Participants received instructions both verbally by the experimenter and visually onscreen. Participants were instructed to maintain their focus at the center of the display, guided by four diagonally oriented fixation bars (Figure 1). 700 ms after the start of each trial, the target was presented for 150 ms (in line with previous implementations of the flanker paradigm [4,5,6]), together with two flankers that were either the same word or the unrelated control word. These flankers were positioned left and right of the target, or above and below the target, with the word centers at a distance of 90 pixels from the screen center. After the target and flankers disappeared, participants had a maximum of 2000 ms to respond whether the target was a word or non-word, with a right- or left-sided button press respectively. Feedback was provided after each response (green or red dot for correct and incorrect answers, respectively). The experimental trials were preceded by a block of eight practice trials, and participants were offered a break halfway through the experiment. The experiment lasted approximately 20 minutes.

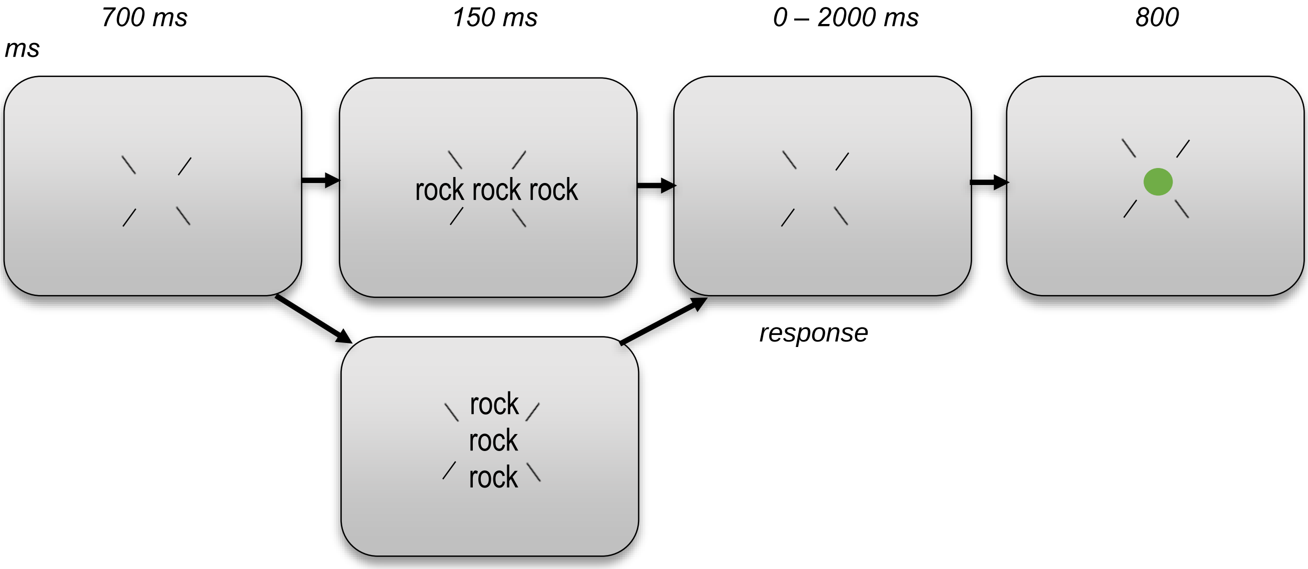

Figure 1. Experiment 1 trial procedure. After a 700 ms fixation display, the target was presented for 150 ms, flanked either by the same word (shown here) or the unrelated control word, left and right of the target (top panel) or above and below the target (bottom panel). Participants then had a maximum of 2000 ms to respond whether the target was a word or non-word, after which feedback was provided (green or red dot for correct and incorrect answers respectively). Note that this figure only shows two of the four possible target displays. The size of stimuli relative to the display is exaggerated in this figure.

Results

For the analysis of response times (RTs) we included all correctly answered trials (with a word target) for which the RT was no further than 2.5 standard deviations from the grand mean (86.91% of trials). The latter criterion led us to exclude 2.12% of trials for the analysis of error rates.

For the RT analysis we used a linear mixed-effects model (LMM) with items and participants as crossed random effects[26]. We determined the maximal random effects structure permitted by the data, leading us to include the interaction of flanker relatedness × position as by-item and by-participant random slopes alongside random intercepts[27]. We report regression coefficients (b), standard errors (SE) and t-values, with values $|t|>1.96$ deemed significant[26]. A logistic LMM was used to analyze the error rates, (here we report z-values instead of t-values). In this particular model, a failure to converge under the maximal random effects structure led us to include only the by-item and by-participant random intercepts. The models were fitted with the lme4 package[28] in the R statistical computing environment.

Table 1. Experiment 1 condition means.

	RTs (ms)		*Error rates*	
Flanker type	*related*	*unrelated*	*related*	*unrelated*
Left/right of target	403.94 (114.37)	443.33 (122.33)	.08 (.06)	.15 (.06)
Above/below target	403.26 (120.81)	404.93 (113.60)	.08 (.06)	.09 (.05)

Note: values in between parentheses indicate standard deviations.

Table 1 shows the mean RTs and error rates for all conditions. There was a significant main effect of flanker relatedness (Table 2), with unrelated flankers leading to longer RTs than related flankers. There were also significantly more errors in the unrelated flanker conditions. Meanwhile, there was no main effect of flanker position.

In line with our hypothesis, there was a significant interaction of flanker relatedness and flanker position (Table 2). As can be clearly seen in Table 1, the effect of flanker relatedness was strongly expressed for the left- and right-positioned flankers ($\Lambda \approx 40$ ms), while it was virtually absent when flankers were positioned above and below the target.

Table 2. Analyses of RTs and error rates (ref.: left/right related flankers).

	RTs			*Error rates*		
	b	SE	*t*	*b*	SE	*z*
(intercept)	404.69	13.95	**29.01**	2.75	0.20	13.99
Relatedness (R)	42.48	7.35	**5.78**	0.77	0.17	**4.63**
Position (P)	0.37	6.51	0.06	0.04	0.18	0.20
R × P	40.58	9.79	**4.14**	0.73	0.25	**2.95**

Note: significant values are shown in bold.

Discussion

Our aim with Experiment 1 was to see whether various degrees of parafoveal-foveal integration could be obtained using horizontally vs. vertically aligned stimuli. We predicted that flankers positioned to the left and right of the target would have a stronger impact on target processing than flankers positioned above and below the target, in line with the idea that attention would be mainly distributed along the horizontal axis during reading. This hypothesis was confirmed, as we found a strong effect of flanker relatedness with horizontally aligned stimuli but not with vertically aligned stimuli (Table 1).

Although these results support the conception that parafoveal-on-foveal effects are driven by parallel graded attention, they do not provide conclusive evidence. Within the alternative line of reasoning – i.e., that foveal letter detectors may be connected to parafoveal feature detectors (causing parafoveal information to have an impact without the necessity of being attended)[3] – it is possible that the letter detectors are mainly connected parafoveal feature detectors in the horizontal dimension rather than the vertical dimension. This would in turn lead to an influence of words to the left and right of fixation, but not of words above and below fixation.

Following the rationale outlined in the Introduction, if the parafoveal-on-foveal effects of horizontally aligned flankers were driven by parallel graded attention, then the reader's pupil size should be contingent with the brightness of the locations of those flankers. In contrast, the pupil size should not be contingent with the brightness of flankers that did not impact on target processing—that is, the vertically aligned flankers.

This prediction was put to the test in Experiment 2. In a setting similar to that of Experiment 1, we manipulated the brightness of flanker locations (i.e., flankers had either a black or white background), and hypothesized a flanker brightness × position interaction effect on the pupil size, with black flanker backgrounds causing a dilated pupil compared to white flanker backgrounds, specifically for horizontally but not vertically aligned flankers. Masks (‘####’) were presented at the flanker positions perpendicular to that of the word flankers, with a background color opposite to that of the word flankers, so that the overall luminance of the display was equal across conditions. The eye position and pupil size were tracked during a fixed 2450 ms interval, allowing us to assess the pupillary light response in full. As in Experiment 1, targets, flankers and masks were shown for 150 ms on each trial, while the flanker backgrounds were kept onscreen throughout the 2450 ms interval.

Experiment 2: Flanker brightness and the pupillary light response

Method

Participants

Twenty-four students from the Aix-Marseille University gave informed consent to participate in this study, carried out at the Laboratoire de Psychologie Cognitive (Marseille, France) for €5,- or its equivalent in course credit. All participants were native to the French language, non-dyslexic and naïve to the purpose of the study. Further, all participants reported to have normal vision (and thus did not use glasses or contact lenses, which tend to disturb the eye-tracker signal).

Materials

As we extended the experimental design with an additional factor (flanker brightness), we increased the total amount of target words (and non-word targets) to 100. The selection criteria were left unchanged from Experiment 1. The average frequency of targets and control words was 5.40 and 5.46 Zipf, respectively.

Design

Experiment 2 followed a 2 × 2 × 2 design, with flanker relatedness (related / unrelated to the target), position (left and right of the target / above and below the target) and background brightness (white / black flanker background) as factors. Participants were Latin-squared into four groups, such that all targets were shown across all conditions, but only twice per participant (with the same flanker identity and position shown in both the dark and the bright background setting). The experiment thus totaled 400 trials (including non-word trials), and these were presented in random order.

Apparatus

The PyGaze back-end[29] was used to process eye movement data online. The participant’s right eye position was recorded with an EyeLink 1000 (SR Research, Mississauga, ON, Canada), a video-based eye tracker sampling at 1000 Hz with a spatial resolution of 0.01°. We acknowledge that the tracking of a single eye prevented us from taking a potential degree of binocular disparity (which may have induced minor noise in the data) into account. Participants were seated at a 90 cm distance from the display, so that each character space subtended 0.35 degrees of visual angle. A chin-rest was used to facilitate a stable head position. Responses were collected with a keyboard instead of a gamepad this time.

Procedure

Prior to the start of the experiment, the participant's right eye position was calibrated using a 9-point calibration grid. The trial display (Figure 2) differed from that of Experiment 1 in two respects: firstly, the flankers were presented in luminance-neutral (gray) color on either a white or black square background, while the target was presented in luminance-neutral green. Secondly, gray masks ('####') were shown at the flanker positions perpendicular to the word flanker positions, with a background color opposite to that of the word flankers, to ensure that the overall display luminance was equal across conditions.

Participants were instructed to maintain their focus at the screen center. As in Experiment 1, targets and flankers were presented for 150 ms. Unlike Experiment 1 however, in Experiment 2 participants were instructed to provide their response during a response display that was presented 2300 ms after the target- and flanker offset. As can be seen in Figure 2, the flanker backgrounds stayed onscreen throughout this interval. The pupil size was measured during 2450 ms after stimulus onset, thus allowing us to assess the pupillary light response in full. Note that participants were not instructed to respond as fast as possible in this experiment because the response itself influences the pupil size; a difference in RT between conditions would as such lead to a cofounded pupil size effect. Participants were offered a break halfway through the experiment. The experiment lasted approximately 45 minutes.

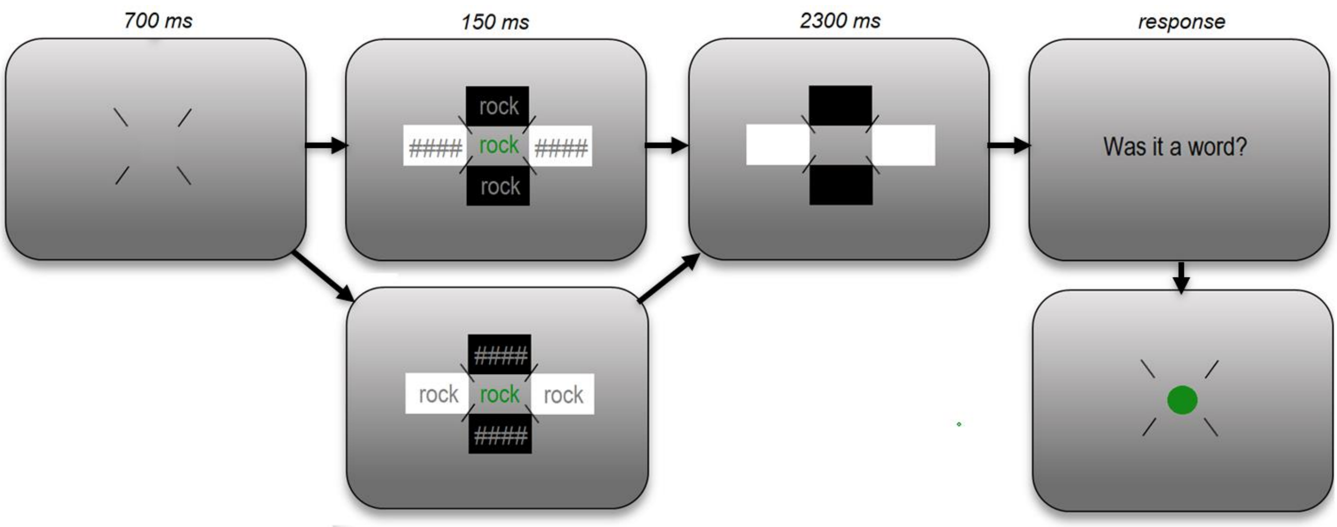

Figure 2. Experiment 2 trial procedure. The target was presented in a luminance-neutral green color on a gray background. Flankers were presented in luminance-neutral gray on a white or black square-shaped background, with masks being presented on backgrounds of the opposite color. Note that this figure only shows two of the eight possible target displays. The size of stimuli relative to the screen is exaggerated in this example.

Results

Since participants were instructed only to respond when the response display was presented (2300 ms after stimulus offset), we do not report RTs for Experiment 2. The error rates are presented in Table 3. In line with previous results, significantly more errors were made in the unrelated flanker conditions compared to the related flanker conditions (b = 0.61, SE = 0.23, z = 2.61), but this effect was not modulated by flanker position (b = 0.32, SE = 0.33, z = 0.96).

Table 3. Experiment 2 error rates.

Flanker type	*Related*	*Unrelated*
Left/right of target	.02 (.02)	.04 (.04)
Above/below target	.03 (.03)	.04 (.02)

Note: values in between parentheses indicate standard deviations.

Pupil size

Prior to the analyses of pupil size, the pupil size was normalized and baseline-corrected (taking the average of 300 ms prior to stimulus onset) such that the pupil size at stimulus onset was equal to 0. The data was downsampled by factor 10 (i.e., taking the average of every 10 ms of data), such that each 2450 ms interval was represented by 245 datapoints per subject. An LMM with flanker brightness, flanker position and flanker relatedness as factors and subjects and items as crossed random effects was run over the course of 245 cycles, each representing 10 ms of pupil size data.

Figure 3 shows pupil size deflections for horizontally aligned flankers (top two panels) and vertically aligned flankers (bottom two panels), both in the related conditions (left panels) as well as the unrelated conditions (right panels). It is apparent that the pupil size was contingent with the location brightness of horizontally aligned flankers, given the decreased pupil size in the presence of a white background compared to a black background in these conditions (top two panels). In contrast, the pupil size deflection was quite similar for white and black background flankers when these were vertically aligned (bottom two panels).

As can be seen in Figure 4, a significant main effect of flanker location brightness was established around the 250 ms mark (with marginal significance showing as early as 200 ms after stimulus onset). Crucially, we also established a significant interaction of flanker location brightness and flanker position, such that the pupil size was contingent with the brightness of horizontally but not vertically aligned flankers, in line with our hypothesis. It is apparent that the data is auto-correlating: i.e., intervals of significance are preceded and followed by trends towards significance, indicating that the observed effects are no false positives (which are by default likely to occur in multiple-comparison analyses).

Interestingly, a significant interaction of flanker location brightness and flanker relatedness emerged around the 500 ms mark, such that the effect of location brightness was enhanced if the flankers were orthographically related to the target. Post-hoc analyses revealed that this effect was driven by the conditions with horizontally aligned word flankers, as the pattern persisted when viewing these conditions in isolation, but not when viewing vertically aligned word flanker trials in isolation. Neither a main effect of flanker relatedness (Figure 4), nor a three-way interaction of location brightness, relatedness and position was established.

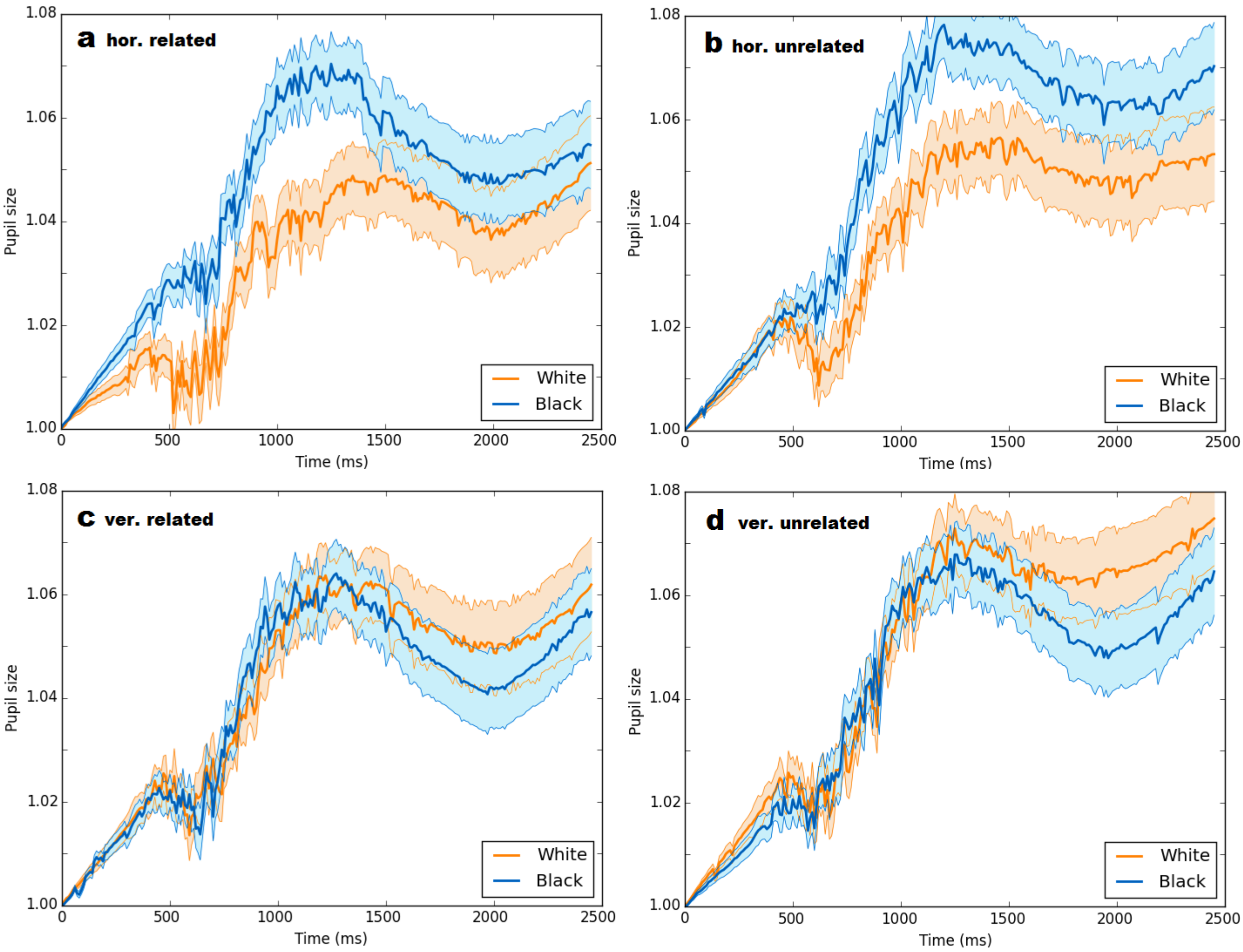

Figure 3. Average pupil size during trials in which the word flankers had black backgrounds (blue lines) versus white backgrounds (orange lines), when the flankers and target were horizontally aligned (panels *a* and *b*) and vertically aligned (panels *c* and *d*) respectively. Panels *a* and *c* show the related flanker conditions, whereas panels *b* and *d* show the unrelated flanker conditions. The blue and orange shaded areas indicate standard errors.

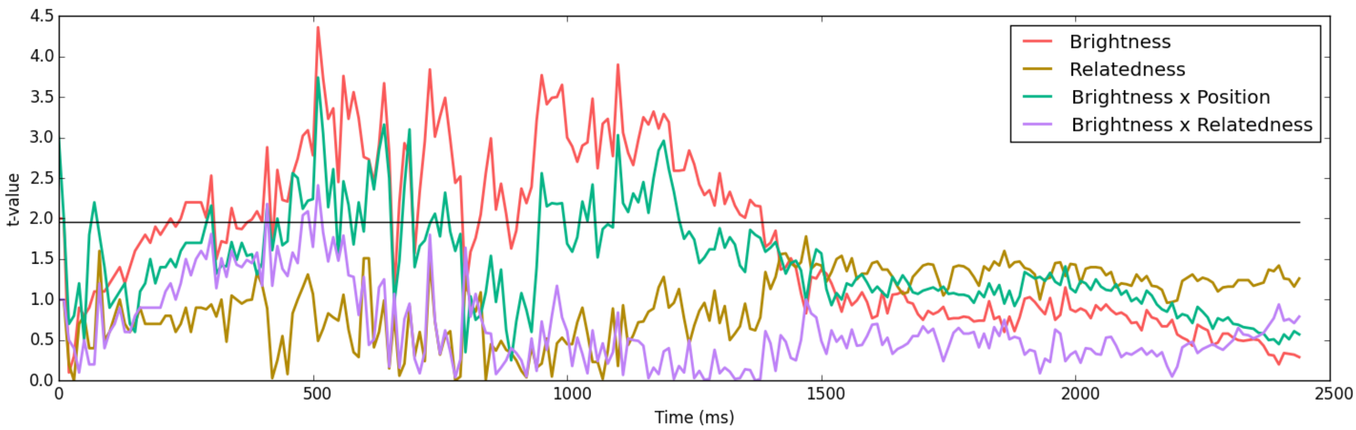

Figure 4. Experiment 2 analysis outcomes. The y-axis represents statistical significance, with the significance threshold (| *t* | = 1.96) being indicated by the black horizontal line. The analysis was carried out on each 10 ms time step, amounting to a total of 245 analyses for the 2450 ms interval.

Discussion

In Experiment 1 we found that information is integrated across horizontally but not vertically aligned stimuli. Crucially, as established in Experiment 2, this asymmetry is also expressed in the pupil size – i.e., we obtained a pupil size effect with horizontally aligned flankers, but not with vertically aligned flankers, indicating that attention was directed specifically to the locations of those stimuli involved in the orthographic parafoveal-on-foveal effect. The flanker location brightness effect started manifesting itself as early as 200 ms after stimulus onset (Figure 4), which is the fastest latency associated with the pupillary light response[14]. This immediacy of effects thus suggests that a portion of attention was directed to the flanking stimuli *during* – rather than *after* – processing of the target, in line with accounts of reading that assume a parallel graded distribution of attention.

We further established an interaction of flanker location brightness and flanker relatedness around the 500 ms mark. Post-hoc analyses (not reported above) revealed that this effect persisted when analyzing horizontally aligned word flanker trials in isolation, but not when analyzing vertically aligned word flanker trials in isolation. This particular finding is quite remarkable in that it reflects how the reading process is driven by a continuous interaction of various cognitive levels. Specifically, parafoveal orthographic processing led to stronger lexical activation in the related flanker conditions, subsequently leading to enhanced low-level visual processing and as such an increased pupillary light response.

General Discussion

In this paper we have presented a novel methodology to track the distribution of visuo-spatial attention in a controlled reading setting. This methodology was employed to address a key phenomenon driving the debate about whether words are processed serially or in parallel during reading: namely, orthographic parafoveal-on-foveal effects. The principal finding is that words are recognized slower when surrounded by orthographically unrelated information (e.g., '*st rock ep*') compared to related information ('*ro rock ck*'). At present, two lines of theory exist to account for these effects. One line assumes that visuo-spatial attention is allocated to multiple words in parallel, as such allowing for the integration of information across the fovea and parafovea[4,5,8,9,10]. The other line assumes that attention is strictly allocated to single words, but that letter detectors are connected to both foveal and parafoveal feature detectors, as such causing word processing to be influenced by parafoveal orthographic information[3].

To provide a direct test of the theory that attention can be allocated to multiple words in parallel, we made use of the principle that the pupil responds to the brightness of covertly attended locations in the visual field[19,20,21]. We predicted that if orthographic parafoveal-on-foveal effects are indeed driven by parallel distributed attention, then the pupil size should be contingent with the brightness of those stimuli that are involved in the parafoveal-on-foveal effect. We therefore aimed to create a setting in which we could obtain various degrees of parafoveal-foveal integration while manipulating the brightness of parafoveal stimuli – expecting the latter to cause a pupil size effect specifically in those conditions that produced an orthographic parafoveal-on-foveal effect.

In Experiment 1 we found that foveal word processing was impacted by horizontally but not vertically aligned adjacent words, as lexical decisions about foveal target words were slowed and less accurate when those targets were flanked by orthographically unrelated words on the left and right (compared to being flanked by a repetition), but not when the same word was presented above and below the target. Following the above rationale, we thus hypothesized that a manipulation of the brightness of flanker locations would cause a pupil size effect with horizontally aligned flankers, but not with vertically aligned flankers.

The results of Experiment 2 were perfectly in line with this hypothesis: an interaction of flanker location brightness and flanker position was established, such that the pupil size was affected by the brightness of words located left and right of the target, but not words located above and below the target. Of crucial importance here is that the pupil size effect started manifesting itself as early as 200 ms after stimulus onset (Figure 4), which is the minimal latency associated with the pupillary light response[14], thus indicating immediate processing whereby portions of attention were allocated to the parafoveal words *during*, rather than *after*, target processing.

Remarkably, the pupillary light response was modulated by the orthographic relatedness of flankers as well – albeit only for a short time window approximately 500 ms after stimulus onset. Our account of this particular finding is that increased lexical activation due to integration of orthographically related information in turn leads to increased activation in earlier visual areas through recurrent processing, consequently enhancing the pupillary light response. It must be acknowledged however, that this interpretation of effects is as of yet merely speculative.

A question that remains, is whether the present findings speak to the reading system in general, and sentence reading in particular. Indeed, one could argue that attention may be strictly directed to single words during sentence reading, while it would be distributed across multiple words in 'unnatural' reading settings such as the experiments reported in this paper. Undoubtedly these settings differ in nature, as is attested by the fact that higher-order (e.g. syntactic, semantic) parafoveal-on-foveal integration was observed in the flanker paradigm[8,9] but not in sentence reading[3,8,9] (however, see [30]). It is possible, however, that such differences are driven by how the reading system organizes incoming information in each respective setting, rather than by how attention is distributed. Indeed, given that readers are not able to effectively focus attention on single words in the flanker paradigm, it stands to question how readers could then manage to do so during sentence reading, considering that (i) parafoveal words are more relevant and important during sentence reading, (ii) parafoveal words are available longer during sentence reading (compared to 150 ms in the flanker paradigm), and (iii) the visual input during sentence reading is more complex and dynamic due to eye-movements. It may in this light be more sensible that the reading system in principle processes multiple words at once, but that mechanisms driving sentence-level comprehension prevent cross-leakage of higher-order information between words during sentence reading; (see [8,9] for a detailed discussion of this possibility).

In sum, the present results suggest that orthographic parafoveal-on-foveal effects are driven by parallel processing of multiple words through a widespread distribution of visuo-spatial attention. We conclude that the pupillary light response is a fruitful means to addressing the role of attention in reading.

Acknowledgements

We thank Svenja Krabbe for her help in carrying out the experiments. This research was supported by grants ANR-11-LABX-0036 (BLRI), ANR-16-CONV-0002 (ILCB), and ANR-15-CE33-0002-01 from the French National Research Agency (ANR).

References

[1] Inhoff, A., Radach, R., Starr, M. & Greenberg, S. Allocation of visuospatial attention and saccade programming during reading in *Reading as a perceptual process* (eds. Kennedy, A., Radach, R., Heller, D. & Pynte, J.) 221-246 (Elsevier, 2013).

[2] Vitu, F., Brysbaert, M., & Lancelin, D. A test of parafoveal-on-foveal effects with pairs of orthographically related words. *European Journal of Cognitive Psychology,* **16,** 154-177 (2004).

[3] Angele, B., Tran, R. & Rayner, K. Parafoveal–foveal overlap can facilitate ongoing word identification during reading: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, **39,** 526-538 (2013).

[4] Dare, N. & Shillcock, R. Serial and parallel processing in reading: Investigating the effects of parafoveal orthographic information on nonisolated word recognition. *The Quarterly Journal of Experimental Psychology,* **66,** 417-428 (2013).

[5] Snell, J., Vitu, F. & Grainger, J. Spatial integration of parafoveal orthographic information: Beyond the sub-lexical level? *Quarterly Journal of Experimental Psychology,* **70,** 1984–1994 (2017a).

[6] Grainger, J., Mathôt, S. & Vitu, F. Test of a model of multi-word reading: Effects of parafoveal flanking letters on foveal word recognition. *Acta Psychologica,* **146,** 35-40 (2014).

[7] Reichle, E., Liversedge, S., Pollatsek, A. & Rayner, K. Encoding multiple words simultaneously in reading is implausible. *Trends in Cognitive Science,* **13,** 115-119 (2009).

[8] Snell, J., Meeter, M. & Grainger, J. Evidence for simultaneous syntactic processing of multiple words during reading*. PLoS ONE,* **12,** (2017b). e0173720. doi:10.1371/journal.pone.0173720.

[9] Snell, J., Declerck, M. & Grainger, J. Semantic processing in the parafovea revisited: Effects of translation equivalents in bilingual readers. *Language, Cognition & Neuroscience,* (in press).

[10] Engbert, R., Nuthmann, A., Richter, M. & Kliegl, R. SWIFT: A dynamical model of saccade generation during reading. *Psychological Review,* **112,** 777-813 (2005).

[11] Kennedy A. & Pynte, J. Parafoveal-on-foveal effects in normal reading. *Vision Research*, **45,** 153-168 (2005).

[12] Reilly, R. & Radach, R. Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research,* **7,** 34-55 (2006).

[13] Engbert, R. & Kliegl, R. Parallel graded attention models of reading in *The Oxford Handbook of Eye Movements* (eds. Liversedge, S., Gilchrist, I. & Everling, S.) 787-800 (Oxford University Press, 2011).

[14] Loewenfeld, I. Mechanisms of reflex dilatation of the pupil. *Documenta Ophthalmologica*, **12,** 185-448 (1958). doi:10.1007/BF00913471

[15] Naber, M., Frassle, S. & Einhauser, W. Perceptual rivalry: Reflexes reveal the gradual nature of visual awareness. *PLoS ONE*, **6,** (2011). e20910. doi:10.1371/journal.pone.0020910

[16] Binda, P., Pereverzeva, M. & Murray, S. Pupil constrictions to photographs of the sun. *Journal of Vision*, **13,** (2013). doi:10.1167/13.6.8

[17] Laeng, B. & Sulutvedt, U. The eye pupil adjusts to imaginary light. *Psychological Science*, **25,** 188-197 (2014).

[18] Mathôt, S., Strijkers, K. & Grainger, J. Pupillary responses to words that convey a sense of brightness or darkness. *Psychological Science,* **28,** 1116-1124 (2017).

[19] Mathôt, S. & van der Stigchel, S. New light on the mind's eye: The pupillary light response as active vision. *Current Directions in Psychological Science,* **24,** 374-378 (2015).

[20] Mathôt, S., van der Linden, L., Grainger, J. & Vitu, F. The pupillary response to light reflects the focus of covert visual attention. *PLoS ONE*, **8,** (2013). e78168. doi:10.1371/journal.pone.0078168

[21] Mathôt, S., Melmi, J., van der Linden, L. & van der Stigchel, S. The mind-writing pupil: A human-computer interface based on decoding of attention through pupillometry. *PLoS ONE,* **11,** (2016). e0148805. doi:10.1371/journal.pone.0148805

[22] Rayner, K. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin,* **124,** 372-422 (1998).

[23] Van Heuven, W., Mandera, P., Keuleers, E. & Brysbaert, M. SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology,* **6**, 1176-1190 (2014).

[24] Ferrand, L., et al. The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods,* **42,** 488-496 (2010).

[25] Mathôt, S., Schreij, D. & Theeuwes, J. OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods,* **44,** 314-324 (2012).

[26] Baayen, R. *Analyzing Linguistic Data: A pratical introduction to statistics*. Cambridge: Cambridge University Press (2008).

[27] Barr D., Levy R., Scheepers C. & Tily H. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, **68,** 255-278 (2013).

[28] Bates, D., Maechler, M., Bolker, B. & Walker, S., Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, **67,** 1-48 (2015).

[29] Dalmaijer, E., Mathôt, S. & Van der Stigchel, S. PyGaze: An open-source, crossplatform toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research Methods,* **46,** 913-921 (2014).

[30] Snell, J. & Grainger, J. The sentence superiority effect revisited. *Cognition,* **168,** 217-221 (2017).

Parallel word processing in the flanker paradigm has a rightward bias

Joshua Snell & Jonathan Grainger

Citation: Snell, J. & Grainger, J. (2018). Parallel word processing in the flanker paradigm has a rightward bias. *Attention, Perception & Psychophysics,* doi: 10.3758/s13414-018-1547-2.

Abstract

Reading research is exhibiting growing interest in employing variants of the flanker paradigm to address several questions about reading. The paradigm is particularly suited for investigating parallel word processing, parafoveal-on-foveal influences and visuo-spatial attention in a simple but constrained setting. However, this methodological deviation from natural reading warrants careful assessment of the extent to which cognitive processes underlying reading operate similarly in these respective settings. The present paper investigated whether readers' distribution of attention in the flanker paradigm resembles that observed during sentence reading; that is, with a rightward bias. Participants made lexical decisions about foveal target words while we manipulated parafoveal flanking words. In line with prior research we established a parafoveal-on-foveal repetition effect, and this effect was increased for rightward flankers compared to leftward flankers. In a second experiment we found that, compared to a no-flanker condition, rightward repetition flankers facilitated target processing, while leftward flankers interfered. Additionally, the repetition effect was larger for right- than leftward flankers. From these findings we infer that attention in the flanker paradigm is indeed biased toward the right, and that the flanker paradigm thus provides an effective analogy to natural reading for investigating the role of visuo-spatial attention. The enhanced parafoveal-on-foveal effects within the attended region further underline the key role of attention in the spatial integration of orthographic information. Lastly, we conclude that future research employing the flanker paradigm should take the asymmetrical aspect of the attentional deployment into account.

1. Introduction

Since decades, cognitive psychology has employed the flanker paradigm (pioneered as the Eriksen flanker task by Eriksen and Eriksen (1974)) to investigate a multitude of things, such as attentional control, parallel stimulus processing and response conflict (e.g., Eriksen & Eriksen, 1974; Schaffer & Laberge, 1979; Gratton, Coles & Donchin, 1992; Lamers & Roelofs, 2011). The principal finding is that responses to target stimuli are influenced by task-irrelevant surrounding stimuli (e.g., a *left* response to a target arrow pointing left, is slowed when this target is flanked by arrows pointing right), hence revealing human's inability to effectively focus processing resources on the stimulus of interest without attending irrelevant stimuli. Classically, those so-called 'noise stimuli' were thought to impact on the stage of decision-making, with response conflict stemming from different stimuli relating to opposing responses (Eriksen & Eriksen, 1974). However, in a variant of the flanker paradigm wherein participants had to respond to the semantic category of target words, Schaffer and Laberge (1979) found that the semantic congruency rather than the response congruency of flankers influenced task performance (i.e., flanking words belonging to a different semantic category but to the same response nonetheless increased response times). This result indicates that information may be integrated across words at the lexical or semantic level.

Now, more than three decades later, reading research is exhibiting renewed interest in the flanker paradigm, driven by the ongoing debate on whether words are processed serially (one-by-one) or in parallel during reading (Engbert, Nuthmann, Richter & Kliegl, 2005; Reichle, Pollatsek & Rayner, 2006; Reilly & Radach, 2006; Reichle, Liversedge, Pollatsek & Rayner, 2009; Engbert & Kliegl, 2011; Dare & Shillcock, 2013; Snell, Vitu & Grainger, 2017a; Snell, Meeter & Grainger, 2017b; Snell, Declerck & Grainger, 2018a). It should be noted that classic implementations of the flanker paradigm did not provide evidence for parallel processing. Following a serial processing rationale, participants could process the flankers after processing of the target but before the final decision (given that flankers and target stayed onscreen until a response was given), thus leaving the flankers ample time to influence the response. However, more recent implementations trade on the use of very brief (150–170 ms) stimulus presentation times. Given that it takes around 150–250 ms to recognize a single word, and that the flankers will logically have disappeared once processing of the target is completed, here it can be argued that the flankers must be processed *during*, rather than *after* target processing, if they are to have any influence on the response.

This is precisely what Dare and Shillcock (2013) demonstrated: using a 150 ms stimulus presentation time, lexical decisions about central target stimuli were influenced by the orthographic relatedness of adjacent letters, such that '*rock*' was recognized faster in '*ro rock ck*' than in '*st rock ep*' (see Grainger, Mathôt, & Vitu, 2014; Snell et al., 2017a; Snell, Bertrand, Meeter & Grainger, 2018b, for similar findings). These so-called parafoveal-on-foveal influences have also been established at higher cognitive levels. Snell et al. (2017b) found that the syntactic categorization of target words was influenced by the syntactic congruency of flanking words, and similarly, that the semantic categorization of target words was influenced by the semantic relatedness of flanking words (2018a). Further, Declerck, Snell and Grainger (2017) found that lexical decisions about target words were made faster when the targets were flanked by unrelated words of the same language, compared to unrelated words of a different language.

Point of controversy is the fact that similar higher-order parafoveal-on-foveal influences have been elusive in more natural reading settings, such as sentence reading. While some studies have reported semantic parafoveal-on-foveal effects in Chinese (Yan, Richter, Shu & Kliegl, 2009;

Yan & Sommer, 2015), similar effects were not found in Roman-alphabetic languages (Angele, Tran & Rayner, 2013; Snell et al., 2017b, 2018), save for one study that did not control for the orthographic overlap between targets and adjacent words (Inhoff, Radach, Starr & Greenberg, 2000).[24]

This discrepancy between the flanker paradigm and sentence reading is likely caused by the different natures of the measures of interest in each respective setting. Sentence reading research is concerned with word viewing times and eye movements (as observed with eye-tracking apparatus), which are thought to provide more direct measures of word recognition speed (e.g. Rayner, 1998). In contrast, a syntactic or semantic categorization decision, as measured in the flanker paradigm, likely takes place post-lexically (Snell et al., 2017b, 2018a). The fact that higher-order cross-word influences seem to occur exclusively at this stage, points to the possibility that words are truly recognized in parallel—i.e., without influencing one another at earlier levels of semantic or syntactic processing. Independently recognized words would then move on to jointly influence responses in the flanker paradigm, hence explaining the effects in RTs but absence of effects in word viewing times.

This reconciliation of contrasting behavioral outcomes in respectively the flanker paradigm and sentence reading, posits that these settings may be describing two sides of the same theoretical coin. Whereas sentence reading research provides a good description of natural reading behavior, the flanker paradigm is suited to assess the reading system at a more fundamental level, hence revealing what the system is in principle capable of (e.g., parallel word processing).

Nonetheless, it would be sensible to assume that behavioral outcomes may not only differ due to the fact that the respective measures of interest (word viewing times versus lexical decision times) represent different things, but also due to the fact that the reader approaches each respective task with different objectives. Whereas the primary goal when reading normal text is to achieve context comprehension, a lexical decision task is likely to induce a strategy that relies more on bottom-up processing of the visual input and less on top-down expectations (e.g., word identities being constrained by preceding context). Furthermore, sentence reading demands an efficient coupling of lexical processing and oculomotor control (e.g., Rayner, 1998), whereas a lexical decision task does not. It remains to be seen to which extent such factors dictate how other components of the reading system operate, and indeed, whether the interplay of cognitive processes such as visual processing, working memory and attention, is fundamentally different across tasks. In short, methodological differences between the flanker paradigm and sentence reading necessitate clarification of the extent to which cognitive processes underlying reading operate similarly in these respective settings.

1.1 Investigating attention in the flanker paradigm

Considering the above, it is noteworthy that parafoveal-on-foveal effects of orthographic relatedness are observable in the flanker paradigm and sentence reading alike. In sentence reading, word viewing times are shorter when followed by an orthographically related word, compared to an unrelated word (Dare & Shillcock, 2013; Angele et al., 2013; Snell et al., 2017a).

[24] Note that this is not to say that word viewing times are not affected by post-lexical processes (e.g. context comprehension) during normal reading. Rather, we specifically note that eye-movement measures relating to word *n* (i.e., fixation durations on *n*, skipping rate of *n*, the number of refixations on *n* and regressions to *n*) are largely unaffected by higher-order (semantic, syntactic) aspects of word *n*+1.

One recent implementation of the flanker paradigm points to a widespread distribution of visuo-spatial attention as the key factor driving orthographic parafoveal-on-foveal effects. In their study, Snell, Mathôt, Mirault and Grainger (2018b) manipulated the brightness of flanker locations, making use of the principle that the pupil size responds to the brightness of covertly attended (i.e., without looking) locations in the visual field (e.g., Mathôt & van der Stigchel, 2015). They found that target processing was influenced by flankers presented left and right of the target, but not by flankers presented above and below the target. Perfectly in line with this pattern, the pupil size was contingent with the brightness of horizontally but not vertically aligned flankers. This suggested that attentional resources were spent on the flankers involved in the parafoveal-on-foveal effect, hence evidencing a key role of attention. The widespread distribution of attention is in line with how parallel processing models (e.g. Engbert et al., 2005; Snell et al., 2017b) conceptualize attention in sentence reading, and further illustrates the flanker paradigm's merit in addressing questions about reading.

To test whether the flanker paradigm indeed provides an effective analogy to natural reading with respect to the attentional distribution, here we reason the other way around: given that attention during sentence reading is biased toward the right in scripts that read from left to right (e.g., McConkie & Rayner, 1976; Rayner, 1998), there should similarly be a stronger influence from right- than leftward flankers. This is not an obvious prediction: in classic variants of the flanker paradigm using single letter targets and flankers, a leftward bias has been observed (Harms & Bundesen, 1983; Hommel, 1995, 2003). However, Hommel (2003) showed that when flanker letters are mirrored, the bias is mirrored as well. This indicates that the attentional distribution is heavily dependent on the nature of the stimuli. In this light, a rightward attentional bias in a flanker task using linguistic stimuli could suggest that the reading system is triggered to engage in 'real reading'. Preliminary evidence for such a rightward bias was provided by Snell, Bertrand and Grainger (2018c), who found that RTs in a lexical decision task decreased when the bigram frequency of rightward flankers increased, while the bigram frequency of leftward flankers had no influence.

Experiment 1 provides a more complete test of the hypothesis that attention should be biased toward the right in the flanker paradigm. We measure lexical decision times for foveal targets as a function of seven flanker conditions that allow both for a replication of previously reported parafoveal-on-foveal effects, as well as an assessment of location-specific (left/right) flanker influences. All data reported in this paper are openly accessible on https://osf.io/dbjvh/.

2. Experiment 1

2.1 Methods

35 students (F = 27) from Aix-Marseille University, ranging in age between 18–24 years old, gave informed consent to participating in this study for 5 euros. All participants reported to be non-dyslexic, native to the French language, and having normal or corrected-to-normal vision.

From the French Lexicon Project database (Ferrand et al., 2010) we retrieved 60 4-letter target words that contained no diacritics. For each target we retrieved an orthographically unrelated word (to be used as unrelated flanker) with the same length and frequency. In a similar fashion we retrieved 60 non-word targets and unrelated non-word flankers from the Pseudo-word Lexicon (Ferrand et al., 2010).

Each target was presented across seven conditions, visualized in Table 1. Each participant saw all 60 targets in all conditions, meaning that there were 420 word target trials per participant. For non-word targets (which were included solely to induce the lexical decision task, and data for which are thus not reported) we implemented the same flanker conditions, thus bringing the total amount of trials per participant to 840. These were presented in random order.

Table 1. Experiment 1 conditions.

Condition	Left flanker	Target	Right flanker
No flankers		rock	
Repetition flankers	rock	rock	rock
Unrelated flankers	step	rock	step
Left repetition flanker	rock	rock	
Right repetition flanker		rock	rock
Mixed flankers, left rep.	rock	rock	step
Mixed flankers, right rep.	step	rock	rock

Participants were seated in a comfortable chair in a dimly-lit room. The experiment, implemented with OpenSesame (Mathôt, Schreij & Theeuwes, 2012), was shown on a 17-inch, 1024x768 pixel, 150 Hz display, at such a distance from the participant that each character space in the stimulus subtended 0.3º of visual angle. Participants were given task instructions both by the experimenter as well as visually onscreen.

The trial procedure is shown in Figure 1. Each trial started with a 500 ms fixation display containing centrally positioned vertical fixation bars, separated from one another by 0.60º of visual angle. A target stimulus (with/without flankers being presented on the left/right, depending on the condition as specified in Table 1) was shown for 150 ms in between the fixation bars, after which participants had a maximum of 2000 ms to respond. Flankers were separated from the target by a single character space. Responses were given by means of a joystick trigger-button press with the right or left index finger, to indicate *word* or *non-word* respectively. Providing response feedback, a green or red fixation dot was then shown at the center of the screen for 500 ms, for correct and incorrect responses respectively (non-responses were also counted as incorrect). The display then returned to the beginning state, 500 ms after which a new trial would commence. The experiment lasted approximately 35 minutes. The participants were offered a break after every 210 trials.

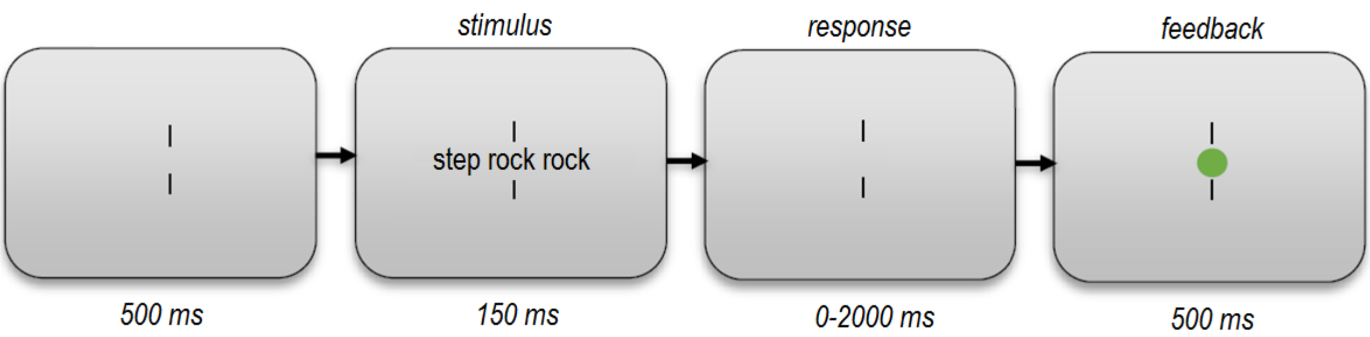

Figure 1. Trial procedure. The size of stimuli relative to the screen is exaggerated in this example.

2.2 Results

The analysis of RTs excluded incorrectly answered trials (6.69%). Additionally, both the analysis of RTs and the analysis of error rates excluded trials with an RT beyond 2.5 standard deviations from the grand mean (2.71%).

Data were analyzed using linear mixed-effect models (LMMs) with items and participants as crossed random effects (Baayen, Davidson & Bates, 2008). The maximal random effects structure that converged, was one including the by-subject random slope alongside by-participant and by-item random intercepts. However, a likelihood-ratio test pointed out that this model differed non-significantly from a model including only random intercepts. Following the recommendation of Baayen et al. (2008) that models not be over-parameterized, we therefore did not include random slopes in the model. We report *b*-values, standard errors (SEs) and *t*-values (RTs) or *z*-values (errors), with *t*- and *z*-values beyond |1.96| deemed significant.[25]

Mean RTs per condition are presented in Figure 2. We replicated previous observations of parafoveal-foveal integration, as the unrelated flanker condition yielded significantly longer RTs than did all other conditions (Table 2). Our hypothesis that rightward flankers should have a stronger influence was confirmed: participants benefitted more from a repetition flanker on the right (the 5th and 7th column in Figure 2) than on the left (the 4th and 6th column). Between the mixed flanker conditions, a rightward repetition flanker yielded shorter RTs than a leftward repetition flanker, with b = -17.72, SE = 3.13, t = -7.40.[26] When using a unilateral repetition flanker (the 4th and 5th column), RTs were again shorter for flankers presented on the right than on the left, with b = -24.66, SE = 3.05, t = -8.08.

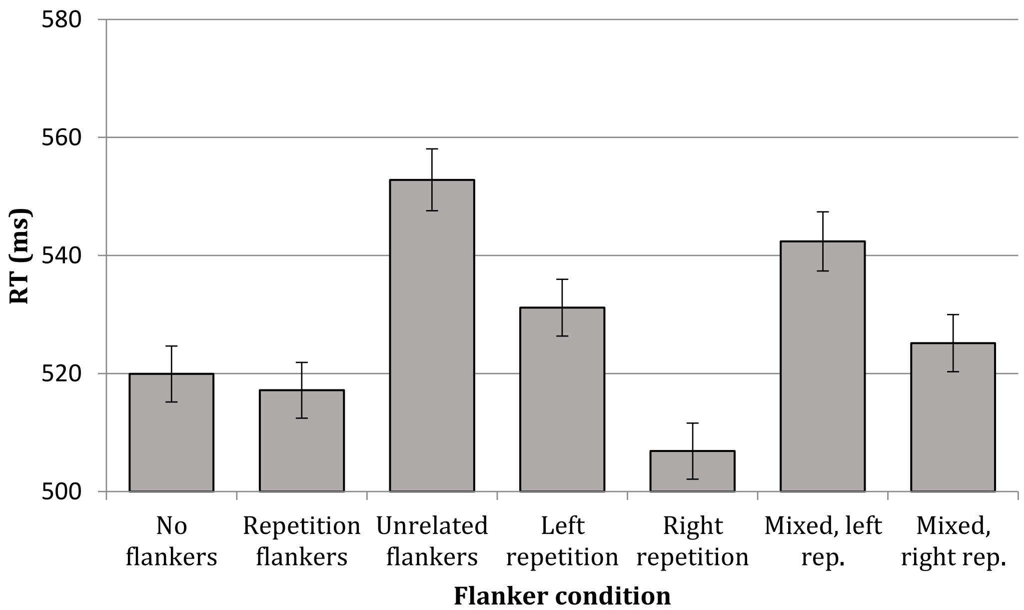

Figure 2. Mean RTs per condition in Experiment 1. Error bars depict 95% confidence intervals.

[25] While the statistical analyses reported below were performed on normal RTs, some have argued that log-transformed RTs would better fit the distributional assumptions of the model (e.g., Kliegl, Wei, Dambacher, Yan & Zhou, 2011). Following advice by a reviewer of this work, we have performed the same analyses on log-transformed RTs and established virtually equal significance values across the two datasets.

[26] Note that these values were obtained by choosing the leftward repetition flanker condition as the reference in the original model. Values reported elsewhere in the text were similarly obtained by re-referencing the model.

Table 2. Analyses of RTs and errors in Experiment 1, with unrelated flankers as reference. Significant values are shown in bold.

	RTs			***Errors***		
Condition	*b*	SE	*t*	*b*	*SE*	*z*
(Intercept)	554.27	10.29	**39.39**	2.89	0.17	**17.28**
No flankers	-32.72	3.10	**-10.57**	0.12	0.12	0.96
Repetition flankers	-35.69	3.08	**-11.60**	-0.50	0.13	**-3.72**
Left repetition	-22.11	3.10	**-7.13**	-0.03	0.12	-0.21
Right repetition	-46.68	3.07	**-15.19**	-0.54	0.14	**-3.98**
Mixed flankers, left rep.	-9.38	3.12	**-3.01**	0.15	0.11	1.23
Mixed flankers, right rep.	-27.24	3.09	**-8.81**	-0.27	0.13	**-2.14**

Strikingly, compared to the condition without flankers, a leftward repetition flanker yielded longer RTs (b = 10.70, SE = 3.08, t = 3.48) while a rightward repetition flanker yielded shorter RTs (b = -13.96, SE = 3.05, t = -4.58). This indicates that the recognition process may be truly facilitated by information on the right, while information on the left generally interferes. This would then also explain why the repetition flanker condition, which is basically a combination of the two unilateral flanker conditions, sits neatly in between these conditions. Indeed, no difference was observed between the repetition flanker condition and the condition without flankers (b = 2.97, SE = 3.07, t = 0.97), suggesting that the flanker on the right compensated for the flanker on the left.

It should be noted that while rightward parafoveal-foveal integration is clearly enhanced, there is nonetheless leftward integration, as the leftward repetition in the mixed flanker condition led to shorter RTs than the unrelated condition: b = -9.38, SE = 3.12, t = 3.01.

Further illustration of the stronger rightward influence is the fact that compared to the unrelated flanker condition, errors were significantly reduced if a repetition flanker was presented on the right, regardless of what was presented on the left (Table 2).

2.3 Discussion

Experiment 1 provides clear evidence for a rightward attentional bias in a flanker paradigm using linguistic stimuli. While readers benefitted from a repetition flanker on the left compared to an unrelated flanker (evidenced by contrasting the mixed flanker condition to the unrelated flanker condition), this difference was clearly enhanced for rightward flankers.

Compared to a condition without flankers, a unilateral repetition flanker further led to better performance when being presented on the right, while its presentation on the left led to worse performance. This particular finding is quite striking, as it illustrates that parafoveal related information can truly facilitate foveal word processing—on the premise that the related information is located in the favored hemifield. Below, we report a second experiment with which we intended to replicate this finding. Experiment 2 tested the two unilateral repetition flanker conditions and the no-flanker condition, alongside two conditions with unilateral unrelated flankers.

3. Experiment 2

3.1 Methods

Of the 35 participants that were recruited for Experiment 2 (F = 32, age range 18—29), 10 had also participated in Experiment 1.[27] Aside from testing a different set of conditions (presented in Table 3 below), the entire methodology for Experiment 2 was unchanged from that of Experiment 1. Including non-word targets, Experiment 2 comprised 600 trials, spanning approximately 25 min of testing time per participant.

Table 3. Experiment 2 conditions.

Condition	Left flanker	Target	Right flanker
No flankers		rock	
Left repetition flanker	rock	rock	
Right repetition flanker		rock	rock
Left unrelated flanker	step	rock	
Right unrelated flanker		rock	step

3.2 Results and discussion

All criteria for data selection and analysis were equal to those of Experiment 1. 7.41% of trials were excluded from the analysis of RTs due to either an incorrect response or an RT beyond 2.5SD from the grand mean. The latter criterion led to the exclusion of 2.44% of trials from the error rate analysis.

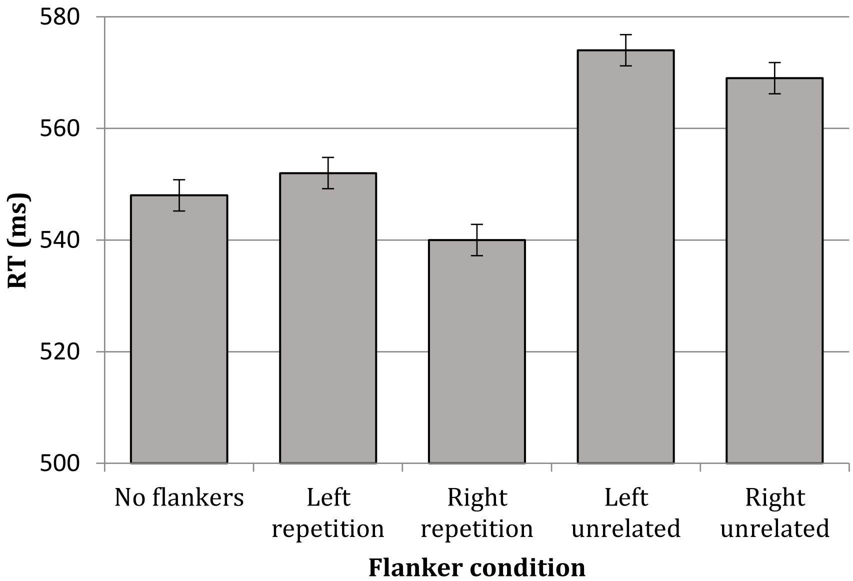

Figure 3. Mean RTs per condition in Experiment 2. Error bars depict 95% confidence intervals.

[27] A model including subjects' prior experience (*did* or *did not participate in Experiment 1*) as factor pointed out that subjects' prior experience did not modulate the pattern of effects reported below. There was no main effect of prior experience (b = -28.88, SE = 23.02, t = -1.25), no interaction with flanker relatedness (b = -1.70, SE = 6.53, t = -0.26) or flanker laterality (b = -0.36, SE = 6.45, t = -0.06) and no three-way interaction of these three factors (b = 4.20, SE = 9.19, t = 0.46).

Table 4. Analyses of RTs and errors for Experiment 2, with the no-flanker condition as reference. Significant values are shown in bold.

	RTs			***Errors***		
Condition	*b*	SE	*t*	*b*	*SE*	*z*
(Intercept)	553.34	10.80	**37.36**	3.13	0.17	**18.35**
Left repetition flanker	3.23	2.96	1.09	0.06	0.13	0.45
Right repetition flanker	-9.89	2.93	**-3.38**	-0.56	0.15	**-3.68**
Left unrelated flanker	24.51	2.96	**8.29**	0.05	0.13	0.37
Right unrelated flanker	20.23	2.96	**6.85**	0.02	0.13	0.18

As can be seen in Figure 3, we replicated effects observed in Experiment 1, with a rightward repetition flanker leading to shorter RTs than the no-flanker condition. However, whereas the leftward repetition flanker yielded longer RTs than the no-flanker condition in Experiment 1, we observed no significant difference between these conditions this time around (Table 4). That the presence of a leftward flanker induced a smaller cost in Experiment 2 compared to Experiment 1, is possibly caused by the fact that a greater portion of trials presented flankers exclusively on the left in Experiment 2 compared to Experiment 1 (i.e., 40% of trials in Experiment 2, versus 14.29% of trials in Experiment 1), which may have led participants to attend the left visual hemifield to a greater degree. Nonetheless, Experiment 2 clearly replicated the observation of a processing asymmetry indicative of a rightward attentional bias.

Noteworthy is that the attentional bias does not only lead to overall decreased RTs with rightward flankers compared to leftward flankers, but also enhances the parafoveal-on-foveal effect of orthographic relatedness for information in the attended region. To support this notion, we analyzed the four unilateral flanker conditions in a separate model that included flanker relatedness (*repetition, unrelated*) and flanker laterality (*left, right*) as factors. In addition to the main effects of relatedness (b = -21.20, SE = 2.97, t = -7.15) and laterality (b = -13.18, SE = 2.94, t = -4.49), we indeed also established an interaction between these factors: b = 8.97, SE = 4.17, t = 2.15, such that the effect of orthographic relatedness was increased for rightward flankers. This further supports the claim of Snell et al. (2018b) that the spatial integration of orthographic information is driven by attention.

The error rate was again only influenced by the relatedness of rightward flankers (Table 4). Indeed, the interaction between relatedness and laterality that was established for RTs, was also established for error rates: b = 0.59, SE = 0.20, z = 2.98, such that the effect of relatedness on the error rate was greater for rightward flankers than for leftward flankers.

4. General discussion

Across two experiments, we have observed a strong rightward processing bias in a flanker paradigm using linguistic stimuli. Flanker relatedness being equal, leftward flankers consistently led to longer RTs than rightward flankers. As Snell et al. (2018b) have shown that spatial orthographic integration effects are driven by attention, we therefore conclude that attention in a flanker paradigm with word stimuli is biased toward the right. This processing asymmetry contradicts the leftward bias observed in classic implementations of the Eriksen flanker task (e.g.

Harms & Bundesen, 1983; Hommel, 1995, 2003) and instead parallels the rightward bias observed in natural (sentence) reading (McConkie & Rayner, 1976; Rayner, 1998).

Evidenced by both Experiment 1 and 2, word processing is facilitated by the presence of the same word when the latter is located within the attended region of the visual field. This benefit is negated (Experiment 2) or even turns into a cost (Experiment 1) when the repetition is presented outside the locus of attention. We reason that a leftward shift of attention, triggered by the leftward stimulus, shifted some processing resources away from the target, hence leading to longer RTs.

This implies that parafoveal words can impact on foveal word processing in two ways: first, information is integrated across words, such that related parafoveal information leads to faster foveal word recognition compared to unrelated parafoveal information (a pattern observed for both left- and rightward flankers). Building on the study of Snell et al. (2018b), which established that attention is a key factor driving the spatial integration of orthographic information, here we observed that an attentional bias enhances integration effects at the attended region (Section 3.2).

The second way in which parafoveal words can impact on foveal word processing, is purely attentional: when presented outside the attended region of the visual field, exogenous attentional capture by the flanker shifts some processing resources away from the target. Following this logic, if a flanker is presented within the attended region, then the spatial integration of information should be the only factor determining the flanker's impact on target processing; hence explaining the facilitatory influence of rightward related flankers. In contrast, a leftward flanker induces a shift of attention, and therefore less efficient target processing—although this may be compensated to some extent by the spatial integration process.

The above scenario generates predictions that may be addressed in future research. If attention is biased toward the left *prior* to the onset of the target and flankers—for instance by a directional/spatial cue (e.g., Posner, 1980)—then the pattern of effects observed here should be mirrored. We thus predict facilitation of repetition flankers presented at the cued location, and interference from (both repetition- and unrelated) flankers presented opposite to the cued location. The effect of parafoveal-foveal relatedness should further be enhanced at the cued side.

It must be acknowledged that the rightward bias of attention may not be the only factor driving the observed processing asymmetry. Prior research has shown that the recognition of isolated words is better in the right- than in the left parafovea (e.g., Ducrot & Grainger, 2007), and one factor that likely contributes here is the fact that language is typically processed in the left hemisphere. As argued by Brysbaert (2004), asymmetries in visual language processing may stem from the fact that information in the right visual hemifield is processed by a shorter neural pathway (i.e., from the left visual cortex to the left-hemispheric language centers) than information in the left visual hemifield, the neural pathway for which additionally comprises interhemispheric transfer through the corpus callosum. However, Ducrot and Grainger (2007) did find that spatial cues affected the rightward processing bias, thus suggesting that attention nonetheless plays a role as well. To pinpoint the respective contributions of the attentional bias and hemispheric asymmetry to the observed processing asymmetry, future research may compare conditions with unilateral unrelated flankers to conditions with unilateral non-linguistic flankers (e.g. mask flankers: '####').

The present results have three implications for reading research. First, it is apparent that readers' distribution of attention can be quite typical of natural reading even outside a natural reading setting. This advances the conception that the flanker paradigm is suited to test aspects of the reading process—at least when pertaining to attention—in a simple and controlled setting that does not necessitate the use of eye-tracking apparatus.

Secondly, the present results generate certain predictions for the realm of sentence reading. In specific, these results predict that orthographic parafoveal-on-foveal effects from word *n*-1 on *n* should be observable in sentence reading, even if influences from word *n*+1 should nonetheless be stronger.

Thirdly, particular care is warranted when employing the flanker paradigm to test within-word letter processing. For instance, Snell et al. (2018c) turned the principle that information is integrated across words into an asset for investigating how letter position is encoded (using various configurations of the target's letters as flankers to see how these affect the recognition process; see also Grainger et al., 2014). In such an experiment, flanker letters may be expected to yield different outcomes depending on laterality.

At the same time, taking the processing asymmetry into account can lead one to gain additional information. As an illustration, consider the studies of Dare and Shillcock (2013) and Grainger et al. (2014), both of which found that word processing was facilitated not only by related bigram flankers presented on the correct side of the target (e.g., '*ro rock ck*'), but equally as much by switched bigram flankers ('*ck rock ro*'). Knowing that the parafoveal-on-foveal effect will largely have been effectuated by different letters in the former condition ('*ck*') compared to the latter condition ('*ro*'), one may conclude that within words, beginning and ending letters bear equal importance. Note that one could not have concluded this without knowing about the processing asymmetry: if flanker processing were symmetrical, then the summed facilitation of '*ro*' and '*ck*' would always be equal to the summed facilitation of '*ck*' and '*ro*', even if '*ro*' bore more weight than '*ck*' (forasmuch as 5+1 equals 1+5), implying that one could not retrieve each bigram's respective importance in this case.[28]

In sum, having revealed a rightward attentional bias in the flanker paradigm, the present study attests to the conception that the paradigm provides an effective analogy to natural reading. Future research employing the flanker paradigm should take this asymmetry into account.

Acknowledgements

This research was funded by the Brain & Language Research Institute (ANR-11-LABX-0036) and the European Research Council (ERC742141). We thank Cécile Mahnich for her help in conducting the experiments.

References

Angele, B., Tran, R. & Rayner, K. (2013). Parafoveal–foveal overlap can facilitate ongoing word identification during reading: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 526-538.

Baayen, R., Davidson, D. & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390-412.

[28] Note that the studies of Grainger et al. (2014) and Dare and Shillcock (2013) are only used here for illustrational purposes; after all, it is possible that participants did not have a rightward attentional bias in these studies, because they saw two-letter flankers rather than whole-word flankers (and were thus possibly not as much engaged in 'real reading').

Brysbaert, M. (2004). The importance of interhemispheric transfer for foveal vision: A factor that has been overlooked in theories of visual word recognition and object perception. *Brain & Language*, *88,* 259-267.

Dare, N. & Shillcock, R. (2013). Serial and parallel processing in reading: Investigating the effects of parafoveal orthographic information on nonisolated word recognition. *The Quarterly Journal of Experimental Psychology, 66*, 417-428.

Declerck, M., Snell, J. & Grainger, J. (2017). On the role of language membership information during word recognition in bilinguals: Evidence from flanker-language congruency effects. *Psychonomic Bulletin & Review,* doi: 10.3758/s13423-017-1374-9.

Ducrot, S. & Grainger, J. (2007). Deployment of spatial attention to words in central and peripheral vision. *Perception & Psychophysics, 69,* 578-590.

Engbert, R. & Kliegl, R. (2011). *Parallel graded attention models of reading.* In Liversedge, S., Gilchrist, I. & Everling, S. (Eds.), The Oxford Handbook of Eye Movements (pp. 787-800). Oxford, England: Oxford University Press.

Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). Swift: A dynamic model of saccade generation during reading. *Psychological Review, 112,* 777-813.

Eriksen, B. & Eriksen, C. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16,* 143-149.

Ferrand, L., New, B., Brysbaert,M., Keuleers, E., Bonin, P.,Méot, A., Augustinova, M. & Pallier, C. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods, 42,* 488-496.

Grainger, J., Mathôt, S. & Vitu, F. (2014). Tests of a model of multi-word reading: Effects of parafoveal flanking letters on foveal word recognition. *Acta Psychologica*, *146,* 35-40.

Gratton, G., Coles, M. & Dunchin, E. Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General, 121,* 480-506.

Harms, J. & Bundesen, C. (1983). Color segregation and selective attention in a nonsearch task. *Perception & Psychophysics, 33,* 11-19.

Hommel, B. (1995). Attentional scanning in the selection of central targets from multi-symbol strings. *Visual Cognition, 2,* 119-144.

Hommel, B. (2003). Spatial asymmetries in the flanker-congruency effect: Attentional scanning is biased by flanker orientation. *Psychological Science, 45,* 63-77.

Inhoff A., Radach R., Starr M. & Greenberg S. (2000). Allocation of visuospatial attention and saccade programming during reading. In Kennedy A., Radach R., Heller D. & Pynte J. (Eds.). *Reading as a perceptual process.* Oxford, UK: Elsevier.

Lamers, M. & Roelofs, A. (2011). Attentional control adjustments in Eriksen and Stroop task performance can be independent of response conflict. *The Quarterly Journal of Experimental Psychology, 64,* 1056-1081.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An opensource, graphical experiment builder for the social sciences. *Behavior Research Methods, 44,* 314-324.

McConkie, G. & Rayner, R. (1976). Asymmetry of the perceptual span in reading. *Bulletin of the Psychonomic Society, 8,* 365-368.

Posner, M. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32,* 3-25.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372-422.

Reichle, E., Liversedge, S., Pollatsek, A. & Rayner, K. (2009). Encoding multiple words simultaneously in reading is implausible. *Trends in Cognitive Sciences, 13,* 115-119.

Reichle, E., Pollatsek, A., & Rayner, K. (2006). E-Z reader: A cognitive-control, serial-attention model of eye movement behavior during reading. *Cognitive Systems Research, 7,* 4-22.

Reilly, R. & Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research, 7,* 34-55.

Schaffer, W., & LaBerge, D. (1979). Automatic semantic processing of unattended words. *Journal of Verbal Learning and Verbal Behavior, 18,* 413-426.

Snell, J., Bertrand, D. & Grainger, J. (2018c). Parafoveal letter-position coding in reading. *Memory & Cognition,* doi: 10.3758/s13421-017-0786-0.

Snell, J., Declerck, M. & Grainger, J. (2018a). Parallel semantic processing in reading revisited: effects of translation-equivalents in bilingual readers. *Language, Cognition & Neuroscience*, *33*, 563-574.

Snell, J., Mathôt, S., Mirault, J. & Grainger, J. (2018b). Parallel graded attention in reading: A pupillometric study. *Scientific Reports, 8,* doi: 10.1038/s41598-018-22138-7.

Snell, J., Meeter, M. & Grainger, J. (2017b). Evidence for simultaneous syntactic processing of multiple words in reading. *PLoS ONE, 12,* e0173720.

Snell, J., Vitu, F. & Grainger, J. (2017a). Integration of parafoveal orthographic information during foveal word reading: Beyond the sub-lexical level? *Quarterly Journal of Experimental Psychology*, *70,* 1984-1996.

Yan, M., Richter, E., Shu, H. & Kliegl, R. (2009). Chinese readers extract semantic information from parafoveal words during reading. *Psychonomic Bulletin & Review, 16,* 561-566.

Yan, M. & Sommer, W. (2015). Parafoveal-on-foveal effects of emotional word semantics in reading Chinese sentences: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory & Cognition, 41,* 1237-1243.

Chapter 3: Higher-order integration: Eye movements versus decisions

Integration of parafoveal orthographic information during foveal word reading: Beyond the sub-lexical level?

Joshua Snell, Françoise Vitu & Jonathan Grainger

Citation: Snell, J., Vitu, F. & Grainger, J. (2017). Integration of parafoveal orthographic information during foveal word reading: Beyond the sub-lexical level? *Quarterly Journal of Experimental Psychology*, *70*, 1984-1996.

Abstract

Prior research has shown that processing of a given target word is facilitated by the simultaneous presentation of orthographically related stimuli in the parafovea. Here we investigate the nature of such spatial integration processes by presenting orthographic neighbors of target words in the parafovea, considering that neighbors have been shown to inhibit, rather than facilitate, recognition of target words in foveal masked priming research. In Experiment 1, we used the gaze-contingent boundary paradigm to manipulate the parafoveal information subjects received while they fixated a target word within a sentence. In Experiment 2, we used the Flanking Letters Lexical Decision paradigm to manipulate parafoveal information while subjects read isolated words. Parafoveal words were either a higher-frequency orthographic neighbor of targets words (e.g., *blue-blur*) or a high-frequency unrelated word (e.g., *hand-blur*). We found that parafoveal orthographic neighbors facilitated, rather than inhibited, processing of the target. Thus, the present findings provide further evidence that orthographic information is integrated across multiple words, and suggest that either the integration process does not enable simultaneous access to those words' lexical representations, or that lexical representations activated by spatially distinct stimuli do not compete for recognition.

Introduction

It is apparent that multiple words are visually available to us simultaneously when we read. However, a longstanding question in reading research has been whether we actually process information from multiple words in parallel (e.g. Inhoff, Radach, Starr & Greenberg, 2000; Radach & Kennedy, 2004; Rayner, 1998; Schotter, Angele & Rayner, 2012; Vitu, Brysbaert & Lancelin, 2004). Recently, two studies have reported parafoveal-on-foveal effects of orthographic relatedness in sentence reading (e.g. Angele, Tran & Rayner, 2013; Dare & Shillcock, 2013), indicating that information is gathered and integrated at least to some extent across multiple words in parallel. Specifically, these studies have shown that the fixation duration on the target word (n) in a sentence decreases if that target word is repeated at n+1, as compared to an orthographically unrelated n+1. These two studies also found a facilitatory influence of orthographically related non-word stimuli at n+1, either formed by transposing two letters of the target word (e.g., cheap-cehap) or by changing one letter of the target word (e.g., news-niws). Furthermore, using the novel Flanking Letters Lexical Decision (FLLD) paradigm, Dare and Shillcock (2013) found that when readers have to make lexical decisions on isolated target words and non-words, response times for targets are decreased when they are flanked by their constituent letters (e.g. ro rock ck) as compared to unrelated letters (e.g. di rock sh). Interestingly, this facilitatory parafoveal-on-foveal effect persisted when the left-and right-side letters of the target occurred as right- and left-side flankers respectively (e.g. ck rock ro).

The findings of Dare and Shillcock (2013) prompted Grainger et al. (2014) to propose a novel theoretical framework for orthographic processing spanning multiple words (see also Grainger, Dufau, & Ziegler, 2016). In this model, location-specific letter detectors operate in parallel across multiple words to activate ordered pairs of contiguous and non-contiguous letter combinations (open-bigrams; see e.g. Grainger & Van Heuven 2003), which in turn activate whole-word representations (Figure 1). Grainger et al. (2014) then proceeded to test their model using the FLLD paradigm of Dare and Shillcock (2013). They found that the facilitatory parafoveal-on-foveal effect was reduced when the letters within flanking letter pairs were reversed (e.g. or rock kc). This result is in line with the “bag-of-bigram” approach of their model, according to which letter order within bigrams is important but bigram order is not. The model thus accounts for the findings of Dare and Shillcock (2013) and Angele et al. (2013) in that orthographic information from multiple words is integrated at the sub-lexical orthographic level.

Angele et al. (2013) proposed a similar account of parafoveal-on-foveal orthographic effects, whereby foveal and parafoveal information is assumed to interact on the letter level, while no interaction takes place at the level of whole-word representations. As was argued by Angele et al., such an account can be harmonized within a serial processing framework, such as the E-Z Reader model (Reichle et al., 2006), because actual lexical access would require attention and therefore occur serially, whereas uptake of sub-lexical orthographic information from a parafoveal word could proceed pre-attentively. Indeed, higher-order (e.g., semantic) parafoveal-on-foveal effects have been more controversial. Angele et al. did not find a facilitatory influence of parafoveal words on semantically related foveal targets (e.g. ‘*chair table*’), while Inhoff et al. (2000) did find such a semantic parafoveal-on-foveal effect. It has been argued by Angele et al. that the results of Inhoff et al. may have been caused by increased sub-lexical orthographic relatedness between the foveal target and the parafoveal word (e.g. ‘*mother father*’). Yet, research has also suggested that the establishment of semantic parafoveal effects is dependent on the strength of the relationship between the prime and target stimulus. Using a parafoveal preview paradigm, where a preview word at position n+1 is replaced with a target word once the eyes move from n to n+1, Schotter (2013) found that fixation durations on the target (e.g. '*begin*') decrease after a synonym preview (e.g. '*start*') but not a (weaker) semantically related word (e.g. '*ready*'). It is conceivable, therefore, that the establishment of a semantic parafoveal-on-foveal effect would similarly depend on the strength of the semantic relation. However, the language in which sentences are presented may play an important role as well. The semantic parafoveal-on-foveal effect reported by Inhoff et al. (2000) was based on German sentence reading, and indeed, Hohenstein and Kliegl (2014) have also reported a parafoveal preview benefit with semantically related stimuli in German sentence reading using pairs of words that were not synonyms.

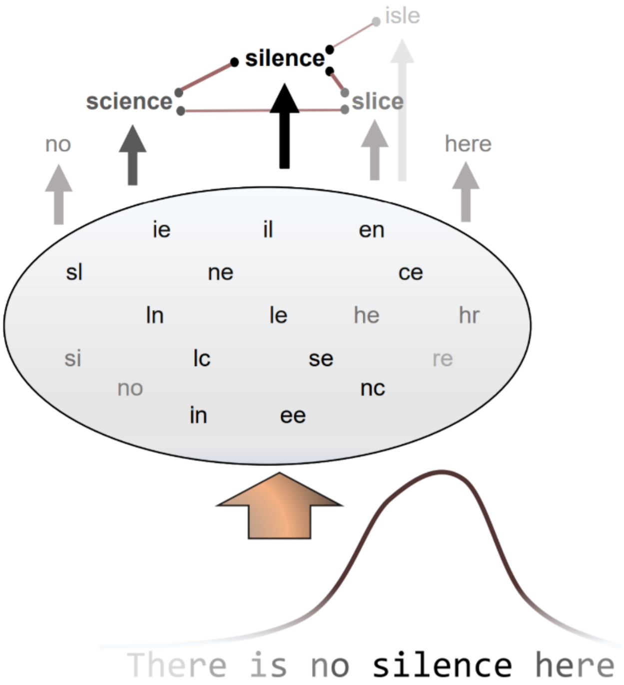

Figure 1. Account of the spatial integration of orthographic information across multiple words during reading, proposed by Grainger et al. (2014). Visual input activates orthographic nodes (sub-lexical, bigram-based), with bigrams in the fovea gaining more activity than bigrams in the parafovea (due to visual acuity and modulated by spatial attention). The attentional gradient (depicted by the curved line) is mainly determined by where the eyes fixate (here on '*silence*'). The orthographic nodes in turn activate word representations. Among activated words, there is competition (i.e. mutual inhibition) so that only one word reaches recognition. The extent to which two words inhibit each other is determined by the amount of orthographic overlap; in the figure, for example, it can be seen that the word nodes for '*no*' and '*here*' do not exert inhibition. Additionally, the more active a word node is, the stronger its inhibitory influence on neighboring words.

Given that reports on semantic parafoveal-on-foveal effects have been ambivalent, in the current article we propose an alternative means to finding signs of higher-order parafoveal-on-foveal information integration. According to the Grainger et al. (2014) model, multiple words may be activated by a given pool of orthographic information, and these words would interact (i.e. compete for recognition) through word-to-word inhibitory connections (see also McClelland & Rumelhart, 1981). In this light it can be argued that parafoveal-on-foveal inhibitory, rather than facilitatory influences, would be indicative of information integration at the level of whole-word representations.

Word-to-word inhibitory effects are traditionally shown through the use of orthographic neighbors, i.e., words that have all but one letter in common with each other, (e.g. rock & sock). Using a masked priming paradigm, Segui and Grainger (1990) found that the recognition of targets was slowed down by higher frequency neighbor primes (see also Davis & Lupker, 2006; De Moor & Brysbaert, 2000; De Moor, Van der Herten & Verguts, 2007). It is conceivable that such neighborhood effects also occur when two words are viewed simultaneously (in the fovea and parafovea), rather than sequentially. However, previous investigations of parafoveal-on-foveal neighborhood effects have not yielded clear-cut results. In their seminal study, Vitu et al. (2004) paired up target words with parafoveal neighbors in a sentence reading-like situation, such that participants had to move their eyes from an x-letter string to the target, to the parafoveal neighbor, to a final x-letter string. Using different indexes of fixation time on the target, Vitu et al. (2004) found that high-frequency parafoveal words tended to facilitate processing of low-frequency targets when the two differed by an outer-positioned letter. When they differed by an inner-positioned letter (e.g. *rock* & *rook*), facilitation only occurred for high-frequency targets with low-frequency parafoveal words. In contrast, no significant effect (but a numerical tendency for inhibition with inner-letter word pairs) was found for low-frequency targets with high-frequency parafoveal word neighbors. It is possible that lexical inhibition did occur, in line with the findings of Segui and Grainger (1990), but that this effect was neutralized by facilitation due to relatedness on the sub-lexical level. Furthermore, as was argued by Vitu et al. (2004), the absence of a clear inhibition effect may have been due to the nature of their task. Participants had to read pairs of semantically unrelated words, which may have encouraged a word-by-word reading strategy where visual attention would be more narrowly distributed as compared to how it may be distributed in everyday reading. In this regard, it may be worthwhile investigating the relationship between foveal and parafoveal neighbors in a more natural sentence reading setting similar to that used in the studies of Angele et al. (2013) and Dare & Shillcock (2013).

It is important to note that orthographic neighbors have been used before in parafoveal processing research, albeit with a parafoveal preview paradigm similar to that used in the studies of Schotter (2013) and Hohenstein and Kliegl (2014) discussed above. Williams, Perea, Pollatsek and Rayner (2006) used the boundary technique (Rayner, 1975) to change a preview word at position n+1 into a target word when the eyes moved from n to n+1, with previews being either identical to the target, a neighbor of the target, or a related non-word. The authors found that fixation durations on low-frequency targets were decreased after identical previews and high-frequency neighbor previews as compared to non-word previews, while there was no difference between identical previews and high-frequency neighbor previews. However, when the frequency status of the neighbor and target was reversed (into low and high respectively), the influence of the neighbor was equal to that of the non-word preview, with significantly increased fixation durations as compared to the identical preview condition. The fact that neighbor previews were able to facilitate (rather than inhibit) target processing suggested that their influence was mainly sub-lexical. On the other hand, the fact that the frequency of the neighbor heavily affected this influence suggests that the preview benefit was also to some extent lexical in nature. As Williams et al. noted, early word processing (which begins in the parafovea) may involve the excitation of letter identities as well as associated lexical entries. Testing the influence of word n+1 on word n rather than on a target at position n+1 (i.e., parafoveal-on-foveal effects rather than parafoveal preview effects) may clarify to what extent these processes occur across multiple words in parallel.

The aim of the present study was twofold: we firstly sought further validation of the Grainger et al. (2014) model, specifically by assessing whether foveally processed words may be inhibited by parafoveal orthographic neighbors. We did so by employing the gaze-contingent boundary paradigm of Rayner (1975) in an otherwise natural sentence reading setting for Experiment 1, and the FLLD paradigm of Dare and Shillcock (2013) for Experiment 2. Secondly, based on the outcomes of these experiments we sought to gain insight concerning the level of processing at which information is integrated across multiple words in reading. Specifically, in the case of facilitatory parafoveal-on-foveal effects we would replicate the findings of Dare and Shillcock (2013) and Angele et al. (2013), strengthening the conception that multiple words are processed in parallel at the sub-lexical orthographic level. Inhibitory parafoveal-on-foveal effects, on the other hand, would signal the possibility that information across multiple words is integrated at the level of whole-word representations, in line with the model of Grainger et al. (2014).

Experiment 1: Parafoveal orthographic neighbors

In our first experiment, we extended previous research of Angele et al. (2013) and Dare and Shillcock (2013) on parafoveal-on-foveal orthographic influences in sentence reading. Whereas these studies have shown that a repetition of the target or orthographically related nonwords at position n+1 have a facilitatory influence on the processing of the target at position n, here we explored the possibility that an orthographically related neighbor word at position n+1 may exert an inhibitory influence on target word processing. This would be in line with the idea that multiple words have to compete for recognition through word-to-word inhibitory connections (McClelland & Rumelhart, 1981), and moreover, that this competition is fueled by information from multiple words in parallel during everyday reading.

Method

Participants

Thirty students (20 female) from Aix-Marseille University were paid €10 each to participate in this experiment, carried out at the Laboratoire de Psychologie Cognitive in Marseille, France. All participants were native French speakers between 18 and 30 years old, and had normal- or corrected-to-normal vision. Further, all participants were naive with regard to the purpose of the experiment.

Materials

The stimulus set consisted of 62 experimental sentences in French, each fitting on a single line with length ranging between 35 and 63 characters ($M = 48.1$, $SD = 6.7$). Each sentence contained a 4- or 5-letter target word (29 and 33 occurrences respectively), followed by a post-target word of the same length. Using the gaze-contingent boundary technique (Rayner, 1975), we manipulated the information available from the right parafovea (i.e., the post-target word) when participants were fixating on the target. Each of the 62 sentences was used in three experimental conditions with the post-target word varying per condition, as can be seen in Figure 2 below. Additionally, we used 62 filler sentences, each one of which was followed by a comprehension question, to be answered with a left / right button response.

Targets were chosen from the French Lexicon Project (Ferrand et al. 2010), and coupled to both an orthographically related neighbor post-target and an unrelated (control) post-target. We used the lexical decision time (LDT) rather than the frequency of these words as a selection criterion, in light of the idea that the former measure provides a more direct estimate of the time needed to process a given word. Targets and post-targets were coupled such that the LDT of the target was at least 50 ms higher than the neighbor- and unrelated post-targets; meanwhile, the LDT of the neighbor was (nearly) identical to that of the unrelated post-target. It is important to note that this strategy is similar to pairing a low-frequency target with a high-frequency neighbor, as was done in the study of Segui and Grainger (1990) that found an inhibitory neighborhood effect. Indeed, although we selected words based on their LDT value, we made sure that the frequency of the target never exceeded that of the post-targets; (mean frequencies of targets and post-targets were 4.12 and 5.04 Zipf, respectively. For more on the Zipf frequency unit, see Van Heuven, Mandera, Keuleers & Brysbaert (2014)). Further, we made sure that the target and neighbor only differed in an inner-positioned letter, similar to the condition wherein Vitu et al. (2004) found a tendency for parafoveal-on-foveal inhibition. Target word types included nouns, verbs and adjectives. The neighbor and unrelated post-target were always from the same syntactic category.

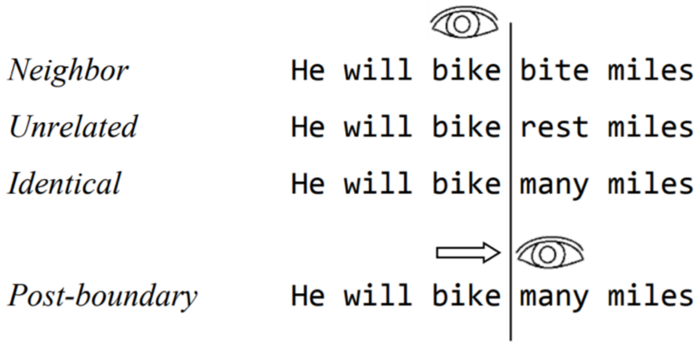

Figure 2. Example of a trial with target word '*bike*', across three different conditions. The upper three sentences show what a stimulus could look like in three conditions before the eyes cross the boundary (vertical line). We used a neighbor condition, where the post-target is an orthographic neighbor of the target with a lower LDT. In the unrelated condition the post-target has an LDT value equal to that of the neighbor post-target, and no orthographic overlap with the target. In the identical condition the pre-boundary and post-boundary post-targets are identical.

Design

Three experimental conditions were used: a neighbor condition, an unrelated condition, and an identical condition. All experimental sentences were tested in all conditions, making 186 experimental trials in total, and meaning that all targets were viewed three times by each participant. Additionally, a filler condition (containing 62 sentences, each followed by a comprehension question) was used to ensure that participants paid attention to the task at hand (i.e., that they truly read for meaning, rather than just moved the eyes from left to right). The total of 248 trials was run in randomized order.

Apparatus and software

The stimuli and experimental design were implemented with OpenSesame (Mathôt et al., 2012), with the PyGaze back-end (Dalmaijer et al., 2014) to process eye movement data online. The reader‟s right eye position was recorded with an EyeLink 1000 (SR Research, Mississauga, ON, Canada), a video-based eye tracker sampling at 1000 Hz with a spatial resolution of 0.01°. Stimuli were presented on a gamma-calibrated 21-inch ViewSonic p227f CRT monitor (1024x768 px, 150 Hz). Participants were seated at a distance of 100 cm from the display, so that each character space subtended 0.35 degrees of visual angle. A chin-rest was used to facilitate a stable head position.

Procedure

Before commencing the experiment, the right eye was calibrated using a 9-point calibration grid with fixation points appearing in random order. In case of a sufficient match between the calibration grid and fixation grid, a validation was carried out to double-check the accuracy of the initial fixations. Prior to the actual experiment, a set of five practice trials (including a filler trial) was used to allow the participant to become acquainted with the procedure. At the start of each trial, a drift correction dot was shown in the center of the screen, on which participants had to fixate before pressing the spacebar. This allowed for an automatic drift correction before the start of every trial – and in the case of failing to align the eye with the fixation point, the initiation of a full recalibration. In case of a successful alignment, a fixation mark in the shape of a forward slash (/) was presented on half a sentence‟s length to the left of the screen center.

When the eyes had stabilized on this fixation mark (within a 40 pixel range) for 700ms, the experimental sentence appeared with its center aligned to the center of the screen, meaning that the beginning of the sentence was aligned to the fixation position.

The position of the eyes was tracked online as participants read the sentence. When the eyes moved beyond the target word, i.e., beyond an invisible boundary indicated by the x-coordinate of the pixels right next to the end of the target word, the post-boundary sentence changed into its baseline form. This display was then kept onscreen until the eyes had reached the end of the sentence (50 pixels to the left of the last character, because readers do not usually move their eyes to the very last pixel of the sentence). At this point, a "*bon*!" (good) message was displayed a little below the end of the sentence, shortly after which the screen was cleared. In the baseline, neighbor and control condition, the drift correction dot would then reappear to begin the next trial. In case of a filler sentence however, a comprehension question was displayed first, with the two possible answers displayed in the lower left and right corner of the display, contingent with the left / right button response. A random side was chosen for the correct answer every trial. The response was met with a green "*bon*!" or red "*non*!" message, depending on whether the answer was correct or incorrect respectively. Shortly hereafter, the screen was cleared to start the next trial. Participants were encouraged to blink before fixating on the slash mark, and to not blink during the presentation of a sentence, because the temporary loss of corneal reflection would cause imprecise gaze estimations. The experiment lasted approximately 45 minutes. After the experiment, a short debriefing was carried out to check if participants noticed any boundary changes.

Results

From the total of 5364 trials, 1383 trials (25.78%) were discarded due to eye blinking or target skipping. Trials where the display-change occurred during a fixation (e.g., due to landing too close to the boundary) were discarded as well, leading to the exclusion of 192 (3.57%) trials. Target fixations outside the 50ms– 1000ms range were considered outliers, leading to the exclusion of an additional 42 (0.78%) trials. In the debriefing, three out of thirty participants reported noticing a word change; all three reported seeing this change two or three times. Given that their data followed general trends, we assumed that it was safe to include them in our analyses. As all participants answered comprehension questions correctly in more than 80% of the filler trials, no participant was excluded.

From the eye tracker data we computed the first fixation duration (FFD), gaze duration (GD) and total viewing time (TVT) on target words. Here, FFD refers to the first fixation duration on the target, regardless of whether there were subsequent target fixations. GD refers to the sum of all fixation durations during first pass, i.e., excluding fixation times following an inter-word regression back to the target. TVT refers to the sum of all fixation durations on the target, both during first pass and following an inter-word regression. We further calculated the probability that a target was skipped, the probability that the post-target was skipped, the probability for a refixation on the target during first pass, and lastly, the probability that the target was re-inspected by means of an inter-word regression.

For the duration measures we used linear mixed models (LMMs) with items and subjects as crossed random effects (Baayen, 2008). The models were fitted with the lmer function from the lme4 package (Bates, Maechler, Bolker & Walker, 2015) in the R statistical computing environment. We report regression coefficients (b), standard errors (SE) and t-values for all factors. Fixed effects were deemed reliable if | t | > 1.96 (Baayen, 2008). Generalized (logistic) LMMs were used to analyze skipping, refixation and regression rates. The z-values can be interpreted in the same way as the t-values.

Fixation duration measures

We expected the fixation duration measures to be increased for the neighbor condition as compared to the unrelated control condition, due to competition for recognition between the target and post-target on the level of whole-word representations. However, we found that all fixation time measures except TVT were significantly shorter in the neighbor condition as compared to the unrelated condition, with a marginal significance in TVT (Tables 1 and 2). All fixation time measures were also significantly shorter for the identical condition as compared to the unrelated condition.

Table 1. Mean fixation times and probabilities.

	FFD	*GD*	*TVT*	*n+1 skip*	*Refixation*	*Regression*
Neighbor	250 (113)	277 (141)	374 (197)	0.16 (0.13)	0.08 (0.08)	0.34 (0.11)
Unrelated	260 (144)	289 (176)	378 (197)	0.15 (0.12)	0.07 (0.06)	0.36 (0.12)
Identical	250 (98)	273 (130)	350 (184)	0.25 (0.15)	0.06 (0.07)	0.26 (0.11)

Note: values in between parentheses indicate standard deviations.

Table 2. Fixation time analyses for Experiment 1.

	FD			GD			TVT		
	b	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*
(intercept)	260.76	9.09	**28.68**	289.59	10.55	**27.45**	441.56	26.16	**16.91**
Neighbor	-12.04	4.91	**-2.45**	-12.54	5.72	**-2.19**	-25.63	14.64	-1.75
Identical	-11.31	4.93	**-2.29**	-18.67	5.75	**-3.25**	-57.89	12.36	**-4.69**

Note: significant values are shown in bold.

Table 3. Saccade type probabilities.

	n+1 skip			*Refixation*			*Regression*		
	b	*SE*	*z*	*b*	*SE*	*z*	*b*	*SE*	*z*
(intercept)	-1.94	0.22	**-8.62**	-2.50	0.18	**-13.62**	-0.62	0.14	**-4.32**
Neighbor	0.13	0.12	1.04	0.14	0.14	1.01	-0.06	0.09	-0.68
Identical	0.74	0.12	**6.37**	-0.25	0.15	-1.63	-0.44	0.10	**-4.59**

Note: significant values are shown in bold.

Skipping, refixation, and regression probabilities
For all probability measures, there were no significant differences between the neighbor and unrelated condition (Table 3). The post-target skipping rate was significantly lower for the neighbor and unrelated condition as compared to the identical condition, (between identical (ref.) and neighbor: b = -0.61, SE = 0.12, z = - 5.45; between identical (ref.) and unrelated condition, b = -0.74, SE = 0.12, z = -6.37). It might have been that the identity of the (pre-boundary) post-target in the neighbor- and unrelated condition alarmed participants that something was wrong, prompting more eye-movements to this location. After all, the post-target in these conditions was of the same syntactic type as the target, and therefore not compatible with the sentence (see the discussion below).

Compared to the identical condition, the rate of regressions was significantly higher for the neighbor and unrelated conditions, (between identical (ref.) and neighbor: b = 0.38, SE = 0.09, z = 3.96; between identical (ref.) and unrelated condition, b = 0.44, SE = 0.10, z = 4.59). The boundary change might have subconsciously alarmed participants that something was different when they were focusing on the target, prompting more regressions to the target (as a means to 'double-check'). A same trend was found for the rate of refixations, although here only the difference between the neighbor and identical condition reached significance. It is difficult to conclude with regard to refixations, given that the refixation rate was very low in general – possibly due to the use of short target word lengths.

Discussion

In Experiment 1, we tested whether parafoveal orthographic neighbors can inhibit foveal word processing through competition for recognition at the lexical level. Our findings do not point in this direction, as all first-pass fixation time measures on the target were decreased with an orthographically related post-target as compared to an unrelated post-target. On the contrary, the results of Experiment 1 are in line with the prior observation that orthographically related non-words in the parafovea facilitate processing of the foveal target word (Angele et al., 2013; Dare & Shillcock, 2013). A direct comparison of the non-word neighbor condition tested by Angele et al. (2013) and the word neighbor condition tested in Experiment 1 (effect sizes of 8.21 ms and 12.04 ms respectively), suggests that the lexical status of the parafoveal neighbor has had little impact on processing of foveal targets. Thus, the results of Experiment 1 point to the spatial integration of orthographic information across foveal and parafoveal stimuli up to the level of sub-lexical orthographic representations, but not up to the level of lexical representations.

We also found significantly increased fixation durations on the target in the unrelated condition as compared to the baseline condition. It might be that preview of the post-target during the fixation on the target caused a disruption, known as saccadic inhibition (Reingold & Stampe 2004). This phenomenon has traditionally been linked to the perception of visual transients, such as very brief changes or disturbances in the visual field. In our experiment, a premature boundary change could have been such a transient. Indeed, in cases where the fixation location would have been on the right end of the target, the eyes could have briefly crossed the boundary to trigger the change, considering that fixations are never really stable. Here this scenario provides an unlikely explanation, however, since all such trials (192 occurrences, equaling 3.57%) were discarded from the analyses. An alternative explanation may be that preview of the post-target signaled „syntactical incorrectness" (targets and pre-boundary post-targets were always of the same syntactic category), causing a disruption in the current fixation.

In support of this theory, we also found that the post-target was significantly more often fixated in the neighbor- and unrelated condition as compared to the identical condition. Importantly, this does not necessitate parallel processing on a semantic or syntactic level. In line with a serial processing account, it could be that visual attention was directed, ahead of the eyes, to the post-target after the target was recognized (see also Angele & Rayner, 2013; Reichle et al., 2003).

With regard to the finding that parafoveal orthographic neighbors decreased fixation durations on the target, at least two scenarios are possible: (i) there are no word-to-word inhibitory processes for simultaneously presented words, or (ii) word-to-word inhibition exists, but is overshadowed by the facilitating effect of orthographic overlap at the sub-lexical level. Importantly, the neighbor condition in our experiment was similar to the condition for which Vitu et al. (2004) found no significant parafoveal-on-foveal effect and even a numerical trend toward inhibition (specifically, targets and neighbors differed in an inner-positioned letter and neighbors were of a higher frequency). It is noteworthy that the fixation duration measures were increased in their experiment as compared to ours, possibly due to the different nature of the respective tasks. Specifically, in the current experiment participants had to read normal sentences, whereas in the study of Vitu et al. (2004) the target and post-target were positioned in between two x-letter strings (e.g. *xxxx thus this xxxx*), with participants having to respond upon identifying an animal name. This is arguably closer to a single word reading task than a sentence-reading task. Alluding to the possibility that lexical influences take place at a later point in time, the longer fixation durations may have granted enough time to integrate information not just on the sub-lexical level, but also on the level of whole-word representations. There was also considerably less information surrounding the target in their experiment (besides the post-target), possibly leading to deeper parafoveal processing of the post-target.

In light of these possibilities, for Experiment 2 we decided to investigate parafoveal-on-foveal influences in a lexical decision task on isolated words that were surrounded by flanking letters, similar to the FLLD task used by Dare and Shillcock (2013) and Grainger et al. (2014). This time we compared parafoveal-on-foveal influences of word neighbors against that of non-word neighbors. We hypothesized that if target processing is facilitated in both conditions, this facilitation should be attenuated in the word neighbor condition due to competition at the lexical level.

Experiment 2: Word vs. non-word neighbor flankers

In Experiment 2 we used the Flanking Letters Lexical Decision (FLLD) paradigm (Dare & Shillcock 2013; Grainger et al. 2014) to assess parafoveal-on-foveal influences in isolated-word reading. In this paradigm, we presented 4-letter target words that were flanked by two letters on the left and two letters on the right. Participants had to indicate by means of a button response whether they perceived a word or pseudo-word at the central location, while we again manipulated the parafoveal information they received. Specifically, we compared the influence of parafoveal related words on foveal target processing against that of parafoveal related non-words. The central idea here was that the facilitatory influence of related words should be attenuated by word-to-word inhibition. As targets were flanked by two letters on the left and two letters on the right, these flanking letters were essentially a neighbor word (or non-word) split in two (e.g. ‘rock’ in ‘ro rook ck’). A potential advantage of this strategy over one where whole-word neighbors would be presented to the left or right of the target, is that the flanking information stays closer to the fovea, reducing constraints of visual acuity associated with parafoveal processing (e.g. Schotter et al., 2012; Vitu et al., 2004).

Method

Participants
Twenty-two students (15 female) from the Aix-Marseille University were paid €2.50 each to participate in this experiment. Three of these students had participated in Experiment 1 as well. All criteria were similar to those stated in Experiment 1.

Materials
For this experiment we generated a list of thirty 4-letter French target words from the French Lexicon Project (Ferrand et al. 2010), all within an LDT of 600–900 ms. Each target word was paired with an orthographic neighbor with an LDT below 600 ms and at least 50 ms shorter than its target counterpart.

This was done with the idea that faster processed words (which are typically more frequent and 'easier'), exert more inhibition on low-frequency neighbors (e.g. Segui & Grainger, 1990). The average Zipf frequencies of targets and neighbors were 5.28 and 5.44, respectively. For each target we further found an orthographically unrelated word which had an LDT equal to the neighbor (deviating no more than 25 ms). In a similar fashion, we gathered a related and unrelated pseudo-word from the French Lexicon Project pseudo-lexicon (Ferrand et al. 2010). There were thus four different flanking stimuli per target word, representing four conditions. A typical trial display can be seen in Figure 3 below. We also generated a list of pseudo-word targets. These filler stimuli were necessary for the lexical decision task and were thus not used in data analyses.

Design
We used a 2 × 2 factorial design, with orthographic relatedness of the flanker (related, unrelated) and lexicality of the flanker (word, non-word) as conditional variables. Thus, besides the related word condition displayed in Figure 3 ('*wa barn rn*'), there was a flanker for the unrelated word condition ('pill'; same frequency as 'warn'), a flanker for the related non-word condition ('*karn*'), as well as a flanker for the unrelated non-word condition ('*lirt*'). All targets were seen four times by each participant, with different flankers for every condition, making a total of 120 trials. We further added 120 trials with a pseudo-word target. The total of 240 trials was presented in randomized order.

Apparatus and software
The stimuli and experimental design were implemented with OpenSesame (Mathôt et al., 2012) and presented on an Asus 17-inch laptop display (1024x768, 150 Hz). Participants were seated in a dimly lit room at a 50cm distance from the display, so that each character space subtended 0.30 degrees of visual angle.

Procedure
Centralized vertical fixation bars were presented throughout the experiment. Every trial, a target stimulus with flanking letters was briefly presented (150ms) at the center of the screen, in between two vertically aligned fixation bars that were previously displayed for 1000ms. Participants then had a maximum of 1800ms to respond with a left ('z') or right ('/') button response (qwerty keyboard layout). For correct responses, a green dot was briefly shown; for

incorrect responses, this was a red dot (see figure 3). The duration of the experiment was approximately 10 minutes.

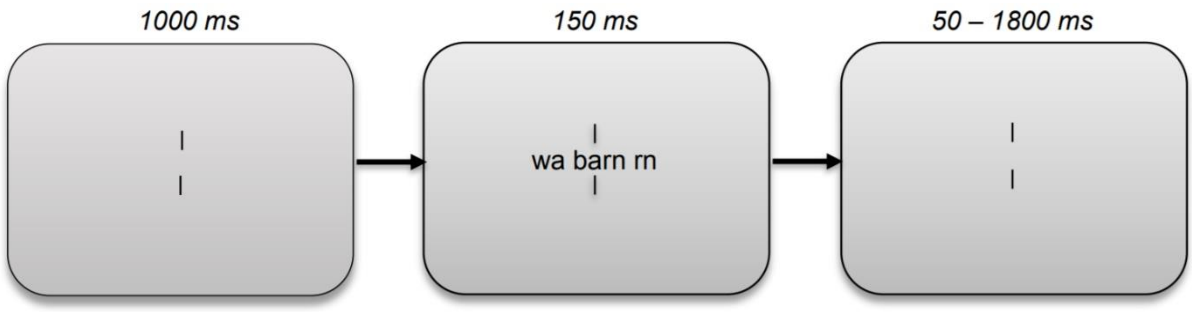

Figure 3. Description of the procedure of Experiment 2. In this example, the target word '*barn*' is flanked by the higher frequency orthographic neighbor '*warn*', with its first and second letter presented on the left side, and the third and fourth letter on the right side. The stimulus appears onscreen for 150 ms, in between vertical fixation bars. After the stimulus disappears, participants have a maximum of 1800 ms to indicate whether they perceived a word or non-word.

Results

From the 22 participants, one participant's data was excluded due to having a 55% error rate. Trials with response times at more than 2.5 SD from the participant's mean (5.4%) were discarded. Only trials with a word target were included in the analyses. To analyze response times, we again employed LMMs with items and subjects as crossed random effects (Baayen et al., 2008). Generalized (logistic) LMMs were used to analyze the error rates. Here we again provide logits back-transformed into probabilities by means of the inverse logit formula.

Table 4. Mean RTs and error rates for Experiment 2.

	RTs		*Error rates*	
	Related	*Unrelated*	*Related*	*Unrelated*
Word flanker	418 (83)	432 (91)	.09 (.07)	.12 (.09)
Non-word flanker	421 (87)	431 (97)	.09 (.06)	.11 (.08)

Table 5. Analyses of RTs and error rates for Experiment 2.

	RTs			Error rates		
	b	SE	*t*	*b*	SE	*z*
(intercept)	432.60	23.64	**18.30**	2.22	0.29	**7.72**
Lexicality	2.91	7.40	0.39	-0.07	0.19	0.17
Relatedness	25.03	7.44	**3.36**	-0.25	0.18	-1.38
L × R	-16.87	10.54	-1.60	0.04	0.25	0.16

Response times and error rates

In line with our findings from Experiment 1, we found a main effect of orthographic relatedness with related flankers yielding shorter response times than unrelated flankers (b = 19.55, SE = 7.41, $|t|$ = 3.04). Our hypothesis that the effect of relatedness should be bigger for word flankers as compared to non-word flankers was not confirmed, as there was no significant interaction of flanker lexicality (word, non-word) and flanker relatedness (related, unrelated), with b = 16.87, SE = 10.54, | t | = 1.60. No significant differences were found in the error rates (Table 5).

Number of potential neighbors per condition

It could have been the case that the combination of flanker and target in one condition potentially activated more orthographically related representations (i.e., words besides the flanker and target) than the flanker and target of another condition. To illustrate this idea, it could be that 'wa barn rn' activated 'war' besides 'warn' and 'barn', while in the related non-word condition, e.g. 'ka barn rn', no additional neighbor could have been activated. As such, RTs in the first example might have been increased as compared to the latter. To test whether such a difference between conditions might have unintentionally taken place, all possible letter combinations (between length 2 and 6 and controlling for position) were determined for all pairs of flankers and targets. Subsequently we checked how many of those letter combinations actually occurred in the French Lexicon Project lexicon. As it turned out, there was no remarkable difference among conditions, with on average approximately one additional activated word per stimulus in each condition, (related word flankers 1.10; unrelated word flankers 0.94; related non-word flankers 0.94; unrelated non-word flankers 1.07). If anything, based on these values the most inhibition could be expected in the related word flanker condition – but the results clearly did not point in this direction.

Discussion

In Experiment 2, we found results similar to what we found in Experiment 1: orthographically related information in the parafovea facilitated processing of the foveal word. In the current experiment however, we assessed in more detail whether this facilitation may be attenuated by word-to-word inhibition of parafoveal orthographic word neighbors, as opposed to orthographically related non-words. This was not the case, as we could not establish an interaction between flanker relatedness and flanker lexicality. Thus, the current results seem to suggest that parafoveal information is only integrated at a sub-lexical orthographic level.

General discussion

In two experiments, we examined the extent to which parafoveal and foveal information may be processed simultaneously. We anticipated two scenarios: one that would replicate findings of previous research, in that parafoveal information facilitates foveal word processing through relatedness at the sub-lexical orthographic level (Angele et al, 2013; Dare & Shillcock, 2013; Grainger et al., 2014; Inhoff et al., 2000; Vitu et al., 2004). The second scenario would entail an integration of foveal and parafoveal information on the level of whole-word orthographic representations, leading to competition for recognition and inhibited target processing, analogous to how foveally presented masked neighbor primes were shown to inhibit target processing in the studies of Segui and Grainger (1990) and Davis and Lupker (2006). Importantly, both facilitatory sub-lexical and inhibitory lexical influences among orthographically related words are predicted by our model of reading, as presented in Grainger et al. (2014), and here in Figure 1. Within this theoretical framework, investigating the nature of effects of parafoveal neighbors provided a means to evaluate the extent to which parafoveal words activate lexical representations that compete for identification within a single processing channel.

In Experiment 1 we found that parafoveal word neighbors had a facilitatory influence on foveal target processing, strengthening the conception that information across multiple words is integrated at the sub-lexical level. Interestingly, we used an experimental condition similar to one for which Vitu et al. (2004) found no effect, (i.e., low-frequency targets and high-frequency parafoveal neighbors that differed in an inner-positioned letter), presumably because the sub-lexical orthographic facilitation was neutralized by lexical inhibition. A critical difference between the current Experiment 1 and that of Vitu et al. was that we opted for a natural sentence reading setting whereas Vitu et al. presented the target and post-target in between two x-letter strings (e.g. xxxx thus this xxxx). It is possible that the different nature of their task (i.e., identifying animal names rather than reading a normal sentence) led to increased fixation durations, granting enough time to integrate parafoveal information at the lexical level. There was also more parafoveal information surrounding the target in our experiment, possibly leading to weaker parafoveal processing. On the other hand, the results of Experiment 1 are in line with the results of experiments that used the same paradigm (sentence reading with the boundary technique) while testing for parafoveal repetition effects and effects of orthographically related parafoveal nonwords (Angele et al., 2013; Dare & Shillcock, 2013), and in line with the interpretation that Angele et al. (2013) proposed for their results. A comparison of the parafoveal nonword neighbor condition tested by Angele et al. (2013) and the parafoveal word neighbor condition tested in Experiment 1 revealed similar parafoveal-on-foveal influences in these two conditions.

In Experiment 2 we compared the influence of parafoveal neighbors against that of related parafoveal non-words, this time using the FLLD paradigm introduced by Dare and Shillcock (2013). According to our model, if there would again be parafoveal-on-foveal facilitation, then this facilitation should be attenuated for the related words as compared to the related non-words. However, this prediction was not supported by our findings, as we could not establish an interaction between flanker lexicality and flanker relatedness. The data thus suggest that information can be extracted from multiple words simultaneously, but that this does not enable simultaneous access to the lexical representations of these words.

With regard to the serial vs. parallel processing debate, we would argue that the current findings do not align with one side more so than the other. A key difference between explanations provided by proponents of serial and parallel processing respectively, seems to be in the role that visuo-spatial attention is assumed to play in parafoveal word processing. According to the serial processing account proposed by Angele et al. (2013), parafoveal feature- / letter-detectors may influence the state of foveal feature- / letter-detectors pre-attentively, i.e., without needing attention, and as such not influencing foveal words‟ lexical representations (see also Reichle et al., 2006). Parallel processing accounts (e.g. Engbert,

Nuthmann, Richter & Kliegl, 2005; Reilly & Radach, 2006) on the other hand, assume that attentional resources are distributed across multiple words as a gradient (mainly in the fovea and partly in the parafovea), and that it is the portion of attention allocated to the upcoming word that underlies information uptake from this location. As can be seen in Figure 1, our model assumes a similar attentional distribution. Our results do not make clear whether parafoveal words were indeed attended during foveal word processing, or that it was rather pre-attentional activity of parafoveal letter detectors that influenced foveal letter detectors, like Angele et al. proposed. One possible way to investigate this would be to use the FLLD paradigm to compare conditions with repetition flankers (e.g. ro rock ck) and unrelated flankers (e.g. le rock ap) against a no-flanker condition. If parafoveal letter detectors indeed influence foveal letter detectors without needing attention, then the repetition flanker condition should yield lower response times than the no-flanker condition. If, on the other hand, the repetition flanker condition yields higher response times than the no-flanker condition (but lower response times than the unrelated flanker condition), this would suggest that parafoveal information demands some attentional resources during foveal word processing.

Our model of reading has thus far been able to account for inhibition from foveal masked neighbor primes, as well as facilitation from parafoveal orthographically related information, on the processing of foveal words. Why then, did we not find inhibition from parafoveal neighbors in our current study? One possibility is that the information that is extracted from a word will only activate its lexical representation when the information is of sufficient quality. In parafoveal word processing, this quality may not be guaranteed due to constraints of acuity and visual attention (see also Schotter et al., 2012; Vitu et al., 2004), and as a consequence, parafoveal information integration would be restricted to the sub-lexical orthographic level. This is basically the same explanation as offered by Williams et al. (2006) for their findings concerning parafoveal preview effects seen with orthographic neighbors. In masked priming studies on the other hand (Davis & Lupker, 2006; Segui & Grainger, 1990), neighbors were presented foveally, where the quality of extracted information was likely to be sufficient for lexical access, and subsequently, competition for recognition with the target. An alternative possibility is that parafoveal words do activate their corresponding lexical representations, but that there is no competition between lexical representations that are activated by spatially distinct stimuli (i.e., the foveal and parafoveal stimuli).

Future research could examine whether or not a parafoveal neighbor might inhibit foveal word processing when the visual quality of the parafoveal neighbor is enhanced. Increasing the size of parafoveal letters to compensate for the loss of visual acuity in the parafovea would be a fair try, although here a problem is that letters will be pushed farther towards the periphery as their size increases, negating the beneficial effect of their increased size. Another possibility would be to use 3-letter targets and post-targets instead of 4- or 5-letter words. Lastly, the allocation of (covert) attention could be manipulated such that it is directed less to the fovea and more to the parafovea. A potential way to do this would be to decrease the saliency of foveal targets whilst increasing the saliency of their parafoveal neighbors. Alternatively, flanking information could be presented a little before target onset, with latencies similar to those used in the masked priming studies of Segui and Grainger (1990).

In summary, the findings of the present study suggest, in line with previous research, that multiple words can be processed simultaneously in everyday reading. Whereas the integration of information across multiple words was assumed to take place at the sub-lexical level, we explored the possibility that information is also integrated at the level of whole-word representations, as described by the model of Grainger et al. (2014). Within this theoretical framework, the results of the present study point to two possibilities that are not mutually exclusive. Either lexical activation generated by parafoveal stimuli is limited due to the constraints imposed by visual acuity, crowding, and spatial attention, or lexical representations activated by spatially distinct stimuli do not compete for recognition.

Supplementary materials

All stimuli used in the present experiments are available in a textfile on the Quarterly Journal of Experimental Psychology website.

References

Angele, B., Tran, R. & Rayner, K. (2013). Parafoveal–foveal overlap can facilitate ongoing word identification during reading: Evidence from eye movements. Journal of Experimental Psychology: Human Perception and Performance. 39(2), 526–538

Angele, B. & Rayner, K. (2013). Processing the in the Parafovea: Are Articles Skipped Automatically? Journal of Experimental Psychology: Learning, Memory and Cognition. 39 (2), 649–662

Baayen, R. (2008). Analyzing Linguistic Data: A pratical introduction to statistics. Cambridge University Press.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015), Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. 67(1), 1–48

Dalmaijer, E., Mathôt, S. & Van der Stigchel, S. (2014). PyGaze: An open-source, crossplatform toolbox for minimal-effort programming of eyetracking experiments. Behavior Research Methods. 46(4), 913–921

Dare, N. & Shillcock, R. (2013). Serial and parallel processing in reading: Investigating the effects of parafoveal orthographic information on nonisolated word recognition. The Quarterly Journal of Experimental Psychology. 66, 417–428

Davis, C. & Lupker, S. (2006). Masked inhibitory priming in English: Evidence for lexical inhibition. Journal of Experimental Psychology: Human Perception and Performance. 32(3), 668–687

De Moor, W., & Brysbaert, M. (2000). Neighborhood frequency effects when primes and targets have different lengths. Psychological Research. 63, 159–162

De Moor, W., Van der Herten, L., Verguts, T. (2007). Is masked neighbor priming inhibitory? Evidence using the incremental priming technique. Experimental Psychology. 54(2), 113–119

Engbert, R., Nuthmann, A., Richter, E. & Kliegl, R. (2005). SWIFT: A dynamic model of saccade generation during reading. Psychological Review. 112(4), 777–813

Ferrand, L., et al. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. Behavior Research Methods. 42, 488–496

Grainger, J., Dufau, S., & Ziegler, J.C. (2016). A vision of reading. Trends in Cognitive Sciences. 20, 171-179

Grainger, J. & van Heuven, W. (2003). Modeling letter position coding in printed word perception. In P. Bonin (Ed.) *The mental lexicon: Some words to talk about words*. New York: Nova Science Publishers.

Grainger, J., Mathôt, S. & Vitu, F. (2014). Test of a model of multi-word reading: Effects of parafoveal flanking letters on foveal word recognition. Acta Psychologica. 146, 35–40

Hohenstein, S. & Kliegl, R. (2014). Semantic preview benefit during reading. Journal of Experimental Psychology: Learning, Memory, and Cognition. 40, 166–190

Inhoff, A., Radach, R., Starr, M. & Greenberg, S. (2000). Allocation of visuospatial attention and saccade programming during reading. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), Reading as a perceptual process. Oxford, UK: North-Holland/Elsevier.

Mathôt, S., Schreij, D. & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. Behavior Research Methods 44(2). 314–324

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

Radach, R. & Kennedy, A. (2004). Theoretical perspectives on eye movements in reading: Past controversies, current issues, and an agenda for future research. European Journal of Cognitive Psychology. 16, 3–26

Rayner, K. (1975). The perceptual span and peripheral cues in reading. Cognitive Psychology. 7, 65–81

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

Reichle, E., Pollatsek, A., & Rayner, K. (2006). E-Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. Cognitive Systems Research. 7, 4–22

Reingold, E. & Stampe, D. (2004). Saccadic inhibition in reading. Journal of Experimental Psy-chology: Human Perception and Performance. 30, 194–211

Reilly, R., Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. Cognitive Systems Research.

Schotter, E., Angele, B. & Rayner, K. (2012). Parafoveal processing in reading. Attention Perception & Psycho-physics. 74, 5–35

Schotter, E. (2013). Synonyms provide preview benefit in English but other semantic relationships do not. Journal of Memory and Language. 69, 619–633

Segui, J. & Grainger, J. (1990). Priming word recognition with orthographic neighbors: Effects of relative prime-target frequency. Journal of Experimental Psychology: Human Perception and Performance. 16(1), 65–75

Van Heuven, W., Mandera, P., Keuleers, E. & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. The Quarterly Journal of Experimental Psychology. 67(6), 1176–1190

Vitu, F., Brysbaert, M. & Lancelin, D. (2004). A test of parafoveal-on-foveal effects with pairs of orthographically related words. European Journal of Cognitive Psychology. 16, 154–177

Williams, C., Perea, M., Pollatsek, A., & Rayner, K. (2006). Previewing the neighborhood: the role of orthographic neighbors as parafoveal previews in reading. Journal of Experimental Psychology: Human Perception and Performance. 32, 1072–1082

Evidence for simultaneous syntactic processing of multiple words during reading

Joshua Snell, Martijn Meeter & Jonathan Grainger

Citation: Snell, J., Meeter, M. & Grainger, J. (2017). Evidence for simultaneous syntactic processing of multiple words during reading. *PloS ONE, 12*, e0173720.

Abstract

A hotly debated issue in reading research concerns the extent to which readers process parafoveal words, and how parafoveal information might influence foveal word recognition. We investigated syntactic word processing both in sentence reading and in reading isolated foveal words when these were flanked by parafoveal words. In Experiment 1 we found a syntactic parafoveal preview benefit in sentence reading, meaning that fixation durations on target words were decreased when there was a syntactically congruent preview word at the target location (*n*) during the fixation on the pre-target (*n-1*). In Experiment 2 we used a flanker paradigm in which participants had to classify foveal target words as either noun or verb, when those targets were flanked by syntactically congruent or incongruent words (stimulus on-time 170 ms). Lower response times and error rates in the congruent condition suggested that higher-order (syntactic) information can be integrated across foveal and parafoveal words. Although higher-order parafoveal-on-foveal effects have been elusive in sentence reading, results from our flanker paradigm show that the reading system can extract higher-order information from multiple words in a single glance. We propose a model of reading to account for the present findings.

Through decades of reading research, much insight has been gained into how properties of fixated (i.e., foveal) words and upcoming (i.e., parafoveal) words influence eye movement behavior as well as processes of word recognition and sentence comprehension. However, the depth of processing of parafoveal words remains a hotly debated issue. Nevertheless, multiple lines of research have suggested that this is likely to depend on several factors, such as the language at hand (e.g. Deutsch, Frost, Pollatsek & Rayner, 2005; Bertram & Hyönä, 2007; Juhasz et al., 2009; Paterson et al., 2014; Yen, Tsai, Tzeng & Hung, 2008; Schotter, Angele & Rayner, 2012 for a review) and inter-individual differences (e.g. Veldre & Andrews, 2014).

There is already considerable evidence that upcoming words are processed sub-lexically in alphabetic languages, and moreover, that orthographic information is integrated across foveal and parafoveal words such that words are recognized faster when they are orthographically related to adjacent words (e.g. Angele, Tran & Rayner, 2013; Dare & Shillcock, 2013; Grainger, Mathot & Vitu, 2014; Inhoff, Radach, Starr & Greenberg, 2000; Radach & Kennedy, 2004; Snell, Vitu & Grainger, 2017; Vitu, Brysbaert & Lancelin, 2004). However, higher-order processing of upcoming words is more controversial (e.g. Schotter et al., 2012). Although one study showed for Chinese that semantic information can be extracted from upcoming words (Yan, Zhou, Shu & Kliegl, 2012), similar investigations have yielded equivocal results for alphabetic languages. Hohenstein, Laubrock and Kliegl (2010) reported a semantic parafoveal preview benefit in German sentence reading, using a fast-priming paradigm where parafoveal previews would change into semantically related targets shortly after a fixation on the pre-target. This finding was later replicated by Hohenstein and Kliegl (2014) with a standard boundary paradigm, where previews changed into targets during the saccade from the pre-target to the preview/target location. On the other hand, using the same paradigm, Rayner, Schotter and Drieghe (2014) could not establish a semantic preview benefit in English sentence reading. Whether upcoming words are semantically processed might thus depend on the language and its users. Interestingly, Schotter (2013) did find a semantic preview benefit in English sentence reading when using synonym previews (e.g. *start – begin*) rather than associative previews (e.g. *ready – begin*), suggesting that in some languages finding a higher-order preview benefit may depend on the semantic relationship between preview and target.

Addressing syntax instead of semantics

Given that research on the semantic access of parafoveal words in sentence reading has yielded equivocal results (but see Hohenstein & Kliegl (2014) for a review concluding that the effect is not so controversial), in the current article we turn to a different form of higher-order processing, namely the *syntactic* classification of words. It is clear that throughout decades of reading research, syntactic parafoveal processing has received far less attention than semantic parafoveal processing. One notable exception is a study reporting a morphological preview benefit for syntactically congruent morphemes in Hebrew (Deutsch, Frost, Pollatsek & Rayner, 2005). In a recent study of our own, we found that the fixation duration on a given word increased if it was followed by a word of the same syntactic category (e.g. *noun noun,* implying an incorrect continuation of the sentence), as compared to a syntactically legal continuation (e.g. *noun verb*) (Snell et al., 2017). A possible explanation for this finding is that the upcoming word signaled 'incorrectness', leading to a disruption known as saccadic inhibition (see e.g. Reingold & Stampe, 2004).

To our knowledge, the only other investigation of syntactic parafoveal processing is a study by Brothers and Traxler (2016), the results of which provide further evidence that readers process the syntactic category of upcoming words. They found that English readers were less likely to skip an upcoming word if it violated syntactic rules (e.g. *noun noun*). However, they did not find the syntactic equivalent of the semantic preview benefit as reported by Hohenstein and Kliegl (2014) and Schotter (2013); that is, fixation durations on target words were not increased after syntactically invalid previews, as compared to syntactically valid previews (preview effects were found when the valid preview was a repetition of the target word, but these effects were likely to be orthographic in nature). Furthermore, unlike Snell et al. (2017), Brothers and Traxler (2016) did not find increased fixation durations on the pre-target word (*n-1*) when the preview (*n*) was syntactically invalid.

Under which conditions syntactic parafoveal-on-foveal effects occur is thus not yet clear, but it seems likely, at least, that these effects are different from the sub-lexical parafoveal-on-foveal effects discussed above, in the sense that they are not facilitatory in nature (e.g., such that processing of a noun type would be sped up by an adjacent noun type). While this may seem logical, it must be noted that previous studies have dismissed the possibility of higher-order parallel processing precisely on the basis of an absence of semantic parafoveal-on-foveal integration (see e.g. Angele et al., 2013). Here we argue, however, that parallel processing does not equal parafoveal-on-foveal integration. Instead, we propose that the brain can keep track of which word has what role in the sentence being read, meaning that higher-order information is kept separate, rather than being integrated, across words. For example, based on sentence constraints we often know that the upcoming word should be a noun, or that a verb will appear two positions to the right. Furthermore, readers can very accurately make regressions to those points in sentences that are critical for resolving syntactic ambiguity (e.g. MacDonald, Pearlmutter & Seidenberg, 1994), suggesting that some representation of the syntactic structure of a sentence is retained in memory (see Boston, Hale, Vasishth & Kliegl, 2011, for a discussion of the role of syntax in parallel word processing).

It is not inconceivable, then, that the reading process is perturbed when the words that are being recognized are syntactically incoherent, explaining why effects of higher-order parafoveal-foveal integration (e.g. lexical, semantic; Snell et al., 2017 and Angele et al., 2013, respectively) in sentence reading have been elusive. At the same time, in light of this scenario we may predict that higher-order parafoveal-on-foveal integration might take place in a setting where readers do not set out to read sentences, that is, a setting where readers do not create sentence-level representations. One example of such a setting would be a flanker paradigm similar to that used in the studies of Dare and Shillcock (2013), Grainger et al. (2014) and Snell et al. (2017), but now using syntactically related flankers rather than orthographically related flankers.

Here we report two experiments that test the hypothesis that words are syntactically categorized in the parafovea, and that are aimed at further exploring the nature of higher-order parallel processing. In Experiment 1, we used the gaze-contingent boundary paradigm (Rayner, 1975) to see if the recognition of a target word (*n*) would be facilitated by a syntactically congruent preview at the target location when readers were fixating the pre-target (*n-1*). In Experiment 2, we used a flanker paradigm in which participants had to indicate in each trial whether a foveally presented target word was noun or verb, while it was flanked by parafoveal words that were syntactically congruent / incongruent with the target (e.g. *cloud horse cloud* vs. *kneel horse kneel*), or words that formed a correct / incorrect sentence with the target (e.g. *young horse jumps* vs. *jumps horse young*).

Experiment 1: Syntactic preview benefit in sentence reading

Methods

Ethics statement

Given that Experiment 1 and Experiment 2 consisted of non-invasive, low-demanding behavioral experiments, ethical approval was deemed unnecessary. Nevertheless, all participants gave written informed consent to their participation in this study. Participants were further given the option to opt out of the study, but none of the participants made use of this option. For administrative (payment) purposes, participants gave their name, address and student number. The experiments were carried out by the authors of this work.

Participants

30 students (19 female, age 18–26) from the VU University (Amsterdam) participated in this study for €4,- or its equivalent in course credit. All participants were native Dutch speakers and had normal or corrected-to-normal vision. All participants reported to be non-dyslexic. Further, all participants were naïve to the purpose of the experiment.

Materials

From the *Dutch Lexicon Project* lexicon (Keuleers, Diependaele & Brysbaert, 2010) we retrieved 150 5-letter target words and 150 5-letter preview words that were noun or verb (75 occurrences of each). Every target was paired with two previews, one of which was syntactically congruent with the target and one of which was incongruent. All previews were thus used twice; once in a congruent condition and once in an incongruent condition. The amount of orthographic overlap with the target was equal for the two preview types, at an average of one letter. We further assigned a 5-letter pre-target to every target, with the rule that the congruent preview would also be a syntactically correct follow-up of this pre-target, whereas the incongruent preview would be an incorrect follow-up. Specifically, when the pre-target was a noun ('*horse*'), the incongruent preview was also a noun ('*table*') and the congruent preview a verb ('*bites*'). When the pre-target was an adjective or verb ('*great*' or '*bites*'), the incongruent preview was a verb ('*walks*') and the congruent preview a noun ('*grass*'). As such we had the possibility to not only find a preview benefit in the fixation on the target, but also saccadic inhibition during the fixation on the pre-target, replicating the finding of Snell et al. (2016) discussed above.

A sentence varying between 29 and 57 characters (including spaces) was constructed for every target. All pre-targets and previews had a lexical decision time (LDT) value between 500 ms and 650 ms, and within each sentence the correct and incorrect preview had an equal LDT.

We used the gaze-contingent boundary technique (Rayner, 1978) to manipulate the identity of the preview while participants focused on the pre-target. Using an eye-tracker to carefully track the eye position, we changed the preview into the target as soon as an eye movement was made from the pre-target to the preview/target location, (see Figure 1). Besides the congruent and incongruent condition, there was also an identical condition where the target was presented throughout the trial.

To check that participants read for meaning, we created a 'quiz-question' for one out of every five items (totaling 10 per condition). Each question was displayed immediately after participants had finished the sentence that it belonged to, and was to be answered with a left / right button response; (two possible answers were displayed in the left- and right-bottom corner of the screen, with the side of the correct answer randomized). All text was displayed in black on a light-grey background, and all stimuli were presented in randomized order.

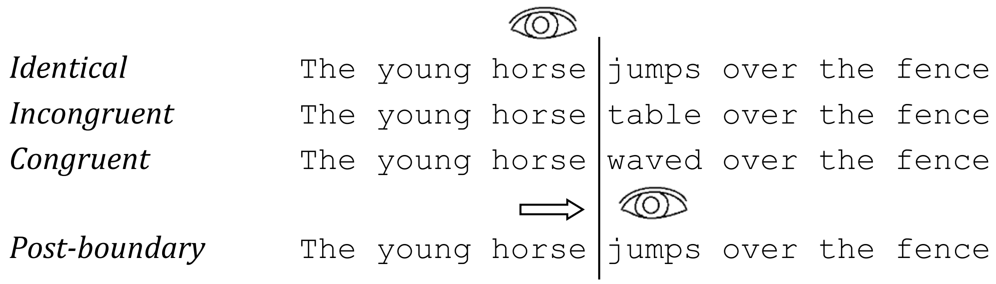

Figure 1. The upper three sentences show what a stimulus could look like in three conditions *before* the eyes cross the boundary (vertical line). We used an identical condition (with target '*jumps*' already visible prior to its fixation), a condition with an incongruent preview ('*table*') and a condition with a congruent preview ('*waved*'). As soon as the eyes move beyond the boundary, the preview changed into the target.

Design

We used a Latin square design to present all stimuli in all three conditions (*congruent preview, incongruent preview, baseline*), but only once per participant. Thus, for each participant, there were 50 items per condition, amounting to 150 experimental trials which were presented in random order.

Apparatus and software

The stimuli and experimental design were implemented with OpenSesame (Mathôt, Schreij & Theeuwes, 2012), with the PyGaze back-end (Dalmaijer, Mathôt & Van der Stigchel, 2014) to process eye movement data online. The participant's right eye position was recorded with an EyeLink 1000 (SR Research, Mississauga, ON, Canada), a video-based eye tracker sampling at 1000 Hz with a spatial resolution of 0.01°. Stimuli were presented on a 1024x768 px, 150 Hz computer monitor. Participants were seated at a distance of 90 cm from the display, so that each character space subtended 0.35 degrees of visual angle. A chin-rest was used to facilitate a stable head position.

Procedure

Before commencing the experiment, the right eye was calibrated using a 9-point calibration grid with fixation points appearing in random order. In case of a sufficient match between the calibration grid and fixation grid, a validation was carried out to double-check the accuracy of the initial fixations. Prior to the actual experiment, a set of five practice trials (including a catch question) was used to allow the participant to become acquainted with the procedure.

At the start of each trial, a drift correction dot was shown in the center of the screen, on which participants had to fixate before pressing the spacebar. This allowed for an automatic drift correction before the start of every trial – and in the case of failing to align the eye with the fixation point, the initiation of a full recalibration.

In case of a successful alignment, a fixation mark in the shape of a forward slash (/) was presented on half a sentence's length to the left of the screen center. When the eyes had stabilized on this fixation mark (within a 40 pixel range) for 700 ms, the experimental sentence appeared with its center aligned to the center of the screen, meaning that the beginning of the sentence was aligned to the fixation position.

The position of the eyes was tracked online as participants read the sentence. When the eyes moved beyond the pre-target, the preview changed into the target. This display was kept onscreen until the eyes had reached the end of the sentence (with a maximum distance of 30 pixels to the left of the last character of the sentence). A green dot was displayed a little to the right of the sentence's end when this end was reached. Shortly thereafter, the drift correction screen was displayed again to begin the next trial. However, for sentences with a comprehension question (30 out of 150 trials), the quiz display was presented first, with the two possible answers displayed below in the left / right corner, contingent with the left / right button response. A random side was chosen for the correct answer every trial. The response was met with a green '*goed!*' (*good* in Dutch) or red '*fout!*' (*wrong* in Dutch) message, depending on whether the answer was correct or incorrect respectively. Shortly hereafter, the screen was cleared to start the next trial.

Participants were encouraged to blink before fixating on the slash mark, and to not blink during the presentation of a sentence, because the temporary loss of corneal reflection would be misinterpreted by the eye-tracker (i.e., it would temporarily pass on wrong fixation coordinates to the computer, causing for instance a premature boundary change).

The experiment lasted approximately 25 minutes. After the experiment, a short debriefing was carried out to check if participants noticed any display changes.

Results

From the total of 4500 trials, 493 trials (10.96%) were discarded due to eye-blinking or the occurrence of a display-change during a fixation (e.g., due to landing too close to the boundary). As all participants answered more than 90% of the catch trials correctly, no participant was excluded. From the 30 participants, 26 reported to have seen a display change. One of these participants reported to have seen a display change at least ten times; all the other participants reported to have seen the display change two or three times.

From the eye tracker data we computed the first fixation duration (FFD), gaze duration (GD) and total viewing time (TVT) on the pre-target and target. Here, FFD refers to the mean first fixation duration on a word, regardless of whether there were subsequent fixations. GD refers to the mean sum of all fixation durations during first pass, i.e., excluding fixation times following a regression back to the word. TVT refers to the mean sum of all fixation durations on a word, both during first pass and following a regression. We further calculated the skipping probability, the probability for a refixation during first pass, and lastly, the probability for a refixation by means of an inter-word regression.

For the duration measures we used linear mixed-effect models (LMMs) with items and participants as crossed random effects (Baayen, 2008). We followed the procedure suggested by Barr, Levy, Scheepers and Tily (2013) to determine the maximal random effects structure permitted by the data. This led us to include by-item and by-participant random intercepts for probability analyses, and by-participant random slopes alongside by-item and by-participant intercepts for the analyses of duration measures. The models were fitted with the lmer function from the lme4 package (Bates, Maechler, Bolker & Walker, 2015) in the R statistical computing environment. We report regression coefficients (b), standard errors (SE) and t-values. Fixed effects were deemed reliable if $|\ t\ | > 1.96$ (Baayen, 2008). Logistic LMMs (fitted with the glmer function) were used to analyze the skipping, refixation and regression probabilities. Here, fixed effects were deemed reliable if | z | > 1.96. In all analyses, values beyond 2.5 SD from the mean, (on average 1.7% of the trials), were marked as outliers and excluded.

Pre-target fixations

In our previous study we found that the fixation duration on word *n* was increased with a syntactically similar word at position *n+1,* as compared to a condition where word *n+1* was a normal follow-up of *n* (Snell et al., 2017). Based on this finding, we expected for the current experiment that the syntactically incongruent preview would lead to increased pre-target fixation durations as compared to the congruent preview, due to saccadic inhibition.

As it turned out, we did not find a significant difference between the congruent and incongruent preview conditions (Tables 1 and 2). A likely cause for this discrepancy is that the fixation durations were in general much lower in the current study as compared to our previous study (mean FFDs of 209 ms and 253 ms, respectively; see the Discussion for a possible explanation of this difference).

For the pre-target there was no significant difference in the skipping and refixation rates among conditions (Tables 1 and 3). The regression rate was increased for the congruent and incongruent condition as compared to the identical condition (with b = 1.02, SE = 0.15, z = 6.97 for the congruent condition and b = 1.17, SE = 0.14, z = 8.09 for the incongruent condition).

Table 1. Pre-target means in Experiment 1. Values in between parentheses indicate standard deviations.

	FFD	*GD*	*TVT*	*Skip*	*Refix*	*Regress*
Congruent	203.8 (72.1)	256.4 (124.5)	291.9 (163.8)	0.12 (0.14)	0.33 (0.26)	0.03 (0.07)
Incongruent	207.2 (76.6)	259.7 (124.1)	299.6 (171.6)	0.12 (0.13)	0.31 (0.24)	0.03 (0.07)
Identical	203.3 (72.1)	254.6 (121.9)	268.3 (139.4)	0.12 (0.11)	0.31 (0.24)	0.02 (0.07)

Table 2. Pre-target analyses of duration measures. Significant values are shown in bold.

	FFD			*GD*			*TVT*		
	b	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*
(Intercept)	200.53	5.62	**35.68**	256.66	10.59	**24.23**	293.74	13.46	**21.83**
Incongruent [a]	3.66	2.38	1.54	3.31	4.57	0.72	5.89	5.72	1.03
Identical [a]	0.25	2.39	0.10	-0.38	4.57	-0.08	20.06	5.71	**-3.52**

Note:
[a] ref.: congruent preview. Significant values are indicated in bold. Abbreviations: SE, standard error; FFD, first fixation duration; GD, gaze duration; TVT, total viewing time.

Table 3. Pre-target analyses of probabilities. Significant values are shown in bold.

	Skip			*Refix*			*Regress*		
	b	*SE*	*z*	*b*	*SE*	*z*	*b*	*SE*	*z*
(Intercept)	-2.45	0.21	**-11.64**	-1.46	0.20	**-7.19**	-2.07	0.16	**-12.58**
Incongruent [a]	-0.09	0.12	-0.75	-0.06	0.09	-0.66	0.17	0.11	1.57
Identical [a]	-0.08	0.12	-0.62	0.00	0.09	0.01	-0.76	0.13	**-6.07**

Note:
[a] ref.: congruent preview. Significant values are indicated in bold. Abbreviations: SE, standard error.

Target fixations

We expected that a syntactically congruent preview at the target location during the fixation on the pre-target would yield a preview benefit during subsequent target processing. This hypothesis was confirmed, as all fixation duration measures except TVT were significantly increased after an incongruent preview as compared to a congruent preview (with marginal significance for TVT; Tables 4 and 5). All fixation durations were also significantly lower for the identical condition as compared to the two preview conditions. This was to be expected, as the target was already visible during the fixation on the pre-target in the identical condition. Nonetheless, our results show that the cost of having a different word at the target location prior to its fixation, is smaller when that word is syntactically compatible with the sentence.

Furthermore, replicating the results of Brothers and Traxler (2016), we found that the target skipping rate was significantly higher after a congruent preview than after an incongruent preview (Tables 4 and 6). This fits quite well with our hypothesis, as it suggests that the incongruent preview signaled that something was wrong at the target location, prompting more fixations to this location (see also Hyönä, 1995).

Table 4. Target means in Experiment 1. Values in between parentheses indicate standard deviations.

	FFD	*GD*	*TVT*	*Skip*	*Refix*	*Regress*
Congruent	239.8 (93.9)	269.3 (122.5)	312.2 (149.4)	0.14 (0.16)	0.14 (0.11)	0.04 (0.05)
Incongruent	249.6 (96.0)	284.9 (128.8)	320.8 (156.5)	0.12 (0.16)	0.15 (0.12)	0.04 (0.05)
Identical	220.9 (79.6)	241.3 (99.1)	260.0 (119.2)	0.15 (0.16)	0.10 (0.10)	0.03 (0.03)

Table 5. Target analyses of duration measures. Significant values are shown in bold.

	FFD			*GD*			*TVT*		
	b	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*
(Intercept)	208.46	11.62	**17.94**	230.24	13.31	**17.30**	268.02	16.71	**16.04**
Incongruent [a]	16.75	4.99	**3.36**	19.73	5.42	**3.64**	10.90	6.14	1.78
Identical [a]	-21.42	4.77	**-4.49**	-28.30	5.59	**-5.06**	-51.38	7.91	**-6.50**

Note:
[a] ref.: congruent preview. Significant values are indicated in bold. Abbreviations: SE, standard error; FFD, first fixation duration; GD, gaze duration; TVT, total viewing time.

Table 6. Target analyses of probabilities. Significant values are shown in bold.

	Skip			*Refix*			*Regress*		
	b	*SE*	*z*	*b*	*SE*	*z*	*b*	*SE*	*z*
(Intercept)	-2.35	0.25	**-9.38**	-2.18	0.18	**-12.17**	-1.55	0.12	**-13.43**
Incongruent [a]	-0.40	0.14	**-2.89**	0.17	0.12	1.45	-0.16	0.10	-1.58
Identical [a]	0.18	0.13	1.15	-0.40	0.13	**-3.16**	-0.98	0.12	**-8.15**

Note:
[a] ref.: congruent preview. Significant values are indicated in bold. Abbreviations: SE, standard error.

Discussion

The results from Experiment 1 suggest that readers can acquire syntactical information from upcoming words during sentence reading, as all first pass fixation duration measures on the target were decreased when it was preceded by a syntactically congruent preview during the fixation on the pre-target. We also expected increased fixation durations on the pre-target in the incorrect / incongruent condition, caused by a disruption by the incorrect parafoveal preview. Although we found that pre-target fixation durations were indeed numerically increased in this condition, this effect did not reach significance (b = 3.66, SE = 2.38, t = 1.54), contrary to our previous study (Snell et al., 2017). This could have been because fixations were generally shorter in the current study.[29] Yet, we also found that targets were skipped more often after a congruent (correct) preview than after an incongruent (incorrect) preview. In line with our hypothesis, this finding suggests that the incorrect preview signaled to readers that something was wrong, prompting more eye-movements to its location. It is further evident that the parafoveal preview could be processed syntactically during the short time window of the pre-target fixation, as we found a clear preview effect at the target location.

As was mentioned in the introduction, Brothers and Traxler (2016) did not find a syntactic preview benefit in their study. Drawing an analogy to the equivocal results generated by investigations of the semantic parafoveal preview benefit (e.g. Hohenstein & Kliegl, 2014; Schotter, 2013; Rayner et al., 2014), it may be the case that higher-order preview effects –at least in fixation times– are more stable in Dutch and German than in English. At the same time it should not be forgotten that Brothers and Traxler (2016) did find increased target skipping rates after syntactically valid previews as compared to invalid previews.[30] Hence, while parafoveal processing effects might manifest themselves differently across languages, the increasing body of results is consistent with the view that higher-order processing can occur for parafoveal words during reading.

As Experiment 1 results provided evidence for higher-order processing of parafoveal words, we set out to investigate the time-course of these processes in Experiment 2 – in particular with respect to whether higher-order processing of parafoveal words may occur *during* or *after* foveal word processing. Indeed, a highly debated issue in reading research concerns the question whether lexical processing can occur across multiple words simultaneously (e.g. Engbert, Nuthmann, Richter & Kliegl, 2005; Reichle, Pollatsek & Rayner, 2006; Risse, Hohenstein, Kliegl & Engbert, 2014). While higher-order parafoveal preview effects show that upcoming words can be processed lexically, it has been argued by Schotter, Lee, Reiderman and Rayner (2015) that such a finding can still be reconciled with a serial processing account if one assumes that attention moves, ahead of the eyes, to word *n+1* when word *n* is recognized to a certain extent. From that moment on, processing of word *n+1* may allow for lexical access – although it must be acknowledged that the time window within which that should happen is considerably short under the assumption of serial processing, considering that the largest portion of the fixation duration would already be spent on the processing of word *n*.

A more effective measure to assess parallel processing may be that of *parafoveal-on-foveal* information integration. It has already been shown in multiple studies that foveal words are recognized faster when they are surrounded by orthographically related information, both in single word reading and in sentence reading (Angele et al., 2013; Dare & Shillcock, 2013; Grainger et al., 2014; Snell et al., 2017), indicating that sub-lexical processing occurs across multiple words in parallel. However, as was stated in the introduction, higher-order parafoveal-on-foveal effects are more controversial. In their study, Angele et al. (2013) did not find that foveal words were recognized faster when they were semantically related to upcoming words. Similarly, in the current study we did not find that foveal words were recognized faster when they were

[29] The difference in FFD between the two studies might have been caused by the fact that we used 5-letter words in Experiment 1, whereas we used 4-letter words in Snell et al. (2017). The longer word length led participants to sometimes make two short fixations instead of a single longer fixation, as evidenced by the fact that approximately one third of the words were refixated in Experiment 1 (Table 1a), as compared to approximately 7% in our previous study.

[30] There may be exceptions to the rule, however. Some studies have shown that the word '*the*' is consistently skipped, even when its position in the sentence is syntactically wrong (e.g. Angele & Rayner, 2013; Abbot, Angele, Ahn & Rayner, 2015). It may be that readers automatically skip '*the*' based on orthographic cues, rather than syntactic cues.

syntactically related to upcoming words. Indeed, from a theoretical standpoint it can be argued that the reading system would not benefit from integrating syntactical information across multiple words: rather, readers would have to keep track of which word has what role in a sentence in order to understand it properly, implying a fairly strict separation of the multiple word identities it contains.

On the other hand, it can be argued that the syntactic categorization of words may be influenced by sentence-level constraints (see e.g. Jordan & Thomas, 2002). Based on the first part of a sentence, for example, readers may have clear expectations about upcoming words, both semantically and syntactically. It is possible that the syntactic recognition of word *n+1* constrains processing of word *n* in a similar way. In this sense, higher-order information integration would not entail the gathering of all available syntactic or semantic information into a single mixture, but rather the construction of a sentence-level representation that interacts with its constituent word identities through feedback connections – a process that can be fundamentally harmonized with a parallel processing account.

In Experiment 2 we employed a flanker paradigm to find out, firstly, whether higher-order (syntactic) information could be integrated across multiple words in parallel, and secondly, whether the nature of this integration process would be one that relies on sentence-level constraints, or one that relies on the integration of syntactic information in a more general sense. To this end, we presented target words in the fovea that were either noun or verb, flanked by words on the left and right which corresponded to either one of four conditions: two conditions where target and flankers would form a grammatically correct or incorrect sentence respectively (to test the hypothesis of sentence-level constraints), and two conditions where target and flankers were syntactically congruent or incongruent (to test the hypothesis of general information integration). As participants had to indicate on each trial whether they read a noun or verb, the expectation was that response times would increase for the incorrect / incongruent flankers as compared to the correct / congruent flankers.

Experiment 2: Syntactic parafoveal-on-foveal influences

Methods

Participants

22 students (16 female, age 18–23) from the VU University (Amsterdam) gave written informed consent to participate in this study, carried out by the authors at the VU University. Participants earned €4,- or its equivalent in course credit for their participation. None of these participants had participated in Experiment 1. Participation criteria were similar to those used in Experiment 1. As in Experiment 1, participants had the option to opt out of the study, but none did so.

Materials

From the *Dutch Lexicon Project* lexicon (Keuleers et al. 2010) we retrieved 50 noun targets and 50 verb targets with a length of 4 or 5 letters (39 and 61 occurrences respectively), from an LDT range of 550 – 750 ms. Each target was coupled to a syntactically congruent flanker and an incongruent flanker with an equal LDT value. Both flankers were of the same length as the target and had no orthographic overlap with the target. For every target we further chose two words with an equal length ranging between 3 and 5 letters, that would form a correct sentence with the target when the one flanker was on the left and the other on the right (e.g. *young horse jumps*), and an incorrect sentence when the flankers would be switched around (*jumps horse young*).

Design
There were four experimental conditions, two of which would test for an effect of syntactical congruency (*congruent* vs. *incongruent* flankers), and two of which would test for an effect of sentence-level constraint (*correct* vs. *incorrect* sentences). Condition examples are shown in Table 7 below. All targets were repeated across all conditions, amounting to 400 experimental trials. We further retrieved stimuli for 12 practice trials, which were not included in the final data analyses. All trials were presented in randomized order.

Apparatus and software
All apparatus and software was similar to those used in Experiment 1, albeit without the use of an eye-tracker.

Procedure
Participants were seated in a comfortable office chair in a dimly lit testing room. The distance from the participants' eyes to the computer screen was 90 cm, so that every character space subtended 0.35 degrees of visual angle. Centralized vertical fixation bars were presented throughout the experiment. Every trial, a target stimulus with flanking words (separated by one character space from the target) was presented for 170 ms at the center of the screen, in between the fixation bars, after which participants had a maximum of 2300 ms to respond with a left ('z') or right ('/') button response (qwerty keyboard layout) whether the target was a noun or verb. Responses for 'noun' were always matched to the right ('/') button. After correct responses, a green dot was briefly shown; for incorrect responses, this was a red dot. Participants were offered a break halfway through the experiment. The duration of the experiment was approximately 20 minutes.

Table 7. Experiment 2 condition examples.

	Noun target	***Verb target***
Congruent	cops rack cops	hear went hear
Incongruent	been rack been	cops went cops
Correct	this rack fell	they went here
Incorrect	fell rack this	here went they

Results

Trials where the response time (RT) was beyond 2.5 standard deviations from the mean (3.14% of all trials) were discarded. Only correctly answered trials were included in the analysis of RTs, leading to the exclusion of another 7.59% of trials. To analyze RTs, we again employed LMM models with items and participants as crossed random effects (including random intercepts and the by-participant random slope). Generalized (logistic) LMM models with by-item and by-participant random intercepts were used to analyze the error rates.

Table 8. Experiment 2 mean RTs (ms) and error rates. Values in parentheses indicate SDs.

	RT	***Error***
Congruent (cops rack cops)	500.46 (150.26)	.062 (.003)
Incongruent (been rack been)	520.51 (152.54)	.094 (.004)
Correct (this rack fell)	504.85 (147.10)	.075 (.003)
Incorrect (fell rack this)	505.83 (150.34)	.077 (.004)

Table 9. Analyses of RTs and error rates: congruent versus incongruent flankers.

	RT			***Error***		
	b	***SE***	***t***	***b***	***SE***	***z***
(Intercept)	505.52	17.00	**29.73**	3.20	0.18	**17.96**
Incongruent	17.45	4.06	**4.30**	0.53	0.12	**4.32**

Note: ref.: congruent flankers. Significant values are shown in bold.

Table 10. Analyses of RTs and error rates: correct versus incorrect sentence flankers.

	RT			***Error***		
	b	***SE***	***t***	***b***	***SE***	***z***
(Intercept)	508.93	17.00	**29.93**	2.96	0.17	**17.09**
Incorrect	-0.27	4.04	-0.07	-0.02	0.12	-0.19

Note: ref.: correct sentence flankers. Significant values are shown in bold.

Congruent vs. incongruent flankers

RTs as well as error rates (Table 8) were significantly lower in the congruent flanker condition as compared to the incongruent flanker condition (Table 9), supporting the hypothesis that lexical information can be gathered and integrated across multiple words simultaneously.

Correct vs. incorrect sentence flankers

Whereas we found a difference between the congruent vs. incongruent flanker conditions, there was no difference between the correct- and incorrect sentence conditions (Table 10).

Discussion

In Experiment 2 we set out to investigate whether syntactic processing may occur across multiple words in parallel. We assessed two scenarios, one of which assumes that information from multiple words would culminate into one syntactic signal (e.g. “this is a noun”, “this is a verb”). The other scenario assumes that the syntactic categorization of words is dependent on constraints at the sentence level, meaning that readers would be more likely to expect (and thus faster to recognize) a noun at position *n* when it is flanked by syntactically compatible words, such as an adjective at position *n-1* and a verb at position *n+1* (in English), as compared to incompatible words. The results of Experiment 2 support the first scenario, as we found significant differences in RTs and errors between the syntactically congruent and incongruent conditions, while no differences were found between the correct- and incorrect sentence conditions.

The current results support the idea that higher-order processing can occur for multiple words in parallel. Indeed, considering that the time it takes to recognize a word is in the range of 150 – 250 ms (e.g. Rayner, 1998), we reckon that there could not have been abundant time to process the target and two flankers serially in the 170 ms that they were presented. It should further be noted that flankers may interfere rather than facilitate – even when the flankers are congruent with the target with respect to the task at hand (e.g., Flowers & Wilcox, 1982). In a recent study, we found that target processing was faster in a no-flanker condition than in the congruent flanker condition (Snell, Declerck & Grainger, in review). This suggests that parafoveal stimuli invariably demand attentional resources, in principle leading to slower target recognition. Crucially, this does not affect the implications of the current results: the fact that syntactically congruent flankers interfered less than syntactically incongruent flankers, provides evidence that the syntactic information of these flankers was available during target processing.

While we found syntactic parafoveal-on-foveal information integration in Experiment 2, we did not find such an effect in Experiment 1. If anything, in Experiment 1 the fixation duration on the pre-target (n-1) was more likely to be increased with a syntactically similar preview (n), as compared to a syntactically different preview. As argued in the Introduction, a potential explanation for this discrepancy is that different tasks might engage different cognitive processes. Specifically, it could be that sentence reading engages the maintenance of a sentence-level representation in working memory, from where top-down feedback would ensure that various syntactic categories are mapped onto the different word positions available in the visual field, to optimize higher-order sentence comprehension. It is conceivable that such a mechanism is not required, and thus not engaged, in a single-word reading paradigm such as the flanker task used in Experiment 2. This would also explain why there was no difference between the conditions where the flankers and target formed a correct vs. an incorrect sentence. Thus, while multiple words can be lexically processed in parallel, it is possible that higher-order mechanisms prohibit the parafoveal-on-foveal integration of higher-order information in sentence reading.

One important issue remains for future research investigating parafoveal processing of syntactic information. That is the potentially asymmetrical nature of such processing, being stronger in the direction of reading (i.e., in the right visual field for the current study). Indeed, there is abundant evidence for asymmetrical processing in the parafovea as concerns various types of information (e.g. Grabbe & Allen, 2013) in line with the fact that the span of effective vision in reading extends further in the direction of reading (e.g., McConkie & Rayner, 1975; Jordan et al., 2014). The present study was not designed to address this issue, but we suspect that evidence for such asymmetrical processing of syntactic information in the parafovea will be conditioned by the task used to investigate this (i.e., sentence reading vs. flanker paradigm), as should become obvious from the comparison of processing involved in Experiments 1 and 2 of the present study in the following discussion.

General discussion

Multiple lines of research have alluded to the possibility that lexical access can occur for parafoveal (upcoming) words in sentence reading, although supporting evidence has been scarce (Hohenstein et al., 2010; Hohenstein & Kliegl, 2014; Rayner et al., 2014; Schotter, 2013). While these studies have focused on semantic parafoveal preview effects, in the current research we focused on syntactic parafoveal effects as an alternative type of higher-order processing.

In Experiment 1 we found that the recognition of a target word was facilitated by a syntactically related preview at the target location during the fixation on the pre-target, suggesting that the reading system extracts higher-order information from parafoveal words. We then set out to investigate whether higher-order parafoveal word processing can take place *during* or rather *after* foveal word processing in Experiment 2, touching upon the highly debated question whether lexical processing occurs serially or rather across multiple words in parallel (e.g. Engbert et al., 2005; Reichle et al., 2006). Multiple lines of research have shown that words can indeed be processed in parallel, by establishing *parafoveal-on-foveal* effects rather than parafoveal preview effects, but these effects were mainly of a sub-lexical orthographic nature (e.g. faster recognition of *n* due to orthographically related *n+1*; see Angele et al., 2013; Dare & Shillcock, 2013; Grainger et al., 2014; Snell et al., 2017). Higher-order parallel processing has been more controversial; for instance, Angele et al. (2013) did not find parafoveal-on-foveal facilitation with semantically related stimuli (but see also Inhoff et al., 2000). As it turned out, in Experiment 2 we found evidence in support of the idea that multiple words can be syntactically processed in parallel, as the syntactic categorization of (foveal) target words was facilitated by syntactically congruent flanking words. Meanwhile, we did not find that words were categorized faster when they were syntactically compatible (i.e., formed a correct sentence) with flanking words.

There is an apparent contradiction between the results of Experiment 1 and Experiment 2 that needs addressing. While we found that word processing was facilitated by syntactically similar words in the parafovea in Experiment 2, this was not the case in Experiment 1. Specifically, we did not find that fixation durations on the pre-target were decreased with a syntactically similar preview, (i.e., the incongruent condition). On the contrary, fixation durations on the pre-target were numerically increased when it was followed by a word of the same syntactic type (see also Snell et al., 2017), alongside an increased fixation rate on the target, suggesting that the reading process was perturbed by the upcoming word. This is most likely due to the fact that word processing is influenced by sentence-level constraints in sentence reading, the underlying mechanisms of which are not engaged in a single-word reading task.

In this scenario, we argue that readers generally process multiple words simultaneously, with each activated word form also activating higher-order semantic and syntactic features (see Figure 2). In our flanker paradigm of Experiment 2, this resulted in the integration of higher-order information across foveal and parafoveal stimuli, such that parafoveal words influenced the decision about the foveally presented target word. During sentence reading, however, activated word forms would append to a sentence-level representation, from where feedback to individual word positions would constrain the recognition process for these words (e.g. through mapping various syntactic categories onto the multiple word positions available in the visual field).

Importantly, this scenario implies that readers are able to keep track of multiple words separately, explaining why previous research has not managed to establish higher-order parafoveal-on-foveal effects in sentence reading. If upcoming words produce a mismatch, for example because they are of an impossible grammatical category (as in the incongruent condition of Experiment 1), the recognition process would be slowed. Indeed, the results of Experiment 1 showed that target words were fixated longer and more often after an incongruent preview.

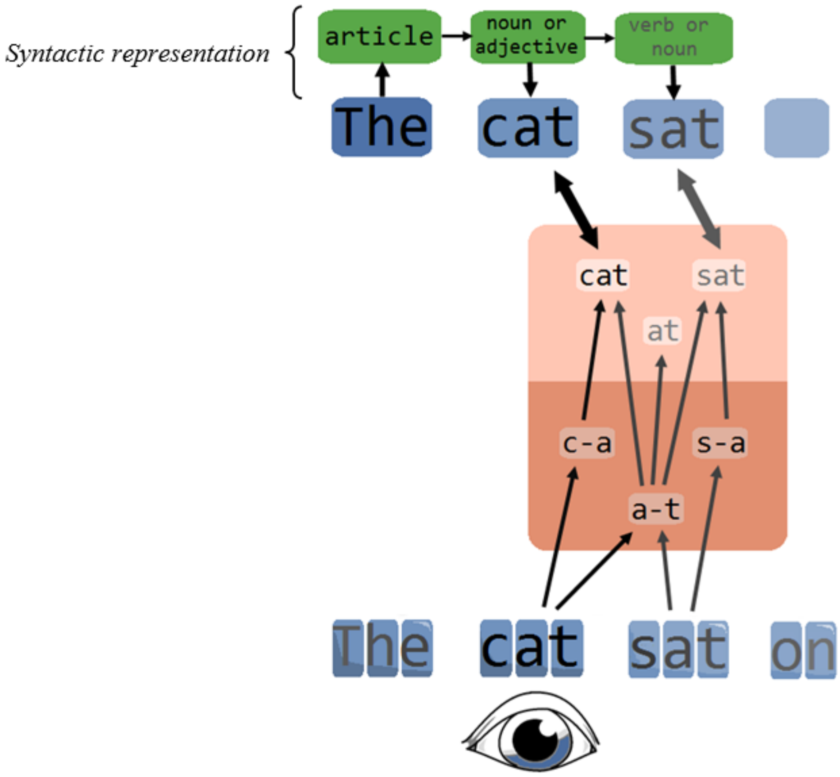

Figure 2. Our conceptualization of the reading system. Sub-lexical orthographic information is gathered across multiple words, with stronger activation of letters in the fovea (here *'cat'*) than letters in the parafovea. Sub-lexical information activates word representations and, importantly, parafoveal information may help to activate the word representation belonging to the fovea if there is orthographic overlap, accounting for the orthographic parafoveal-on-foveal effects reported in the literature. Activated word representations are projected onto a plausible location in a spatiotopic representation, based on visual features such as word length and shape. From here, recognized words append to a sentence-level representation that follows syntactic rules: for instance, if word *n* is recognized as an article, word *n+1* is expected to be a noun or adjective (in English). Feedback from the syntactic level to the individual word positions constrains the recognition process while allowing for the simultaneous recognition of multiple words.

In sum, we propose that multiple words can be processed in parallel beyond the sub-lexical level, leading to higher-order parafoveal-foveal integration when readers are set out to recognize no more than one word (Experiment 2). During sentence reading, however, sentence-level feedback to individual word positions would constrain the recognition of these words, counteracting integrative effects (e.g. faster syntactic categorization due to syntactically congruent adjacent words) and thus explaining why higher-order parafoveal-on-foveal effects in sentence reading have been elusive.

References

Abbot, M., Angele, B., Ahn, Y. & Rayner, K. (2015). Skipping syntactically illegal the previews: The role of predictability. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *41*, 1703–1704.

Angele, B. & Rayner, K. (2013). Processing the in the parafovea: Are articles skipped automatically? *Journal of Experimental Psychology: Learning, Memory and Cognition*. *39*. 649–662.

Angele, B., Tran, R. & Rayner, K. (2013). Parafoveal–foveal overlap can facilitate ongoing word identification during reading: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 526–538.

Baayen, R. (2008). *Analyzing Linguistic Data: A pratical introduction to statis*tics. Cambridge University Press.

Barr, D., Levy, R., Scheepers, C. & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015), Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67,* 1–48.

Bertram, R. & Hyönä, J. (2007). The interplay between parafoveal preview and morphological processing in reading. In van Gompel, R., Fischer, M., Murray, W. & Hill, R. (Eds.). *Eye movements: A window on mind and brain*. Oxford, UK: Elsevier.

Boston, M., Hale, J., Vasishth, S. & Kliegl, R. (2011). Parallel processing and sentence comprehension difficulty. *Language and Cognitive Processes*, 26, 301–349.

Brothers, T. & Traxler, M. (2016). Anticipating syntax during reading: Evidence from the boundary change paradigm. *Journal of Experimental Psychology: Learning, Memory & Cognition.* (in press)

Dalmaijer, E., Mathôt, S. & Van der Stigchel, S. (2014). PyGaze: An open-source, crossplatform toolbox for minimal-effort programming of eyetracking experiments*. Behavior Research Methods, 46,* 913–921.

Dare, N. & Shillcock, R. (2013). Serial and parallel processing in reading: Investigating the effects of parafoveal orthographic information on nonisolated word recognition. *The Quarterly Journal of Experimental Psychology, 66*, 417–428.

Deutsch, A., Frost, R., Pollatsek, A., & Rayner, K. (2005). Morphological parafoveal preview benefit effects in reading: Evidence from Hebrew. *Language and Cognitive Processes, 20,* 341–371.

Engbert, R., Nuthmann, A., Richter, E. & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112,* 777–813.

Flowers, J. & Wilcox, N. (1982). The effect of flanking context on visual classification: The joint contribution of interactions at different processing levels. *Perception & Psychophysics, 32*, 581–591.

Grabbe, J. W., & Allen, P.A. (2013). Age-related sparing of parafoveal lexical processing. *Experimental Aging Research, 39*, 419–444.

Grainger, J., Mathôt, S. & Vitu, F. (2014). Test of a model of multi-word reading: Effects of parafoveal flanking letters on foveal word recognition. *Acta Psychologica, 146*, 35–40.

Hohenstein, S., Laubrock, J. & Kliegl, R. (2010). Semantic preview benefit in eye movements during reading: A parafoveal fastpriming study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1150–1170.

Hohenstein, S. & Kliegl, R. (2014). Semantic preview benefit during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 166–190.

Hyönä, J. (1995). Do irregular letter combinations attract readers' attention? Evidence from fixation locations in words. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 68–81.

Inhoff, A., Radach, R., Starr, M. & Greenberg, S. (2000). Allocation of visuospatial attention and saccade programming during reading. In Kennedy, A., Radach, R., Heller, D. & Pynte, J. (Eds.). *Reading as a perceptual process*. Oxford, UK: Elsevier.

Jordan, T. et al. (2014) Reading direction and the central perceptual span: evidence from Arabic and English. *Psychonomic Bulletin & Review, 21*, 505–511.

Juhasz, B., Pollatsek, A., Hyönä, J., Drieghe, D. & Rayner, K. (2009). Parafoveal processing within and between words. *Quarterly Journal of Experimental Psychology*, *62*, 1356–1376.

Keuleers, E., Diependaele, K. & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology, 1*, 1–15.

MacDonald, M., Pearlmutter, N. & Seidenberg, M. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review, 101,* 676–703.

McConkie, G. & Rayner, K. (1975) The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics, 17*, 578–586.

Mathôt, S., Schreij, D. & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*, 314–324

Paterson, K., McGowan, V., White, S., Malik, S., Abedipour, L. & Jordan, T. (2014). Reading direction and the central perceptual span in Urdu and English. *PLOS ONE, 9*: e88358. doi: 10.1371/journal.pone.0088358

Radach, R. & Kennedy, A. (2004). Theoretical perspectives on eye movements in reading: Past controversies, current issues, and an agenda for future research. *European Journal of Cognitive Psychology, 16*, 3–26.

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

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

Rayner, K., Schotter, E. & Drieghe, D. (2014). Lack of semantic parafoveal preview benefit revisited. *Psychonomic Bulletin & Review, 21,* 1067–1072.

Reichle, E., Pollatsek, A., & Rayner, K. (2006). E-Z Reader: A cognitive-control, serial-attention model of eye movement behavior during reading. *Cognitive Systems Research, 7,* 4–22.

Reingold, E. & Stampe, D. (2004). Saccadic inhibition in reading. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 194–211.

Risse, S., Hohenstein, S., Kliegl, R. & Engbert, R. (2014). A theoretical analysis of the perceptual span based on SWIFT simulations of the n+2 boundary paradigm. *Visual Cognition, 22*, 283–308.

Schotter, E.. Angele, B. & Rayner, K. (2012). Parafoveal processing in reading. *Attention Perception & Psychophysics, 74*, 5–35.

Schotter, E. (2013). Synonyms provide semantic preview benefit in English. *Journal of Memory and Language, 69*, 619–633.

Schotter, E., Lee, M., Reiderman, M. & Rayner, K. (2015). The effect of contextual constraint on parafoveal processing in reading. *Journal of Memory and Language, 83,* 118–139.

Snell, J., Vitu, F. & Grainger, J. (2017). Spatial integration of parafoveal orthographic information during foveal word reading: Beyond the sub-lexical level? *The Quarterly Journal of Experimental Psychology.*

Snell, J., Declerck, M. & Grainger, J. (in review). Semantic processing in the parafovea revisited: Effects of translation equivalents in bilingual readers. *Language, Cognition & Neuroscience*, (in review).

Veldre, A. & Andrews, S. (2014). Parafoveal preview benefit is modulated by the precision of skilled readers' lexical representations. *Journal of Experimental Psychology: Human Perception & Performance, 41,* 219–232.

Vitu, F., Brysbaert, M. & Lancelin, D. (2004). A test of parafoveal-on-foveal effects with pairs of orthographically related words. *European Journal of Cognitive Psychology, 16,* 154–177.

Yan, M., Zhou, W., Shu, H. & Kliegl, R. (2012). Lexical and sublexical semantic preview benefits in Chinese reading. *Journal of Experimental Psychology: Learning, Memory and Conginition, 38,* 1069–1075.

Parallel semantic processing in reading revisited:
Effects of translation equivalents in bilingual readers

Joshua Snell, Mathieu Declerck & Jonathan Grainger

Citation: Snell, J., Declerck, M. & Grainger, J. (2018). Semantic processing in the parafovea revisited: Effects of translation equivalents in bilingual readers. *Language, Cognition & Neuroscience, 33,* 563-574.

Abstract

Previous research has failed to establish semantic parafoveal-on-foveal effects during reading. As an explanation, we theorize that sentence reading engages a sentence-level representation that prevents semantic parafoveal-foveal integration. Putting this account to the test, we examined parafoveal-foveal influences both in- and outside a sentence reading setting. Optimizing chances of establishing parafoveal-on-foveal effects, we used translation-equivalent word pairs with French-English bilingual participants. Experiment 1 provided no evidence for semantic parafoveal-on-foveal integration during sentence reading, but some evidence that semantic information had been extracted in parallel from multiple words. Experiments 2 and 3 employed a flanker paradigm in which participants semantically categorized English foveal target words, while these were flanked by the French translation or an unrelated French word (stimulus on-time 170 ms). Performance was drastically better with translation flankers, suggesting that readers can integrate semantic information across multiple words when the task does not require a strict separation of higher-order information.

While decades of reading research have yielded much insight into how foveal isolated words are processed and recognized, the mechanisms of parafoveal word processing are still far from understood. To gain understanding in the exact nature of parafoveal word processing, the field is required to answer two fundamental questions: (i) how deeply are parafoveal words processed, and (ii) how does the extraction of information from the parafovea influence foveal word processing and vice versa? It is unlikely that investigations into these matters will yield singular answers, as multiple lines of research have suggested that they are dependent on several factors, such as the language at hand (e.g. Deutsch, Frost, Pollatsek & Rayner, 2005; Bertram & Hyönä, 2007; Juhasz et al., 2009; Yan, Zhou, Shu & Kliegl, 2012; Schotter, Angele & Rayner, 2012 for a review) and inter-individual differences (e.g. Veldre & Andrews, 2014). Perhaps unsurprisingly then, the field has not yet come to a consensus, and years of research has divided scientists into, roughly, two schools of theorizing. One of these schools assumes that higher-order (e.g., lexical, semantic, syntactic) word processing occurs serially, i.e., for one word at a time, with attention moving from one word to the next when the former has been recognized (e.g. Angele, Tran & Rayner, 2013; Rayner, Schotter & Drieghe, 2014; Reichle, Pollatsek & Rayner, 2006). This means that if an upcoming word (*n+1*) can be lexically accessed at all, this would happen only after recognition of the foveal word (*n*), implying that the processing of word *n* should not be influenced by higher-order features of word *n+1.* The other school assumes that multiple words may be processed in parallel, as visuo-spatial attention would be distributed across multiple words as a gradient (Engbert, Nuthmann, Richter & Kliegl, 2005; Radach & Kennedy, 2004; Reilly & Radach, 2006). Consequently, it may be that the lexical properties of upcoming words are processed simultaneously with foveal words, and that higher-order features from upcoming words influence the foveal word recognition process.

For written languages that use an alphabetic script, there is considerable evidence that multiple words can be processed in parallel at the sub-lexical level. The principal finding is that foveal words are recognized faster when they are orthographically related to adjacent (upcoming) words – a so-called *parafoveal-on-foveal* effect (e.g. Angele et al., 2013; Dare & Shillcock, 2013; Grainger, Mathot & Vitu, 2014; Inhoff, Radach, Starr & Greenberg, 2000; Radach & Kennedy, 2004; Snell, Vitu & Grainger, 2017a) – suggesting not only that upcoming words are processed to some extent prior to being fixated, but also that this happens *during* rather than *after* foveal word processing.

For higher-order parallel processing, matters seem to be more complicated. There is evidence that both semantic and syntactic information can be extracted from upcoming words prior to these words being fixated (see below); however, this is not direct evidence for parallel processing. Under the assumption of serial processing, for example, it is possible that higher-order processing of the upcoming word occurs during the interval in which the foveal word has been recognized but the eyes have not yet moved to the upcoming word (e.g. Schotter, Reichle & Rayner, 2014). It is important to note, however, that while evidence for higher-order parafoveal processing is compatible both with serial and parallel processing, an *absence* of such evidence would argue directly against parallel processing. Using the gaze-contingent boundary paradigm (Rayner, 1975) in German sentence reading, Hohenstein and Kliegl (2014) found that a target word (*n*) was recognized faster when there was a semantically related preview word at the target location when readers were fixating the pre-target (*n-1*). However, using the same paradigm, Rayner et al. (2014) could not establish such a semantic parafoveal preview benefit in English sentence reading. Similarly, Altarriba, Kambe, Pollatsek, and Rayner (2001) failed to find a preview benefit from translation equivalents during sentence reading by Spanish-English bilinguals. On the other hand, Schotter (2013) did find a semantic parafoveal preview effect in

English sentence reading, but only when previews and targets were synonyms (e.g. *start – begin*), and not when they were associatively related (e.g. *ready – begin*), indicating that the likelihood of finding such an effect might depend on the strength of the relationship between words in some languages. More recent work by Veldre and Andrews (2016) further suggests that the plausibility of the preview word within the sentence context is a key factor, with semantic preview benefits being largely eradicated when the preview was not contextually plausible.

Finally, two recent studies have found evidence for syntactic processing of upcoming words. Snell, Meeter, and Grainger (2017b) found that target processing was facilitated when previews were syntactically congruent with the target (i.e., from the same syntactic category), as compared to when previews were from a different syntactic category. In a similar vein, Brothers and Traxler (2016) found that English readers were less likely to skip an upcoming word if it violated syntactic rules (e.g. noun followed by a noun).

While an absence of parafoveal preview effects would argue directly against parallel processing (and thus in favor of serial processing), higher-order *parafoveal-on-foveal* effects would argue directly against serial processing—the rationale being that processing of the upcoming word could only influence recognition of the fixated word if the latter was still being processed. Here, however, reports are again inconsistent. Snell et al. (2017a) showed that when word *n+1* was a high-frequency orthographic neighbor of word *n* (e.g. *rock rack*), processing of word *n* was facilitated, and not inhibited, as might be expected following lexical competition between orthographic neighbors (e.g. Davis & Lupker, 2006; Segui & Grainger, 1990). Furthermore, Angele et al. (2013) showed that processing of word *n* was not facilitated by a semantically related word *n+1* in English sentence reading. On the other hand, Inhoff et al. (2000) did find such a facilitatory effect in German sentence reading. Yet, as has been pointed out by Angele et al. (2013), it is unclear whether the effect reported by Inhoff et al. (2000) was purely semantic, or possibly orthographic in nature (e.g., processing of *'mother'* might be facilitated by '*father*' due to semantic relatedness, but also due to orthographic relatedness).

The study of Inhoff et al. (2000) taken aside, the general absence of higher-order parafoveal-on-foveal effects has been taken as evidence against parallel processing (e.g. Angele et al., 2013). However, here it should be noted that the premise that higher-order parafoveal-on-foveal effects evidence parallel processing, does not logically imply that an absence of such effects disproves parallel processing. As argued by Snell et al. (2017b), if multiple words are processed in parallel, it would be quite problematic if higher-order information is integrated across these words, given that each word has a distinct role in contributing to sentence comprehension. It would therefore seem more likely that if words are indeed processed in parallel, there would be a mechanism at play that allows readers to keep track of separate word identities, thus enabling independent extraction of semantic and syntactic information from words in the visual field. The mechanism that was theorized by Snell et al. (2017b) is a spatiotopic sentence-level representation in working memory, onto which activated lexical representations would be mapped (Figure 1). The location that an activated word is appointed in the sentence-level representation is determined by low-level visual cues (e.g. expectations about word length) and top-down grammatical constraints (e.g., 'I have recognized an article at position *n*, so I expect an adjective or noun at position *n+1*'). Similarly, top-down feedback from the sentence-level representation to individual words may attenuate or enhance their activation (e.g., the visual input '*This beer tastes good*' may activate both '*beer*' and '*been*', but the latter would be rejected due to grammatical constraints). Semantic and syntactic information associated with specific word identities at specific positions in the sentence would then provide the essential ingredients for computing sentence meaning. In sum, this account proposes that higher-order semantic and syntactic information can be processed in parallel across multiple words, but that the constraints associated with sentence reading ensure that such information is only integrated at the sentence level; hence the absence of higher-order parafoveal-on-foveal influences during sentence reading.

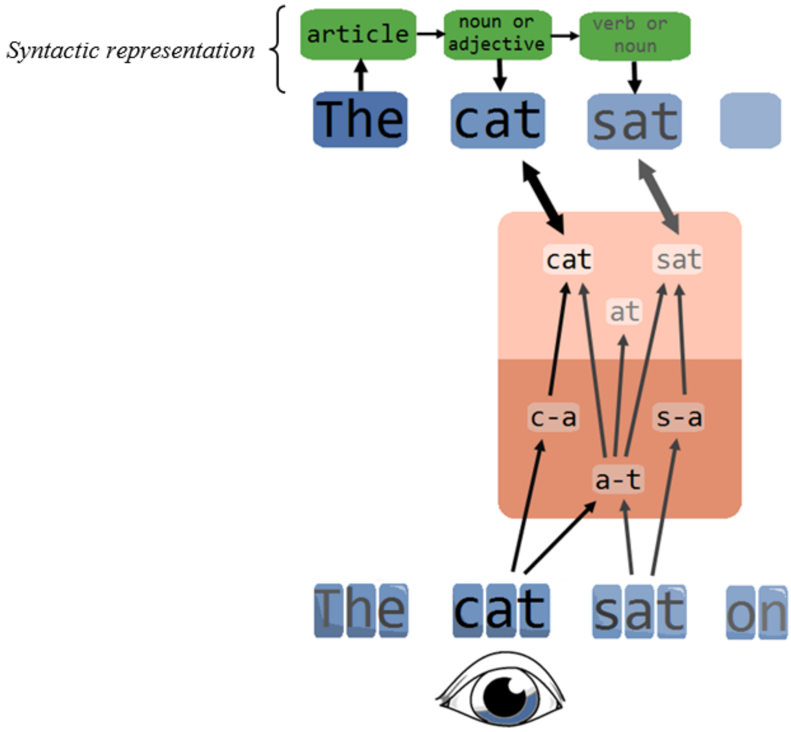

Figure 1. Our conceptualization of the reading system as proposed in Snell et al. (2017b). Sub-lexical orthographic information is gathered across multiple words, with stronger activation of letters in the fovea (here *'cat'*) than letters in the parafovea. Activated word representations are projected onto a plausible location in a spatiotopic representation, based on visual features such as word length and shape. From here, recognized words append to a sentence-level representation that follows syntactic rules: for instance, if word *n* is recognized as an article, word *n+1* is expected to be a noun or adjective (in English). Feedback from the syntactic level to the individual word positions constrains the recognition process for these words.

One key prediction of this model is that parafoveal-on-foveal effects as evidence for parallel processing should be observable in a paradigm that does not require sentence-level comprehension. In line with this prediction, we found that readers were faster and more accurate to classify Dutch foveal targets as noun or verb, when these were flanked by syntactically congruent flankers as compared with incongruent flankers (Snell et al., 2017b).

Further in line with the predictions of the model, in sentence reading, we found that syntactic information was extracted in parallel from– but not integrated across, the fixated and upcoming word: syntactically congruent word *n*+1's caused a tendency for *increased* rather than decreased fixation durations on word *n,* along with an increased fixation rate on word *n*+1 (Snell et al., 2017b). This is likely because the reading process was disturbed by readers' awareness of the grammatically incorrect continuation of the sentence (see also Snell et al., 2017a).

In the present study we provide a further test of higher-order parallel processing in both a sentence reading paradigm and a flanker paradigm, and this time for semantically rather than syntactically related parafoveal words, thus potentially unveiling differences in the natures of semantic parallel processing and syntactic parallel processing as investigated in Snell et al.

(2017b). Indeed, the sentence-level feedback mechanism that was theorized here and in Snell et al. (2017b) mainly revolves around grammatical constraints, and it is possible that the behavioral patterns that were established in our previous study do not hold when using semantic relatedness.

We chose to test for effects of translation equivalents in bilingual participants, since translation equivalence arguably provides the strongest semantic relation between two words (e.g. Grainger & Frenck-Mestre, 1998; Perea, Duñabeitia & Carreiras, 2008; Duñabetia, Perea & Carreiras, 2010), thus maximizing the chances of obtaining semantic parafoveal-on-foveal effects. This provides a strong test of our model that predicts no such effect in sentence reading, accompanied by evidence for parallel processing of semantic information in the flanker paradigm.

Experiment 1

Methods

Participants

30 French-English bilingual students (11 female) from the Aix-Marseille Université (Marseille, France) gave written informed consent to their participation in this experiment. Participants earned €5 each for participating. All participants reported to be native to the French language, non-dyslexic and had normal or corrected-to-normal sight. The participants were naive with regard to the purpose of the experiment.

Prior to the actual experiment, the participant's proficiency in both French and English was tested with the *LexTALE* language proficiency test (Brysbaert, 2013; Lemhöfer & Broersma, 2012). From the initial 30 participants, 25 succeeded the English test with a score of at least 60%. We then recruited more participants to bring the total amount of proficient bilingual participants back to 30. The average score for French and English was 84.0% and 67.8% respectively.

	Organic	*Inorganic*
Target	w<u>ol</u>f	h<u>o</u>le
Translation	<u>lo</u>up	tr<u>o</u>u
Control	<u>lo</u>ge	pi<u>o</u>n

Figure 2. Stimuli examples. As can be seen, if any of a target's constituent letters appear in its French translation word (underlined), these letters would also appear at the same position in the French control word.

Materials

We generated a set of 100 English target words with a length ranging between 3–6 letters. The French translation equivalent for each of these targets was a non-cognate word (e.g. *mist – brume*) with a length ranging between 3–7 letters. For every translation pair we retrieved a French control word from the French Lexicon Project database (Ferrand et al., 2010) that was similar to the translation word in terms of length and orthographic overlap with the target (see Figure 2).

We also made sure that the average frequency of translations and controls was equal, at 4.72 and 4.82 Zipf, respectively.[31]

For every target we constructed an English sentence, fitting on a single line with length ranging between 23 and 47 characters (including spaces). Sentences contained 7.76 words on average (min. 5, max. 13 words). During stimulus presentation, we manipulated the post-target (n+1) prior to and during the fixation on the target (n), such that the post-target was either the French translation or the French control word. When participants moved their eyes from the target to the post-target, the latter would change into an English word that formed a logical continuation of the sentence (Figure 3). The average target/post-target boundary location was 15 characters from the start of the sentence (min. 8, max. 23 characters). For one out of every five sentences we also created a quiz question, shown directly after the sentence and to be answered with a two-button response. These served as catch trials to make sure that participants were reading for meaning.

Translation	`The strong wind	vent blow her away`
Control	`The strong wind	gant blow her away`
Post-boundary	`The strong wind	will blow her away`

Figure 3. The upper two sentences show what a stimulus could look like in respectively the translation and control condition *before* the eyes crossed the invisible boundary (here marked by the vertical line). As soon as the eyes moved beyond the boundary, the post-target changed into a grammatically correct continuation of the sentence.

Design

Our experimental design consisted of two post-target word conditions (*translation / control*). A Latin-square design was used to ensure that all 100 sentences were presented in all conditions but only once per participant. The experiment thus consisted of 100 trials (50 with post-target translation word and 50 with post-target control word), and these were presented in randomized order.

Apparatus and software

The stimuli and experimental design were implemented with OpenSesame (Mathôt, Schreij & Theeuwes, 2012), with the PyGaze back-end (Dalmaijer, Mathôt & Van der Stigchel, 2014) to process eye movement data online. With an EyeLink 1000 (SR Research, Mississauga, ON, Canada), a video-based eye tracker sampling at 1000Hz with a spatial resolution of 0.01°, the reader's right eye position was recorded. Stimuli were presented on a 1024x768 px, 150 Hz computer monitor. Participants were seated at a distance of 90 cm from the display, so that each character space subtended 0.35 degrees of visual angle. Manual responses were collected with a keyboard. A chin-rest was used to stabilize the head position.

[31] For more on the Zipf frequency scale, see Van Heuven, Mandera, Keuleers and Brysbaert (2014).

Procedure

Before commencing the experiment, the right eye was calibrated using a 3-point horizontal calibration grid with fixation points appearing in randomized order. In case of a sufficient match between the calibration grid and fixation grid, a validation was carried out to double-check the accuracy of the initial fixations. Participants then received instructions both verbally by the experimenter and visually on screen.

A drift correction was performed before the start of every trial. In case of a successful calibration, a forward slash (/) was presented as a fixation point, at a location that matched the start of the sentence when it appeared. As soon as the eyes had stabilized on the slash (within a 0.70 degrees range) for 700ms, the sentence stimulus appeared with the first letter aligned to the fixation location.

As participants read the sentence, the position of the eyes was tracked online. As soon as the eyes moved beyond an invisible boundary, the x-coordinate of which marked the exact middle between the target and the post-target word, the latter changed into a logical continuation of the sentence. When participants reached the end of the sentence, a green dot would briefly appear to the right of the sentence as a means to give recognition that the sentence was read. Shortly afterwards, the next trial would commence. However, if the sentence belonged to one of the 20 sentences for which we created a quiz question, participants would first see a display with that question and two possible answers in the left and right bottom corner of the screen. The participants had to choose one of these answers with respectively a left- or right-handed keyboard button response.

Participants were asked not to blink while reading the sentences but rather in-between the trials, because the temporary loss of corneal reflection could cause imprecise gaze estimations. The experiment lasted about 20 minutes and participants were allowed to take a break at their own convenience.

Results

From the total of 3000 trials across participants, 234 trials (7.80%) were discarded due to eye-blinking or a premature boundary change (due to landing too close to the boundary). We computed three fixation duration measures: the first fixation duration (FFD), gaze duration (GD) and total viewing time (TVT). Here, FFD refers to the first fixation on a target word, regardless of whether there is more than one fixation on this word. GD refers to the sum of all first-pass fixation durations on a target word, while TVT refers to the sum of all fixations on a target word, that is, including fixations after a regression. We also calculated three probability values: the probability that the target was skipped, the probability that the target was refixated during first-pass, and the probability that the target was refixated by means of an inter-word regression.

For the duration measures we used linear mixed-effect (LMM) models with items and subjects as crossed random effects (Baayen, 2008). The models were fitted with the lmer function from the lme4 package (Bates, Maechle, Bolker & Walker, 2015) in the R statistical computing environment. We report regression coefficients (b), standard errors (SE) and t-values for all factors. Fixed effects were deemed reliable if $| t | > 1.96$ (Baayen, 2008). Logistic LMM models (fitted with the glmer function) were used to analyze the skipping, refixation and regression probabilities. Here, the z-values can be interpreted in the same way as the t-values. In all analyses, values beyond 2.5 SD from the condition mean, (on average 2.06% of the trials), were marked as outliers and excluded.

Table 1. Mean fixation times and probabilities for Experiment 1.

	Condition means					
	FFD	GD	TVT	Skip	Refix	Regress
Translation	229.7 (87.7)	394.4 (217.1)	611.6 (425.9)	.08 (.08)	.47 (.16)	.40 (.14)
Control	233.2 (88.5)	393.4 (224.3)	614.8 (408.8)	.08 (.07)	.47 (.13)	.44 (.15)

Note: values in between parentheses indicate standard deviations. Abbreviations: FFD, first fixation duration; GD, gaze duration; TVT, total viewing time.

As it turned out, there were no significant effects in the fixation duration measures (Tables 1 and 2). The relatively low skipping rate, high refixation and regression rates and high GD and TVT values indicate that our participants' eye-movement behavior consisted of many fixations connected by short saccades, which, as has been argued by Rayner (1998), is typical of low-proficiency reading behavior.[32] The rate of regressions was significantly increased for the control condition as compared to the translation condition (Table 3). Although this may be an indication that comprehension of the target word was better in the translation condition, this was not reflected in the TVT. There were no significant effects in the other probability values.

Table 2. Analyses of fixation duration measures for Experiment 1 (ref: control).

	FFD			*GD*			*TVT*		
	b	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*
(Intercept)	233.06	6.44	**36.21**	390.57	18.59	**21.01**	624.74	41.46	**15.07**
Translation	−3.26	3.16	−1.03	5.26	7.51	0.70	−13.05	12.24	−1.07

Note: significant values are shown in bold.

Table 3. Analysis of probability values for Experiment 1 (ref: control).

	Skip			Refix			Regress		
	b	SE	z	b	SE	z	b	SE	z
(Intercept)	−2.90	0.20	**−14.25**	−0.03	0.14	−0.23	0.04	0.23	0.16
Translation	0.00	0.14	−0.03	0.04	0.08	0.45	−0.26	0.09	**−2.86**

Note: significant values are shown in bold.

A Bayesian analysis of the null-hypothesis (i.e., that target words are not influenced by the semantic relatedness of post-target words in sentence reading) was carried out for the FFD, GD and TVT measures (Rouder, Speckman, Sun, Morey & Iverson, 2009). We found positive evidence favoring the null-hypothesis in all these measures, with BF_{01} = 3.17 for FFD, BF_{01} = 4.11 for GD, and BF_{01} = 3.05 for TVT (Kass & Raftery, 1995).

Lastly, we assessed whether the participant's L2 (English) proficiency had an impact on the (absence of an) effect of our manipulation. L2 proficiency was entered in a separate LMM as a 2-level factor, determined by whether the participant's LexTale score was above or below the group median. In none of the measures L2 proficiency modulated the effect of our manipulation (FFD: b = 5.80, SE = 6.43, t = 0.90; GD: b = 7.12, SE = 15.25, t = 0.47; TVT: b = 3.00, SE = 24.91, t = 0.12; skips: b = 0.04, SE = 0.30, z = 0.12; refixations: b = 0.11, SE = 0.17, z = 0.67; regressions: b =

[32] These behavioral patterns were not restricted to the target word: for instance, overall, 27% of the saccades were regressions in our study, whereas 10-15% would be typical of normal reading behavior (Rayner, 1998).

0.14, SE = 0.19, z = 0.74). There was nonetheless a marginally significant main effect of L2 proficiency on the fixation duration (FFD: b = 21.62, SE = 12.02, t = 1.80). Interestingly however, the direction of this effect was such that greater proficiency led to longer fixation durations.

Discussion

Experiment 1 failed to find evidence for semantic parafoveal-on-foveal effects even in conditions where the semantic relation between the parafoveal and foveal word was maximal (i.e., translation equivalents). The strength of the semantic relation between translation equivalents is attested by research using the masked priming paradigm (Forster & Davis, 1984) showing robust effects of non-cognate translation primes in conditions where other types of semantic relation typically do not exhibit priming (e.g., Grainger & Frenck-Mestre, 1998; Perea et al., 2008; see Duñabetia et al., 2010, for a review). These translation priming effects are particularly robust when primes are in the first language (L1) and targets in the second language (L2), which corresponds to the parafoveal word in L1 and target in L2 in Experiment 1.

One could argue that the absence of a parafoveal-on-foveal influence of translation equivalents in Experiment 1 is due to the fact that participants were reading in a strictly monolingual context, and therefore that words from the other language would not be activated. Here, however, it is important to note that there is clear evidence that translation equivalents are activated in a strictly monolingual context (e.g., Thierry & Wu, 2007) in line with non-selective accounts of lexical access in bilinguals (e.g., Grainger & Dijkstra, 1992). Furthermore, the fact that there were fewer regressions in the translation condition than in the control condition suggests that semantic information was extracted from the parafoveal word, but that this information did not impact on foveal word processing, as predicted by our model.

Experiment 2 provides a test of the other side of the theoretical coin described in the Introduction. That is, that evidence for parallel semantic processing across multiple words should be observable in a task that does not require a strict separation of higher-order information. This task is the flanker task, with horizontally arranged flanker words placed left and right of a central target word, thus mimicking stimulus presentation in sentence reading. Crucially, and contrary to previous semantic flanker studies showing facilitatory effects of semantic relatedness (Shaffer & Laberge, 1979) and effects of translation equivalents in bilinguals (Guttentag, Haith, Goodman & Hauch, 1984)[33], target and flankers were presented very briefly (170 ms) to prevent the possibility that there would be enough time to process the flankers after the target was recognized, (i.e., there would only be time to process flankers *during* target processing). The short presentation time further prevented that any eye movements were made to the flanking stimuli. Using the same stimuli as in Experiment 1, we tested for effects of L1 translation equivalent flanker words on the semantic categorization of L2 target words in bilingual participants.

Experiment 2

Method

Participants

Twenty French-English bilingual students (12 female) from the Aix-Marseille University (Marseille, France) gave written informed consent to their participation in this experiment.

[33] This prior research used vertically arranged flankers positioned above and below centrally located targets, and target and flankers remained on-screen until participants responded.

Participants earned €4,- or its equivalent in course credit. All participants reported to be native to the French language, non-dyslexic, and had normal or corrected-to-normal vision. Further, all participants had *LexTALE* language proficiency test scores (Brysbaert, 2013; Lemhöfer & Broersma, 2012) of at least 60% (average scores for French and English were 86.6% and 65.0%, respectively).

Materials
The targets and post-targets from Experiment 1 were used as targets and flankers in Experiments 2 and 3. Of these targets, 50 corresponded to a natural object (e.g. *duck, neck, king*) and 50 corresponded to an artifactual object (e.g. *bridge, skirt, pen*).

Design
Our experimental design consisted of two flanker type conditions (*translation / control*). Every target was presented twice to each participant: once flanked by the translation, and once flanked by the control. The experiment thus consisted of 200 trials (100 with *natural* as correct response and 100 with *artifactual* as correct response), and these were presented in randomized order.

Apparatus and software
The stimuli and experimental design were implemented with OpenSesame (Mathôt, Schreij & Theeuwes, 2012). Stimuli were presented on a 1024x768 px, 150 Hz computer monitor. Participants were seated at a distance of 90 cm from the display, so that each character space subtended 0.35 degrees of visual angle. Manual responses were collected with a keyboard.

Procedure
After taking the language proficiency test, participants received instructions both verbally by the experimenter and visually on screen. Every trial would start with two centrally positioned vertical fixation bars (see Figure 4). After 600 ms, the target word appeared in between these fixation bars, with the French translation or control word flanking its left and right side, (targets and flankers were separated by one character space on each side). After 170 ms, the target and flankers disappeared, and participants had 2000 ms to indicate whether they had recognized the target as being a natural or artifactual object. This was done with a left- or right-sided button press ('w' and '!' respectively on an azerty layout keyboard), with the right button always corresponding to 'natural'. A green or red dot was then briefly shown at the center of the screen, depending on whether the participant's response was correct or incorrect respectively, shortly after which the next trial would commence. Before the start of the experiment, a set of twelve practice trials was run to allow participants to become acquainted with the procedure. A break was offered halfway through the experiment. The experiment lasted approximately 20 minutes in total.

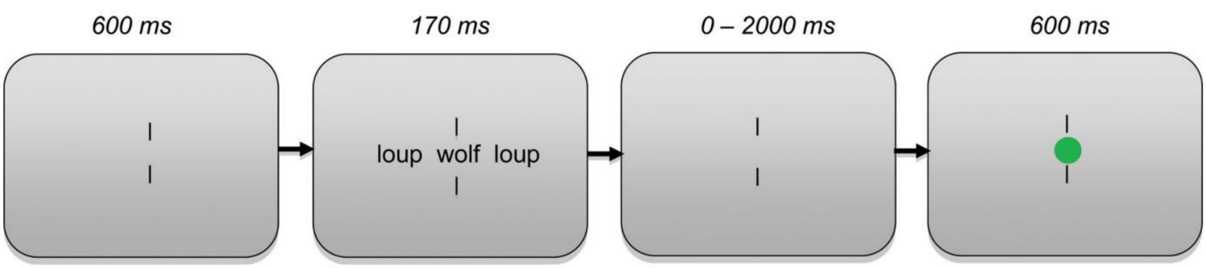

Figure 4. Overview of the trial procedure. The size of stimuli relative to the screen is exaggerated in this example.

Results

Trials where the response time (RT) was beyond 2.5 standard deviations from the condition mean (3.63% of all trials) were discarded. Only correctly answered trials (78.15% of all trials) were included in the RT analyses. For our RT analyses we again used LMMs with items and subjects as crossed random effects, fitted with the lmer function from the lme4 package (Bates et al., 2015) in the R statistical computing environment. We report regression coefficients (b), standard errors (SE) and t-values. Logistic LMMs (fitted with the glmer function) were used to analyze the error rates.

Table 4 shows the mean RTs and error rates for the translation and control condition. RTs were significantly lower in the translation condition than in the control condition, with b = 40.25, SE = 7.17, t = 5.62. This effect of flanker type did not differ significantly between trials where the target was a natural object and trials where the target was an artifactual object: b = 18.36, SE = 14.34, t = 1.28. The error rate was significantly lower in the translation condition as well, with b = 0.40, SE = 0.09, z = 4.41.

Table 4. Mean RT's and error rates for the translation and control condition of Experiment 2.

	RT	Error
Translation	625.03 (232.04)	.166 (.060)
Control	662.91 (246.38)	.216 (.071)

Discussion

In Experiment 2, we set out to investigate whether readers can semantically process multiple words in parallel. To maximize our chances of finding evidence for such parallel processing of semantic information, we used non-cognate translation equivalents with bilingual participants, as in Experiment 1. The results of Experiment 2 suggest that readers indeed process semantic information across foveal and parafoveal stimuli, as target processing was strongly facilitated by translation flankers, compared to control flankers. This effect could not have been caused by sub-lexical factors such as orthographic overlap, as the orthographic overlap (letter identity and position) with the target was identical for translations and controls. Moreover, the stimulus presentation time was considerably short (170 ms), suggesting that there would not have been enough time to process the flankers after the target was recognized, as would otherwise be possible under the assumption of serial processing. The size of the effect (b = 40.25) is considerably larger than those found in previous implementations of this paradigm, such as an earlier study in which we used syntactically related stimuli (b = 17.45; Snell et al., 2016b) and studies that examined sub-lexical parafoveal-on-foveal effects (Dare & Shillcock, 2013; Grainger et al., 2014; Snell et al., 2016a; all b-values <20.00), indicating that stronger higher-order relations between foveal and parafoveal words may indeed lead to greater parafoveal-on-foveal influences.

The primary aim of Experiment 3 was to replicate the results of Experiment 2, while including a third *no-flanker* condition. According to our model, attentional resources are distributed across multiple words. As a result, orthographically unrelated flanker stimuli, such as the translation and unrelated flankers used in Experiment 2, should interfere with the orthographic processing of the target. This should lead to increased difficulty in target processing in the presence of flanking words compared with a no-flanker condition.

Experiment 3

Method

The methodology for Experiment 3 was nearly identical to that of Experiment 2, the only difference being the inclusion of a third, *no-flanker* condition. The 100 targets from Experiment 2 were also used in Experiment 3. This time, each target was presented three times to every participant, corresponding to the three flanker conditions, thus making the total amount of experimental trials 300. We again recruited twenty students (15 female), none of whom had participated in the previous experiments. This group of participants had average *LexTALE* language proficiency scores of 92.0% and 63.8% for French and English, respectively. The experiment lasted approximately 25 minutes.

Results

We applied criteria identical to those used in Experiment 2 for the exclusion of trials in Experiment 3. We ended up with data from 79.87% of all trials for the analysis of RTs. We again employed LMMs for the analyses of RTs and error rates. Condition means for Experiment 3 are shown in Table 5. Importantly, we replicated our results from Experiment 2, as RTs in the translation condition were again significantly decreased as compared to the control condition: $b = 35.59$, SE = 6.87, $t = 5.18$. We also hypothesized that orthographically unrelated flanker words should interfere with the orthographic processing of target words hence slowing target word processing compared with the no-flanker condition. This was indeed the case. The no-flanker condition yielded even lower RTs than the translation condition: $b = 28.83$, SE = 6.84, $t = 4.22$. As in Experiment 2, the flanker effect did not differ between trials where the target was a natural object and trials where the target was an artifactual object: $b = 9.01$, SE = 6.89, $t = 1.31$. Although the error rate was again numerically lower in the translation condition than in the control condition, the difference did not reach significance this time around: $b = 0.14$, SE = 0.09, $z = 1.46$. The error rates did not differ between the translation and no-flanker condition either: $b = 0.10$, SE = 0.09, $z = 1.06$.

Table 5. Mean RT's and error rates for the no-flanker, translation and control condition of Experiment 3.

	RT	Error
No-flanker	608.44 (219.50)	.178 (.081)
Translation	638.95 (224.08)	.168 (.073)
Control	672.47 (232.76)	.179 (.092)

Note: values in between parentheses indicate standard deviations.

Discussion

We replicated our results from Experiment 2 in Experiment 3. Together, these experiments provide clear evidence that readers cannot effectively focus their attention on single words, causing higher-order processing to take place in parallel for foveal and parafoveal words. The fact that the no-flanker condition yielded even lower RTs than the translation condition is in line with our model, according to which orthographically unrelated flankers will interfere with target word processing. Thus, in the absence of grammatical constraints, the inevitable processing of multiple words leads to cross-leakage of information both at the sub-lexical level, where orthographically unrelated flanking stimuli (i.e., both the non-cognate translation equivalents and the unrelated control words) interfere with orthographic processing of the target word, as well as beyond the lexical level, where semantic information extracted from the translation equivalents facilitates semantic categorization.

General discussion

In the present study we examined semantic parafoveal-on-foveal effects by testing bilingual participants with parafoveal words that were translation equivalents of the foveal target word. Our prior research investigating syntactic parafoveal-on-foveal effects (Snell et al., 2017b) pointed to parallel independent extraction of syntactic information from multiple words during sentence reading. The independent nature of such parallel processing, induced by top-down grammatical constraints, was taken as the explanation for why no syntactic parafoveal-on-foveal effects were observed. However, when integration of syntactic information was beneficial for the task at hand, we found evidence for parallel processing of syntactic information. This task involved syntactic classification of a central target word flanked by unrelated flanker words that could be from the same or a different syntactic category. Syntactic information extracted from multiple words in parallel could then be pooled into a single response channel, hence the facilitation from syntactically congruent flankers. The present study built directly on this prior work, but now testing for semantic parafoveal-on-foveal effects, and using what is arguably the strongest semantic relation between two words – translation equivalence. We predicted an absence of semantic parafoveal-on-foveal effects in sentence reading, accompanied by evidence for parallel processing of semantic information in the flanker paradigm.

Experiment 1 employed a parafoveal-on-foveal manipulation with the boundary technique during sentence reading. The presence of a translation equivalent at position *n*+1 was found to have no significant influence on the processing of word *n*, as revealed by fixation duration measures. Some evidence for parallel semantic processing of words *n*+1 and *n* was found, however, in the form of a decreased rate of regressive saccades to word *n* when *n*+1 was a translation of *n*. Yet, the decreased regression rate was not reflected in the total viewing time (TVT) on the target, suggesting that parallel semantic processing may influence higher levels of reading comprehension, rather than processing of individual words. Furthermore, given the recent results of Veldre and Andrews (2016), the absence of a semantic parafoveal-on-foveal effect in Experiment 1 could be due to the fact that the translation equivalent did not fit plausibly into the sentence context – a scenario that fits well with how the spatiotopic sentence-level representation (Figure 1; Snell et al., 2017b) would operate.

Experiments 2 and 3 used a flanker paradigm with horizontally arranged flanker words located to the left and to the right of a central target word. Flanker words could be the translation equivalent in L1 of the target word in L2, or completely unrelated L1 words. Semantic categorization of target words was facilitated by the translation flankers relative to the unrelated flankers. The stimulus presentation time was considerably short (170 ms), such that there would not have been enough time to process the flankers after recognition of the target, as would otherwise be possible if words were processed serially rather than in parallel. Furthermore, in Experiment 3 we found that target identification was faster when there were no flanker stimuli compared with the translation flanker condition. This is captured in our model by the interference generated from orthographically unrelated flanking stimuli, following an attentional distribution spanning multiple words. Semantically related flankers help reduce this flanker interference by providing congruent semantic input into the mechanism that decides whether the target word is a living thing or not.

One might wonder why Altarriba et al. (2001) failed to find an influence of parafoveal translation equivalents in a paradigm (their Experiment 1) that shares certain similarities with the flanker paradigm used in Experiments 2 and 3 of the present study. In the Altarriba et al. study, bilingual participants had to fixate a central fixation cross while a parafoveal word was presented at 2° of eccentricity (fixation cross to beginning of word) to the right of fixation. Participants made an eye movement to the parafoveal stimulus, and during that eye movement the preview word was replaced with the target word that participants had to read aloud as rapidly as possible. One major difference with respect to our flanking paradigm is that the parafoveal word and foveal word were presented sequentially at the same location in Altarriba et al.'s study, as opposed to the parallel presentation at different locations in our study. Thus, any benefit of a semantically related preview in the Altarriba et al. study might have been cancelled by the interference caused by having orthographically different stimuli appear at the same location (see also e.g. Kliegl, Hohenstein & McDonald, 2013; Marx, Hawelka, Schuster & Hutzler, 2017, regarding the interplay of preview benefit versus preview cost).

Overall, the results of the present study are in line with the predictions of a new model of parallel word processing and reading (Snell et al., 2017b). According to this model, orthographic information spanning several words is integrated into a single processing channel (Grainger et al., 2014; 2016), hence explaining orthographic parafoveal-on-foveal effects. Orthographic word identities continue to be processed in parallel, nevertheless, with each word identity being associated with a particular position in the sentence that is being read. This parallel, independent, location-specific processing of word identities enables parallel independent activation of semantic and syntactic information from multiple words, which then feed information into higher-level sentence comprehension processes (Figure 1). The independent nature of word-level processing means that neither semantic nor syntactic parafoveal-on-foveal effects should be observed. On the other hand, in paradigms where this information can be pooled in order to generate a response, one can demonstrate parallel processing of semantic and syntactic information across multiple words.

To finish on a methodological note, a common criticism of reading research using static paradigms such as the flanker paradigm, is that these paradigms do not reflect normal reading. In response to this, we would point out that such paradigms provide theoretical leverage that cannot be achieved in sentence reading paradigms, and that what is crucial here is the possibility to create fundamental connections between processing involved in the simplified paradigms and processing involved in the more complex, and naturalistic, sentence-reading paradigms. Our model of parallel word processing allows us to establish such connections and to use the data obtained from multiple paradigms to inform the general mechanisms involved in everyday reading. Crucially, the flanker paradigm shows that readers are not able to effectively focus attention on single words. While it is evident that the flanker paradigm is different from sentence reading, serial processing accounts are challenged by the question of how readers would then be better at focusing attention on single words during sentence reading, given that (i) the visual input during sentence reading is more complex than in the flanker paradigm, and more dynamic due to eye-movements, (ii) parafoveal information is of interest during sentence reading, and (iii) parafoveal information is available longer during sentence reading.

Finally, the theoretical framework that is discussed in this paper may be tested in ways other than those employed here. For instance, Snell and Grainger (2017) have recently employed a paradigm that seems to sit neatly in between the flanker paradigm and natural sentence reading. It was demonstrated that the identification of a target word in a briefly presented sequence of four words is facilitated when the words form a grammatically correct sentence, compared to when the same words are presented in a shuffled agrammatical order. Future research may further endeavor to create a 'sentence reading' setting in which readers do not engage the sentence-level representation: for instance, higher-order parafoveal-on-foveal influences may be found during the reading of random, agrammatical word sequences.

In sum, results from the present experiments suggest that readers can extract semantic information from multiple words at once. In sentence-reading, this information is appended to a sentence-level representation rather than integrated as a whole (Figure 1), explaining why parafoveal-on-foveal effects in sentence reading have been elusive. As seen in Experiments 2 and 3, however, higher-order information can be integrated across multiple words when readers do not engage this sentence-level representation.

Acknowledgements

The authors would like to thank Margot Roisin and Svenja Krabbe for their help in conducting this research. This research was funded by the French National Research Council (ANR-11-LABX-0036, ANR-15-CE33-0002-01).

References

Altarriba, J., Kambe, G., Pollatsek, A. & Rayner, K. (2001). Semantic codes are not used in integrating information across eye fixations in reading: Evidence from fluent Spanish-English bilinguals. *Perception & Psychophysics, 63*, 875–890. doi: 10.3758/BF03194444

Angele, B., Tran, R. & Rayner, K.(2013). Parafoveal–foveal overlap can facilitate ongoing word identification during reading: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 39*, 526–538. doi: 10.1037/a0029492

Baayen, R. (2008). *Analyzing Linguistic Data: A pratical introduction to statistics*. Cambridge: Cambridge University Press.

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

Bertram, R. & Hyönä, J. (2007). The interplay between parafoveal preview and morphological processing in reading. In van Gompel, R., Fischer, M., Murray, W. & Hill, R. (Eds.). *Eye movements: A window on mind and brain*. Oxford, UK: Elsevier. doi: 10.1075/ml.6.1.04ber

Brothers, T., & Traxler, M. (2016). Anticipating syntax during reading: Evidence from the boundary change paradigm. *Journal of Experimental Psychology: Learning, Memory & Cognition, 42,* 1894–1906. doi: 10.1037/xlm0000257

Brysbaert, M. (*2013*). LEXTALE_FR: A fast, free, and efficient test to measure language proficiency in French. *Psychologica Belgica, 53*, 23–37. doi: 10.5334/pb-53-1-23

Dalmaijer, E., Mathôt, S. & Van der Stigchel, S. (2014). PyGaze: An open-source, crossplatform toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research Methods, 46*, 913–921. doi: 10.3758/s13428-013-0422-2

Dare, N. & Shillcock, R. (2013). Serial and parallel processing in reading: Investigating the effects of parafoveal orthographic information on nonisolated word recognition. *Quarterly Journal of Experimental Psychology, 66,* 417–428. doi: 10.1080/17470218.2012.703212

Davis, C. & Lupker, S. (2006). Masked inhibitory priming in English: Evidence for lexical inhibition. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 668–687. doi: 10.1037/0096-1523.32.3.668

Deutsch, A., Frost, R., Pollatsek, A. & Rayner, K. (2005). Morphological parafoveal preview benefit effects in reading: Evidence from Hebrew. *Language and Cognitive Processes, 20*, 341–371. doi: 10.1080/01690960444000115

Duñabetia, J.A., Perea, M. & Carreiras, M. (2010). Masked translation priming effects with highly proficient simultaneous bilinguals. *Experimental Psychology*, *57*, 98-107. doi: 10.1027/1618-3169/a000013

Engbert, R., Nuthmann, A., Richter, E. & Kliegl, R. (2005). SWIFT: A dynamic model of saccade generation during reading. *Psychological Review, 112*, 777–813. doi: 10.1037/0033-295X.112.4.777

Ferrand, L., et al. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods, 42*, 488–496. doi: 10.3758/BRM.42.2.488

Forster, K., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*, 680–698. doi: 10.1037/0278-7393.10.4.680

Grainger, J. & Dijkstra, T. (1992). On the representation and use of language information in bilinguals. In R.J. Harris (Ed.) *Cognitive processing in bilinguals*. Amsterdam: North Holland.

Grainger, J., Dufau, S. & Ziegler, J.C. (2016). A vision of reading. *Trends in Cognitive Sciences*, *20*, 171-179. doi: 10.1016/j.tics.2015.12.008

Grainger, J. & Frenck-Mestre, C. (1998). Masked translation priming in bilinguals. *Language and Cognitive Processes, 13*, 601–623. doi: 10.1080/016909698386393

Grainger, J., Mathôt, S. & Vitu, F. (2014). Test of a model of multi-word reading: Effects of parafoveal flanking letters on foveal word recognition. *Acta Psychologica, 146*, 35–40. doi: 10.1016/j.actpsy.2013.11.014

Guttentag, R., Haith, M., Goodman, G. & Hauch, J. (1984). Semantic processing of unattended words by bilinguals: A test of the input switch mechanism. *Journal of Verbal Learning and Verbal Behavior, 23*, 178–188. doi: 10.1016/S0022-5371(84)90126-9

Hohenstein, S. & Kliegl, R. (2014). Semantic preview benefit during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 166–190. doi: 10.1037/a0033670

Hyönä, J. & Häikiö, T. (2005). Is emotional content obtained from parafoveal words during reading? An eye movement analysis. *Scandinavian Journal of Psychology, 46*, 475–483. doi: 10.1111/j.1467-9450.2005.00479

Inhoff, A., Radach, R., Starr, M. & Greenberg, S. (2000). Allocation of visuospatial attention and saccade programming during reading. In Kennedy, A., Radach, R., Heller, D. & Pynte, J. (Eds.). *Reading as a perceptual process*. Oxford, UK: Elsevier.

Juhasz, B., Pollatsek, A., Hyönä, J., Drieghe, D. & Rayner, K. (2009). Parafoveal processing within and between words. *Quarterly Journal of Experimental Psychology, 62*, 1356–1376. doi: 10.1080/17470210802400010

Kass, R. & Raftery, A. (1995). Bayes Factors. *Journal of the American Statistical Association, 90,* 773–795. doi: 10.1080/01621459.1995.10476572

Kliegl, R., Hohenstein, S. & McDonald, S. (2013). How preview space/time translates into preview cost/benefit for fixation durations during reading. *Quarterly Journal of Experimental Psychology, 66,* 581-600. doi: 10.1080/17470218.2012.658073

Lemhöfer, K. & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods, 44*, 325–343. doi: 10.3758/s13428-011-0146-0

Marx, C., Hawelka, S., Schuster, S. & Hutzler, F. (2017). Foveal processing difficulty does not affect parafoveal preprocessing in young readers. *Scientific Reports, 7,* e41602. doi: 10.1038/srep41602

Mathôt, S., Schreij, D. & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*, 314–324. doi: 10.3758/s13428-011-0168-7

Perea, M., Duñabeitia, J. & Carreiras, M. (2008). Masked associative/semantic priming effects across languages with highly proficient bilinguals. *Journal of Memory and Language, 58,* 916–930. doi: 10.1016/j.jml.2008.01.003

Radach, R. & Kennedy, A. (2004). Theoretical perspectives on eye movements in reading: Past controversies, current issues, and an agenda for future research. *European Journal of Cognitive Psychology, 16,* 3–26. doi: 10.1080/09541440340000295

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology, 7,* 65–81. doi: 10.1016/0010-0285(75)90005-5

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422. doi: 10.1.1.294.4262

Rayner, K., Schotter, E. & Drieghe, D. (2014). Lack of semantic parafoveal preview benefit revisited. *Psychonomic Bulletin & Review, 21*, 1067–1072. doi: 10.3758/s13423-014-0582-9

Reichle, E., Pollatsek, A. & Rayner, K. (2006). E-Z Reader: A cognitive-control, serial-attention model of eye movement behavior during reading. *Cognitive Systems Research, 7,* 4–22. doi: 10.1016/j.cogsys.2005.07.002

Reilly, R. & Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research, 7,* 34–55. doi: 10.1016/j.cogsys.2005.07.006

Rouder, J., Speckman, P., Sun, D. & Morey, R. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16,* 225–237. doi: 10.3758/PBR.16.2.225

Schadder, W. & Laberge, D. (1979). Automatic semantic processing of unattended words. *Journal of Verbal Learning and Verbal Behavior, 18*, 413–426. doi: 10.1016/S0022-5371(79)90228-7

Schaffer, W., & LaBerge, D. (1979). Automatic semantic processing of unattended words. *Journal of Verbal Learning & Verbal Behavior, 18,* 413–426. doi: 10.1016/S0022-5371(79)90228-7

Schotter, E. (2013). Synonyms provide semantic preview benefit in English. *Journal of Memory and Language, 69*, 619–633. doi: 10.1037/a0036763

Schotter, E., Angele, B. & Rayner, K. (2012). Parafoveal processing in reading. *Attention Perception & Psychophysics, 74*, 5–35. doi: 10.3758/s13414-011-0219-2

Schotter, E., Reichle, E. & Rayner, K. (2014). Rethinking parafoveal processing in reading: Serial attention models can account for semantic preview benefit and n+2 preview effects. *Visual Cognition, 22*, 309–333. doi: 10.1080/13506285.2013.873508

Schwanenflugel, P. & Rey, M. (1986). Interlingual semantic facilitation: Evidence for a common representational system in the bilingual lexicon. *Journal of Memory and Language, 25*, 605–618. doi: 10.1016/0749-596X(86)90014-8

Segui, J., & Grainger, J. (1990). Priming word recognition with orthographic neighbors: Effects of relative prime-target frequency. *Journal of Experimental Psychology: Human Perception and Performance*, *16,* 65–76. doi: 10.1037/0096-1523.16.1.65

Snell, J. & Grainger, J. (2017). The sentence superiority effect revisited. *Cognition, 168,* 217-221. doi: 10.1016/j.cognition.2017.07.003

Snell, J., Vitu, F. & Grainger, J. (2017a). Spatial integration of parafoveal orthographic information: Beyond the sub-lexical level? *Quarterly Journal of Experimental Psychology, 70,* 1984–1994. doi: 10.1080/17470218.2016.1217247

Snell, J., Meeter, M. & Grainger, J. (2017b). Evidence for simultaneous syntactic processing of multiple words during reading*. PLoS ONE, 12,* e0173720. doi:10.1371/journal.pone.0173720

Thierry, G. & Wu, Y. J. 2007. Brain potentials reveal unconscious translation during foreign language comprehension. *Proceeding of National Academy of Sciences*, *104*, 12530–12535. doi: 10.1073/pnas.0609927104

Van Heuven, W., Mandera, P., Keuleers, E. & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology, 67*, 1176–1190. doi: 10.1080/17470218.2013.850521

Veldre, A. & Andrews, S. (2014). Parafoveal preview benefit is modulated by the precision of skilled readers' lexical representations. *Journal of Experimental Psychology: Human Perception & Performance, 41*, 219–232. doi: 10.1037/xhp0000017

Yan, M., Zhou, W., Shu, H. & Kliegl, R. (2012). Lexical and sublexical semantic preview benefits in Chinese reading. *Journal of Experimental Psychology: Learning, Memory and Congnition, 38,* 1069–1075. doi: 10.1037/a0026935

Chapter 4: Top-down expectations

The sentence superiority effect revisited

Joshua Snell & Jonathan Grainger

Citation: Snell, J. & Grainger, J. (2017). The sentence superiority effect revisited. *Cognition, 168,* 217-221.

Abstract

A sentence superiority effect was investigated using post-cued word-in-sequence identification with the rapid parallel visual presentation (RPVP) of four horizontally aligned words. The four words were presented for 200 ms followed by a post-mask and cue for partial report. They could form a grammatically correct sentence or were formed of the same words in a scrambled agrammatical sequence. Word identification was higher in the syntactically correct sequences, and crucially, this sentence superiority effect did not vary as a function of the target's position in the sequence. Cloze probability measures for words at the final, arguably most predictable position, revealed overall low values that did not interact with the effects of sentence context, suggesting that these effects were not driven by word predictability. The results point to a level of parallel processing across multiple words that enables rapid extraction of their syntactic categories. These generate a sentence-level representation that constrains the recognition process for individual words, thus facilitating parallel word processing when the sequence is grammatically sound.

1. Introduction

James McKeen Cattell, the first Professor of Psychology in the USA, is well known for having reported a "*word superiority effect"* whereby word naming latencies for monosyllabic words are faster than single letter naming times, and more letters can be reported from briefly presented words than scrambled letter strings. What is less well known, however, is that Cattell also discovered a "*sentence superiority effect*". That is, he reported that sentences containing up to seven words could be recalled correctly after a single exposure, while the corresponding number of recalled words when these were unrelated was three to four (Cattell, 1886; reported in Scheerer, 1981). Can Cattell's sentence superiority effect be taken as evidence in favor of rapid parallel processing of syntactic and semantic information across multiple words during sentence reading? As was the case with the word superiority effect, the methodology used in Cattell's experiments left open possible roles for memory and guessing. Indeed, Cattell's sentence superiority effect is very likely to have been influenced by memory factors, such that a correct sentence is easier to maintain in working memory compared with an unrelated list of words. In line with this, more recent research has shown that recall of grammatically correct sentences is superior to recall of scrambled lists of the same words (e.g., Baddeley, Hitch, & Allen, 2009; Toyota, 2001).

The work of Reicher (1969) and Wheeler (1970) renewed interest in the word superiority effect by showing that it can be observed with a post-cued partial report procedure following brief presentations of words and nonwords. The research spurred by this methodological development provided the empirical foundations for current theoretical accounts of letter and word processing, and in particular, the parallel, cascaded, interactive nature of such processing (McClelland & Rumelhart, 1981). Crucially, superior report of letters in words compared with pseudoword stimuli (e.g., B in TABLE vs. PABLE) suggests that whole-word representations can be activated before identification of the constituent letters (i.e., parallel, cascaded transmission of information from letters to words), and once a word representation is activated it can influence letter identification either via feedback and/or via decision making processes (Grainger & Jacobs, 1994; 2005). In the present study, we apply the same logic to word and sentence processing. If post-cued partial report of words is more accurate in grammatically correct sentences compared with scrambled sequences of the same words, then this would imply parallel, cascaded transmission of information from word identities to sentence-level representations.

Asano and Yokosawa (2011) were the first to use a partial report procedure to study the sentence superiority effect. They employed a post-cued 4-alternative forced choice procedure with brief presentation of sentences and agrammatical lists of words. Sentences were on average 12-14 Japanese Kanji characters long and were presented for 200 ms and followed by a continuous pattern mask. Targets were 2-character words randomly located at different positions in the sentence, and the four response alternatives were presented vertically aligned below the location of the target word in the sentence. Asana and Yokosawa found that semantic relatedness, rather than syntactic structure, facilitated word identification. However, this might be due to the specific nature of Kanji text that imposes less grammatical constraint compared with other writing systems. In line with this reasoning, it should be noted that in an earlier study, Toyota (2001) found an effect of syntactic structure on the recall of Japanese words written in Hiragana, an alphabetic script. Furthermore, in a sequential variant of the word-in-sequence identification paradigm, with target words being presented after a sentence context and not simultaneously with the context, Jordan and Thomas (2002) found evidence for effects of syntactic coherence rather than between-word semantic priming in English.

Asano and Yokosawa's (2011) findings are important in two respects. First, to our knowledge this is the first published report using rapid parallel visual presentation (RPVP) of word sequences accompanied with a post-cued partial report procedure. Second, the effect of semantic relatedness reported by Asano and Yokosawa is further evidence in favor of parallel processing of words up to the semantic level, in line with the growing evidence for semantic processing of words in the parafovea[34] (e.g., Engbert & Kliegl, 2011; Hohenstein & Kliegl, 2014; Schotter, 2013; Veldre & Andrews; 2015; 2016; 2017). There is also growing evidence for syntactic processing of multiple words in parallel. Snell, Meeter, and Grainger (2017a) found that readers were faster to categorize foveal target words as noun or verb, when those targets were flanked by syntactically congruent words. In sentence reading, on the other hand, syntactically similar adjacent words seemed to interfere with, rather than facilitate, foveal word processing (see also Snell, Vitu & Grainger, 2017b; Brothers & Traxler, 2017).

If higher-order processing can indeed occur for multiple words in parallel, then it is possible that the semantic and syntactic properties of words can mutually influence their identification. Snell et al. (2017a) proposed that during sentence reading, feedback from a sentence-level representation to individual word positions constrains the identification of these words via semantic and syntactic constraints. For example, if word *n* was recognized as a noun, word *n+1* could be expected to be a verb (in English) – and word *n+1* may constrain word *n* in a similar way. Crucially, higher-order processing can occur for multiple words in parallel via position-specific processes that interact only via sentence-level representations. This account of word identification and sentence reading therefore predicts that a syntactic sentence superiority effect should be observed in a post-cued RPVP word-in-sequence identification paradigm, where the sequence of words can either be a grammatically correct sentence or an agrammatical re-arrangement of the same words.

2. Methods

2.1 Participants

Thirty students from Aix-Marseille University were paid €5 to participate in the experiment. All participants reported to be native speakers of French, non-dyslexic, and having normal or corrected-to-normal vision.

2.2 Materials

We constructed 200 sentences that consisted of four words each. Word length was from three to five letters, and the average word frequency (Ferrand et al., 2010) was 6.48 (SD = 1.02) Zipf (log10 per billion occurrences: Van Heuven, Mandera, Keulers & Brysbaert, 2014). We constructed grammatically correct sentences that were as semantically neutral as possible. This was evaluated by obtaining cloze probability measures for the last word in each sentence.[35] The

---

[34] We acknowledge that semantic parafoveal preview effects are not necessarily evidence for parallel processing of semantic information across multiple words (e.g., Schotter, Reichle, & Rayner, 2014), but note that the RPVP paradigm might offer stronger evidence for parallel processing, albeit in the context of reading without eye movements.

[35] Cloze probability values were obtained via an on-line experiment containing 50 questions. Each question consisted of the first three words of each sentence of the main experiment, and participants were asked to type in the first word that came to mind as a likely continuation of the sentence. The sentences were presented in a different randomized order for each participant. 128 volunteers (74 female; age 17–26 years; reporting to be native speakers of French) were recruited via the RISC network of departments in the Cognitive Sciences in France.

average cloze probability of these words was .005 (SD = .009). In every sentence one of the four words was marked as the target word, at varying positions, so that the total of 200 sentences yielded 50 targets for each word position. For every sentence we further constructed a scrambled version in which all words but the target switched positions (see Figure 1). We made sure that these scrambled sentences were syntactically incorrect.

2.3 Design

A 2 × 4 factorial design was used, with Context (normal vs. scrambled) and Target Position (1-4) as variables. Participants were Latin-squared into two groups so that all sentences were shown in both context conditions, but only once per participant. The experiment thus consisted of 200 trials, and sentences were presented in a different randomized order for each participant.

	position 1	**position 2**
normal	our fox can fly	that was not red
scrambled	our can fly fox	not was red that
	position 3	**position 4**
normal	she can work now	the guy did this
scrambled	now she work can	guy did the this

Figure 1. Sentence examples across conditions. One of the four words was marked as the target word in every sentence. As can be seen in these examples, the target was at the same location in the scrambled condition. Targets were not underlined during actual stimulus presentation. The sentences in this figure are in English rather than French for illustration purposes.

2.4 Software and Apparatus

The experiment was implemented with OpenSesame (Mathôt, Schreij & Theeuwes, 2012). Stimuli were presented on a gamma-calibrated 21-inch screen (1024x768 px, 150 Hz). Responses were collected with an azerty-layout keyboard. Participants were seated at a 80 cm distance from the display, so that each character space subtended 0.35 degrees of visual angle.

2.5 Procedure

Participants were seated in a comfortable chair in a dimly lit testing room, and received instructions verbally by the experimenter and visually onscreen. Each trial started with a display of two vertical bars positioned at the screen center, and participants were instructed to fixate between these two bars. A sequence of four words was then briefly presented at the screen center, so that two words appeared to the left of fixation and two words to the right of fixation. This display was followed by a backward mask, consisting of hash marks ('#') at all prior letter locations, and a post-cue for the target, consisting of a dot above the target location. Participants could type in their response at this point, and their response appeared in a box located slightly below the string of hash marks. Responses were finalized by pressing the return key. Feedback was given in the form of a briefly presented green (correct) or red (incorrect) dot. Prior to the

Cloze probability was calculated as the number of answers that corresponded to the word at position 4 in the original sentence (ignoring diacritics) divided by the number of participants.

main experiment, participants received eight practice trials. The experiment lasted approximately 25 minutes.

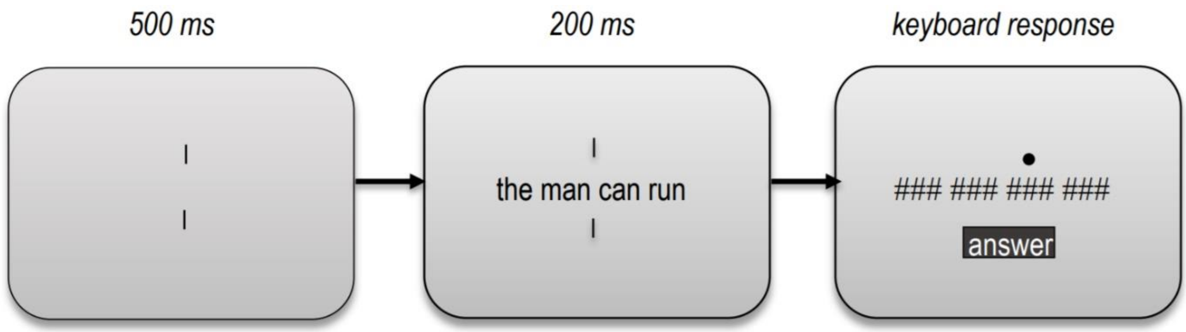

Figure 2. The post-cued partial report Rapid Parallel Visual Presentation (RPVP) procedure used in the present experiment.

3. Results

Probability correct responses were analyzed using generalized linear mixed models (LMMs) with items and participants as crossed random effects (including by-item and by-participant random intercepts; Baayen, 2008). The models were fitted with the glmer function from the lme4 package (Bates, Maechler, Bolker & Walker, 2015) in the R statistical computing environment. We report regression coefficients (b), standard errors (SE) and z-values. Fixed effects were deemed reliable if $|z| > 1.96$ (Baayen, 2008).

Our hypothesis that the word identification accuracy would be higher in the normal sentence condition than in the scrambled sentence condition was confirmed, with b = 1.19 (corresponding to a probability difference of 0.20), SE = 0.17, z = 6.84. As can be seen in Figure 3, the effect was virtually equal across all target positions.

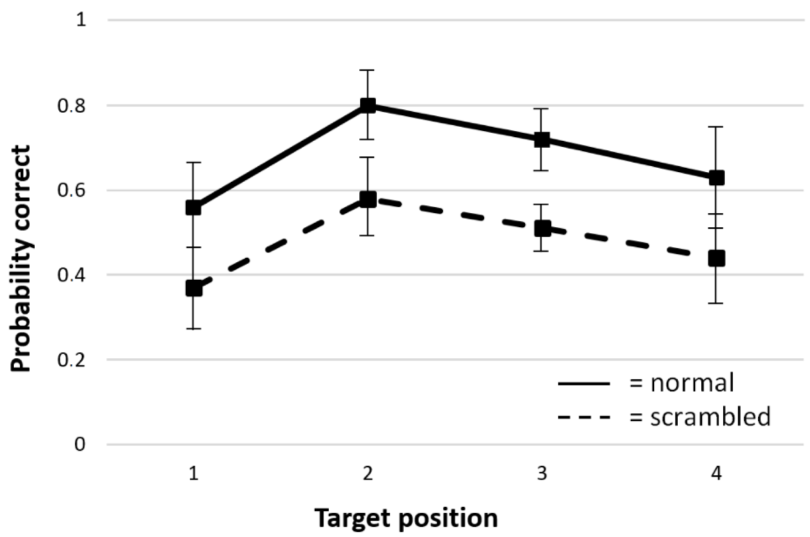

Figure 3. Mean probability correct identification at the different target positions (1-4) in the normal sentence condition (solid line) and scrambled condition (dashed line). Error bars indicate 95% confidence intervals.

The overall performance did vary across positions, and aiming to pinpoint the factors underlying this variance we ran a separate LMM that included Context (normal vs. scrambled), Target Frequency (continuous) and Eccentricity (central vs. peripheral targets) as factors. We found that both Frequency and Eccentricity had a significant influence (Frequency, b = 0.12, SE = 0.05, z = 2.47; Eccentricity, b = 1.13, SE = 0.31, z = 3.68), but importantly, neither factor modulated the effects of Context (Frequency, b = 0.07, SE = 0.17, z = 0.41; Eccentricity, b = 0.18, SE = 0.32, z = 0.56). Moreover, post-hoc testing revealed that performance did not differ between leftward and rightward positions (b = 0.01, SE = 0.06, z = 0.11), suggesting that targets were not processed in a serial, left-to-right fashion.

An explanation for our Context effect alternative to parallel processing, may be that of target predictability. In this scenario, recognition of one word may allow readers to predict the target, and this would be easier in the normal sentence condition than in the scrambled condition. Putting this scenario to the test, we entered the cloze probability values that were obtained for the final word of each sentence as a factor in a separate LMM testing for the effects of Context at that position. Cloze probability did not significantly influence identification accuracy (b = 13.27, SE = 7.93, z = 1.67), and did not interact with Context (b = 14.57, SE = 9.43, z = 1.55), suggesting that predictability did not drive our effect.[36]

Lastly, for all incorrect trials we assessed how often the given response was one of the other words in the sequence (i.e., location errors). This occurred on only 10.68% of the incorrect responses, 77.3% of which were in the scrambled condition. Furthermore, participants were not more likely to respond with centrally positioned words, as 55.76% of location errors were from positions 1 and 4. The low occurrence of this type of error suggests that word identities are strictly tied to specific locations, and particularly so in a sentence context.

4. Discussion

An experiment tested post-cued identification of one word within a sequence of four briefly presented (RPVP) horizontally aligned words. The four words could either represent a grammatically correct sequence (e.g., *she can work now*) or an agrammatical sequence (*now she work can*) and the probability of a correct identification of the same word (“work” in these examples) at the same position was compared across the two contexts. We reasoned that if word-in-sequence identification is more accurate in grammatical sequences compared with agrammatical ones, a “sentence superiority effect”, then this would be evidence for parallel, cascaded transmission of information from word identities to sentence-level representations. Such sentence-level representations could then be used to improve post-cued word identification via feedback from sentence-level representations to word representations.

[36] To further test a “guessing” account of these results we ran a two-alternative forced-choice version of the main experiment. Thirty participants were instructed to decide which of two alternative words had been present at the cued location. Foil words were matched in length with target words, and formed a syntactically correct sentence in the normal sentence condition. The experiment was otherwise identical to the main experiment. Crucially, there was a significant effect of context on RT, with lower RTs in the intact condition compared to the scrambled condition: b = 36.30, SE = 12.66, t = 2.87. This effect was significant at positions 2 and 3, respectively b = 62.94, SE = 25.15, t = 2.50 and b = 49.89, SE = 23.95, t = 2.08), but not at positions 1 and 4, respectively b = 3.18, SE = 24.36, t = 0.13 and b = 30.02, SE = 25.03, t = 1.20. The effect of context was overall not significant in the accuracy data (b = 0.04, SE = 0.06, z = 0.55), but participants were significantly more accurate in the normal sentence condition when the target was at position 3 (b = 0.25, SE = 0.13, z = 1.96), while accuracy was virtually equal across conditions for the other target positions. See supplementary materials for a complete description of this experiment and the results.

The results are clear-cut. Word identification was more accurate in grammatically correct word sequences than in agrammatical scrambled sequences. This sentence superiority effect was obtained in conditions that minimized any potential influence of between-word semantic relatedness or predictability (confirmed by our cloze probability measures), thus pointing to syntactic representations as the source of the effect. Partial parallel identification of words would generate an elementary syntactic representation when the sequence of words forms a sentence, and this syntactic representation then constrains on-going identification processes through feedback. In line with this interpretation is the fact that context effects were equivalent across all positions in the sequence. This was not the case, however, in a follow-up experiment using a two-alternative forced choice (2AFC) procedure (see footnote 3 and Supplementary Materials). Although there was a significant main effect of context in the RT data, post-hoc analyses revealed that the effects were only significant at the two central positions. We think this is because partial information about letter identities in words can be used to respond with the 2AFC procedure (e.g., having noticed the presence of the letter '*c*', the correct response would be '*cat*' and not '*dog*'), and that use of such letter-level information will dominate responding at the positions where word identification is hardest.

The syntactic sentence superiority effect found in the present study contrasts with the results of the only prior research (Asano & Yokosawa, 2011) to have investigated a sentence superiority effect with the post-cued RPVP paradigm. Asano and Yokosawa tested identification of 2-character Kanji words embedded in sequences of 12-14 characters, and failed to find an effect of grammatical structure while observing a significant influence of between-word semantic relatedness. As argued in the Introduction, these differences are likely due to the different scripts used in the two studies, logographic Kanji in the Asano and Yokosawa study and the Roman alphabet in our study. Logographic scripts provide less syntactic constraints than alphabetic writing systems. Related to this, it is interesting to note that the clearest evidence for semantic parafoveal processing has been obtained with Chinese logographic stimuli (Yan, Zhou, Shu & Kliegl, 2012; Zhang, Li, Wang & Wang, 2015). Thus, taken together, the results of Asano and Yokosawa (2011) and the present results point to both semantic and syntactic influences in the post-cued RPVP paradigm. This is in line with the growing evidence, summarized in the Introduction, that both semantic and syntactic information can be extracted from words in the parafovea. The results of the present study align particularly well with Veldre and Andrews' (2016) finding that semantic preview effects are eliminated when the preview is a syntactically implausible continuation of the sentence. Both sets of results suggest that the processing of words in the parafovea is modulated by how well the parafoveal word fits with the sentence being processed.

In sum, the current results are in accordance with an account of reading that assumes parallel word processing guided by sentence-level constraints. Future research applying the RPVP paradigm should help further elucidate the mechanisms involved in parallel word processing and reading.

5. Acknowledgments

This research was supported by grant ANR-15-CE33-0002-01 from the French National Research Agency (ANR).

6. References

Asano, M. & Yokosawa, K. (2011). Rapid extraction of gist from visual text and its influence on word recognition. *The Journal of General Psychology*, 138, 127-154.

Baayen, R. (2008). *Analyzing Linguistic Data: A pratical introduction to statistics*. Cambridge, UK: Cambridge University Press.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015), Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, 67, 1-48.

Baddeley, A., Hitch, G. & Allen, R. (2009). Working memory and binding in sentence recall. *Journal of Memory and Language*, 61, 438-456.

Brothers, T. & Traxler, M. (2017). Anticipating syntax during reading: Evidence from the boundary change paradigm. *Journal of Experimental Psychology: Learning, Memory & Cognition,* in press.

Cattell, J.M. (1886). The time it takes to see and name objects. *Mind*, 11, 53-65.

Engbert, R. & Kliegl, R. (2011). Parallel graded attention models of reading. In Liversedge, S., Gilchrist, I. & Everling, S. (Eds.), *Oxford handbook of eye movements* (pp. 787–800). Oxford, UK: Oxford University Press.

Ferrand, L., et al. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods,* 42, 488–496.

Grainger, J., & Jacobs, A. (1994). A dual read-out model of word context effects in letter perception: Further investigations of the word superiority effect. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 1158-1176.

Grainger, J. & Jacobs, A. (2005). Pseudoword context effects on letter perception: The role of word misperception. *European Journal of Cognitive Psychology*, 17, 289-318.

Hohenstein, S. & Kliegl, R. (2014). Semantic preview benefit during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 166-190.

Jordan, T. & Thomas, S. (2002). In search of perceptual influences of sentence context on word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 28, 34-45.

McClelland, J. & Rumelhart, D. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*, 88, 375-407.

Reicher, G. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 81, 274–280.

Scheerer, E. (1981). Early German approaches to experimental reading research: The contributions of Wilhelm Wundt and Ernst Meumann. *Psychological Research*, 43, 111-130.

Schotter, E. (2013). Synonyms provide semantic preview benefit in English. *Journal of Memory and Language*, 69, 619-633.

Schotter, E., Reichle, E. D., & Rayner, K. (2014). Rethinking parafoveal processing in reading: Serial-attention models can explain semantic preview benefit and N+2 preview effects. *Visual Cognition*, 22, 309-333.

Snell, J., Meeter, M. & Grainger, J. (2017). Evidence for simultaneous syntactic processing of multiple words during reading. *PLoS ONE, 12,* e0173720.

Snell, J., Vitu, F., & Grainger, J. (in press). Integration of parafoveal orthographic information during foveal word reading: Beyond the sub-lexical level? *Quarterly Journal of Experimental Psychology, 70,* 1984-1994.

Toyota, H. (2001). Changes in the constraints of semantic and syntactic congruity on memory across three age groups. *Perceptual and Motor Skills*, 92, 691-698.

Van Heuven, W., Mandera, P., Keuleers, E. & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, 67, 1176-1190.

Veldre, A. & Andrews, S. (2015). Parafoveal preview benefit is modulated by the precision of skilled readers' lexical representations. *Journal of Experimental Psychology: Human Perception & Performance*, 41, 219-232.

Veldre, A. & Andrews, S. (2016). Is semantic preview benefit due to relatedness or plausibility? *Journal of Experimental Psychology: Human Perception & Performance*, 42, 939-952.

Veldre, A. & Andrews, S. (2017). Parafoveal preview effects depend on both preview plausibility and target predictability. *Quarterly Journal of Experimental Psychology*, in press.

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, 1, 59-85.

Yan, M., Zhou, W., Shu, H. & Kliegl, R. (2012). Lexical and subexical semantic preview benefits in Chinese reading. *Journal of Experimental Psychology : Learning, Memory and Cognition*, 38, 1069-1075.

Zhang, W., Li, N., Wang, X. & Wang, S. (2015). Integration of sentence-level semantic information in parafovea : Evidence from the RSVP flanker paradigm. *PLoS One*, 10(9), e0139016.

Word position coding in reading is noisy

Joshua Snell & Jonathan Grainger

Citation: Snell, J. & Grainger, J. (2018). Word position coding in reading is noisy. *Psychonomic Bulletin & Review* (in review).

Abstract

The present paper outlines a cognitive process, explicitly, for the first time: *word position coding*. The question of how readers perceive word order is not trivial: recent research suggests that readers associate activated word representations with plausible locations in a sentence-level representation. Rather than simply being dictated by the order in which words are recognized, word position coding may be influenced by bottom-up visual cues (e.g., word length information) as well as top-down expectations. Here, we assessed how flexible word position coding is. We let readers make grammatical judgments about four-word sentences. Incorrect sentences were constructed by transposing two words in a correct sentence (e.g., *'the man can run'* became *'the can man run'*). The critical comparison was between two types of incorrect sentence: one with a transposition of the inner two words, and one with a transposition of the outer two words (*'run man can the'*). We reasoned that under limited (local) flexibility, it should be easier to classify the outer-transposed sentences as incorrect, as words are further away from their plausible locations in this condition. Under maximum (global) flexibility, on the other hand, there should be no difference between these two conditions. As it turned out, we observed longer response times and higher error rates for inner-transposed sentences than outer-transposed sentences, indicating that local flexibility and top-down expectations can jointly lead the reader to confuse the locations of adjacent words. We conclude that word position coding is subject to local noise.

Introduction

No prior research has explicitly investigated how humans encode the order of words during reading, and a lightning-speed summary of the history of reading research readily reveals why. Firstly, a large chunk of reading research has focused on the recognition of single, isolated words (e.g., Grainger, 2008, for a review). Secondly, research on sentence- and text-reading has been dominated by the view that words are recognized in a serial, one-by-one fashion, so that knowledge of word order is simply implicated by the order in which recognized words are appended to sentence representations in memory. In short, the encoding of word order has not been investigated because it was considered a given.

Presently however, reading research is facing accumulating evidence that words are to some extent processed in parallel rather than serially. This necessitates, more than ever, answering of a 10-year old question: *How would a parallel processing system keep track of word order?* (Reichle, Pollatsek, Liversedge & Rayner, 2009). In their opinion article, Reichle et al., authors of the serial processing E-Z Reader model, noted that a parallel processing system is likely to recognize words out of order, for instance when an upcoming word is much easier to recognize than the fixated word, and that it is unclear how the system would handle such occurrences. They then noted that "one possibility is that a buffer maintains word meanings, and that some mechanism re-orders out-of-order words"; however, a "[…] problem with this solution is how such mistakes are detected without using comprehension difficulty to signal such occurrences" (pp. 116-117). Indeed, this 'parallel processing problem' (Snell, van Leipsig, Grainger & Meeter, 2018a) has long been a key argument in favor of serial processing—and perhaps rightfully so, if for no better reason than Occam's razor. Surely the reading system could not be so complex as to engage in parallel processing?

The reading system engages in parallel processing

Syntactic and semantic categorization decisions to foveal target words are made faster when these are flanked by syntactically or semantically congruent parafoveal words, as compared to incongruent words (Snell, Meeter & Grainger, 2017; Snell, Declerck & Grainger, 2018b). Crucially, these effects are established when the target and flankers are shown (simultaneously) for only 170 ms, which is shorter than the average time needed to recognize single words (Rayner, 1998). The fact that the syntactic and semantic characteristics of adjacent words nonetheless have an impact indicates that they must have been processed *during*, rather than *after* target processing.

Using the novel Rapid Parallel Visual Presentation (RPVP) paradigm, Snell and Grainger (2017) found that when viewing four-word sentences for only 200 ms, readers were able to recognize any word with an accuracy of ~70% if the four words were syntactically coherent. A scrambled sequence of the same words, with the same target word having been presented at the same position, led to accurate recognition about ~50% of the time (Snell & Grainger, 2017). This *sentence superiority effect* was perfectly equal across the four word positions, indicating that syntactic information was picked up from all words during the 200 ms presentation time, which in turn constrained lexical identification of individual words.

How then, would a parallel processing system keep track of word order? The answer offered by Snell and colleagues (2017; 2018a; 2018b) is that a first glance at a sentence generates a spatiotopic sentence-level representation that comprises information about the number of to-be-recognized words in the perceptual span, as well as low-level visual information (e.g. word shapes, word lengths). Sub-lexical processing across multiple words would lead to the activation of multiple lexical representations irrespective of location. These would subsequently be mapped onto plausible locations in the sentence-level representation, based on bottom-up visual cues as well as top-down expectations (e.g., having recognized an article at position 1, one may expect an adjective or noun at position 2; and a word at position 3 may constrain lexical identification for position 2 in a similar way).

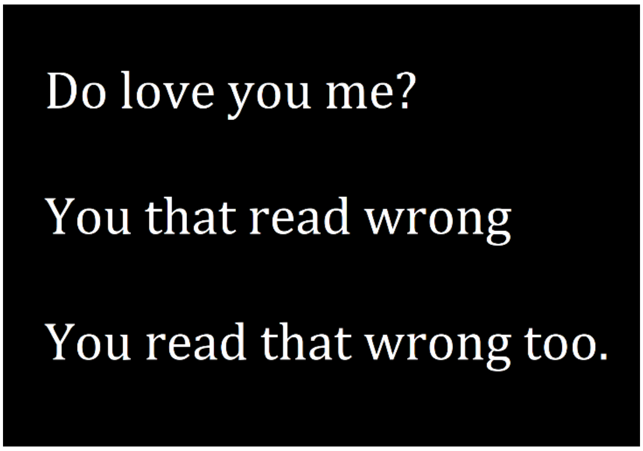

Figure 1. Illustration of the transposed-word effect.

Recently, we have investigated a phenomenon that is illustrated in Figure 1. The figure shows that the reading system is in fact capable of doing what Reichle et al. (2009) were afraid of: that is, to recognize words out of order—which, as noted by themselves, should not be possible under the assumption of serial processing. Testing the so-called *transposed word* phenomenon in an experimental setting, Mirault, Snell and Grainger (2018) let readers make grammatical judgments about sentence stimuli. The crucial comparison was between two types of ungrammatical sentences: one that could be 'corrected' by the system through the transposition of two words (e.g., '*the ran dog slowly*', which could be resolved into '*the dog ran slowly*'), and one that could not be corrected (e.g., '*the was dog slowly*'). Mirault et al. (2018) observed that readers had a much harder time classifying the former as incorrect, suggesting that the reading system retains some flexibility in the encoding of word position.

Some uncertainties remained, however. Firstly, although the Mirault et al. study tested the same words in the two conditions across different sentences and different participants using a Latin-square design, different word combinations necessarily occurred in the two conditions, and these could have differed in terms of semantic coherence (e.g., '*ran*' and '*slowly*' may be experienced as more semantically coherent than '*was*' and '*slowly*', thus biasing the reader to classify the sentence as grammatically correct). Secondly, it is not yet clear how flexible the process of word position coding truly is: either words in the perceptual span may be recognized completely irrespective of location (meaning that words are freely assigned to spatial locations), or alternatively, words are to some extent inherently tied to spatial locations, for instance through the reader's knowledge about which letters belong to which spatial location (e.g., Grainger, Dufau & Ziegler, 2016). The latter scenario might allow for some flexibility, such that the positions of two adjacent words may be mixed up, while no confusion would occur for words that are further apart.

The present study: local or global flexibility in word position coding?

Here we report an experiment that tests the flexibility of word position coding while avoiding the confounds that have potentially occurred in the study of Mirault et al. (2018), hence providing a first explicit investigation of word position coding. We let readers make grammatical judgments about four-word sentences and, similar to the design reported by Mirault et al. (2018), compared performances between two types of ungrammatical sequences: one with a transposition of the inner two words (‘*the can man run*’), and one with a transposition of the outer two words (‘*run man can the*’).

We anticipated that, under limited positional flexibility, it should be harder to classify the inner-transposed sentences as grammatically incorrect than the outer-transposed sentences. Alternatively, if the reading system were organized such that activated words were completely freely assigned a spatial location, then there would be no difference between the inner- and outer-transposed sentences.

Methods

Participants

24 participants from Aix-Marseille University gave informed consent to their participation in this study. All participants reported to be non-dyslexic, native to the French language, and to have normal or corrected-to-normal vision. Participants received 5 euros as monetary compensation.

Stimuli and design

From the 200 French four-word sentences used in the RPVP study of Snell and Grainger (2017), we used a subset of 120 stimuli in the present study. These stimuli have been tested on their semantic neutrality, as reflected by words’ Cloze-probabilities deviating non-significantly from zero (Snell & Grainger, 2017). Word lengths ranged between three and five letters. For each stimulus, we made sure that both a transposition of the inner two words, as well as a transposition of the outer two words, rendered the stimulus ungrammatical.

All 120 stimuli were shown four times to all participants: once with an inner transposition, once with an outer transposition, and twice as the grammatically correct base sentence (to induce the grammatical judgment task with an equal occurrence of grammatical and ungrammatical sequences). The 480 experimental stimuli were presented in random order.

Apparatus and software

The experiment was implemented with OpenSesame (Mathôt, Schreij & Theeuwes, 2012). Stimuli were presented on a gamma-calibrated 21-inch screen (1024 × 768, 150 Hz). Responses were collected with an azerty-layout keyboard. Participants were seated at an 80 cm distance from the display, so that each character space subtended 0.35 degrees of visual angle.

Procedure

Participants were seated in a comfortable office chair in a dimly lit room, where they received task instructions from the experimenter, as well as visually onscreen. The trial procedure is shown in Figure 2. Each trial started with a 500 ms fixation display comprising two centrally positioned vertical fixation bars on a luminance-neutral gray background. The fixation display was succeeded by a stimulus display that stayed onscreen until the participant responded, with a right- or left-handed button press, for grammatical and ungrammatical sequences, respectively. Participants had a maximum of 3000 ms to respond. Upon the participant's response, a 400 ms feedback screen was shown, with a red or green dot to indicate incorrect or correct responses, respectively. Participants were offered a break on two occasions. The 480 experimental trials were preceded by 8 practice trials for which we did not collect data. The experiment lasted approximately 25 mins.

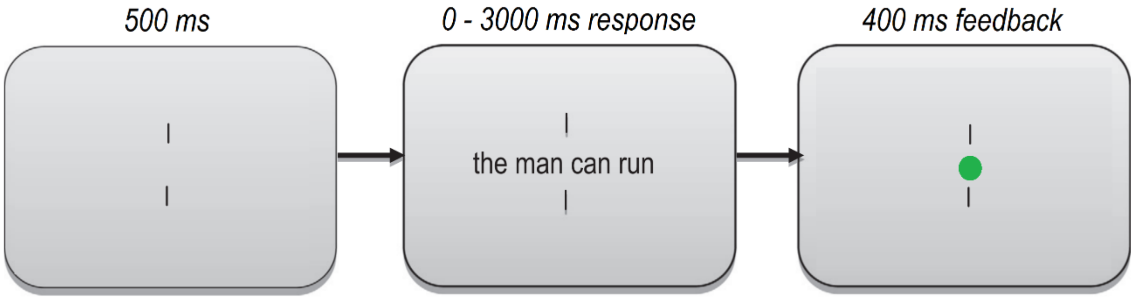

Figure 2. Trial procedure. The size of stimuli relative to the screen is exaggerated in this example.

Results

In the analyses of response times (RTs) as well as errors, we excluded trials with an RT beyond 2.5 SD from the grand mean (2.66%). For the analysis of RTs, we also excluded incorrectly answered trials (9.97%).

Data were analyzed with linear mixed-effects models (LMEs) with items and participants as random effects. We isolated trials with inner- and outer-transposed sentences, and added Condition as a two-level fixed factor to the models. The LMEs successfully converged when including by-item and by-participant random slopes as well as random intercepts. We report *b*-values, SEs and *t*-values (RTs) or *z*-values (errors), with $|t|$ and $|z| > 1.96$ deemed significant (Baayen, 2008).

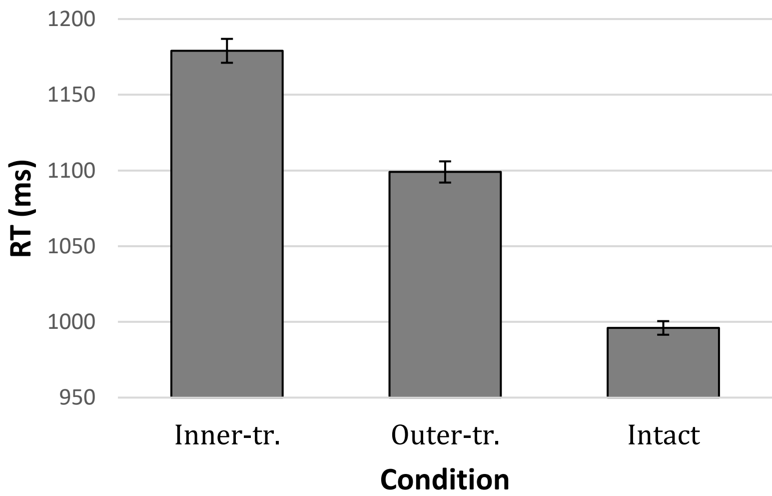

Figure 3. Average RTs per condition. Error bars indicate standard errors (SEs).

Average RTs are plotted in Figure 3. A significant difference was observed between the inner- and outer-transposed sentences, such that readers were slower to classify inner-transposed sentences as grammatically incorrect: b = 90.36, SE = 17.77, t = 5.09. This effect was also expressed in the error rate, with more errors in the classification of inner-transposed sentences: b = 1.72, SE = 0.22, z = 7.75.

We established that the critical point of ungrammaticality was considerably earlier in the outer- compared to the inner-transposed sentences. If readers would process words in a serial, left-to-right fashion, then this would lead readers to respond quicker to outer- than inner-transposed sentences. To make sure that our effects were not driven by such a confound, we isolated all items wherein the critical point of ungrammaticality was not before the second word.

When analyzing this subset of items (which had an equal average point of ungrammaticality in both conditions: word 2.14 and word 2.15 for inner- and outer-transposed sentences, respectively), our effects remained intact: for RTs, b = 55.30, SE = 17.27, b = 3.20; for errors, b = 1.40, SE = 0.33, z = 4.18. What is more, the critical point of ungrammaticality (which on each trial was set to 2, 3 or 4) was included in the LME as covariate, and was not found to have a significant influence on the RT (b = -16.36, SE = 16.72, t = -0.98) or error rate (b = -0.01, SE = 0.19, z = -0.07), hence suggesting that words were not processed in a serial left-to-right fashion.

Discussion

The present results are remarkably clear-cut: we found that readers had more difficulty classifying inner-transposed sentences as incorrect, compared to outer-transposed sentences. Given that stimuli comprised the same words across conditions, and that the critical point of ungrammaticality was found to have no influence, we surmise that our effects were caused by differences in the distance between words' true locations on the one hand, and words' plausible locations on the other. More specifically, the reading system seems capable of confusing adjacent words that are one position away from their plausible location, while a three positions' distance is considerably less plausible. From these results we infer that word position coding is subject to a moderate amount of local noise.

As a confusion of word order was deemed impossible under the assumption of serial processing (Reichle et al., 2009), this study strengthens the conception that words are processed in parallel. At the same time, it should be acknowledged that the process of word position coding is possibly more intricate than was previously theorized by Snell and colleagues (2017; 2018a; 2018b). In specific, lexical representations are not activated irrespective of location; instead, location information seems to be, to some extent, a component inherent to the recognition of individual words. A question that boasts immediate pertinence, then, is to which extent this location information is dictated by bottom-up visual cues on the one hand (e.g., readers may estimate which letters belong to which location in space; Grainger et al., 2016), or top-down expectations on the other (e.g., upon starting to read a sentence, activated words of one syntactic category may be strongly favored for a given location over words of other syntactic categories; Snell et al., 2017; 2018a; 2018b).

Having opened the investigation of word position coding, we welcome endeavors that will lead to the answering of such questions. For instance, to pinpoint the relative contribution of bottom-up and top-down processes, paradigms similar to the one reported here may manipulate the extent to which readers can rely on bottom-up and top-down cues, respectively. One hypothesis would be that the effect of bottom-up cues should be attenuated as the number of letters shared by the two transposed words increases. Similarly, bottom-up cues should be weaker if the two transposed words are equal in length. Both these manipulations affect the discriminability between the two transposed words, and the rationale is that a decrease in discriminability should lead to an increase in flexibility (and thus more difficulty in classifying a transposed-word sentence as incorrect) if bottom-up visual cues play a key role.

Investigating the contribution of top-down influences may prove to be a bigger challenge, and may necessitate the employment of different paradigms. One possibility is to briefly present two consecutive sentences (note that the RPVP setup discussed in the Introduction has already shown that readers can generate sentence-level representations within 200 ms of sentence presentation time), and have the reader make same/different matching decisions. The critical comparison would be between two conditions where the second sentence is a transposed-word version of the first sentence: one where the transposition of words leads to a change in sentence structure (e.g., '*with these fine ladies*' — '*with fine these ladies*'), and one where the transposition does *not* lead to a change in sentence structure (e.g. '*with pretty fine ladies*'— '*with fine pretty ladies*'). If word position coding is influenced by top-down processes, then same/different matching decisions should be harder to make in the latter condition than in the former.

Finally, while we here claim to have provided a first investigation of word position coding, it is worth acknowledging the closely related research applying rational analysis to the study of language comprehension. Specifically, prior research has indicated that readers may maintain and act on uncertainty in past linguistic input (Levy, Bicknell, Slattery & Rayner, 2009). Bergen, Levy and Gibson (2012) found that the ongoing generation of sentence-level representations can be subject to syntactic re-analyses (for instance in anticipating or resolving garden-path constructions) that are incompatible with the language input. However, given that sentences in this seminal study were presented one word at a time (both in the auditory and visual modality), and that syntactic uncertainty thus pertained to past rather than present linguistic information, the observed effects appear to have been driven by post-hoc operations altering representations in working memory. In contrast, in the present study sentences were displayed until the reader provided a response. In line with the transposed word phenomenon illustrated in Figure 1, this suggests that reader's on-line interpretation of the relative positions of individual words is flexible.

In sum, in the present paper we have shown that word position coding in reading is noisy. We are confident that future research will help further elucidate the process of mapping activated words onto sentence-level representations.

Acknowledgements

We thank Cécile Mahnich for her help in conducting this experiment. This research was funded by the Brain & Language Research Institute (ANR-11-LABX-0036) and the European Research Council (ERC742141).

References

Baayen, R. (2008). *Analyzing Linguistic Data: A practical introduction to statistics.* Cambridge: Cambridge University Press (2008).

Bergen, L., Levy, R. & Gibson, E. (2012). Verb omission errors: Evidence of rational processing of noisy language inputs. *Proceedings of the Annual Meeting of the Cognitive Science Society, 34.* Retrieved from: https://escholarship.org/uc/item/9x62j2cw.

Grainger, J. (2008). Cracking the orthographic code: An introduction. *Language and Cognitive Processes, 23*, 1-35.

Grainger, J., Dufau, S. & Ziegler, J. (2016). A vision of reading. *Trends in Cognitive Sciences*, *20,* 171-179.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An opensource, graphical experiment builder for the social sciences. *Behavior Research Methods, 44,* 314–324.

Mirault, J., Snell, J. & Grainger, J. (2018). You that read wrong again! A transposed-word effect in grammaticality judgments. *Psychological Science,* (in review).

Levy, R., Bicknell, K., Slattery, T. & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences, 106,* 21086.

Reichle, E., Liversedge, S., Pollatsek, A. & Rayner, K. (2009b). Encoding multiple words simultaneously in reading is implausible. *Trends in Cognitive Sciences, 13,* 115-119.

Snell, J., Declerck, M. & Grainger, J. (2018b). Semantic processing in the parafovea revisited: Effects of translation equivalents in bilingual readers. *Language, Cognition & Neuroscience, 33,* 563-574.

Snell, J. & Grainger, J. (2017). The sentence superiority effect revisited. *Cognition, 168,* 217-221.

Snell, J., Meeter, M. & Grainger, J. (2017). Evidence for simultaneous syntactic processing of multiple words in reading. *PLoS ONE, 12,* e0173720. doi:10.1371/journal.pone.0173720

Snell, van Leipsig, Grainger, J. & Meeter, M. (2018a). OB1-reader: A model of word recognition and eye movements in text reading. *Psychological Review,* (in press).

Chapter 5: OB1-Reader: A successful parallel processing system

OB1-Reader: A model of word recognition and eye movements
in text reading

Joshua Snell, Sam van Leipsig, Jonathan Grainger & Martijn Meeter

Citation: Snell, J., van Leipsig, S., Grainger, J. & Meeter, M. (2018). OB1-Reader: A model of word recognition and eye movements in text reading. *Psychological Review* (in press).

Abstract

Decades of reading research have led to sophisticated accounts of single word recognition and, in parallel, accounts of eye-movement control in text reading. While these two endeavors have strongly advanced the field, their relative independence has precluded an integrated account of the reading process. To bridge the gap, we here present a computational model of reading, OB1-reader, which integrates insights from both literatures. Key features of OB1 are: (i) parallel processing of multiple words, modulated by an attentional window of adaptable size; (ii) coding of input through a layer of open bigram nodes that represent pairs of letters and their relative position; (iii) activation of word representations based on constituent bigram activity, competition with other word representations and contextual predictability; (iv) mapping of activated words onto a spatiotopic sentence-level representation to keep track of word order; and (v) saccade planning, with the saccade goal being dependent on the length and activation of surrounding word units, and the saccade onset being influenced by word recognition. A comparison of simulation results with experimental data shows that the model provides a fruitful and parsimonious theoretical framework for understanding reading behavior.

1. Introduction

For decades, reading research has advanced along two relatively independent lines. One of these lines has focused on orthographic processing of single words, spurring various accounts of how readers may code for letter identity and –position (e.g. McClelland & Rumelhart, 1981; Davis, 1999; Whitney, 2001; Grainger & van Heuven, 2003). The other line, meanwhile, has made key contributions to our knowledge of eye-movement control in text reading (e.g. Engbert, Nuthmann, Richter & Kliegl, 2005; Reichle, Rayner & Pollatsek, 1998, 2003; Reilly & Radach, 2006). These two endeavors have led to large advances in our understanding of the reading system. Word recognition research has enabled us to predict fairly accurately how long it takes to recognize a given word and to describe how orthographic information is integrated over time (e.g. Grainger, 2008; 2018, for reviews). Meanwhile, research on eye-movements in text reading has enabled us to predict temporal and spatial eye movement parameters, based on properties of the text being read (e.g. Rayner, 1998; 2009, for reviews).

At the same time it is clear that not all pending issues concerning reading can be answered by these domains in isolation. For example, research on word recognition has generally ignored how recognition processes may be influenced by surrounding words and context. Meanwhile, the major focus of research on text reading has been at the lexical level (but see e.g. Hyönä, 1995; White & Liversedge, 2004), making it difficult to account, for instance, for how and when text reading may go awry.

Following the recommendations of Grainger (2003) and Grainger, Dufau, and Ziegler (2016), here we integrate the two domains of reading research in a computational model of reading, OB1-reader, developed with the aim to solve aforementioned issues. In the following section we provide a theoretical background, comprised of a short overview of models of word recognition on the one hand, and models of eye-movement control on the other, highlighting those elements that are, in our view, key to achieving an integrative account of reading. In subsequent sections we summarize the main assumptions and implementation of our model. We conclude with a comparison of simulation results to experimental data.

2. Theoretical background: Models of reading

2.1 Models of word recognition

While a substantial body of visual word recognition research has been dedicated to phonological, morphological and semantic processing (e.g. Frost, Grainger & Rastle, 2005), orthographic processing, i.e. the process of coding for the identities and positions of letters, is generally thought to lie at the heart of the word recognition process. In the early 80's, McClelland and Rumelhart (1981; 1982) provided what has likely been the most influential account of the word recognition process to date. According to their Interactive-Activation model (IAM), letters in the visual input activate position-coded letter nodes. These in turn activate nodes for words with letters at matching positions (e.g. a letter node coding for '*e*' at position 2 would activate '*best*', '*leave*', '*see*', et cetera), until one of the word nodes reaches an activity threshold that marks the point of recognition. Importantly however, activated words provide feedback activation to those letter representations that match their respective location in the word (e.g. '*best*' would activate the letter node coding for '*s*' at position 3). This mechanism accounts for the classic word-superiority effect reported by Cattell (1886), whereby recall of individual letters is better when those letters form a word, as compared to a non-word string. The model also correctly predicted that recognition of low-frequency words is hampered when they have at least one high-frequency orthographic neighbor (sharing all but one letter while respecting letter positions; e.g., blur-blue) (Grainger, O'Regan, Jacobs, & Segui, 1989; Grainger, 1990).

One of the major drawbacks of McClelland and Rumelhart's seminal model, however, was that letter processing took place in a rigid slot-based fashion, meaning that a stimulus with a certain letter at a certain position would only activate words that have the same letter at the same position. Since the IAM's first appearance, there has been a wealth of evidence against such absolute letter position coding, and in favor of a more flexible letter-word interface. Using a paradigm where subjects had to identify two words that were briefly presented together (e.g. *sand lane*), McClelland and Mozer (1986) showed that letter migration errors can occur (e.g. *land sane*). Years later, Davis and Bowers (2004) showed that such illusory identifications do not have to respect position: given a word pair like *step soap,* participants could also respond '*stop*', indicating a migration of the letter 'o' from position two to position three.

Another body of evidence in favor of flexible letter position coding comes from the masked priming paradigm, which tests the influence of briefly presented prime words on the processing of subsequently presented target words. It has been shown that target words are recognized faster after a transposed-letter prime (e.g. *mohter – mother*) as compared to a prime with different letters at the same positions (e.g. *monder – mother*) (e.g., Perea & Lupker, 2004; Perea & Carreiras, 2006). Further, Peressotti and Grainger (1999) found that the processing of 6-letter target words was facilitated by 4-letter relative-position primes (e.g. *mthr – mother*) as compared to unrelated primes (e.g. *lndn – mother*) (see also Grainger et al. 2006).

The need for more flexibility in the word recognition process has led to three major modeling approaches: *noisy slot-based coding, spatial coding* and *relative-position coding*. Noisy slot-based coding refers to the addition of Gaussian noise to the slot-based scheme of the IAM (Davis & Bowers, 2004; Gomez, Ratcliff & Perea, 2008), meaning that each letter of a stimulus word would not only activate the node representing that letter at its specific slot (*s*), but also in slots *s-2, s-1, s+1, s+2* etc., with increasing eccentricity from the letter's true position leading to weaker activation. This Gaussian noise renders the system less efficient, but more flexible, and allows it to account for the transposed-letter priming effect discussed above. Spatial coding, as used in Davis's SOLAR model (1999; 2010a; 2010b) implements flexibility in a fairly similar way, by adding letter position uncertainty to a spatial code of letter representations. Additionally, the SOLAR model adopts flexibility through length-independence, meaning, e.g., that '*stop*' would also activate '*stopwatch*'.

The third modeling approach, relative-position coding, abandons the IAM's slot-based scheme altogether. Instead, orthographic input is assumed to activate nodes that represent the relative position of within-word letter pairs (e.g. the stimulus *'rock'* would activate nodes for *'ro', 'rc', 'rk', 'oc', 'ok'* and *'ck'*; see Whitney, 2001; Grainger & van Heuven, 2003). These so-called *open-bigram* nodes in turn activate all lexical representations that they belong to. The node *'ro'*, for example, would activate *'rock'*, but also *'rose'* and *'ribbon'*. Accounting for the transposed-letter priming effect (e.g. *'rock'* is primed more strongly by *'rcok'* than by *'rduk'*), the lexical representation of *'rock'* would be activated by a larger subset of open-bigram nodes with the former prime (*'rc', 'ro', 'rk', 'ck', 'ok'*) as compared to the latter prime (*'rk'*).

While these three modeling approaches have all done a good job in explaining the experimental findings discussed above, some recent lines of research may slightly favor relative position coding over the other approaches. Specifically, it has been shown that processing of a foveal word is facilitated by simultaneously presented orthographically similar parafoveal words (e.g. *rock rack*) as compared to unrelated words (e.g. *rock dash*) (Angele, Tran & Rayner, 2013;

Dare & Shillcock, 2013; Snell, Vitu & Grainger, 2017a). In a similar vein, using their *Flanking Letters Lexical Decision* (FLLD) paradigm, Dare & Shillcock (2013) found that lexical decisions about foveal target words (i.e., indicating whether the target is a word or non-word) were made faster and more accurately when the target was flanked by two parafoveal related letters on each side (e.g. *'ro rock ck'*) as compared to unrelated letters (*'sp rock it'*) (see also Grainger, Mathôt & Vitu, 2014; Snell et al., 2017a). Crucially, the order of flanker bigrams did not matter (i.e., *'ck rock ro'* facilitated processing as strongly as *'ro rock ck'*), indicating the importance of relative- rather than absolute letter position.[37]

It is difficult to conceive how noisy slot-based coding would account for these findings. Regarding the flanker effects reported by Dare and Shillcock (2013), for example, a noisy slot-based model would have to assume that the letters in each slot of a 4-letter target word would receive additional activation from a letter that is five slots away (given that there are five letter spaces between the '*r*' in '*rock*' and the '*r*' in '*ro*' in the above example). Allowing letters to influence one another at such eccentricities would impair the model's performance greatly, (indeed, simulations with the SOLAR model with high position uncertainty showed that the model was unable to distinguish extreme anagrams, (e.g. '*bnoclay*' - '*balcony*'; Davis, 2010b)). In contrast, open-bigram coding explains these orthographic parafoveal-on-foveal effects quite effectively. Open-bigram units are location-invariant, meaning that both '*rock*' and '*rack*' in '*rock rack*' activate the nodes '*rc*', '*rk*' and '*ck*', thus resulting in faster word recognition.

Considering that the Open Bigram model has only been applied in settings where just one or two words were presented as visual input (e.g. Hannagan & Grainger, 2012), it remains to be seen whether the model would fare well processing normal text. A potential problem is that the visual input during text reading would activate a larger amount of bigrams compared to the visual input during single word reading, subsequently increasing the chance that incorrect words are activated (e.g., '*word*' and '*bonding*' leading to recognition of '*wording*').

2.2 Models of eye-movement control in text reading

Despite the important contribution of the above discussed work to our understanding of the reading system, it is clear that reading is more than word recognition. Reading comprises a delicate interplay of various cognitive mechanisms, involving not only sub-lexical orthographic, lexical and semantic processing, but also memory, visuo-spatial attention and oculo-motor control.

Readers make roughly five saccades (i.e., eye movements) per second to bring words into the fovea, where visual acuity is the highest. In between those saccades, fixation durations (the time spent viewing a word) reflect the time-course of the word recognition process, and can be largely predicted by the length, frequency and predictability of the fixated word (Rayner, 1998,

[37] Note, however, that there may be additional mechanisms at play that allow readers to have some knowledge of absolute letter position—at least to the extent of knowing whether letters are situated to the right or left of fixation. In a study using 6-letter targets and 3-letter flankers, Snell, Bertrand and Grainger (2018a) found that the order of flankers had an influence on target recognition speed (e.g., '*target*' was recognized faster in '*tar target get*' than in '*get target tar*'). Accounting for the discrepancy between these results and results reported by Dare and Shillcock (2013) and Grainger et al. (2014) (i.e., '*ck rock ro*' and '*ro rock ck*' yielding equal response times), Snell et al. (2018a) have posited that stimuli comprising more letters bear more processing weight, causing increased lateral activation at early visual processing stages and consequently allowing higher processing levels to make a stronger distinction between information stemming from the left and right visual hemifields.

2009; Sereno & Rayner, 2000) as well as surrounding (parafoveal) words (Kennedy & Pynte, 2005), implying an interaction between lexical processing and oculo-motor control.

Various models of eye-movements in text reading have been developed in the past few decades, all aiming to understand and predict reading behavior based on properties of the text being read. The primary goal of these models is to provide accounts of *when* and *where* the eyes move during reading (Reichle, Pollatsek, Fischer & Rayner, 1998, 2006; Engbert, Nuthmann, Richter and Kliegl, 2005; Reilly & Radach, 2006). With respect to when the eyes move, the models generally agree that lexical processing has some influence on the decision to move the eyes from one word to another. However, there is much less agreement on the factors driving lexical processing itself – in particular with respect to the role of visuo-spatial attention therein.

There are roughly two schools of theorizing about attention in reading, represented by *sequential attention shift* (SAS) models on the one hand, and *parallel graded processing* (PG) models on the other. Driven by the general principle that serial word order is important for sentence comprehension, SAS models operate on the assumption that attention is allocated to exactly one word at a time, and that attention shifts from one word to the next in strict serial order (Reichle et al., 1998; 2006). The most prominent of these, the E-Z Reader model of Reichle and colleagues (1999; 2006; 2009a) has been able to account for many phenomena in reading behavior, such as the occurrence of word skips (i.e., instances where the eyes move from word *n* to word *n+2*) and refixations (i.e., saccades that update the eye's fixation position within the same word), each representing approximately 20% of all eye-movements in reading (Rayner, 1998).

As has been acknowledged by Reichle et al. (2006) however, not all phenomena were accounted for by their model. In particular, the model does not explain regressive saccades (i.e., backwards eye-movements to earlier points in the text), which make up approximately 10-20% of all eye-movements (Rayner, 1998; Radach, Reilly & Inhoff, 2007). As of version 10 of E-Z Reader, Reichle et al. (2009a) did incorporate a "post-lexical processing stage", whereby recognized words would have a certain chance of 'not fitting with the prior context', thereby prompting a regressive saccade. This process of fitting recognized words with the prior context was not actually modeled in E-Z Reader; rather, regressions were triggered by sampling from a random distribution, intended as "[…] a placeholder for a deeper theory of post-lexical language processing during reading" (Reichle et al., 2009a, p.7). While such an approach would indeed allow a model to generate any desired number of regressions, the result is a model that mimics, rather than explains, reading behavior.

E-Z Reader determines its next saccade target by aiming for the center of the closest unrecognized word. The final saccade amplitude is subject to random and systematic error, the latter of which is a tendency to err towards a standard amplitude of 7 letters, hence accommodating the finding that saccades shorter than the standard amplitude tend to overshoot the target, whereas longer saccades tend to undershoot the target (McConkie, Kerr, Reddix & Zola, 1988).

The theoretical alternative to the SAS approach, represented by PG models, abandons the idea that words are processed in strict serial fashion. Instead, this approach assumes simultaneous processing of multiple words, with the amount of processing per word being modulated by a visuo-spatial attentional gradient (e.g. Engbert & Kliegl, 2011). The most prominent PG framework is the SWIFT model of Engbert et al. (2005). SWIFT is based on dynamic field theory, and assumes that each word in the perceptual span (i.e., the span of effective vision) has a level of activity that represents both the extent to which it is recognized as well as the probability that it is targeted by the next saccade (Engbert & Kliegl, 2011). The Gaussian distribution of visuo-spatial attention causes words near the center of attention to be activated more strongly, making it likely that foveal words are recognized sooner than upcoming words. As soon as a word reaches its recognition threshold (determined by word length and frequency), its activity falls back to zero, so that it will not be targeted by the next saccade. It is possible however, that previously fixated (or skipped) words have not yet been recognized, and that their activation level triggers a backwards saccade. As such, SWIFT has implemented an account of regressions whereas the E-Z Reader model of Reichle et al. (1999; 2006; 2009a) has not (even though the latter model does make regressions, as outlined above).

In SWIFT, the decision of when to move the eyes is determined by sampling from a random distribution, with an inhibition of random saccade timing by the amount of activation of the currently fixated word (Engbert & Kliegl, 2011). Thus, whereas SWIFT and E-Z Reader differ strongly in the decision of where to move the eyes (dynamic field activation versus hardcoded selection of the first unrecognized target, respectively), the decision of when to move the eyes is made in a similar fashion.

With their Glenmore model, Reilly and Radach (2006) provided a PG framework that is fairly similar to SWIFT, in that saccade targets are determined by dynamic field activation. The main difference between the two models is that Glenmore starts operating at the letter-level, whereas SWIFT (like E-Z Reader) operates at the word level only. In Glenmore, the combined activities of a word's constituent letters determine the activity of the word, and the word with the highest activity becomes the next saccade target (this is fairly similar to SWIFT, in which the most salient word has the highest probability of being fixated). Despite operating at the letter-level, Glenmore is not a model of orthographic processing in text reading: each letter unit in the visual field is connected to the appropriate word a priori, meaning that orthographic processing and subsequent lexical selection is assumed rather than modeled. As was acknowledged by Reilly and Radach (2006), the primary focus of their model was on saccade target selection rather than to implement a realistic word recognition module.

Glenmore differs from SWIFT and E-Z Reader in the decision of when to move the eyes. Whereas saccade program initiation in SWIFT and E-Z Reader operates on a random timer, in Glenmore saccades are initiated when the summed activation of all letters in the visual field reaches a certain threshold.

Whether words are processed serially or in parallel is still disputed. Yet, recent research has provided various types of evidence in favor of parallel processing. The first type of evidence is provided by corpus data, which has shown that the time spent viewing word *n* is influenced by the frequency and length of word *n+1* (Kennedy & Pynte, 2005; Kennedy, 2008), which is a natural consequence of parallel but not serial processing. The second type of evidence comes from experiments using the gaze-contingent boundary paradigm (Rayner, 1975) to manipulate word *n+1* during the fixation on word *n.* Such experiments have not only shown that (upcoming) parafoveal words can be lexically processed prior to being fixated (Hohenstein, Laubrock & Kliegl, 2010; Hohenstein & Kliegl, 2014; Schotter, 2013; Veldre & Andrews, 2015, 2016; Snell, Meeter & Grainger, 2017b), but also that processing of the upcoming word can occur simultaneously with processing of the fixated word (Dare & Shillcock, 2013; Angele, Tran & Rayner, 2013; Inhoff, Radach, Starr & Greenberg, 2000; Snell et al., 2017a). Finally, a third body of evidence is provided by the FLLD studies discussed in section 2.1. In these studies, foveal target word processing was shown to be influenced by parafoveal flanking stimuli (Dare & Shillcock, 2013; Grainger et al., 2014; Snell et al., 2017a; 2018a), despite the short stimulus on-time (150 ms).

Proponents of serial word processing have taken these findings to argue that parallel processing may occur on a sub-lexical (letter-) level, but that actual lexical access would still occur on a serial basis. Yet, using higher-order (syntactic, semantic) variants of the FLLD paradigm, it has been shown that syntactic decisions (e.g. classifying targets as *noun / verb*) and semantic decisions (e.g. classifying targets as *natural / artifactual object*) were made faster with respectively syntactically and semantically congruent flankers, as compared to incongruent flankers (Snell et al., 2017b; Snell, Declerck & Grainger, 2018b).

On the other hand, similar higher-order parafoveal-on-foveal effects have been elusive in sentence reading (e.g. Angele et al., 2013; Snell et al., 2017b; 2018b). This absence of higher-order parafoveal-on-foveal effects in sentence reading being largely regarded as evidence in favor of serial processing, relatively little attention has been paid to the possibility that higher-order information is simply not integrated across words during sentence reading, even if parallel processing were true. As argued by Snell et al. (2017b; 2018b), a successful parallel processing model would have to keep track of which information belongs to which word, rather than to integrate everything into a single mixture, given that each word has a unique role in contributing to sentence comprehension. The solution to this 'parallel processing problem', as proposed by Snell et al. (2017b; 2018b), is that activated words would be mapped onto a spatiotopic sentence-level representation, guided by expectations about word length and syntactic structure. Such a mechanism would prevent parafoveal-foveal integration of higher-order information during sentence-reading, while integration of parafoveal and foveal information can still be shown outside a sentence-reading setting—post-lexically—for instance at the level of making decisions, as has been shown in the flanker paradigm.

Interestingly, a key argument against parallel processing has been that a parallel processing system would not be able to recognize words in the correct order (e.g., Reichle, Liversedge, Pollatsek & Rayner, 2009b). The sentence-level representation effectively overcomes this challenge, by positing that activated words are associated with plausible locations. For instance, in "*The scientist is here*", '*is*' may be recognized before '*scientist*', but the former is much more likely to be associated with position three than position two, due to low-level visual cues (a two-letter word is expected at position three but not at position two) and syntactic constraints (given the article at position one, a verb is not expected at position two).

In line with the idea of a spatiotopic sentence-level representation is the finding that readers can make very accurate regressions to earlier points in the text to resolve syntactic ambiguity (e.g., MacDonald, Pearlmutter & Seidenberg, 1994; Inhoff, Weger & Radach, 2005), which suggests that readers must retain some representation of the syntactic structure of the sentence in working memory. Further evidence for the role of a sentence-level representation in parallel processing was provided by Snell and Grainger (2017). In line with Snell et al. (2017b; 2018b), they theorized that feedback from the sentence-level to the level of lexical representations would constrain the recognition process for individual words. A simple prediction that follows from this theory, is that word recognition should be better in a syntactically sound context, versus a syntactically incorrect context. In line with this prediction, using the novel Rapid Parallel Visual Presentation (RPVP) paradigm, Snell and Grainger (2017) found that the recognition of target words in briefly presented (200 ms) four-word arrays was better when those words formed a correct sentence, compared to when the same words were presented in a scrambled, ungrammatical order (with the target being presented at the same location in both conditions). This *sentence superiority effect* did not vary as a function of the target's position in the sequence, indicating that syntactic information was indeed retrieved from all four words during their 200 ms presentation time.

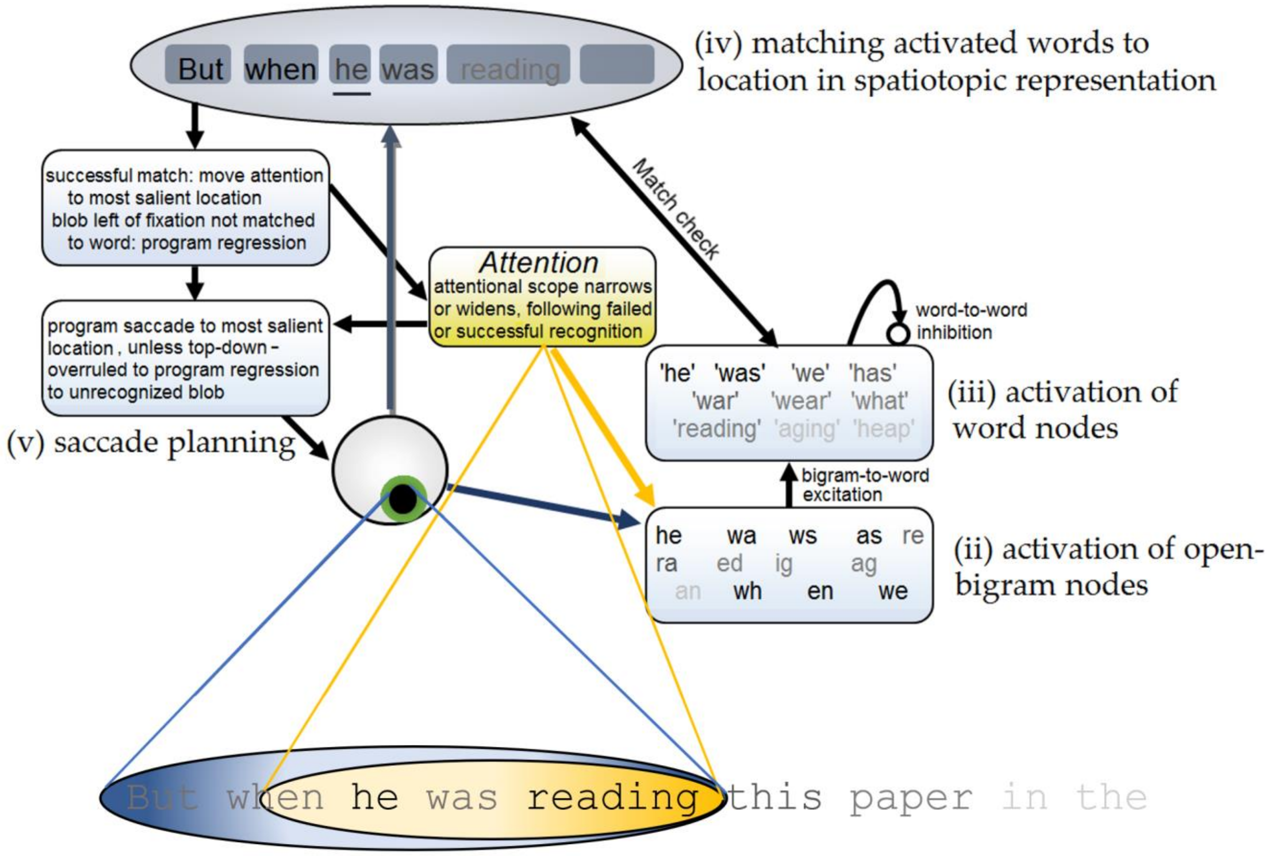

Figure 1. Schematic diagram of OB1-reader. (i) OB1 sees multiple words at a time (two words on either side of the fixated word, "he" in the figure). Letters occupy 0.33 degrees of visual angle. Within the visual input, letter processing is modulated by visual acuity and visuo-spatial attention. The attentional distribution is skewed towards the right. The focus of attention can be shifted independently of the eye's fixation. (ii) Open-bigram nodes are activated by the visual input, with stronger activation of letters that are close to the centers of fixation and attention, but weaker activation of crowded letters. (iii) Word nodes are activated by nodes coding for open-bigrams that occur in the word. Word nodes are inhibited by word nodes that share the same bigrams. (iv) Upon fixating a text, OB1 generates a spatiotopic sentence-level representation that represents expectations about the length of individual words. Word nodes that reach a recognition threshold are matched to locations ('blobs') in the spatiotopic representation based on length. OB1 only recognizes a word when it can be mapped onto a plausible location. Recognized words generate expectations about upcoming words, through feedback activation of word nodes based on Cloze-probability. When a word is successfully recognized, attention moves ahead of the eyes to the most salient location. Each word's saliency is determined by the proximity of its letters to the centers of fixation and attention. (v) Whether a saccade program is initiated is stochastically determined in each 25 ms processing cycle, with successful word recognition increasing the chance of initiation. The center of the most salient word form in the visual input becomes the saccade target. Saliency-based saccade targeting is overruled when a word location to the left of fixation has not yet been marked as recognized. Instead, a regression to that location will be executed. The attentional gradient is widened after each fixation during which a word is successfully recognized, while it is narrowed after each fixation without successful recognition.

3. OB1-reader: Key features and architecture

Summarizing the previous section, the single word recognition literature has spawned several modeling approaches, with recent evidence favoring relative position coding. Similarly, although two competing sets of models exist in the domain of eye-movement control in text reading, parallel graded processing models have received somewhat more support recently. These two approaches, relative position coding for word recognition on the one hand and PG-modeling for text reading on the other, make quite a natural fit. Both approaches assume parallel processing, with relative position coding assuming parallel letter identification in multi-letter strings, and PG-models assuming parallel processing of multiple words.

The question that remains is whether information from multiple words in a text could be successfully processed by a model that integrates these two approaches. Indeed, relative position coding models have only been tested in situations where the visual input was comprised of one or two words, while PG-models have abstracted away from the level of sub-lexical orthographic processing where confusion may occur. Hence, OB1-reader was developed to test whether a combination of relative letter-position coding and parallel graded word processing could yield behavioral patterns that are in accordance both with the literature on word recognition, as well as the literature on eye-movements in text reading.

In the present section we turn to a detailed description of OB1-reader. OB1 has five key components, which are marked in Figure 1, and which are discussed in the sub-sections below. An overview of the model's parameters is presented in Table 1, Section 3.6.

3.1 Spatially distributed processing

During each fixation, the visual input is comprised of the fixated word (*n*), along with words *n-2, n-1, n+1* and *n+2.*[38] Letters are perceived with variable strength depending on the visual acuity at those letters' respective eccentricities, and how attention is distributed across the visual field.

As acuity diminishes with eccentricity, visual input v_i generated by a letter *i* is assumed to be a decreasing function of eccentricity e_i, computed by assuming a letter size of .33 letters per degree of visual angle. Moreover, visual input is a function of the attentional weight a_i (Eq. 2) and a masking factor m_i. The masking factor reflects crowding, which causes letters to be less visible in the word's center than at its edges or in isolation (Marzouki & Grainger, 2014; Grainger, Tydgat & Isselé, 2010; Perea & Gomez, 2012). Following Marzouki and Grainger (2014) who found that the recognition of briefly presented (91 ms) letters at position 2 was approximately half as good as that of letters at position 1, m_i is set to 1 when *i* is an outer-positioned (edge) letter, and 0.5 when it is an inner-positioned letter. Together, these three factors determine visual input v_i in the following way:

$$v_i = a_i * m_i \left[{}^{1}\!\Big/_{c_e} \left(0.018 * e_i + \frac{1}{0.64} \right) \right] \qquad \text{Eq. 1}$$

The factor within square brackets represents eccentricity's influence on the input, and is assumed to be proportional to cortical magnification in V1 (the increase in cortical space devoted to locations closer to fixation). The term within round brackets was taken from Harvey and Dumoulin's (2011) estimate of cortical magnification in humans. Since this formula computes millimeters of cortex, constant c_e rescales magnification so that its maximum value, obtained for letters at the fixation location, equals 1 (c_e=35.56).

Attention (a_i in eq. 1) modulates the input according to a Gaussian distribution centered on the focus of attention. This approach to implementing attention is similar to that used in other parallel graded models of reading, such as SWIFT (Engbert et al., 2005) and Glenmore (Reilly &

---

[38] Clearly, the visual input during text reading normally comprises more than five words. However, we assumed that the influence of words beyond two positions from fixation would be negligible. We therefore limited the amount of visible words to five for computational efficiency.

Radach, 2006). The attentional weight a_i that a letter i receives is a function of f_i, its distance in letters to the focus of attention, and attentional width:

$$a_i = \frac{1}{width} * e^{-\frac{(f_i)^2}{2*(width*asym)^2}} + c_a \qquad \text{Eq. 2}$$

This function describes a Gaussian centered around an attentional focus, with the standard deviation of the Gaussian functioning as a changeable width and a hemispheric asymmetry *asym*. *Asym* is equal to 1 towards the right and 0.25 towards the left, in line with literature suggesting that the span of effective processing is approximately four times greater to the right (14-15 letters) than to the left (3-4 letters) (e.g., Rayner, 1986; 1998). Outside the Gaussian, the attentional weight is set to constant c_a (fixed at 0.25). The width of the attentional distribution is variable, and is determined by recent success and failure in word recognition: it increases after successful recognition, and decreases again after failure so as to produce more precise reading. As such, the attentional distribution adapts to the skill of the reader and the difficulty of the text that is being read (see Ans, Carbonnel & Valdois, 1998, for a similar proposal). Width can vary between 3 and 5 letters. It is increased 0.5 with every forward saccade, and reset to 3 with every regression. This flexible control of attention was not investigated extensively in the current simulations, but will be in future work.

Thus, each letter in the visual input is appointed a weight, set by its eccentricity, its distance from the focus of attention and by whether or not it is adjacent to a space. The model is simulated in cycles of 25 ms, with each saccade lasting one cycle. Visual input is constant during each fixation, and is absent during saccades to simulate saccadic suppression (Erdmann & Dodge, 1898; Campbell & Wurtz, 1978).

3.2 Activation of open-bigrams

All combinations of letters activate open-bigram nodes that respect the relative order of their constituent letters in the visual input. An open-bigram node is only activated when its constituent letters are within the same word (Grainger et al., 2014) and not further apart from one another than three letter positions. As in Grainger and van Heuven (2003), the activation of each open-bigram node is equal to the square root of its constituent letters' multiplied weights (as defined in Eq. 1). Hence, output O_{ij} of such a bigram node is computed as:

$$O_{ij} = \sqrt{v_i v_j} \text{ if } i,j \text{ within same word and max. 3 letters apart; 0 otherwise.} \qquad \text{Eq. 3}$$

Activity of single-letter bigrams, e.g. *#i,* is computed using the same formula, with j substituted for i (resulting in v_i being squared). If an open-bigram occurs multiple times in the visual input (e.g., '*ta'* in '*that task*'), its activation is the sum of the output from each individual occurrence.

3.3 Activation of word nodes

All open-bigram nodes are connected to nodes coding for the words that they occur in, and during reading those word nodes receive input equal to the sum of their constituent bigram activities. Activity of a word w, S_w, is initialized at 0, and is bound to the interval between 0 and 1. It is updated each time step with the following difference equation:

$$\Delta S_w = -\tau S_w + (1 - S_w) * \left[c_1\left(\sum_{i,j \in w} O_{ij}\right) - c_2\left(\sum_k d_{w,k} S_k\right)\right] \qquad \text{Eq. 4}$$

In this equation, the first right-hand term gives passive decay back to a value of 0 (τ =.05). This implements the idea that words' activities should decline in the absence of visual input (e.g., during saccades). The second term describes the input to the word node, multiplied by a factor $(1\text{-}S_w)$ that induces an asymptotic increase towards the maximum activity (equal to 1).

The input, between square brackets, consists of, firstly, excitatory input from the bigram nodes that are part of the word. This sum is multiplied by a constant c_1 (set to 0.0044). The second part of the input comprises inhibition from other word nodes. During word processing, activated lexical representations that share at least one open-bigram are considered to be lexical competitors (Grainger & van Heuven, 2003). To lower the chance that a word in the visual input leads to recognition of more than one word (i.e., instances where multiple word nodes reach the recognition threshold simultaneously), these lexical competitors exert mutual inhibition. The more active a word node is, the stronger its inhibition on competing word nodes will be. Additionally, inhibition is influenced by the amount of orthographic overlap between each word pair: words that share many open-bigrams inhibit each other more than words that do not. This can be thought of as a weight on the inhibitory connection between word nodes. Activity from each word node k is therefore weighted by a factor $d_{w,k}$ that is equal to the number of open bigrams shared between the words. This term is multiplied by constant c_2 (equal to 1.4 divided by the size of the lexicon, since word-to-word inhibition increases with larger lexicons). This architecture is equal to that used in the Open-Bigram model of Grainger and van Heuven (2003).

OB1 recognizes a word when activation of a word node reaches a recognition threshold. This threshold is influenced by the length, frequency and predictability (given preceding words) of each respective word (i.e., longer, less frequent and less predictable words have a higher threshold; e.g. Rayner, 1998; Kennedy & Pynte, 2005; Bicknell & Levy, 2012; Kliegl, Grabner, Rolfs & Engbert, 2004) with the following formula:

$$T = C\,\frac{c_4\,ln(freq_{max}) - ln(freq_w)}{c_4\,ln(freq_{max})} * \frac{c_5\,ln(pred_{max}) - ln(pred_w)}{c_5\,ln(pred_{max})} * \left(1 - \boldsymbol{c_6}^{-c_7\,\boldsymbol{l_w}}\right) \qquad \text{Eq. 5}$$

Here, l_w is the length of word w, $freq_w$ is the frequency with which w occurs within a reference corpus (we used SUBTLEX-DE for German), and $freq_{max}$ is the frequency of the most frequent word within that corpus. $Pred_w$ is the cloze probability of w given the preceding words (values obtained by Kliegl et al., 2006 and Laubrock & Kliegl, 2015), and $pred_{max}$ is the maximum cloze probability within the sentence corpus. Scaling parameters c_4 and c_5 determine the size of the effect of frequency and predictability on the threshold, with lower values indicating stronger effects; these were set to 5.5 and 9.0, respectively. Scaling parameters c_6 and c_7 (values 0.61 and 0.44, respectively) alter the size of the effect of frequency and predictability as a function of word length, to compensate for the stronger excitation of long words. Overall scaling value C was set to .22 (C was not an independent free parameter, since it can be scaled against the input parameters c_1 and c_2). Similar variable recognition thresholds are employed by E-Z Reader and SWIFT (in contrast, in Glenmore, increased frequency and predictability lead to stronger per-cycle activation, rather than a lower recognition threshold).

3.4 The spatiotopic representation

Since open-bigram information is location-independent, OB1 does not inherently know which activated lexical representation belongs to which spatial location. However, low-level visual information allows OB1 to generate expectations about the number of to-be recognized words in the visual field, as well as their approximate length; (as illustrated in Figure 1, to-be recognized words are initially perceived as 'blobs'). This information operates as a spatiotopic sentence-level representation in working memory, to which word identities may be appended, or on the basis of which activated word candidates may be rejected (e.g. Snell et al., 2017b; 2017c). As such, word recognition would be a process of matching activated representations to perceived 'blobs': the length of the activated representation must meet the expected word length, in order for the representation to count as recognized. As an example, the phrase *'sit for dinner'* may lead to erroneous recognition of the word *'sinner'* (which would be mapped onto the third word position since it has a matching length) but not *'beginner*', since the latter representation does not match any of the word lengths occurring in the phrase. Moreover, if the eyes were fixated on the first word, but the node coding for '*for*' reaches its threshold earlier than '*sit*', '*for*' would be erroneously linked to the first word position.

The spatiotopic representation is implemented by means of the creation of an array of word lengths (representing length in number of letters for words *n-2* to *n+2*) upon each fixation. The array's indices represent word positions, and these are marked either as *recognized* or *not-recognized*. Prior to the activation of a given word in OB1's lexicon (as per the mechanisms described in Section 3.3), a check is performed whereby the word has to match one of the values in the array. When the word's length does not approximate any of the values, the word does not receive activation. Similarly, when a word reaches its recognition threshold, its length has to approximate one of the array's *not-recognized* length values in order to be count as recognized. It would not be realistic to assume that OB1 is able to count letters in the periphery, so word length is estimated with a 15% error margin, such that a seven-letter representation might also be matched to a six- or eight-letter word form in the spatiotopic representation.

3.5 Saccade planning

Research has indicated that it takes approximately 125 ms to plan and execute a saccade (e.g. Becker & Jürgens, 1979, Meeter & Van der Stigchel, 2013). Thus, given that the average time spent viewing a word is short (around 200–300 ms; see e.g. Rayner, 1998), the decision to execute a saccade has to take place in the first 100 ms of a fixation (but do note that longer fixations principally allow for more lenient numbers). This has led researchers to argue that word recognition cannot be the sole factor driving eye-movements in reading (e.g. Reichle et al., 2003). In each of OB1's processing cycles, random sampling from a Gaussian distribution $N(\mu, \sigma)$ determines whether a saccade program is initiated (this approach is similar to that employed in SWIFT and E-Z Reader). Lexical processing influences the decision of when to move the eyes in so far that a wider range of values from this distribution is taken as decision to program a saccade when a word is recognized (σ = 125 ms, μ = 50), as compared to when no word has been recognized yet (σ = 95 ms, μ = 50ms.). Whenever the time that has elapsed since the last eye movement is larger than this sample, attention is shifted in the direction of the upcoming saccade (e.g. Baldauf & Deubel, 2008). Processing of upcoming words may thus be enhanced during saccade programming. An eye movement rigidly follows 100 ms after each attentional shift. With the 25 ms motor delay, this aligns with the 125 ms estimate for saccade planning.

Under normal conditions, the saccade target location is determined by the visual salience of word forms in the visual field. This salience is the sum of the weights of words' constituent letters, as determined by crowding, eccentricity and proximity to the focus of attention (see section 3.1). Large words close to the focus of attention are usually most salient, and are thus selected as the target. This approach to saccade target selection is equal to that employed in the Glenmore model of Reilly and Radach (2006), and fairly similar the SWIFT model of Engbert et al. (2005) in which a word's activity determines the probability with which it is fixated; (in contrast, in the E-Z Reader model of Reichle et al. (2003) word activation does not play a role, as the first non-recognized word is selected as the saccade target). The center of the target word is taken as the intended landing location. However, because saccades are imprecise, the final location is affected by both systematic and random error. The systematic error reflects the principle that eye movements tend to overshoot nearby targets and undershoot faraway targets (e.g., Kapoula & Robinson, 1986; McConkie et al., 1988), and is modeled as a tendency to err towards a standard distance, D (set to 7 letters). Random error is assumed to be Gaussian, with a standard deviation increasing as a function of intended saccade size. The number of letters moved, h, given a target distance d, is therefore equal to, rounded to the nearest integer:

$$h = N(\mu, \sigma) \quad with \quad \mu = d + .2\,(d - D), \quad \sigma = .18 + .08\,d \qquad \text{Eq. 6}$$

For the word that is already being fixated, salience is computed only for the portion to the right of the currently fixated letter. This prevents leftward letters from having an influence when the intention is to make a forward (rightward) saccade. Naturally, that is not to say that leftward letters are not visible; rather, we assume that the direction of the saccade (left/right) is determined before the actual goal of the saccade. Given that the two visual hemifields are represented in different hemispheres of the brain, the directional decision entails that saliency is computed for visual input in one hemisphere without interference from the other.

It is possible that saliency-based target selection is overruled by the need to make a regression. This happens when any of the words that the eyes have already gone past has not yet been marked as recognized (i.e., there would still be an unmarked 'blob' to the left of the fixated location in the spatiotopic representation; Section 3.4). In such a case, the unrecognized word is marked as the target, prompting a regressive saccade. An unrecognized word can trigger only one regression: if this does not result in successful recognition, the word is simply left unidentified (which would reveal a weakness in OB1's word identification capabilities).

4. Evaluation of the model

We expect OB1 to account for a range of low-level reading phenomena, such as refixations, regressions, word skips, preview effects and spillover effects. Its word recognition mechanisms should be able to account for orthographic parafoveal-on-foveal effects (e.g. Dare & Shillcock, 2013; Angele et al., 2013; Grainger et al., 2014; Snell et al., 2017a), neighborhood effects (e.g., Grainger, O'Regan, Jacobs & Segui, 1989; Perea & Pollatsek, 1998; Acha & Perea, 2008) and possibly lexical parafoveal-on-foveal effects as reported by Kennedy and Pynte (2005). We also expect OB1 to account for the occasional misreading (i.e., instances of erroneous word recognition); something no previous model of text reading has been able to simulate. Quantitatively, OB1-reader should depict a distribution of word viewing times similar to that obtained in experimental settings. In particular we expect OB1 to show word length, frequency and predictability effects, all of which are well-established in the literature (e.g. Rayner, 1998).

The present section describes how we evaluated these factors. Our assessment consisted of two parts: firstly, we let OB1 read sentences from the Potsdam sentence corpora (PSC; Kliegl et al., 2006; Laubrock & Kliegl, in preparation) and compared simulation results with their experimental data. Secondly, we simulated the experiment of Dare and Shillcock (2013) that obtained an orthographic parafoveal-on-foveal effect using the gaze-contingent boundary technique to manipulate word *n*+1 during the fixation on word *n*. For the sake of consistency, this simulation also made use of the PSC reading materials, rather than the stimuli used by Dare and Shillcock.

4.1 The Potsdam sentence corpus simulation

4.1.1 Reading materials

We used the 577 sentences (4921 words) from the EyetrackR package (Laubrock & Kliegl, in preparation) as reading materials. This package comprises the PSC (Kliegl et al., 2006), PSC2 (Laubrock & Kliegl, 2015) and a sample of the Potsdam Commentary Corpus (PCC; Stede & Neumann, 2014). The experimental data obtained with these materials comprises eye-movement data of 180 participants between 15–80 years old. The log-frequency of each word in these texts was determined with the SUBTLEX-DE database of Brysbaert et al. (2011), and is based on the occurrence of the word in German subtitles for film and television. Kliegl et al. (2006) and Laubrock and Kliegl (2015) obtained the predictability value for each word using the incremental-cloze task.[39]

OB1 was given a mental lexicon comprised of all the words occurring in one or more of the three Potsdam sentence corpora, such that the lexicon contained 701 unique word forms. We further made sure that the 200 highest-frequency words as indicated by the SUBTLEX-DE database of Brysbaert et al. (2011) were part of the lexicon, leading us to add another 75 words to bring the final lexicon size up to 776 words.[40]

4.1.2 Model parameter fitting

An overview of all parameters is presented in Table 1 below. We make a distinction between *free* and *fixed* parameters. Free parameters are parameters for which we could not make reasonable estimations, and which were thus determined through trial and error. These include decay, bigram-to-word excitation and word-to-word inhibition. Fixed parameters were determined a priori, with values being based on reasonable theoretical assumptions (e.g., values reported in the literature).

The process of fitting the model's free parameters (i.e., bigram-to-word excitation, word-to-word inhibition and per-cycle activity decay) consisted of, firstly, taking values representing bigram-to-word excitation and word-to-word inhibition from the Open-Bigram model of Grainger and van Heuven (2003), and secondly, repeatedly letting the model read short texts and making continuous slight parameter adjustments to approximate realistic model output. Here, we mainly focused on word viewing times and fixation type probabilities. It is possible that a more extensive

[39] In the incremental-cloze task, participants start each trial by guessing the first word of a sentence, after which the actual word is displayed and participants have to guess the next word. This process continues until the end of the sentence is reached (see Kliegl et al., 2006).

[40] We acknowledge that the present lexicon size is relatively small compared to the actual number of words known by skilled readers. It should be noted, however, that the present lexicon size is comparable to that used in single word recognition models. Moreover, larger lexicons would slow the simulations exponentially.

parameter search would have yielded better quantitative fits (but see e.g. Roberts & Pashler (2000) for arguments against the importance of quantitative fit). During the heuristic fitting process, we first fitted Dutch and English short texts, before going on to use the PSC2 reading set. Switching text language did not noticeably affect model performance. Note that all final parameter values were fixed during the actual simulation.

Table 1. OB1's parameters

Parameter	*Value*	*Eq.*	*Description*	*Determination*
τ	0.05	4	Decay	Heuristic fitting
c_1	0.0044	4	Bigram-to-word excitation	Heuristic fitting
c_2	0.0018	4	Word-to-word inhibition	Heuristic fitting
c_e	35.56	1	Scaling cortical magnification	Cortical magnification derived from Harvey & Dumoulin (2011), scaled so as to have max. 1
m_i	1 for outer-, 0.5 for inner letters	1	Masking factor describing crowding	Marzouki & Grainger (2014)
Asym	1 toward the right, 0.25 toward the left	2	Asymmetry of attention	Four times greater toward the right than toward the left (Rayner, 1998)
c_a	0.25	2	Residual attentional weight outside of focus of attention	A priori
Max/min attention	5.0 / 3.0	2	Maximum and minimum size of attentional window	A priori
Time step	25 ms	n.a.	Duration of 1 time step	Average duration of a saccade
c_4	5.5	5	Weight of word frequency in threshold setting	A priori
c_5	9.0	5	Weight of predictability in threshold setting	A priori
c_6	0.61	5	Maximum lowering of threshold for short words	A priori
c_7	0.44	5	Scaling of effect	A priori

Note: *A priori* parameters were fixed prior to simulating the Potsdam Sentence Corpus.

4.1.3 Procedure

The 577 sentences were presented to OB1 as one continuous sequence of words. Given that its saccade planning mechanisms make OB1 non-deterministic, we let OB1 read the materials four times, and the simulation results thus represent an average of these four simulations. Because of the size of the stimulus set, each replication yielded similar averages and adding replications did not alter results.

4.1.4 Simulation results

Figure 2 shows the various fixation type probabilities for OB1 compared with the Potsdam experimental data. The simulated fixation type probabilities approach the experimental data quite well, with a slight overestimation of the amount of refixations and regressions and a slight underestimation of the amount of single fixations. Moreover, these probabilities were modulated by word length (Figure 3), frequency (Figure 4) and predictability (Figure 5) in a way very similar to that depicted by the Potsdam experimental data.

Like the experimental data, OB1 shows an effect of word length on word viewing times, with longer words leading to longer viewing times (Figure 6). In both the simulated and experimental results this effect is expressed in gaze duration (GD) and total viewing time (TVT), but not in the single fixation duration (SFD). Here, SFD refers to cases where words were fixated only once. GD refers to the sum of all first-pass fixations, i.e., the sum of first fixations and refixations but not refixations following a regression. TVT refers to the sum of all fixations including regressions. OB1 also depicts an effect of word frequency on word viewing times similar to that of the Potsdam experimental data (Figure 7).

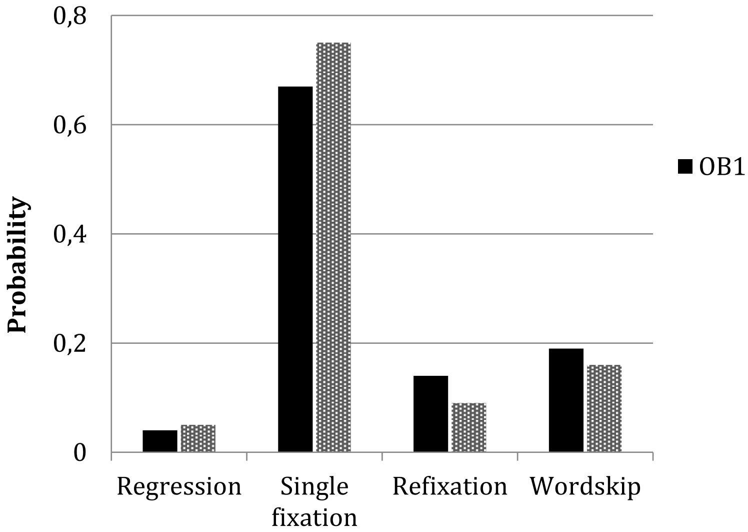

Figure 2. Fixation type probabilities for the simulated data of OB1 and the Potsdam experimental data.

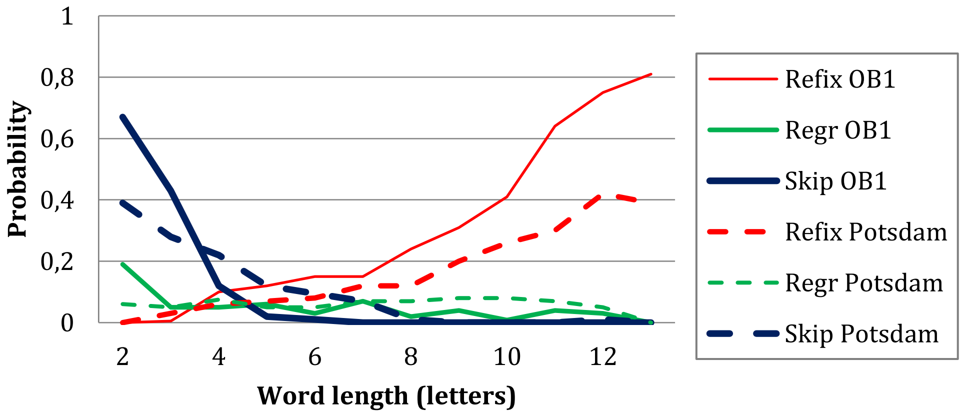

Figure 3. Fixation type probabilities for the simulated data of OB1 (solid lines) and the Potsdam experimental data (broken lines), as modulated by word length. Abbreviations: Refix: refixation; Regr: regression.

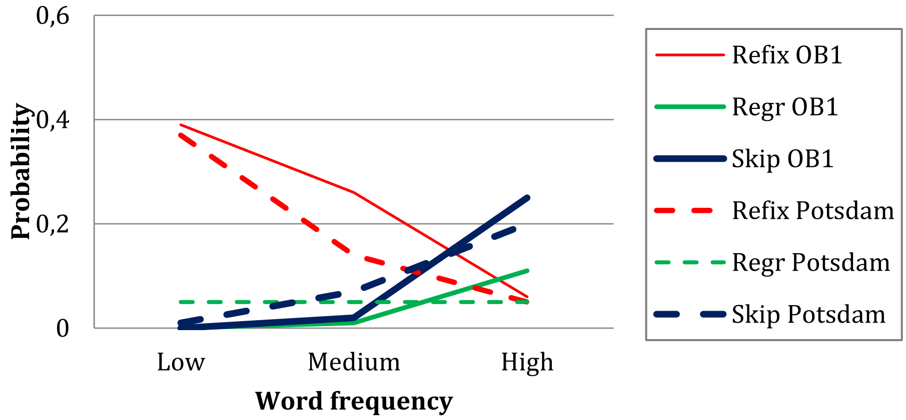

Figure 4. Fixation type probabilities for the simulated data of OB1 (solid lines) and the Potsdam experimental data (broken lines), as modulated by word frequency (data was split into log-frequency tertiles). Abbreviations: Refix: refixation; Regr: regression.

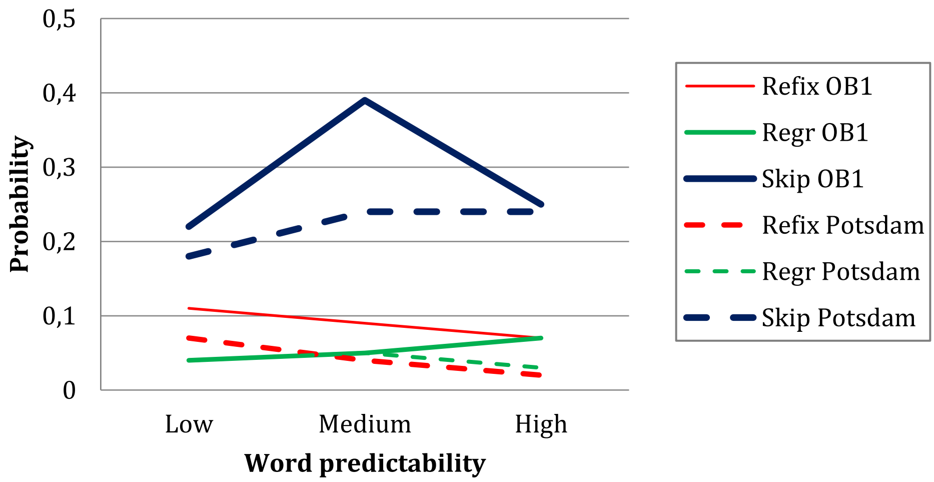

Figure 5. Fixation type probabilities for the simulated data of OB1 (solid lines) and the Potsdam experimental data (broken lines), as modulated by word predictability (data was split into predictability tertiles). Abbreviations: Refix: refixation; Regr: regression.

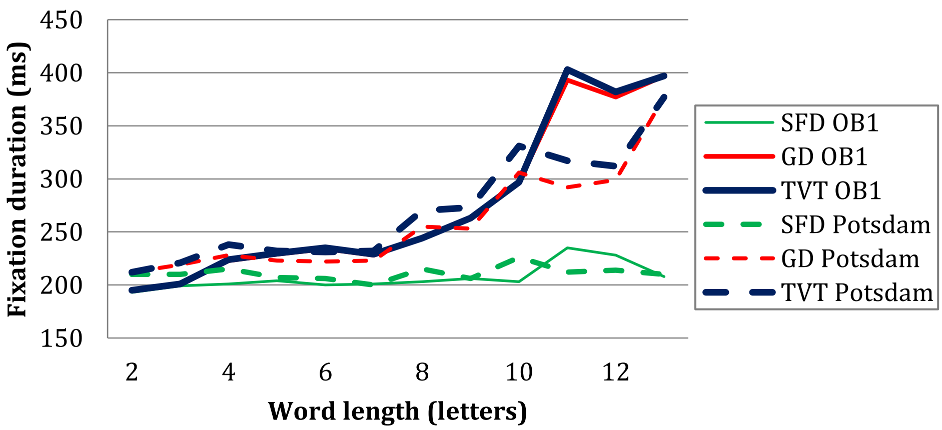

Figure 6. Word viewing times across all word lengths occurring in the corpora, for OB1 (solid lines) and the Potsdam experimental data (broken lines). Abbreviations: SFD: single fixation duration; GD: gaze duration; TVT: total viewing time.

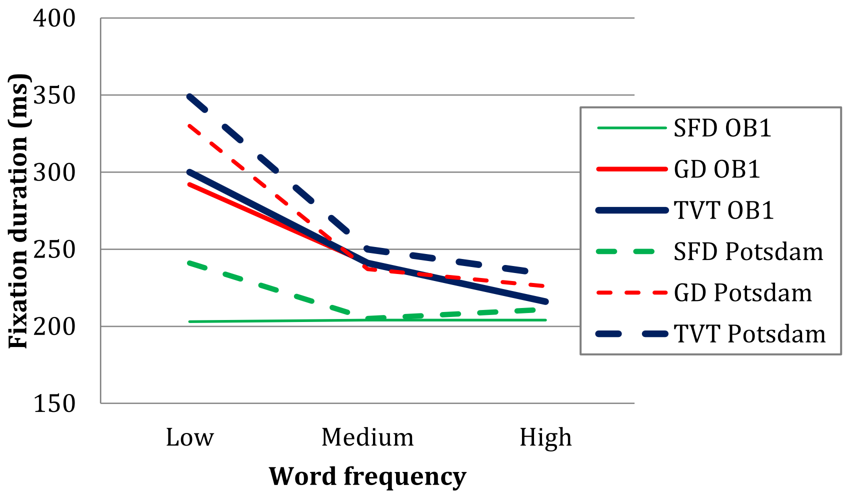

Figure 7. Word viewing times as modulated by word frequency (divided into log-frequency tertiles), for OB1 (solid lines) and the Potsdam experimental data (broken lines). Abbreviations: SFD: single fixation duration; GD: gaze duration; TVT: total viewing time.

The saccade amplitude (i.e., the distance between two consecutive fixations) was slightly less variable in the simulation than in the experimental data (Figure 8). The normal forward saccades (from word *n* to word *n+1*) were slightly shorter in OB1, while word skipping saccades tended to be longer.

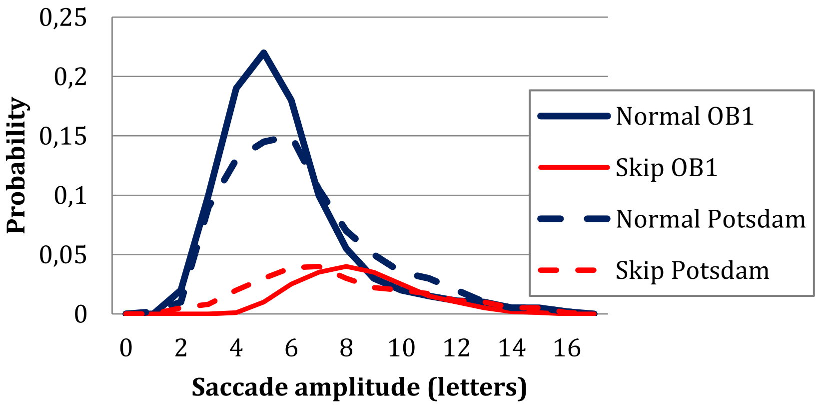

Figure 8. Distribution of saccade amplitudes for normal forward saccades and word skipping saccades.

We also tested for lag- and successor effects, whereby the time spent viewing word *n* is influenced by the frequency and predictability of words *n-1* and *n+1* respectively (Kennedy & Pynte, 2005). While these effects occurred in the experimental data, they could not be captured in the simulation results (Figures 9 and 10). We address this point in the General Discussion.

The last thing that we tested for in this simulation was the neighborhood size effect, whereby word recognition is slowed as the number of existing high-frequency orthographic neighbors (e.g. *blur* – *blue*) increases (Grainger et al., 1989; Perea & Pollatsek, 1998; Acha & Perea, 2008). We plotted the GD against the neighborhood size, and indeed found that GD increased with an increasing amount of high-frequency orthographic neighbors (Figure 11).

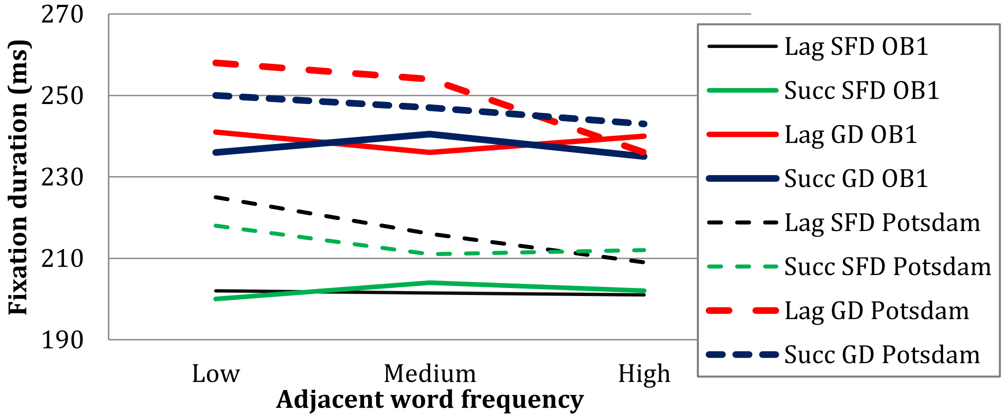

Figure 9. Influence of the frequency words *n-1* (lag) and *n+1* (successor) on word *n* viewing times in OB1 (solid lines) and the Potsdam experimental data (broken lines). Abbreviations: SFD: single fixation duration; GD: gaze duration; Succ: successor.

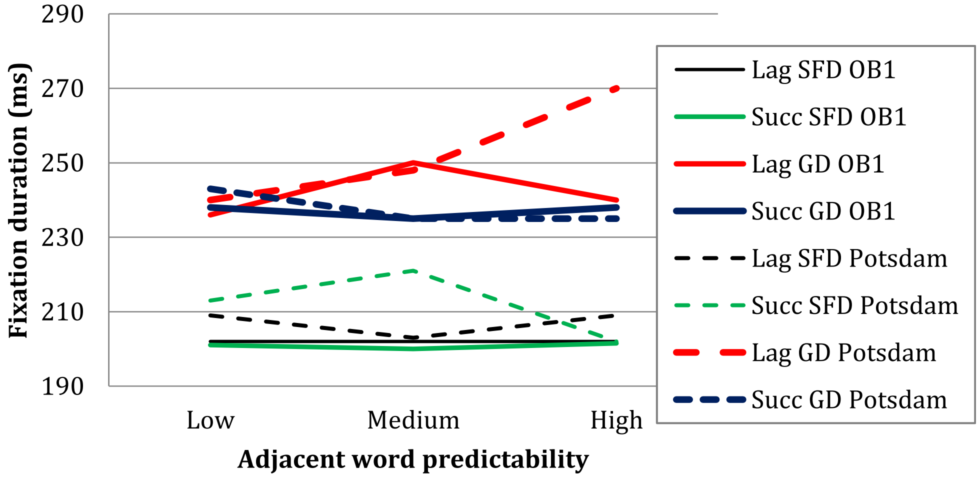

Figure 10. Influence of the predictability of words *n-1* (lag) and *n+1* (successor) on word *n* viewing times in OB1 (solid lines) and the Potsdam experimental data (broken lines). Abbreviations: SFD: single fixation duration; GD: gaze duration; Succ: successor.

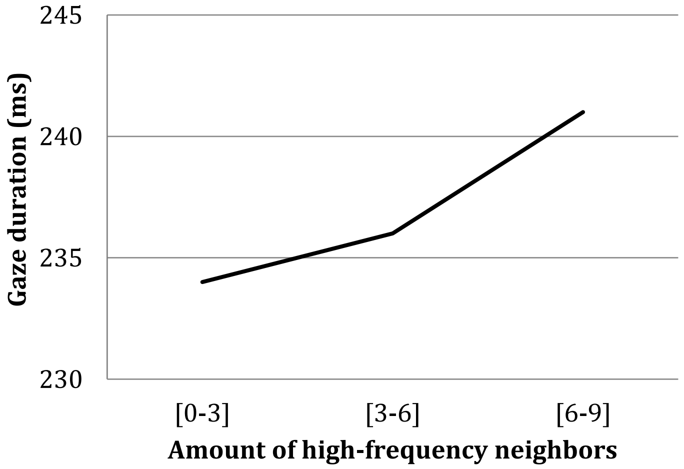

Figure 11. Gaze duration as modulated by the high-frequency orthographic neighborhood size.

4.1.5 Unrecognized words

Approximately 1% of the words were not recognized by OB1. Strikingly, as can be seen in Figure 12, word recognition probability was the lowest for 2-letter words. It would be difficult to establish how many words are misread during normal reading, as post-lexical processes (which are lacking in OB1) would probably correct many errors. For example, Angele and Rayner (2013) the have shown that readers tend to not notice (and skip) the article '*the'* when it is at an incorrect position (such as in this sentence, right after the reference), indicating that in normal reading, high-frequency function words may go unrecognized as well.

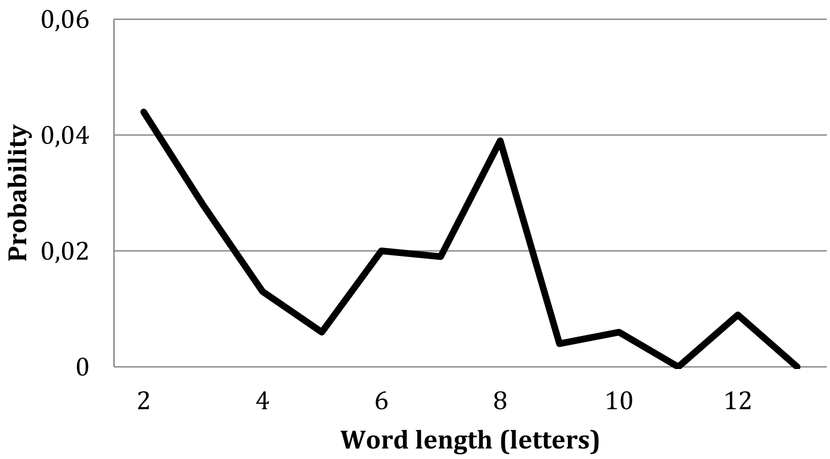

Figure 12. Probability that words were not correctly recognized, as modulated by word length.

4.2 Simulation of the boundary paradigm

4.2.1 Procedure

Next, we simulated the gaze-contingent boundary technique as employed by Dare and Shillcock (2013), Angele et al. (2013) and Snell et al (2017a). For this simulation, we filtered all sentences with the occurrence of two adjacent 4- or 5-letter words from the reading materials of the EyetrackR package (e.g. '*The people <u>stay</u> <u>here</u> now*'). The first of the two adjacent words was marked as the target in all these sentences, and for each target we retrieved a control word from the SUBTLEX-DE database (Brysbaert et al., 2011) that was equal in length, had no orthographic overlap with the target and that had a log-frequency value that deviated from the target's by 1.0 at most (e.g. *stay – jump*).

Each of these sentences was presented three times to OB1: once with the target word (position *n*) being repeated at position *n+1* during the fixation on *n* (e.g. '*The people <u>stay</u> stay now*'; the repetition condition), once with the control word at position *n+1* (e.g. *'The people <u>stay</u> jump now*'; the control condition), and once in the original form (e.g. *'The people stay here now*'; the baseline condition). During the saccade from word *n* to word *n+1*, the latter word was changed into its original form (thus, nothing changed in the baseline condition). The word viewing times on *n* were compared across these three conditions.

4.2.2 Simulation results

In line with results reported by Dare and Shillcock (2013) and Angele et al. (2013), OB1 depicted shorter viewing times on the target (*n*) when the target was repeated in *n+1*, compared to when *n+1* was an orthographically unrelated control word (Table 2). The rate of refixations was decreased in the repetition condition as well.

Table 2. Average fixation duration (in ms.) and refixation probability across conditions.

Condition	*Single fixation duration*	*Gaze duration*	*Refixation probability*
Repetition	200	210	0.13
Control	205	243	0.20
Baseline	203	242	0.19

Further aligning with the studies of Dare and Shillcock and Angele et al., differences were observed between the repetition and baseline condition, while results in the baseline and control condition were virtually equal.[41] With respect to effect size, the difference in SFD between the repetition and control condition was similar to that reported by Angele et al. ($b \approx 7$ ms in their study, versus 5 ms in our simulations), while the difference in GD was more pronounced in OB1 ($b \approx 20$ ms versus 32 ms, respectively).

Hence, these simulation results underline the theoretical plausibility of the idea that orthographic parafoveal-on-foveal effects are driven by location-invariant activation of sub-lexical nodes (e.g. bigrams, letters) by letter information across the visual field.

5. General discussion

In this paper we have described a set of theoretical ideas about word recognition and eye-movement control in reading, along with a computational model that integrates these ideas. OB1-reader is the first model of eye-movements in text reading that incorporates a word recognition module wherein letter information from the visual field activates lexical candidates. At the same time OB1-reader distinguishes itself from word recognition models by moving from isolated word recognition to text reading, taking into account evidence that words are processed not only in– but also beyond the fovea.

Our simulations show that OB1 successfully recognizes most words in the text, and that it reproduces orthographic effects such as that of neighborhood size (e.g. Grainger et al., 1989; Perea & Pollatsek, 1998; Acha & Perea, 2008). OB1 also accounts for a range of text reading phenomena, such as refixations, regressions, word skips and preview effects (e.g. Rayner, 1998). Quantitatively, OB1 produces a distribution of word viewing times, word length, frequency and predictability effects, and landing positions similar to those obtained in experimental settings.

[41] Do note that Snell et al. (2017a) did observe longer word viewing times in the control condition compared to the baseline condition. This difference was ascribed to readers' awareness of the syntactically implausible preview at position *n*+1 during the fixation on *n* in the control condition. Naturally, the lack of post-lexical processes in OB1 prevents the model from capturing such effects.

However, the main advance may be theoretical. OB1 is a descendant of the parallel-graded attention line of models of eye-movement control (e.g. SWIFT, Glenmore) on the one hand, and relative position-coding models of word recognition (e.g. the Open-Bigram model) on the other. In specific, OB1 adopts the successful approach of SWIFT and Glenmore in addressing the question of where to move the eyes during reading (i.e., saliency-based target selection). By adopting the relative position-coding approach to word recognition, OB1 has a clear means to code for letter position across multiple words in parallel.[42] As evidenced by our simulations, the integration of these approaches allows OB1 to account not only for the 'traditional' phenomena mentioned above, but also for sub-lexical parafoveal-on-foveal effects as reported in more recent research (Angele et al., 2013; Dare & Shillcock, 2013; Grainger et al., 2014; Snell et al., 2017a). Moreover, this integration allows us to inspect how and when word identification processes in text reading may go awry; a feature that is not possessed by other models of text reading.

Another theoretical advancement of OB1 is its use of a spatiotopic sentence-level representation. This representation answers the question of how a parallel processing system can successfully identify multiple words without losing track of word order, hence meeting one of the major challenges raised against parallel processing systems by proponents of serial processing (Reichle et al., 2009b). Low-level visual information, which is used to associate activated words with plausible positions, can further be used to constrain word activation. This provides a valuable counterweight to the possibility that open-bigrams activated by multiple words in the visual field are combined to activate an incorrect word (e.g., in *'The butter flies through the room'*, *'butter flies'* would activate '*butterflies*' if not for the guidance of the spatiotopic representation). In accordance with this idea is the finding of Inhoff et al. (2003) that target words are recognized faster after viewing length-accurate parafoveal previews, compared to length-inaccurate previews. Finally, a spatiotopic representation would explain how readers can make accurate long-range regressions to words earlier in the sentence (e.g. Macdonald et al., 1994; Inhoff et al., 2005)

Given the pivotal role of the spatiotopic representation in OB1, one may wonder how readers can be fairly successful in reading unspaced text (e.g., '*youcandefinitelyreadthis*'). Regarding this, it should be noted that while one can indeed read unspaced text, the removal of inter-word spaces undoubtedly has a negative impact on the ease with which the text is read (e.g., Epelboim, Booth & Steinman, 1995; Perea & Acha, 2009; Rayner, Fischer & Pollatsek, 1998; Mirault, Snell & Grainger, n.d.). OB1 would deal with the removal of spaces by activating all words containing present bigrams, regardless of word length. Interestingly, earlier simulations of the model without length constraint did show fairly successful reading behavior, albeit with a 9% rate of unrecognized words (compared to 1% in the present simulations). It is possible that the human reading system would show a similar number of errors during unspaced reading, but that post-lexical processes (e.g., determining whether activated words fit with the prior context) correct for falsely identified words.

Further, while low-level visual information (e.g. word length) may not be available at first glance during unspaced reading, it is conceivable that readers nonetheless engage a sentence-level representation. For instance, Mirault et al. (n.d.) found that, even in unspaced reading, saccade amplitudes were influenced by the length of fixated as well as upcoming words,

[42] In contrast, it is not clear how parafoveal letters should connect to lexical representations if those letters would be coded for their absolute position (e.g., Gomez et al., 2008) – especially given the increased positional noise at increased eccentricities. The SOLAR model of Davis (1999; 2010) forms an exception, as words away from fixation are activated quite similarly as those at fixation. However, it is not clear how the SOLAR model would be able to account for orthographic parafoveal-on-foveal effects.

suggesting that readers mentally parse unspaced text into separate words at quite a rapid pace, conceivably driven by lexical identification processes as well as top-down expectations. In any case, accounting for reading of unspaced text remains an interesting challenge for ongoing model development, especially considering the possibility of accommodating alphabetic writing systems that do not use spaces, such as Thai (e.g., Winskel, Radach & Luksaneeyanawinc, 2009).

5.1 Limitations and future directions

While OB1 approaches experimentally observed reading behavior quite well, OB1 is not a perfect model of reading. One weakness is OB1's inability to recognize some words (Figure 12), with words of length 2 having the lowest recognition rate (95.6%). It may be that certain high-frequency short words have a dedicated representation that generally allows them to be activated directly, without having to rely on bigram-to-word activation. In this regard, it is interesting to note that the most frequent word in the English language, '*the*', tends to be skipped even when it appears at an unpredictable and ungrammatical location (Angele & Rayner, 2013), suggesting that it has special status. In any case, the fact that some words were not recognized is informative in the sense that even skilled readers may err at times – either consciously or subconsciously – and it is likely that post-lexical processes allow human readers to correct such mistakes (whereas OB1 cannot).

The implementation of higher-order feedback processes, involving syntactic and/or semantic constraints, may alleviate this shortcoming of the model. As proposed in Snell et al. (2017b), activated words may be categorized syntactically (e.g. *noun, verb*), and be appended to a syntactic sentence-level representation that follows the grammatical rules of a given language. Feedback from this higher layer to individual word representations (in the form of activation or inhibition, for syntactically legal and illegal words respectively) would constrain the recognition process for those words. For example, if the article '*ein'* is surrounded by a verb at *n*-1 and a noun at *n*+1, syntactic feedback would strongly activate '*ein'* as one of the few plausible articles for position *n*, whilst inhibiting syntactically implausible words such as the verb '*eingestallt'*. Indeed, as discussed in Section 2.2, Snell and Grainger (2017) provided evidence in favor of this theory, as word recognition was found to be better in grammatical than in ungrammatical contexts.

Another imperfection is that OB1 did not capture the lexical lag- and successor effects (e.g. Kennedy & Pynte, 2005) that were present in the Potsdam experimental data. Orthographic overlap is a prerequisite for OB1 to display interactions among words in the fovea and parafovea (i.e., bigram-to-word excitation and word-to-word inhibition). This implies that, if anything, word viewing times should be increased by highly frequent adjacent words, as the nodes belonging to those adjacent words should exert more inhibition on the node belonging to the foveal word. Yet, the experimental data showed a reversed pattern with higher-frequency *n*+1's leading to a shorter gaze duration on word *n*, suggesting that this lexical successor effect is not driven by direct word-to-word dynamics as displayed by OB1. It rather seems that high-frequency successors demand fewer processing resources, subsequently leading to stronger activation of the fixated word.

Applying this conception, future implementations of OB1 may adopt a different approach to how visuo-spatial attention is distributed. In its current form, the model follows the proposal of Ans et al. (1998) with an attentional distribution that has a variable width tuned to recent success and failure in word recognition. An alternative approach would be to let the width of the attentional gradient be influenced by the speed with which parafoveal words become active, with increased activation leading to a narrowed attentional distribution centered on the fixated word.

As such the gradient width would not be determined by failure and success, but rather by *anticipated* failure and success. Future simulations should point out whether such an adjustment allows OB1 to effectively account for lexical lag- and successor effects.

Finally, in the current implementation of the model, the attentional gradient width dynamic is only used to allow the model to find an optimal size of the attentional window for the input it receives. In the future, this dynamic may be explored in more detail, as it possibly accounts for the progression of reading with increased skill – from reading letter-by-letter in beginning readers to reading whole words in experienced readers (e.g. Rayner, 1998). Sub-optimal deployments of attention, such as one where the gradient width is too large given the reader's skill (e.g. Geiger et al., 2008), leading to increased parafoveal interference, may account for cases of poor reading and dyslexia. OB1 provides a suitable theoretical framework for putting such scenarios to the test.

5.2 Concluding remarks

In conclusion, we believe that OB1-reader provides an important step in the convergence of the neighboring domains of single word recognition and eye-movement control in text reading. The model's architecture successfully accounts for a wide range of phenomena, including phenomena that were not explained by other models of text reading (e.g. neighborhood size effects, orthographic parafoveal-on-foveal effects). The architecture further allows one to track not only normal reading, but also reading development (e.g. manipulating the width of the attentional window to simulate differences between beginning and skilled readers), and potentially processes involved in dyslexia. Finally, the connectionist approach of the model allows for easy expansion, such as the implementation of syntactic constraints as discussed above. Future research will reveal how well OB1-reader fares in exploring these various domains and components.

Acknowledgements

This research was supported by grant ANR-11-LABX-0036 from the French National Research Agency (ANR).

References

Acha, J. & Perea, M. (2008). The effects of length and transposed-letter similarity in lexical decision: Evidence with beginning, intermediate, and adult readers. *British Journal of Psychology, 99,* 245-264.

Andrews, S. (1996). Lexical retrieval and selection processes: Effects of transposed letter confusability. *Journal of Memory and Language, 35,* 775-800.

Angele, B. & Rayner, K. (2013). Processing in the parafovea: Are articles skipped automatically? *Journal of Experimental Psychology: Learning, Memory & Cognition, 39,* 649-662.

Angele, B., Tran, R. & Rayner, K. (2013). Parafoveal–foveal overlap can facilitate ongoing word identification during reading: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 526-538.

Ans, B., Carbonnel, S. & Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. *Psychological Review, 105,* 678-723.

Baayen, R. (2010). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon, 5,* 436-461.

Baldauf, D. & Deubel, H. (2008). Properties of attentional selection during the preparation of sequential saccades. *Experimental Brain Research, 184,* 411-425.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015), Fitting Linear Mixed-Effects Models using lme4*.* *Journal of Statistical Software*, 67, 1-48.

Becker, W. & Jürgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research, 19,* 967-983.

Bicknell, K. & Levy, R. (2010). *Word predictability and frequency effects in a rational model of reading.* In Miyake, N., Peebles, D. & Cooper, R. (Eds.), Proceedings of the 34th annual conference of the Cognitive Science Society. Cognitive Science Society, Austin, TX (2012), pp. 126-131.

Bowers, J., Davis, C., & Hanley, D. (2005). Automatic semantic activation of embedded words: Is there a 'hat' in 'that'? *Journal of Memory and Language, 52*, 131-143.

Boden C. & Giaschi, D. (2007). M-stream deficits and reading-related visual processes in developmental dyslexia. *Psychological Bulletin, 133*, 346-366.

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A., Bölte, J., et al. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology, 58,* 412-424.

Campbell F. & Wurtz R. (1978). Saccadic omission: why we do not see a grey-out during a saccadic eye movement. *Vision Research, 10,* 1297-1303.

Catell, J. (1886). The time it takes to see and name objects. *Mind, 11,* 53-65.

Chambers, S. (1979). Letter and order information in lexical access. *Journal of Verbal Learning and Behavior, 18,* 225-241.

Collis, N., Kohnen, S. & Kinoshita, S. (2013). The role of visual spatial attention in adult developmental dyslexia. *Quarterly Journal of Experimental Psychology, 66,* 245-260.

Dare, N. & Shillcock, R. (2013). Serial and parallel processing in reading: Investigating the effects of parafoveal orthographic information on nonisolated word recognition. *The Quarterly Journal of Experimental Psychology, 66*, 417-428.

Davis, C. (1999). The self-organizing lexical acquisition and recognition (SOLAR) model of visual word recognition. Unpublished doctoral dissertation. University of New South Wales, Australia.

Davis, C. (2010). SOLAR versus SERIOL revisited. *European Journal of Cognitive Psychology, 22,* 695–724.

Davis, C., & Bowers, J. (2004). What do letter migration errors reveal about letter position coding in visual word recognition? *Journal of Experimental Psychology: Human Perception and Performance, 30*, 923–941.

Davis, C. (2010). The spatial coding model of visual word identification. *Psychological Review, 117,* 713-758.

Engbert, R., Nuthmann, A., Richter, M. & Kliegl, R. (2005) SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112,* 777-813.

Engbert, R. & Kliegl, R. (2011). *Parallel graded attention models of reading.* In Liversedge, S., Gilchrist, I. & Everling, S. (Eds.), The Oxford Handbook of Eye Movements (pp. 787-800). Oxford, England: Oxford University Press.

Epelboim, J., Booth, J. & Steinman, R. (1995). Much Ado About Nothing : the Place of Space in Text. *Vision Research*, *36*, 461-470.

Erdmann, B. & Dodge, R. (1898). *Psychologische Untersuchung über das Lesen auf experimenteller Grundlage*, Niemeyer: Halle.

Franconeri, S., Scimeca, J., Roth, J., Helseth, S., & Kahn, L. (2012). Flexible visual processing of spatial relationships. *Cognition, 112*, 210-227.

Geiger, G., Cattaneo, C., Galli, R., Pozzoli, U., Lorusso, M., et al. (2008). Wide and diffuse perceptual modes characterize dyslexics in vision and audition. *Perception, 37,* 1745-1764.

Gomez, P., Ratcliff, R. & Perea, M. (2008). The Overlap Model: A model of letter position coding. *Psychological Review, 115,* 577-600.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language, 29,* 228-244.

Grainger, J. (2003). *Moving eyes and reading words: How can a computational model combine the two?* In J. Hyona, R. Radach, & H. Deubel (Eds.) The Mind's Eye: Cognitive and Applied Aspects of Eye Movements. Oxford: Elsevier.

Grainger, J. (2008). Cracking the orthographic code: An introduction. *Language and Cognitive Processes, 23*, 1-35.

Grainger, J. (2018). Orthographic processing: A "mid-level" vision of reading. *Quarterly Journal of Experimental Psychology*, in press.

Grainger, J., Dufau, S. & Ziegler, J. (2016). A vision of reading. *Trends in Cognitive Sciences*, *20,* 171-179.

Grainger, J., Granier, J., Farioli, F., Van Assche, E. & van Heuven, W. (2006). Letter position information and printed word perception: The relative-position priming constraint. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 865-884.

Grainger, J., Mathôt, S. & Vitu, F. (2014). Test of a model of multi-word reading: Effects of parafoveal flanking letters on foveal word recognition. *Acta Psychologica,* 146, 35-40.

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

Grainger, J., Tydgat, I. & Isselé, J. (2010). Crowding affects letters and symbols differently. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 673-688.

Grainger, J., & van Heuven, W. (2003). Modeling letter position coding in printed word perception. In P. Bonin (Ed.), *The mental lexicon* (pp. 1-23). New York: Nova Science.

Hannagan, T. & Grainger, J. (2012). Protein analysis meets visual word recognition: A case for string kernels in the brain. *Cognitive Science, 36,* 575-606.

Hyönä, J. (1995). Do irregular letter combinations attract readers' attention? Evidence from fixation locations in words. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 68-81.

Hohenstein, S., Laubrock, J. & Kliegl, R. (2010). Semantic preview benefit in eye movements during reading: A parafoveal fastpriming study. *Journal of Experimental Psychology: Learning, Memory & Cognition, 36,* 1150-1170.

Hohenstein, S. & Kliegl, R. (2014). Semantic preview benefit during reading. *Journal of Experimental Psychology: Learning, Memory & Cognition, 40*, 166-190.

Inhoff, A., Radach, R., Eiter, B. & Juhasz, B. (2003). Distinct subsystems for the parafoveal processing of spatial and linguistic information during eye fixations in reading. *Quarterly Journal of Experimental Psychology, 56,* 803-827.

Inhoff A., Radach R., Starr M. & Greenberg S. (2000). *Allocation of visuospatial attention and saccade programming during reading*. In Kennedy A., Radach R., Heller D. & Pynte J. (Eds.). Reading as a perceptual process. Oxford, UK: Elsevier.

Inhoff, A., Weger, U., & Radach, R. (2005). *Sources of information for the programming of short- and long-range regressions during reading.* In Underwood, G. (Ed.), Cognitive Processes in Eye Guidance. Oxford: Oxford University Press.

Kapoula, Z. & Robinson, D. (1986). Saccadic undershoot is not inevitable: Saccades can be accurate. *Vision Research, 26,* 735-743.

Kennedy, A. (2008). Parafoveal-on-foveal effects are not an artifact of mis-located saccades. *Journal of Eye Movement Research, 2,* 1-10.

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

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology, 16,* 262-284.

Laubrock, J. & Kliegl, R. (2015). The eye-voice span during reading aloud. *Frontiers in Psychology, 6*. doi: https://doi.org/10.3389/fpsyg.2015.01432

Laubrock, J. & Kliegl, R. (in preparation). EyetrackR: an R package for the analysis of eye movement data. http://read.psych.uni-potsdam.de/index.php?option=com_content&view=article&id=43:eyetrackr&catid=13:r-playground&Itemid=15

MacDonald, M., Pearlmutter, N. & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101,* 676-703.

Marzouki, Y. & Grainger, J. (2014). Effects of stimulus duration and inter-letter spacing on letter-in-string identification. *Acta Psychologica, 148,* 49-55.

Marzouki, Y., Meeter, M. & Grainger, J. (2013). Location invariance in masked repetition priming of letters and words. *Acta Psychologica, 142,* 23-29.

McClelland, J., & Mozer, M. (1986). Perceptual interactions in two-word displays: Familiarity and similarity effects. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 18-35.

McClelland, J. & Rumelhart, D. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*, *88,* 375-407.

McConkie, G., Kerr, P., Reddix, M. & Zola, D. (1988). Eye movement control during reading: I. The location of initial eye fixations on words. *Vision Research*, 28, 1107-1118.

Meeter, M. & Van der Stigchel, S. (2013). Visual priming through a boost of the target signal: Evidence from saccadic landing positions. *Attention, Perception & Psychophysics, 75,* 1336-1341.

Mirault, J., Snell, J. & Grainger, J. (n.d.). Reading without spaces revisited: The roles of word identification and sentence-level constraints. *Scientific Studies of Reading* (manuscript in review).

Norris, D., & Kinoshita, S. (2012). Orthographic processing is universal; it's what you do with it that's different. *Behavioral and Brain Sciences, 35,* 296-297.

O'Connor, R. & Forster, K. (1981). Criterion bias and search sequence bias in word recognition. *Memory and Cognition, 9,* 78-92.

Perea, M. & Acha, J. (2009). Space information is important for Reading. *Vision Research*, *49*, 1994-2000.

Perea, M. & Carreiras, M. (2006). Do transposed-letter similarity effects occur at a syllable level? *Experimental Psychology, 53,* 308-315.

Perea, M. & Gomez, P. (2012). Subtle increases in interletter spacing facilitate the encoding of words during normal reading. *PLoS ONE, 7,* e47568. doi:10.1371/journal.pone.0047568.

Perea, M. & Lupker, S. (2004). Can CANISO activate CASINO? Transposed-letter similarity effects with non-adjacent letter positions. *Journal of Memory and Language, 51,* 231-246.

Perea, M. & Pollatsek, A. (1998). The effects of neighborhood frequency in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 767-777.

Perea, M., Rosa, E. & Gomez, C. (2005). The frequency effect for pseudowords in the lexical decision task. *Perception and Psychophysics, 67,* 301-314.

Peressotti, F. & Grainger, J. (1995). Letter position coding in random consonant arrays. *Perception & Psychophysics,* 57, 875-890.

Peressotti, F., & Grainger, J. (1999). The role of letter identity and letter position in orthographic priming. *Perception and Psychophysics, 61,* 691-706.

Radach, R., Reilly, R., & Inhoff, A. (2007). *Models of oculomotor control in reading: towards a theoretical foundation of current debates.* In R. van Gompel, M. Fischer, W. Murray & R. Hill (Eds.), Eye movements: A window on mind and brain. Oxford: Elsevier.

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology, 7,* 65-81.

Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology, 41,* 211-236.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372-422.

Rayner, K. (2009). The thirty fifth Sir Frederick Bartlett lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62,* 1457-1506.

Rayner, K., Fischer, M. & Pollatsek, A. (1998). Unspaced text interferes with both word identification and eye movement control. *Vision Research*, *38,* 1129-1144.

Reichle, E. (2011). *Serial-attention models of reading*. In Liversedge, S., Gilchrist, I. & Everling, S. (Eds.), The Oxford Handbook of Eye Movements (pp. 787-800). Oxford, England: Oxford University Press.

Reichle, E., Liversedge, S., Pollatsek, A. & Rayner, K. (2009b). Encoding multiple words simultaneously in reading is implausible. *Trends in Cognitive Sciences, 13,* 115-119.

Reichle E., Pollatsek, A., Fischer, D. & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105,* 125-157.

Reichle, E., Rayner, K. & Pollatsek, A. (1999). Eye movement control in reading: Accounting for initial fixation locations and refixations within the E-Z Reader model. *Vision Research, 39,* 4403-4411.

Reichle, E., Rayner, K. & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26,* 445-526.

Reichle, E., Pollatsek, A. & Rayner, K. (2006). E-Z reader A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research, 7,* 4-22.

Reichle, E., Warren, T. & McConnell, K. (2009a). Using E-Z Reader to model the effects of higher-level language processing on eye movements during reading. *Psychonomic Bulletin and Review, 16,* 1-21.

Reilly, R. & Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research, 7,* 34-55.

Roberts, S. & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107,* 358-367.

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

Schotter, E., Angele, B. & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception & Psychophysics*, *74,* 5-35.

Sereno, S. & Rayner, K. (2000). Spelling-sound regularity effects on eye fixations in reading. *Perception & Psychophysics, 62,* 402-409.

Snell, J., Bertrand, D. & Grainger, J. (2018a). Parafoveal letter-position coding in reading. *Memory & Cognition.* doi:10.3758/s13421-017-0786-0

Snell, J., Declerck, M. & Grainger, J. (2018). Semantic processing in the parafovea revisited: Effects of translation equivalents in bilingual readers. *Language, Cognition & Neuroscience, 33,* 563-574.

Snell, J. & Grainger, J. (2017). The sentence superiority effect revisited. *Cognition, 168,* 217-221.

Snell, J., Meeter, M. & Grainger, J. (2017b). Evidence for simultaneous syntactic processing of multiple words in reading. *PLoS ONE, 12,* e0173720. doi:10.1371/journal.pone.0173720

Snell, J., Vitu, F. & Grainger, J. (2017a). Integration of parafoveal orthographic information during foveal word reading: Beyond the sub-lexical level? *Quarterly Journal of Experimental Psychology*, *70*, 1984-1996.

Stede, M. & Neumann, A. (2014). *Potsdam Commentary Corpus 2.0: Annotation for Discourse Research.* In: Proceedings of the Language Resources and Evaluation Conference (LREC*)*, Reykjavik.

Van Assche, E., & Grainger, J. (2006). A study of relative-position priming with superset primes. *Journal of Experimental Psychology: Learning, Memory and Cognition, 32,* 399-415.

White, S. & Liversedge, S. (2004). Orthographic familiarity influences initial eye fixation positions in reading. *European Journal of Cognitive Psychology, 16,* 52-78.

Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin and Review, 8,* 221-243.

Winskel, H., Radach, R. & Luksaneeyanawinc, S. (2009). Eye movements when reading spaced and unspaced Thai and English: a comparison of Thai-English bilinguals and English monolinguals. *Journal of Memory and Language, 61,* 339-351

Synthesis

One of the most hotly debated issues in reading research concerns the question whether readers process words serially or in parallel. Until now, serial processing has been the most dominant view, largely driven by the popularity of the E-Z Reader model (Reichle et al., 2006). However, as argued in Chapter 5, it would seem inappropriate to use E-Z Reader's performance (i.e., its goodness of fit to experimental data) as argument in favor of serial processing, given that parallel processing models (SWIFT of Engbert et al. (2005) and Glenmore of Reilly and Radach (2006)) have provided equally decent fits.

A critical flaw of all these models is that they do not generate any predictions. The key problem, in my view, is that these researchers have modelled data rather than theory. Concretely, when these models were developed, it was already known that the length and frequency of words has an influence on the time it takes to recognize those words. Similarly, it was already known that readers execute word skips, refixations and regressions with a certain probability. One can easily hard-code such behavior in a model, and consequently be guaranteed to observe a good fit to the data; but the fit will not have provided a test of any theory.

Thankfully, as summarized in Chapter 3's articles, the field has also made efforts to evidence serial or parallel processing directly with the help of various experimental designs. Unfortunately however, this effort has largely betted on a single horse: namely, to establish (an absence of) parafoveal-on-foveal effects in sentence reading. As argued in Chapter 3, while it is true that higher-order influences of an upcoming word (word *n*+1) on the fixated word (word *n*) would evidence that the two were processed in parallel, the absence of such influences does not evidence serial processing. After all, it may have been the case that the two words were truly processed in parallel, without cross-leakage of information. Frankly, this can be seen as a flaw of the parallel processing approach, for such a rationale makes it considerably more difficult to falsify parallel processing.

Yet, by means of the theoretical framework outlined in Chapter 3 (p. 121), we were able to generate other ways through which parallel processing could potentially be falsified. One crucial prediction is that while higher-order cross-word influences would not show up in direct reading measures (e.g., fixation durations), such influences should show up at the decision-level when one induces a task. In line with this prediction, we found that syntactic and semantic decisions about foveal target words were influenced by the syntactic and semantic characteristics of adjacent words, even when the stimuli were shown simultaneously for only 170 ms.

Importantly, we have also explored means to directly track visuo-spatial attention during reading. In Chapter 2, it was shown that the pupil responds to the brightness of the locations of stimuli involved in parafoveal-on-foveal effects (Snell, Mathôt, Mirault & Grainger, 2018)—evidencing that these stimuli were covertly attended (e.g. Mathôt & van der Stigchel, 2015). This result undermines serial processing accounts of orthographic parafoveal-on-foveal effects, which state that attention is strictly directed to the foveal word, with foveal letter detectors nonetheless being influenced by parafoveal feature detectors (Angele, Tran & Rayner, 2013). Additional doubt is cast over this serial processing account by the fact that foveal word processing is not only influenced by parafoveal letters (with identical letters causing faster recognition than different letters), but also by the relative order of these letters (Chapter 2: Snell, Bertrand & Grainger, 2018), suggesting that the locus of orthographic parafoveal-on-foveal effects is beyond the level of feature processing. The latter conception is further strengthened by our EEG study in which we established that the locus of orthographic integration effects is in the N250 window (Chapter 2: Snell, Meade, Meeter, Holcomb & Grainger, 2018). The absence of effects prior to 200 ms argues against the involvement of low-level visual processes, such as processing at the level of feature detectors.

The conclusion that attention can indeed be allocated to multiple words at once does not invalidate some of the challenges that were raised against parallel processing systems by Reichle et al. (2009) per se. In particular, they noted that a parallel processing system might recognize words out of their canonical order, and they questioned how the comprehension problems that this would cause are dealt with. The theoretical framework outlined in this thesis provides an answer to this question. The central idea is that activated lexical representations are associated with plausible locations in a sentence-level representation, based on low-level visual cues (word length) and top-down expectations.

Indeed, in Chapter 4 it was shown that readers may in fact recognize words out of order (Snell & Grainger, 2018), which, according to Reichle et al. (2009), should not be possible if serial processing were true. It seems reasonable to claim that readers are not only parallel processors, but also *predictors* that work with good-enough (sentence-level) representations. Those representations are generated quite rapidly upon a first glance at the sentence stimulus. As shown with our Rapid Parallel Visual Presentation (RPVP) paradigm, the recognition of individual words is influenced by surrounding context even when the whole sequence is displayed for only 200 ms (Chapter 4: Snell & Grainger, 2017). Such findings cannot be reconciled within a serial processing framework.

Another challenge raised by Reichle et al. (2009), ("*how are two words processed at the same time given existing models of word recognition?*"; p. 118 in their opinion paper), was invalid from the get-go. All word identification models assume parallel activation of multiple lexical representations (e.g. McClelland & Rumelhart, 1981; Grainger & van Heuven, 2003; Davis, 2010). Indeed, in Chapter 1 it was made clear that even single words can lead to the activation of multiple lexical representations, as we established that the speed of word recognition is influenced by the number, length and frequency of embedded words (e.g. '*red*' in '*predict*'; Snell, Grainger & Declerck, 2018).

In light of the above, it is my sense that serial processing should no longer be the status quo. The results reported in this thesis fit nicely within a parallel processing framework, and, on the contrary, are quite difficult to explain from a serial processing perspective.

Finally, I believe that this thesis points to exciting new avenues for reading research. Firstly, whereas word position coding was considered a given under the assumption of serial processing, it is now clear that the process of associating words with locations is not so straightforward. As noted in Chapter 4 (Snell & Grainger, 2018), it will be interesting to determine how bottom-up cues and top-down expectations interact in the process of word position coding.

Secondly, as noted in the Introduction, *ideally* readers would be serial processors. After all, our research has shown that the recognition of words is generally fastest when those words are viewed in isolation (Chapters 2 and 3). In this light, a question of high prominence is whether certain reading problems (i.e., dyslexia) may be caused by a distribution of attention that is too diffuse, which would lead to weaker foveal processing and stronger interference from parafoveal information. This scenario can be investigated by testing the flanker paradigm, as reported in many forms in this thesis, on dyslexic readers, with the hypothesis that flanker effects should be stronger in this group.

In conclusion, the research reported in this thesis compels me to claim that the reading system is organized as follows: visuo-spatial attention is distributed across multiple words, which causes these words to jointly activate sub-lexical nodes (letters, bigrams). In parallel, a first glance at the sentence leads the reader to generate a spatiotopic representation based on low-level visual cues, such that the reader knows from the get-go that the sentence starts with, say, a short word, followed by a moderately long word, et cetera. Widespread sub-lexical orthographic processing leads to the activation of multiple lexical representations. Those activated lexical representations are associated with a plausible location (e.g., a short word may be associated with the first position, while a moderately long word may be associated with the second position). Such associations are further driven by top-down expectations (e.g., a verb is less likely to be associated with the first position). This process of associating lexical representations with positions generates a tentative sentence-level representation, that in turns constrains the recognition of individual words.

References

Angele, B., Tran, R. & Rayner, K. (2013). Parafoveal–foveal overlap can facilitate ongoing word identification during reading: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 526-538.

Cattell, J.M. (1886). The time it takes to see and name objects. *Mind*, 11, 53-65.

Davis, C. (2010). SOLAR versus SERIOL revisited. *European Journal of Cognitive Psychology, 22,* 695-724.

Engbert, R., Nuthmann, A., Richter, E. & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112,* 777-813.

Erdmann, B. & Dodge, R. (1898). *Psychologische Untersuchungen über das Lesen: Auf experimenteller Grundlage*. Halle: Niemeyer.

Grainger, J. & van Heuven, W. (2003). Modeling letter position coding in printed word perception. In P. Bonin (Ed.) *The mental lexicon: Some words to talk about words*. New York: Nova Science Publishers.

Huey, E. (1908). *The psychology and pedagogy of reading* (5th ed.). Cambridge, MA: MIT Press.

Mathôt, S. & van der Stigchel, S. (2015). New light on the mind's eye: The pupillary light response as active vision. *Current Directions in Psychological Science, 24,* 374-378.

McClelland, J. & Rumelhart, D. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*, *88,* 375-407.

Rayner, K., Pollatsek, A., Ashby, J. & Clifton Jr., C. (2012). *Psychology of reading.* New York, NY: Psychology press.

Reichle, E., Liversedge, S., Pollatsek, A. & Rayner, K. (2009). Encoding multiple words simultaneously in reading is implausible. *Trends in Cognitive Sciences, 13,* 115-119.

Reichle, E., Pollatsek, A., Fisher, D. & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*, 125-157.

Reichle, E., Pollatsek, A., & Rayner, K. (2006). E-Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research, 7,* 4-22.

Reilly, R., Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. Cognitive Systems Research.

Schotter, E., Angele, B. & Rayner, K. (2012). Parafoveal processing in reading. *Attention Perception & Psychophysics, 74,* 5-35.

Snell, J., Bertrand, D. & Grainger, J. (2018). Parafoveal letter-position coding in reading. *Memory & Cognition, 46,* 589-599.

Snell, J. & Grainger, J. (2017). The sentence superiority effect revisited. *Cognition, 168,* 217-221.

Snell, J. & Grainger, J. (2018). Word position coding in reading is noisy. *Psychonomic Bulletin & Review* (in review).

Snell, J., Grainger, J. & Declerck, M. (2018). A word on words in words: How do embedded words affect reading? *Journal of Cognition* (accepted).

Snell, J., Mathôt, S., Mirault, J. & Grainger, J. (2018). Parallel graded attention in reading: A pupillometric study. *Nature Scientific Reports, 8*, 3743. doi: 10.1038/s41598-018-22138-7

Snell, J., Meade, G., Meeter, M., Holcomb, P. & Grainger, J. (2018). An electrophysiological investigation of orthographic integration in reading. *Journal of Cognitive Neuroscience* (in review).