



ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Special issue: Research report

A paradox of apparent brainless behavior: The time-course of compound word recognition

Daniel Schmidtke* and Victor Kuperman

McMaster University, Canada

ARTICLE INFO

Article history:

Received 21 December 2017

Reviewed 23 March 2018

Revised 23 May 2018

Accepted 8 July 2018

Published online xxx

Keywords:

Complex word recognition

Morphological processing

Time-course of word recognition

Survival analysis

Eye-movements during reading

Eye-tracking

ABSTRACT

A review of the behavioral and neurophysiological estimates of the time-course of compound word recognition brings to light a paradox whereby temporal activity associated with lexical variables in behavioral studies predates temporal activity of seemingly comparable lexical processing in neuroimaging studies. However, under the assumption that brain activity is a cause of behavior, the earliest reliable behavioral effect of a lexical variable must represent an upper temporal bound for the origin of that effect in the neural record. The present research provides these behavioral bounds for lexical variables involved in compound word processing. We report data from five naturalistic reading studies in which participants read sentences containing English compound words, and apply a distributional technique of survival analysis to resulting eye-movement fixation durations (Reingold & Sheridan, 2014). The results of the survival analysis of the eye-movement record place a majority of the earliest discernible onsets of orthographic, morphological, and semantic effects at less than 200 ms (with a range of 138–269 ms). Our results place constraints on the absolute time-course of effects reported in the neuro-linguistic literature, and support theories of complex word recognition which posit early simultaneous access of form and meaning.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

In the last two decades, a number of studies have compared the time-course of visual word processing across two classes of experimental paradigms: behavioral (represented, for instance, by the eye-tracking record of eye-movements) and neuroimaging (represented by electro- or magneto-encephalographic records of brain activity; EEG and MEG). Both paradigms afford a high temporal resolution, yet most comparisons of their respective results – from either within-

subject studies employing co-registration of eye-movements and EEG signals (see Dimigen, Sommer, Hohlfeld, Jacobs, & Kliegl, 2011; Kretzschmar, Schlesewsky, & Staub, 2015), or between-subject studies that employ these paradigms separately (Dambacher & Kliegl, 2007; Dambacher, Kliegl, Hofmann, & Jacobs, 2006; Rayner & Clifton, 2009; Schmidtke, Matsuki, & Kuperman, 2017; Sereno & Rayner, 2003; Sereno, Rayner, & Posner, 1998) – lead to the same paradoxical observation: lexical effects on behavior either predate the emergence of these same effects in brain activity, or take place

* Corresponding author.

E-mail address: schmiddf@mcmaster.ca (D. Schmidtke).<https://doi.org/10.1016/j.cortex.2018.07.003>

0010-9452/© 2018 Elsevier Ltd. All rights reserved.

without leaving a corresponding trace of detectable neural activity (Dimigen et al., 2011; Kretzschmar et al., 2015). These findings present a clear challenge to what we consider a foundational assumption of cognitive neuroscience, which is that a human behavioral response to a sensory stimulus must arise as a result of some initial neural activity, whether said activity is detectable by current experimental methods or not (see discussion and references in Krakauer, Ghazanfar, Gomez-Marin, McIver & Poeppel, 2017). As a result, an apparent failure to reconcile behavioral and neural signatures of the same cognitive processes casts doubt on the validity of temporal estimates obtained by current measurement and analytical techniques in neuroscientific research.

In the present paper, we argue that an important step towards achieving credible reports of brain activity is to establish onsets of word recognition processes in the behavioral record. It is important to acknowledge the strong effort that neuroscientific literature puts into ensuring accurate estimation of how early brain activity occurs from EEG/MEG signals (e.g., Hauk, Davis, Ford, Pulvermüller, & Marslen-Wilson, 2006; Pulvermüller, Shtyrov, & Hauk, 2009; Shtyrov & MacGregor, 2016). We add to this effort with the provision of behaviorally-credible temporal upper bounds for when that activity must occur.

Central to our argument is the view that the behavioral onset of a lexical variable must serve as an upper temporal bound for its causal neural activation. A recent study by Schmidtke et al. (2017) employed a non-parametric distributional survival analysis to identify onsets of formal, semantic and morphological effects in the time-course of visual recognition of derived words (e.g., *smoker*, *government* and *cowardly*), on the basis of lexical decision and eye-tracking data. Schmidtke et al.'s (2017) review of the brain imaging research on the topic revealed a familiar paradox, whereby behavioral onsets of effects predate neural ones by as much as 150–200 ms. The behavioral estimates of these lexical onsets were then argued to serve as upper temporal bounds for future neuroscientific exploration of derived word processing. The present paper is a companion piece to Schmidtke et al., (2017): we refer the reader to that previous paper for detailed argumentation and a description of the method. The novelty of the present study is that it addresses a different type of morphological complexity, i.e. compounding (e.g., *catfish*, *whalebone* and *paintbrush*), reviews a different body of the neuroimaging and behavioral literature, and reports a new set of five eye-tracking studies of naturalistic compound word reading. Thus, a central goal of this paper is to pit findings in the literature against new behavioral data and identify reliable temporal benchmarks for future brain imaging research on compound word recognition. The literature review is found in the Introduction below, while the upper temporal bounds are set out in the General Discussion.

Our second goal regards a hotly contested issue of the time-course of the processing of morphologically complex words. Current models of visual recognition of morphologically complex words can be grouped into two broad categories based on their assumptions about the time-course of morphological processing. First, *form-then-meaning* models of complex word

recognition describe a process whereby access to the meaning of a complex word form is granted only after its orthographic form has been decomposed into constituent morphemes, and the subsequently isolated morphemic representations have been recombined (Rastle & Davis, 2008; Solomyak & Marantz, 2010; Taft & Forster, 1975). In contrast, *form-and-meaning* accounts (Feldman, Milin, Cho, Moscoso del Prado Martín, 2015; Feldman, O'Connor, Moscoso del Prado Martín, 2009; Marelli & Luzzati, 2012) assume that after visual uptake of a complex word's orthography, access to semantic properties of the whole word may proceed without intervention from the processes of morphological decomposition and recombination. The distributional survival analysis approach taken in the present paper enables us to evaluate the relative order of formal and semantic effects in the time-course of compound processing. Below we discuss the implications of this relative order for the conflicting theoretical accounts.

1.1. The reported neural record of compound word processing

The EEG and MEG line of inquiry into the time-course of written compound word recognition involves registering brain activity during tasks that require silent inspection of compound words (see an excellent review of literature in Fiorentino, Naito-Billen, Bost, & Fund-Reznicek, 2014). These techniques afford a high level of temporal resolution which renders them especially well-suited to the task of investigating time-course issues such as the study of the temporal flow of information during lexical processing. Our (possibly incomplete) inspection of the literature has identified eight relevant papers which use EEG or MEG for the task of uncovering the dynamics of written compound word processing.

First, Krott, Baayen, and Hagoort (2006) employed a passive reading paradigm and recorded EEG responses to known and novel compounds in Dutch. In this experiment, compound words were manipulated in three ways. In a first condition, linking elements (analogous to *-s-* in the English word *sportsman*) were selected to be either grammatical or ungrammatical in existing compounds; in a second condition, linking elements in novel compounds were chosen to support or contradict the preferred choice of the linking element in similar compounds; and finally, some plural suffixes were added to both existing and novel compounds that would produce either grammatical or ungrammatical forms. Of principal interest for our purposes, Krott et al. observed a significant negative deflection in the EEG waveform for novel compounds at around 350 ms after word presentation, which they argue points to lexicality and constituent frequency effects. Moreover, a left anterior negativity (LAN) was observed in a 400–700 ms window for the processing of existing compounds related to the grammaticality manipulation of the linking element. Krott et al. argue for an association of the LAN component with the morphosyntactic characteristics of a complex word, and the N400 component with a response to the lexicality of the compound.

The remainder of neurophysiological studies under discussion utilized a lexical decision task without overtly

manipulating the grammaticality of morphemes. A study of Italian visual compound word recognition by [El Yagoubi et al. \(2008\)](#) presented participants with existing and novel compounds that had either the first or the second morpheme as their grammatical and semantic head, as well as pseudo-compounds which were constructed by reversing the order of morphemes. Compared with head-initial compounds, they found that head-final compounds exerted a stronger positivity in posterior brain sites at a time window of around 300 ms, with a second positivity observed in the 500–800 ms time window. Their inspection of the 0–275 ms window did not reveal any effect of compound structure or headedness. [El Yagoubi et al. \(2008\)](#) relate this finding to the P300 component and present this, along with other findings of theirs, as evidence for a decompositional process. Furthermore, a similar time-course was reported by [Arcara, Marelli, Buodo, and Mondini \(2014\)](#) in a lexical-decision-plus-EEG study which considered head-initial, head-final, and verb-noun Italian compounds. They observed an effect of compound type (i.e., headedness) in a first time window of 280–400 ms, and these effects carried over to a later 400–600 ms window. In sum, these studies support the hypothesis that the grammaticality of the compound is appraised at around 280 ms and onwards after visual uptake of the compound word form, which is taken as an indicator of early morphological decomposition.

In addition, MEG studies appear to uncover further evidence in favour of the form-then-meaning nature of complex word recognition. In a MEG study in which English compound words and simplex control matches were presented for a lexical decision task, [Fiorentino and Poeppel \(2007\)](#) found that compound words elicited faster response times and an earlier latency of the M350 component ($M = 333$ ms) than simplex words ($M = 360$ ms). Because the two word groups were matched on whole-word properties, Fiorentino and Poeppel reached the conclusion that the facilitation observed for compound words was brought about by activation of the compound's decomposed morphemes. An EEG study conducted by [Fiorentino et al. \(2014\)](#) further supports this interpretation. Their lexical decision experiment expanded on Fiorentino and Poeppel's condition list by including both existing and novel English compounds, as well as existing long simplex words (matched with compounds on a number of whole-word properties) and non-words. ERP responses to compounds, both lexicalized and novel, differed from non-compound stimuli in the 275–400 ms time window, with further observable differences in the topographic distribution of responses emerging between lexicalized and novel compounds in that same time window. Moreover, [Fiorentino et al. \(2014\)](#) observed interactive effects of structure and lexicality in a later 400–700 ms time window. As in [El Yagoubi et al. \(2008\)](#), no morphologically relevant effects were observed in the 0–275 ms window.

A similar timeline of processing effects emerges in [Pylkkänen, Feintuch, Hopkins, and Marantz's \(2004\)](#) study of family size and family frequency (i.e., the number and the summed frequency of distinct words sharing a morpheme) of English compounds and affixed forms. They reported that

greater family size was associated with an M350 latency reduction ($M = 336$ ms for high family size, and $M = 354$ ms for low family size). Additionally, the authors report that family frequency not only influenced the amplitude, but also the peak of the M350 component. That is, [Pylkkänen et al.](#) additionally conducted an analysis of both latency and amplitude at time-points where the source amplitude reached 25%, 50%, 75% and 100% of the maximum source strength at M170 and M250 ms. Thus, even when the net was cast wide, the temporal locus of morphological family effects on both latency and amplitude were confined to the 100% (and occasionally the 75%) point of maximum source strength at around 340 ms and onwards. In sum, even though [Pylkkänen et al.](#)'s analysis did not specifically include responses to compound words (MEG activity was collapsed across both compounds and affixed words), the evidence appears to suggest a time window of 340 ms for the appraisal of morphological effects.

Further MEG evidence for delayed semantic effects during English compound word processing was found by [Brooks and Cid de Garcia \(2015\)](#) in a word naming study which used the partial-repetition priming paradigm. In this study brain activity was recorded and analysis of this data was restricted to the visual processing of compound word primes, which were either opaque (e.g., *hogwash*) or transparent (e.g., *hogwash*). Their results show that, relative to processing of simplex primes, processing of transparent compounds was associated with greater neural activation in the anterior middle temporal gyrus in a time-window of 250–470 ms, with stronger effects occurring in the posterior superior temporal gyrus at 430–600 ms. These results therefore suggest that neural activity associated with semantic processing of compound words occurs from 250 ms onwards, and is strongest between 430 and 600 ms. Furthermore, the study's behavioral word naming effects are argued to point to a differentiation of morphology from semantic processing, and that morphological access occurs relatively early. When considering behavioral results alongside neurophysiological data, the authors conclude that “early activation of constituents via morphological decomposition happens irrespective of semantic transparency” and that “what differentiates transparent and opaque compound must happen, thus, during a later stage of morphemic composition” ([Brooks & Cid de Garcia, 2015](#), p. 6).

Interestingly, some other studies have observed earlier semantic effects in neural responses than the literature reviewed above. For example, [Vergara-Martínez, Duñabeitia, Laka, and Carreiras \(2009\)](#) orthogonally manipulated the frequencies of the left and right constituents of Basque compounds with the aim of evaluating the time-course of access to morphemes. In this study, compounds were embedded in sentences and were presented for silent reading while the EEG signal was recorded. Their results showed an early larger negativity in the EEG waveform for compounds with high-frequency first constituents (100–300 ms time window), and an increase in N400 amplitude for compounds with low-frequency second constituents. [Vergara-Martínez et al.](#) interpret the early negativity

difference as an index of access to potential word candidates, and the N400 effect for the second constituent as the selection of that constituent and its integration with the whole word representation. A potential reason for the observation of the earlier effect of constituent frequency may be that Vergara-Martinez et al. specifically targeted the earlier 100–300 ms epoch in their analysis of the ERP waveform. Though some of the aforementioned studies explored effects at earlier time-windows, it is possible that early morphological effects were missed because only later time windows were examined. For example, in their analysis of constituent frequency effects, Krott et al. (2006) exclusively analysed two later and broader time windows in their study (400–700 ms and 900–1200 ms).

For completeness, we also review findings of auditory compound word recognition. Even though task demands on cognitive linguistic processes of word recognition are very different, neurophysiological results from studies investigating the auditory compound word processing also reveal an inconsistent time-course for the effects of morphological information processing. For example, in semantic and grammaticality judgement studies of acoustically-presented German compound words Koester and colleagues reported increased N400 effects for (1) transparent compounds compared to opaque ones (Koester, Gunter & Wagner, 2007), (2) number-incongruent heads, but not nonhead constituents (Koester, Gunter, Wagner & Friederici, 2004), and (3) less plausible head constituents and second constituents in tri-constituent compounds (Koester, Holle & Gunter, 2009). All of these studies therefore converge on a 400 ms time-point for effects of semantic, grammatical and lexical integration of compound word forms. More recently, MacGregor and Shtyrov (2013) detected effects of semantic transparency and compound frequency on the ERP waveform using the passive-listening oddball ERP paradigm. MacGregor and Shtyrov time-locked ERP responses to the acoustic offset of the first constituent which was consistently 400 ms across all compound word stimuli. Results show that Larger Mismatch Negativity (MMN) brain-responses were elicited by high frequency opaque compound words as early as 150 ms after the acoustic onset of the second constituent of the compound word, and so, at about 550 ms post-onset of the compound word.

To summarize the findings of neurophysiological research on visual and auditory compound word recognition, results appear to converge on a time-point of around 350 ms—with few exceptions – as the earliest temporal locus for any morphological effect. Because this time-point roughly corresponds to the N400 compound in the EEG signal and the M350 component in the MEG signal, these components are often advertised as the timeframe when the processing effort of compound recognition becomes detectible by current experimental methods, see arguments in Fiorentino et al. (2014) for N400 effects and in Pyllkänen, Feintuch, Hopkins, and Marantz (2004) for M350 effects. This temporal estimate is a focus of our first goal, i.e., to establish credible upper temporal bounds for expected brain activity. The neuroscientific literature reviewed above is also relevant for our second goal, i.e., to test the relative order of formal and semantic effects in the

time-course of compound processing. Since most of the neurophysiological effects reported as “early” tend to be driven by morphological properties of complex words, and not their semantic characteristics, the dominant interpretation of neurophysiological results is that they support accounts proclaiming across-the-board semantics-blind decomposition during morphological processing.

Not all neurophysiological studies of complex word recognition demonstrate late effects of morphology and semantics. We wish to point out that several recent studies on visual recognition of another type of morphological complexity (derivation) have reported earlier onsets of morphological effects in neural activity and, importantly, earlier onsets of semantic effects during complex word processing in dyslexic and typical readers in the 100–250 ms time window (see MEG studies by Cavalli et al., 2016, 2017; and an EEG study by Jared, Jouravlev, & Joanisse, 2017), see also Schmidtke et al. (2017) for a review of early effects of word frequency on simple word processing in EEG studies. Thus, it is possible for both morphological structure and semantics to exert an effect on brain activity at an earlier timeframe.

1.2. Evaluating the behavioral record of compound word processing

Studies of eye-movements during reading provide a temporal estimate of the impact of a lexical effect during word processing. Crucially, as was reviewed extensively by Schmidtke et al. (2017), the results of prior eye movement studies of complex (derived and compound) words read silently in context and presented in isolation are not aligned with the timeline of the aforementioned effects reported in the neurophysiological literature. For compound word reading specifically, studies report effects of morphological and semantic characteristics on the very earliest eye-movement measures. For example, semantic transparency of compound words (i.e., the extent to which the meaning of a compound is related to the meaning of its constituents) has been reported to affect first fixation durations as early as an average of 231 ms (Marelli & Luzzatti, 2012), and 256 ms (Marelli, Amenta, Morone, & Crepaldi, 2013; Experiment 1 on isolated words) in Italian, and on first-of-many fixation durations at an average of 246 ms in English (Schmidtke, Van Dyke, & Kuperman, 2018). In addition, Juhasz (2016) observed an effect of sensory experience of the whole compound word on first fixations with a by-item average duration of 259 ms. Moreover, effects of morphological family size, as index of morphological processing, have been observed on gaze durations (the sum of all fixations before the eye proceeds past the target word) at an average of 272 ms in English (Juhasz & Berkowitz, 2011), and also at an average of 270 ms in first fixation durations in Dutch compound word reading (Kuperman, Schreuder, Bertram, & Baayen, 2009): for additional evidence from sentence reading and isolated word reading see Schmidtke et al. (2017). Without exception, these findings, which are based on mean latencies, imply that morphological and semantic processing of compound words is accomplished within a timeframe that precedes the window in which

morphological and semantic effects are expected to emerge in the neurophysiological record, i.e., 275 ms or later, with a peak amplitude around 350–400 ms.

In the current examination of the discrepancy between paradigms, we collected eye-movement data in five naturalistic compound word reading experiments. For each study, we conducted a divergence point analysis procedure (Reingold & Sheridan, 2014), which is a statistical technique designed to pinpoint the earliest point in time at which a predictor has an appreciable impact on chronometric behavioral responses. We considered a broad range of orthographic, morphological and semantic characteristics hypothesized as influential for compound word processing. We briefly reiterate our motivation for using this analytical technique. Firstly, by determining the earliest point in time at which a variable of interest has an effect on the decision to terminate an eye fixation, we are able to set an upper temporal bound for the activation in the neurophysiological record associated with that variable. Secondly, the resulting divergence points provided by the distributional analysis allow one to construct a timeline of the onset of individual variables during compound word processing. For each eye-movement experiment, we cross-check the timeline of effect onsets against the predictions of the form-then-meaning and form-and-meaning accounts of morphological decomposition. In what follows, we present our study and demonstrate that (i) the absolute timeline of morphological and semantic effects in eye-movement data is incompatible with that reported in the brain imaging literature, and (ii) there is little evidence to corroborate the notion that compound processing is initially semantics-blind.

2. Methods

2.1. Participants

2.1.1. Study 1

Forty-five participants (23 female; 22 male) were recruited in Hamilton, Ontario, Canada, within an age range of 18–31 ($M = 23.24$, $SD = 4.21$). Participants were paid \$15–20 CAD/hr and were recruited from the local community in a number of ways, including presentations at local colleges; advertisements placed on local community sections of online classified advertising services (Craigslis, Indeed, and Kijiji); posters/flyers placed on adult school and community college campuses, public transportation hubs, and from referrals from past and current study participants. All participants were non-college bound individuals (formal level of education did not exceed the equivalent of high school level). All were native speakers of English, had normal or corrected-to-normal vision and none had a diagnosed reading or learning disability.

2.1.2. Study 2

Thirty-five undergraduate students from McMaster University (26 female; nine male) within an age range of 18–37 ($M = 20$, $SD = 3.58$) completed the eye tracking study for course credit.

All participants were native speakers of English. All participants had normal or corrected-to-normal vision, and did not report a diagnosed reading or learning disability.

2.1.3. Study 3

Twenty-four undergraduate students from McMaster University (19 female; 5 male) within an age range of 17–28 ($M = 19.98$, $SD = 2.57$) completed the eye tracking study for course credit. All participants were native speakers of English. All participants had normal or corrected-to-normal vision, and did not report a diagnosed reading or learning disability.

2.1.4. Study 4

Thirty-eight undergraduate students from McMaster University (27 female; 11 male) within an age range of 18–30 ($M = 20.72$, $SD = 2.7$) completed the eye tracking study for course credit. All participants were native speakers of English. All participants had normal or corrected-to-normal vision, and did not report a diagnosed reading or learning disability.

2.1.5. Study 5

Fourty-one undergraduate students from McMaster University (34 female; 6 male; one undisclosed) within an age range of 18–26 ($M = 19.94$, $SD = 1.72$) completed the eye tracking study for course credit. All participants were native speakers of English. All participants had normal or corrected-to-normal vision, and did not report a diagnosed reading or learning disability.

2.2. Materials

A list of English concatenated noun–noun compound words was constructed for each study. The same stimuli list was used for Study 1 and Study 4. None of the compound words overlapped across the remaining lists. The lists for Studies 1, 3, and 4 consisted of 200 compounds. The lists for Studies 2 and 5 consisted of 216 compounds. The compound words were extracted from the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995) and the ELP lexical database (Balota et al., 2002). The compounds were each embedded within a single sentence frame (e.g., *Only a bookworm would carry an armful of the latest book releases.*). For all sentences, the sentence context preceding each compound word was neutral and each compound word did not occupy the first or last position of each sentence. All sentences were limited to 90 characters in length and did not exceed one line on the computer screen.

Not all eye movements to stimuli in each list could be used in the survival analyses. This is because not all the independent lexical variables were available for all of the words in each experimental list (see Predictor Variables section). This reduced the number of items in each list to 82 (Studies 1 and 4), 80 (Study 2), 209 (Study 3), 131 (Study 5).

2.3. Apparatus and procedure

All eye tracking studies followed the same procedure. The sentences were displayed on a 17-inch monitor with a resolution of 1,600 × 1,200 pixels, and a refresh rate of 60 Hz. Eye movements during sentence reading were recorded with an

Eyelink 1000 desk-mounted eye tracker (SR Research Ltd., Kanata, Ontario, Canada). The eye tracker is an infrared video-based tracking system combined with hyperacuity image processing. The data were collected at a 1,000 Hz sampling rate from the participants' dominant eye, or the right eye if the dominant eye was not known. Sentences were presented one at a time in Courier New, a monospace font, size 20, in black on a white background, and occupied exactly one line on the screen. Each character subtended .36° of visual angle. A three-point horizontal calibration of the eye tracker and a three-point horizontal accuracy test were performed before the beginning of each experiment, and after any breaks. A chin support and forehead rest was used to stabilize participants' gross head movements.

In order to familiarize participants with the experiment, each experiment began with a practice block consisting of 10 sentences. Participants then silently read sentences containing the target compound words in their sentence context. Participants were instructed to press a button when they had finished reading the sentence, and the sentences remained on the screen until the button was pressed. For each experiment, participants read the target sentences in randomized order. Each sentence trial was preceded by a drift correction, which used a fixation point positioned 20 pixels to the left of the beginning of the sentence, in order to ensure accurate recording of eye movements. Sentences were presented 100 pixels away from the left edge of the screen, and in the middle of the vertical dimension of the screen. Comprehension questions followed 20% of target sentences. Participants were presented with the sentences and were asked to respond whether they were true or false. Participants pressed the "a" key if the sentence was true and the "'" (single quote) key if it was false. 50% of the correct answers were true, and 50% were false. The proportion of correct responses across all three studies was high (>88%).

2.4. Response variables

The dependent variable for the eye movement analyses was the first fixation duration. This dependent measure is based on all trials for which there were no more than two fixations on the critical compound word in the first pass of reading.

2.5. Predictor variables

The goal of this research is to determine the absolute onset and relative order of effects elicited by variables related to a compound word's morphological structure (including frequency-based measures), orthography, and semantics. We introduce these groups of variables in turn.

2.5.1. Frequency characteristics

We extracted frequency statistics from the 50-million token SUBTLEX US corpus (Brysbaert & New, 2009). We considered the frequency of occurrence of the compound (e.g., *cheesecake*), the stand-alone frequency of occurrence of the modifier constituent (e.g., *cheese*), and the stand-alone frequency of the head constituent (e.g., *cake*).

2.5.2. Morphological variables

We took into consideration morphological family sizes of the modifier and head constituents (the number of word types that share the left or right constituent with the target compound) as two indices of morphological complexity. These measures were obtained from the CELEX database (Baayen et al., 1995). To take an example, the compound *campfire* has a left constituent family size of 7, meaning that there are 7 complex word forms (inflected derived and compound) sharing the constituent *camp* (e.g., *camping*, *camper*, *campground* etc.). The same compound has a right constituent family size of 5, indicating that there are 5 complex word forms that share its right constituent *fire* (e.g., *fired*, *wildfire*, *fireball* etc.). The left family size and right family size measures are labelled as Left fam. size and Right fam. size in figures and tables respectively. The motivation for the inclusion of family size follows from Pykkänen et al. (2004)'s MEG study which found that larger family size was associated with an M350 latency reduction.

2.5.3. Orthographic variables

We considered three orthographic variables. First, we measured the orthographic neighborhood of a word (abbreviated to Orth. density in tables and figures). This was computed by calculating the average Levenshtein distance, defined as the mean orthographic distance from 20 nearest orthographic neighbors of the compound word under consideration. This measure was estimated for each target word in our stimulus list, using the library *vwr* (Keuleers, 2013) in the R statistical computing software program (R Core Team, 2014). A list of all unique words from the SUBTLEX-US corpus was used as the lexicon with which to estimate orthographic neighborhood density for all compounds. We also calculated a measure of the transition between the left and the right constituents for all compound words. We refer to this variable as the bigram transition probability (TPB; see Solomyak & Marantz, 2010). TPB is defined as the frequency of the first two letters of the right constituent that the preceding two letters (i.e., the last two letters of the left constituent) appear in their position relative to the end of the word. Finally, we also considered word length (abbreviated to Length in tables and figures) in characters.

2.5.4. Semantic variables

We explored the word recognition time-course of two semantic variable types: "relational" and "atomic", i.e. ones that are only defined as a relation between morphemes and ones that are inherently linked to a single morpheme's meaning (Kuperman, 2013). Three atomic semantic properties were examined: the psychological valence (positivity) of the whole compound word (e.g., *seafood*), its left constituent (e.g., *sea*), and the right constituent (e.g., *food*), for motivation see Kuperman (2013) and Schmidtke et al. (2017). Valence estimates were obtained from a set of norms to 14,000 English lemmas (Warriner, Kuperman, & Brysbaert, 2013): In this corpus, words were rated on a scale of 1–9 (sad to happy) by about 20 raters each. In tables and figures, these three variables are referred to as Left valence (valence of the left constituent), Right valence (valence of the right

constituent) and Whole valence (valence of the compound word). For the purposes of the current paper, all word properties that cannot be solely gleaned from the word's form without some knowledge of word meaning are considered as semantic (e.g., cannot be learned by baboons solely on the basis of orthographic statistics). While early emotional effects in word processing have been attributed to the use of specific cortical pathways (Keuper, Zwitterlood, Rehbein, Eden, Laeger, Junghöfer, Zwanzger, Dobel & Christian, 2013), we still consider these effects to transcend pure form and contain a semantic gist.

We considered two relational semantic properties of compound words. That is, we included in our distributional analysis computational estimates of the semantic similarity of the meaning of the whole compound and the meaning of its left constituent (e.g., *shell* and *shellfish*), and separately, the semantic similarity of the meaning of the whole compound and the meaning of its right constituent (e.g., *fish* and *shellfish*). Traditionally, these variables have been referred to as *semantic transparency*. A semantically transparent compound, such as *earthworm*, has a meaning that is related to that of its constituents. However, a compound such as *ladybird* is opaque because there is no clear semantic similarity between the whole word meaning and the meanings of *lady* and *bird*. Thus, semantic transparency describes to what extent the meaning of a compound is related to the meaning of its constituents. The presence of a relational effect indicates access to both meanings that form the relation, that is, the relation between either of the two constituents and the whole compound word.

We estimated semantic similarity using Latent Semantic Analysis (Landauer & Dumais, 1997). The term-to-term LSA scores for left-whole and right-whole semantic similarity (labelled as LW similarity and RW similarity in plots and tables) for the target compounds were collected from <http://meshugga.ugent.be/snaut-english/>, with a default setting of 300 factors and a window of 6 words (Mandera, Keuleers, & Brysbaert, 2017). LSA scores were calculated over word occurrences in the 201 million-token SUBTLEX-UK corpus of film subtitles (van Heuven, Mandera, Keuleers, & Brysbaert, 2014). The estimate of semantic similarity between the words that those vectors represent is estimated by the cosine of the angle between the vectors, ranging from 0 to 1. We inverted the raw scores (by multiplying by negative one) that were collected from the web interface for comparability with other LSA scores. Values closer to 1 imply a greater semantic similarity between the pair of words under comparison.

Table 1 reports descriptive statistics of all lexical variables and first fixation duration times for Studies 1–5. Appendix A–E report pairwise Spearman correlations between predictor variables and first fixation duration times for Studies 1–5. We report correlations based on raw data, and not on mean fixation times on words. Thus, because of repeated measures of lexical variables, our correlations are weaker. The full list of materials, including sentence frames and lexical variables, for all stimuli used in the survival analysis are reported in the online Supplementary materials (S1).

2.6. Distributional analysis

We implemented the *Divergence Point Analysis* (DPA) procedure (Reingold, Reichle, Glaholt, & Sheridan, 2012), a distributional analysis technique which determines the earliest discernible impact of a variable on a chronometric dependent variable (here first fixation durations) by comparing survival curves across two experimental conditions. In the present paper, we implemented the confidence interval DPA procedure (Reingold & Sheridan, 2014), a modified version of the DPA procedure. This DPA procedure was used by Schmidtke et al. (2017) in order to determine the earliest impact of orthographic, morphological and semantic variables on lexical decision responses and eye-movement fixations to derived words. We encourage readers to refer to Schmidtke et al. (2017) and Reingold and Sheridan (2014) for extensive descriptions of the confidence interval technique and how it builds upon the original DPA procedure.

A central aspect of the DPA survival technique is the comparison of two survival functions, which each represent responses to a level of the same conditioning variable. First, consider that a standard density plot of a distribution of first fixation durations can be revisualized to show the percentage of eye fixations in the distribution that are greater than duration t (see Fig. 1 below). This visualization of a distribution can be expressed in terms of 'survival'. That is, the notion of survival here refers to the proportion of unterminated eye-fixations in a distribution of eye-movement fixation durations as a function of time t . For example, at $t = 1$ ms, 100% of trials still remain in the distribution of eye-movement fixations, because at this time no participant in the sample executed an eye-movement fixation that short. However, as one gradually moves through the response time distribution, progressively more and more eye fixations are terminated, until eventually 0% of trials remain unterminated at the point at which the longest fixation duration has been reached.

In the DPA procedure, two levels of an experimental variable of interest are established, and separate survival functions are generated for each level of the variable. Fig. 1 visualizes the survival curves for high and low left-whole semantic similarity. As can be viewed in the plot, for both conditions of left-whole semantic similarity, their survival curves both decrease monotonically with a large linear decrease in the middle of the curve, followed by a more gradual decline of survival percentage at the end of the distribution. However, first fixations durations on compounds with a high left-whole semantic similarity (e.g., *rainstorm*) start to decline in survival percentage more rapidly at the beginning of the curve (dashed line), as compared to the survival curve (solid line) for first fixations durations on compounds with low left-whole semantic similarity (e.g., *hogwash*). The precise purpose of the DPA procedure is to establish when the earliest significant difference in these two survival curves emerges.

In order to establish the earliest significant effect of a lexical variable on eye-movements to compounds we

Table 1 – Descriptive statistics for lexical variables used in each study.

Study	First fixation duration	Whole freq.	Left freq.	Right freq.	Whole valence	Left valence	Right valence	LW similarity	RW similarity	Left family size	Right family size	Length	Orthographic neighbourhood density (ORTH_DENSITY)	Bigram transition probability (TPB)
Study 1	Min	81	27	82	1.8	1.8	1.9	.36	.36	1	1	6	1.4	.0012
	1st Qu.	176	1432	1201	4.5	5.2	5.1	.57	.56	4	5.5	7	2.5	.0053
	Median	213	3782	4258	5.5	5.9	5.7	.68	.65	9.5	12	8	2.9	.012
	Mean	229	211	8518	15,479	5.2	5.7	5.7	.65	.67	13	22	2.9	.022
	3rd Qu.	268	168	10,961	11,917	6.3	6.7	6.5	.75	.76	17	22	3.4	.032
Study 2	Min	809	3142	108,306	99,890	7.7	7.7	.99	.95	45	155	11	4.5	.13
	1st Qu.	84	20	69	82	1.9	1.9	2.9	0	0	1	6	1.8	.0002
	Median	176	42	1284	1037	4.8	5.3	5	.58	4	3	7	2.5	.0035
	Mean	215	66	2882	4878	5.4	5.8	5.7	.64	.72	7	13	2.8	.012
	3rd Qu.	260	225	276	6407	14,161	5.3	5.8	.69	.65	12	19	2.8	.035
Study 3	Min	649	4958	44,168	102,467	8.3	8.3	.95	.95	42	155	11	4.4	.88
	1st Qu.	81	13	27	21	2.1	2.2	2.2	.37	1	1	6	1.8	.00066
	Median	180	30	1115	1207	4.8	5.2	5.3	.63	3	3	8	2.5	.005
	Mean	215	50	2990	4577	5.4	5.8	5.7	.71	7	10	8	2.8	.012
	3rd Qu.	263	154	7086	16,326	5.4	5.7	5.7	.71	12	21	8.5	2.9	.026
Study 4	Min	956	3881	99,890	204,428	7.8	7.6	.81	.97	16	23	9	3.3	.039
	1st Qu.	81	21	27	82	1.8	1.8	1.9	.36	1	1	6	1.4	.16
	Median	173	41	1432	1201	4.5	5.2	5.1	.56	4	5.5	7	2.5	.0012
	Mean	211	64	3782	4258	5.5	5.9	5.7	.68	9.5	12	8	2.9	.0053
	3rd Qu.	259	211	8518	15,479	5.2	5.7	5.7	.65	13	22	8.4	2.9	.012
Study 5	Min	826	3142	108,306	99,890	7.7	7.7	.99	.95	45	155	11	4.5	.032
	1st Qu.	81	20	24	234	2.2	1.9	1.6	.38	0	0	6	1.8	.13
	Median	179	35	1056	1772	4.7	5	5.2	.64	0	5	8	2.6	.0011
	Mean	216	58	3553	5043	5.5	5.6	5.7	.73	2	9	9	2.9	.0046
	3rd Qu.	261	100	8715	15,623	5.4	5.6	5.7	.73	5.6	19	8.6	2.9	.01
Max	794	555	133,117	108,306	7.7	8.3	7.8	.84	.82	6	24	9	3.4	.027
								1	1	45	155	14	5	.13

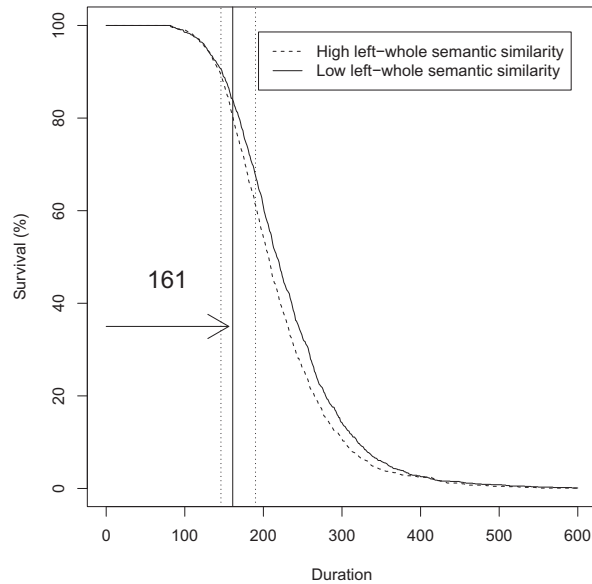


Fig. 1 – The divergence point estimate and its confidence interval for survival curves formed by the effect of left-whole semantic similarity in Study 4.

implemented the confidence interval DPA procedure as proposed by Reingold and Sheridan (2014). The confidence interval procedure runs 1,000 iterations of random resampling (with replacement) of the distribution of first fixation durations. Each iteration of the bootstrap resampling procedure begins by generating, for each individual participant, a survival curve for each level of the experimental variable under consideration. Next, survival percentages are averaged across participants separately for each condition. The average survival percentage is calculated for each 1 ms bin from 1 ms to the maximum latency in the whole distribution of durations (in the present study we used a 1000 ms cut-off). At the final step of the iteration, the divergence point of the two averaged survival curves is estimated. The divergence point is defined as the first of five consecutive 1-ms bins in which the difference between the survival proportion of both conditions is greater than a critical percentage value. As in Schmidtke et al. (2017) we opted for critical divergence point threshold of 3%, which is a more conservative threshold than the 1.5% proposed by Reingold and Sheridan (2014). The entire run of bootstrap resampling produces 1,000 divergence point estimates. From this distribution, the median value is taken as the divergence point estimate of the sample, and the 25th and the 975th values represent the bounds of the 95% confidence interval. Finally, as well as the 3% threshold criterion for establishing the divergence point, we exercised extra caution by only considering a divergence point for a lexical predictor variable if there was a reliable divergence in at least 500 out of the total 1,000 bootstrap iterations of the DPA procedure (see Schmidtke et al., 2017 for further elaboration on these cautionary measures).

As Fig. 1 depicts, the divergence point for left-whole semantic similarity in Study 4 is estimated as 161 ms, and has a confidence interval of 146 ms–190 ms. The median divergence point of 161 ms is the point at which 15% of first fixation durations were terminated. In our study we compute divergence point estimates for each individual lexical variable. Since we are interested in the time-course of compound word processing, our procedure follows that of Schmidtke et al. (2017) for derived word recognition. For all of our continuous lexical variables (see Predictor variables above), we first dichotomize the variable into high and low contrasts by splitting at the median value of that variable. Next, for each individual study, we then enter each dichotomized variable into the DPA confidence interval procedure and record the divergence point and confidence intervals.

Importantly, our stimuli selection was not designed with factorial contrasts (e.g., high versus low semantic transparency) in mind. Therefore, our lexical predictors were selected to represent a naturalistic distribution of that variable and variables were not orthogonally manipulated. It therefore could be the case that what on the surface appears to be an onset of the effect of a given predictor, may in fact be a reflection of a change in the distributions of other predictors at that timepoint. Following Schmidtke et al. (2017, see Supplementary materials S1), we conducted tests for this potential confound by examining, for each permutation of pairs of lexical variables (Lexical Variable A and Lexical Variable B), whether the distribution of Lexical Variable A is significantly different for the data at $t_1 = 1$ ms onwards, i.e. when no response is terminated, compared to the dataset at t_2 and onwards, when the median divergence point for Lexical Variable B is detected. We applied this check for all permutations of lexical variable pairs in our 5 datasets (of which all 172 had an absolute correlation strength of $r > .1$). Thus, none of the pairs showed a potentially confounding effect (after a Bonferroni correction for multiple comparisons). We conclude that it is unlikely that the divergence points are influenced by a confounding “third variable”.

It is important to note that the earliest time-point that an effect exerts a discernible impact on a distribution of responses is complementary to statistical techniques which are designed to explain variance in that distribution (ANOVA or regression), see Van Zandt (2002) and a detailed discussion in Schmidtke et al. (2017). For this reason, it is expected that an effect size is unrelated to how early or late the divergence point is observed in a response time distribution (for verification of this point, see Reingold & Sheridan, 2018). We confirmed this was the case in our study by correlating divergence point estimates with the strength of the correlations of those predictors with first fixation durations. We found no relationship between effect strength and divergence point estimates in all of our studies, thus confirming the dissociation between the onset of a lexical variable and its effect strength as a predictor of fixation durations (Study 1: $\rho = .75$; Study 2: $\rho = .06$; Study 3: $\rho = -.18$; Study 4: $\rho = .6$; Study 5: $\rho = -.2$; all p values $< .05$). Finally, we tested for the possibility that survival analysis produces a divergence point

estimate when there is no effect in an analysis of the central tendencies. As is reported in the online Supplementary materials (S2), we conducted a simulation study with the purpose of checking whether the convergence of the survival analysis is contingent on the size of the contrasts in means across “fast” and “slow” conditions. The results of the simulation showed that for very small effect sizes ($d = .1$), the survival analysis does not produce a reliable divergence point with 3% difference in survival curves. Only 14% of samples with this effect size produce a divergence point estimate, which does not pass the criterion of the survival analysis of the present study. However, for simulated samples with effect sizes of $d \geq .2$, 100% of samples produced divergence point estimates. We therefore conclude that the survival technique is capable of detecting small effect sizes reliably, but crucially it does not “create” an effect when there is not one in a traditional-style analysis of the means. In what follows, we report the results of the confidence interval DPA procedure across five separate eye-movement studies.

2.7. Results and discussion

The same data clean-up procedure was applied to each eye movement data set. We removed trials for which the eye tracking signal was lost, the target word was skipped, was fixated on for more than 6 times in total, was fixated on more than 2 times in the first pass of reading, was fixated on for the first time after gaze proceeded past the target word, was fixated on for less than 80 ms or was fixated on for more than 1000 ms. After clean-up, for each study we also removed (1) participants who made eye-fixations to fewer than 40 experimental stimuli, and (2) stimuli for which there were less than 8 observations. Furthermore, for each data set, we started out with trials for stimuli which had a complete set of values for all lexical predictors (see Materials).

The initial data set for Study 1 consisted of 3,771 trials. The application of data cleaning procedures led to a loss of 1902 (50%) trials from the initial raw data set. The resulting final data set comprised of 1,869 valid trials. For Study 2, we began with a data set of 2,835 trials. The data cleaning steps resulted in a loss of 797 (28%) data points, which produced a final data set of 2,038 valid trials. The initial data set for Study 3 consisted of 5,087 trials. The application of data cleaning steps reduced this data set to 3,446 valid trials (1,641 trials – 32% of data points lost). For Study 4 we started with a total of 3,194 trials. Data cleaning reduced this data set by 917 (29%) data points, which resulted in a final data set consisting of 2,277 trials. Finally, for Study 5 we began with 5,453 trials. The cleaning steps reduced this data set to 4118 (1,335–24% of data points lost). After clean-up procedures were applied, the proportion of trials for which a second fixation was made was 49% in Study 1, 47% in Study 2, 53% in Study 3, 46% in Study 4, and 48% in Study 5. For presentational purposes, we collapse together the reporting of results for each study.

We plot the divergence point estimates in Fig. 2, which visualizes median divergence point estimates for each lexical

variable for each of the five eye-movement studies. With all lexical variables collapsed together across studies, median divergence point estimates ranged from 138 ms to 269 ms. This indicates that, at the very latest, all lexical characteristics (orthographic, morphological and semantic) will have exerted an influence on the decision to terminate a fixation during the reading of a compound by 269 ms.

Focussing on groups of individual lexical variables, we consider first the individual divergence point estimates of frequency-based lexical variables. When present, the divergence point estimates for left constituent frequency emerge at time-points of 169 ms (Study 2) and 161 ms (Study 3). This is complemented by a broader range of values for whole word frequency (Study 1; 156 ms, Study 2; 144 ms, Study 3; 187 ms, Study 4; 151 ms, Study 5; 219 ms). Reliable divergence point estimates for right constituent frequency only emerged in Study 2 (160 ms) and Study 4 (204 ms). Importantly, in four out of five studies (Studies 1, 2, 4 and 5) the divergence point estimates of whole word frequency either preceded those of left constituent frequency, or occurred without the presence of a divergence point for left constituent frequency. Since the whole word frequency measure is taken to index the process by which morphological representations undergo recombination in accounts of semantics-blind obligatory decomposition (Solomyak & Marantz, 2010; Taft, 2004), the current findings suggest that access to the full form representation of a compound word is not necessarily contingent upon prior access to individual constituent representations.

Furthermore, the results indicate that access to the semantic representation of the whole compound word form occurs relatively early in the time course. Specifically, the same individual studies produce reliable divergence point estimates for left whole semantic similarity (Study 1; 142 ms, Study 2; 142 ms, Study 4; 161 ms) and left morphological family size (Study 1; 258 ms, Study 2; 200 ms, Study 4; 170 ms), and all divergence point estimates of left-whole semantic similarity precede access to indices of access to morphological family size. This finding indicates that the individual meanings of the whole compound word and the left constituent (e.g., the meanings of *cheese* and *cheesecake*) are accessed and their semantic relatedness evaluated without the prerequisite of accessing morphological structure, i.e., the family size of the compound's left constituent. The same pattern is true of right-whole similarity and right constituent family size. Across all studies where divergence point estimates are present for these variables, divergence point estimates for right-whole semantic similarity (Study 1; 173 ms, Study 3; 183 ms, Study 4; 167 ms) arrive before those of right constituent family size (Study 1; 177 ms, Study 3; 192 ms, Study 4; 200 ms). Thus the relative order of effects shown in the survival analysis here do not conform with the expected temporal flow of information predicted by form-then-meaning account of obligatory morphological decomposition (Rastle & Davis, 2008; Solomyak & Marantz, 2010; Taft & Forster, 1975; see also discussion in; Schmidtke et al., 2017).

In addition to posing a challenge for the relative order of formal and semantic effects, the findings concerning family size are incompatible with the purported absolute timeline of morphological processing. MEG results (Pylkkänen et al., 2004)

identify family size effects of compound words at 340 ms onwards. Here we show reductions of, at minimum, $340 - 258 = 82$ ms (left family size; Study 1) and $340 - 200 = 140$ ms (right family size; Study 4) in the estimation of the emergence of these effects in the behavioral record. In the General Discussion, we will motivate an even further reduction and an even earlier set of behaviorally informed upper temporal bounds. In sum, the temporal estimates from the results of the five eye-tracking studies regarding morphological effects presented here are out of sync with reports in the MEG or EEG literature.

Finally, with respect to other indices of semantic access, the results indicate that the emotional positivity of the compound word and the independent estimates of the emotional positivity of the left and right constituents occur relatively early in the time-course. For example, the valence of the whole compound word exerts an influence on compound recognition first out of all variables in Study 5 (161 ms) and either precedes or is contemporaneous with divergence point estimates of all morphological family size measures in Study 2 (183 ms) and Study 3 (189 ms). In Study 4, the onset of left family size (170 ms) arrives before the onset of compound word valence (191 ms), and both onsets are succeeded by the onset of right family size (200 ms). In addition, as can be gleaned from Fig. 2 (Studies 3, 4 and 5), it is possible for the valence of a whole compound word to be activated before access of the emotional positivity of its constituents. This might suggest that access to meaning of the compound word is not dependent on access to individual constituents. All together, these onset effects indicate that the emotional positivity of the whole compound word is accessed early in the time course (<200 ms) of

compound word processing, and that the cognitive registering of this semantic variable occurs before full form-based morphological decomposition. We discuss the implications of these findings for models of morphological processing further in the General discussion.

3. General discussion

The present paper expands the scope of a recent body of work that assesses processing stages in word recognition by comparing temporal estimates of the emergence of lexical effects in behavioral versus neuroimaging paradigms. In a series of five eye-tracking studies of sentence reading, we registered eye-movements to compound words (e.g., *airfield*, *catwalk*, *snowman*) and used the non-parametric technique of survival analysis (Reingold & Sheridan, 2014) to establish the earliest time-points at which effects of formal, morphological, and semantic variables emerge in the eye-movement record. One goal was to provide reliable behavioral estimates of upper temporal bounds for when respective brain activity is expected to emerge as a result of the processing of a given lexical variable. We report findings which are relevant to this goal in two ways. First, we compared descriptive statistics of compound processing from the literature and our own new data without resorting to a specific analytical method, to demonstrate the apparent paradox of “brainless behavior”. Second, we recruited evidence from survival analysis of five separate behavioral datasets to refine the temporal estimates of compound recognition in the behavioral record, and set them as benchmarks for brain imaging research. Our second goal was to use this data to adjudicate between competing form-then-

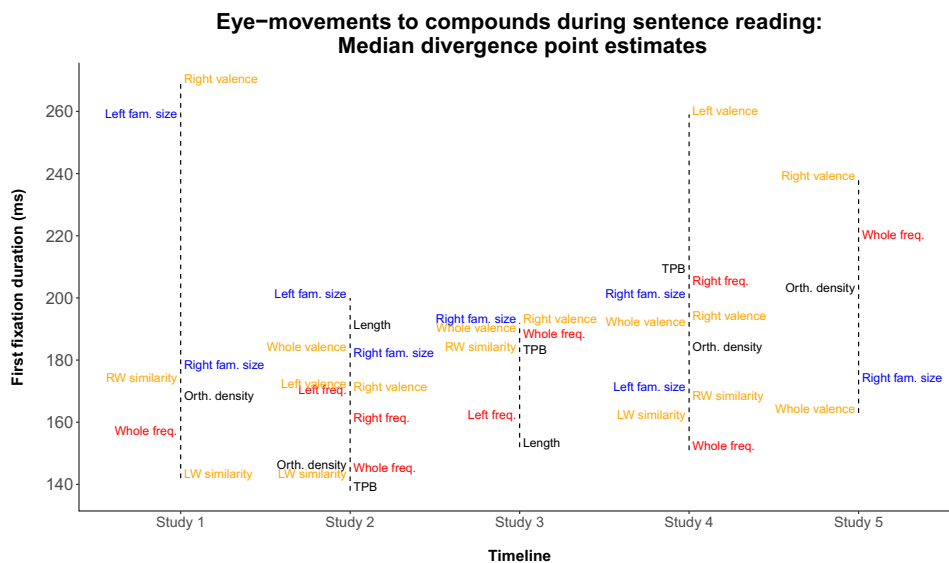


Fig. 2 – Plot of median divergence points for first fixation durations to English compound words presented in sentence context. Frequencies of the whole word and constituents are shown in red; orthographic predictors in black, morphological in blue, and semantic in orange. Only variables with a detectable divergence point in at least 500 out of 1,000 iterations with a 3% minimal contrast are plotted. TPB = bigram transition probability.

meaning and form-and-meaning models of complex word recognition. In what follows we discuss our findings with respect to both research goals.

3.1. Behavior predates brain activity in complex word processing

EEG and MEG experiments on compound word recognition (reviewed in the Introduction) place the earliest temporal estimates of virtually all lexical effects on word recognition at around 300–350 ms (but see Vergara-Martínez et al., 2009 for an earlier timeframe). However, our present findings across five eye-tracking experiments show that the average duration of the first fixation on compound words is 220–230 ms (Table 1). With the overview of the results of the survival analysis in the next section, this average latency will be refined and reduced further. We raise this point here because it is independent of statistical method and serves a comparative purpose.

Moreover, the average first-fixation duration latencies we report in our studies gain further importance when coupled with the observation that first fixation duration was correlated with one or more of the following measures: compound frequency, frequencies of constituents, and both semantic transparency measures, see correlation Tables in Appendix A.E. These behavioral data are fully in line with rich cross-linguistic eye-tracking research that has repeatedly demonstrated compound- and constituent-related lexical effects on first fixation durations (with an average of 230–260 ms), as well as on later or cumulative eye-tracking measures, see Introduction. Considered jointly, the previous and present findings reveal the paradoxical discrepancy whereby neural activity reported in the literature lags behind a host of behavioral effects that it must precede; the same discrepancy that is signalled in most comparative studies of the brain and behavior interface (see Introduction and Schmidtke et al., 2017). Simply put, a synthesis of the current evidence appears to indicate that by the time the brain responds to most formal, morphological or semantic properties of a compound word, on average the eyes will have already completed both the visual uptake and lexical processing of that compound, and will have moved to the next word, possibly even finishing its recognition and moving to yet another word (see Rayner, 1998 for estimates of fixation and saccade latencies).

There are a number of logical possibilities that might be offered to resolve this paradox. One is that the reported neurophysiological studies examine words in isolation (either presented in an RSVP paradigm or outside of a sentence context), while eye-tracking studies examine words in visual and semantic context. This might reduce fixation durations as a result of the benefit of parafoveal preview of upcoming visual information and/or contextual constraint. This possibility can only be ruled out by co-registration ERP-plus-eye-tracking studies with either isolated words or sentence reading as a task: in such a setup, parafoveal preview or contextual constraint (or the lack of both) would equally

influence the eye-movement and the neural record. We argue, however, that the difference in task (reading isolated words vs words in context) is unlikely to underlie the discrepancy. Several eye-tracking studies of isolated (simplex and complex) word recognition have reported fixation durations to words that are nearly identical to those registered in sentence reading (below 270 ms) and are again much shorter than the temporal loci of lexical effects proposed in the EEG and MEG data on complex words (see Kuperman et al., 2009, and a further review of cross-linguistic eye-tracking studies of isolated word recognition in; Schmidtke et al., 2017).

Another possibility for the paradox to arise is the fact that the temporal estimates reported in many EEG and MEG studies reflect an average neural response within a time-window rather than its onset latency. Indeed, as articulated in Fiorentino and Poeppel (2007), theirs and similar studies do not directly pursue the goal of identifying the earliest possible impact of morphology. Thus, the estimates of relevant brain activity may appear inflated and incompatible with behavioral data because they do not target the quantity of interest, i.e., the latency of a neural response. In this case, an argument can be made that neuroscientific studies of the word processing time-course have picked a suboptimal dependent variable. It is often so that validity of conflicting theories under discussion hinges on when certain effects emerge relative to one another; yet no theory, as far as we can tell, hinges on the knowledge of when the difference between the average response across two conditions is reliable within a broad time-window. It is thus possible that onsets of lexical effects on brain activity have latencies that precede the upper temporal thresholds set by behavioral responses. In this case, all that the brain imaging research field would need to do is shift its attention to the onset latency of response.

Possible ways of attending to this methodological issue have been proffered by Pulvermüller et al. (2009). They suggest that even when considering earlier time windows, ERP and MEG studies often sample the neural record within an insufficiently narrow time-window. A consequence of analysing a broad window of time is a failure to detect short-lived and more fine-grained fluctuations in the ERP and MEG signal which may be features of early stages of lexical processing in the neurophysiological record. In turn, they argue that since early effects may be fleeting and weaker, they may also be more sensitive to stimulus variance (see e.g., Assadollahi & Pulvermüller, 2001; 2003 for reports of modulation of onset of frequency effect by word length in MEG and Penolazzi, Hauk, & Pulvermüller, 2007 for a similar effect of word length in an ERP study of cloze probability and lexical processing). In sum, Pulvermüller et al. (2009) recommend that possible remedies for failing to detect earlier effects may involve narrowing the time-window of analysis, employing greater rigour in the matching of stimuli characteristics and taking greater care in minimizing of stimulus variance in factorial experiment designs. Arguably, these solutions developed for word recognition in general are equally

applicable specifically to the neuroscientific study of morphological processing.

However, in the literature on morphological processing so far, these appear to be unlikely remedies. A few EEG and MEG studies reviewed in the Introduction have examined a variety of time-windows or the amplitudes of these components at different levels of source intensity. To take a few examples from the Introduction, Fiorentino et al.'s (2014) EEG study of English compounds and El Yagoubi et al., (2008) EEG study of Italian compounds did not detect an effect of lexicality (and, in the Italian case, of compound headedness) in the 0–275 ms window, and only found a reliable effect in the 275–400 ms window. If these effects had a behavioral signature, they are nearly certain to have affected eye-movements at, on average, 200–250 ms. Indeed, in a manipulation of compound headedness similar to El Yagoubi et al.'s, Marelli and Luzzatti (2012) registered effects of headedness on first fixation duration that lasted, on average, 231 ms (for further converging evidence see eye-tracking studies by Amenta, Marelli, & Crepaldi, 2015; Marelli, Amenta, Morono, & Crepaldi, 2013).

Another approach taken by Pykkänen et al. (2004) was to examine effects of morphological family size and frequency not only at the peak latency of the MEG 350 component but at time-points corresponding to 25, 50, and 75% of the maximum intensity of this source, with the 25% time-point found at about 300 ms post-onset; also, earlier M170 and M250 components were examined. Reliable effects of either morphological predictors were only found at 75% and 100% of maximum source intensity, i.e., at about 330 ms post-onset and later. Latencies or amplitudes of either earlier components or other levels of intensity showed occasional numerical trends in predicted directions but did not show a reliable effect. The resulting temporal estimate of about 330 ms is incompatible with the behavioral timeline stemming from multiple eye-tracking reports of family size or family frequency effects obtained across languages, labs, populations, and tasks (words read in isolation or in context): they routinely associate morphological family-related effects with early eye-movement measures at around 270 ms (e.g., Juhasz & Berkowitz, 2011; Kuperman et al., 2009).

In sum, even when the net is cast wide in analyses of neural data, it does not appear to yield results that resolve the paradox of “brainless behavior”. This conclusion echoes findings of co-registration studies which show that the peak amplitude of the predictability effect in the ERP signal takes place when 96% of fixations on the target word had terminated, and that the onset of the N400 components takes place when 53% of fixations on the word had terminated (Dimigen et al., 2011; see also; Kretzschmar et al., 2015). While our data cannot directly point to the aspect(s) of neurophysiological research responsible for this discrepancy, we believe they clearly demonstrate the need to revisit the current practices of either experimental design, data collection, or data analysis of EEG and MEG studies of word recognition (i.e.,

recommendations offered by Pulvermüller et al., 2009, see above discussion).

3.2. When is the upper temporal bound for neural activity?

The previous section used an average duration of eye-fixation as a gross temporal estimate of when a behavioral effect on eye-movements emerges. This estimate has an advantage of being independent of data-analytical methods, however it is demonstrably inflated. Our use of survival analysis enabled us to considerably reduce the time-window during which lexical effects have their onsets. In five eye-tracking studies, we identified points of divergence between survival curves associated with virtually all lexical variables that prior research proposed as influential for compound word recognition. These points occupied a range of time between 138 ms and 269 ms, with a majority of effects having their corresponding median points of divergence before 200 ms, i.e., when less than 30% fixations terminated. Remarkably, these time-windows dovetail with Schmidtke et al.'s, (2017) survival analysis of three eye-tracking studies of derived words, thus corroborating reliability of the data and the method.

Furthermore, it is important to realize that a divergence point is not a time-point where the effect has its onset. It is a point at which fixations are differentially affected by high and low levels of a predictor (e.g., compound frequency, constituent frequency, semantic transparency etc.) begin to terminate at different rates. The actual influence of word properties on behavior takes place before the fixations terminate. Thus, any divergence point is but an upper temporal bound for an effect that necessarily has an earlier point of onset.

There is another oculomotor reason to treat divergence points in survival analyses as upper bounds. Word-level effects can only influence first fixation duration before saccadic motor programming reaches its nonlabile stage, i.e., when a saccade becomes non-cancellable. This puts the hypothesized onset of an effect some 80 ms before the fixation ends and the next saccade is launched (Becker, 1991; Dimigen et al., 2011; Findlay & Harris, 1984). Thus, it is likely that the actual behavioral response to some lexical variables initiates some 70–80 ms after exposure to the word and, even for the latest effects in Fig. 1, does not go later than 170–180 ms. If one takes 330 ms as the time-point at which EEG and MEG research places the earliest lexical effects in visual compound recognition, behaviorally informed onsets of effects precede this point by an astounding 150–250 ms. It is therefore sensible to expect a successful neuroscientific research program to demonstrate an emergence of lexically-driven brain activity well before 100 ms post-onset. This would be in line with many existing EEG and MEG reports for simplex words (Assadollahi & Pulvermüller, 2003; Hauk et al., 2006; Shtyrov & MacGregor, 2016; Penolazzi et al., 2007; Pulvermüller, 2002; Reichle, Tokowicz, Liu, & Perfetti,

2011; Sereno, et al., 1998), and would resolve this paradox of “brainless behavior”.

3.3. The relative order of morphological and semantic effects

A topic of contention between the form-then-meaning and form-and-meaning models of morphological processing is whether meanings of morphemes are only accessed after non-semantic morpho-orthographic information is used to segment a (complex or a pseudo-complex) word into its morphemes, and whether the meaning of the whole word is only accessed after meanings of morphemes are successfully combined into one representation (e.g., Rastle & Davis, 2008; Solomyak & Marantz, 2010; Taft & Forster, 1975, and EEG and MEG studies in the Introduction). Our findings do not support this form-then-meaning order of complex word recognition in which purely formal and morphological properties of the word are the first to exert their influence, followed by meanings of morphemes, and ultimately, whole word meaning. Instead, our data demonstrate that semantic effects (e.g., semantic similarity between constituents and compounds; or psychological valence of constituents and compounds) are virtually contemporaneous with effects associated with morpho-orthographic compound properties (e.g., frequencies of compounds and constituents, or constituent family sizes). These patterns, replicated over five studies in the present paper and three eye-tracking studies on derived words in Schmidtke et al. (2017), are most consistent with parallel-processing theories which posit that all sources of information are used simultaneously, as soon as even partial information about word form becomes available (Burani & Caramazza, 1987; Grainger & Ziegler, 2011; Kuperman et al., 2009; Schreuder & Baayen, 1995). These findings can also be accommodated under the Naive Discriminative Learning model (Baayen, Milin, Đurđević, Hendrix, & Marelli, 2011), which allows all formal (orthographic) cues to map and provide access to meanings, both for constituents and whole compounds.

In sum, we have demonstrated dramatic, paradoxical discrepancies between the time-courses of complex word processing based on neural and behavioral records of word recognition. Descriptive results demonstrate that these discrepancies are independent of the analytical technique used to describe behavioral data. Moreover, the application of a specific non-parametric survival analysis to eye-tracking data yields reliable and refined estimates of upper temporal bounds for the occurrence of neural activity associated with the processing of both formal and semantic compound properties. These upper temporal bounds indicate that there is a time window within the first 100 ms during which neurophysiological processing onsets of all the lexical properties of interest must happen. We acknowledge that the most accurate comparison of experimental paradigms will come from co-registration studies of eye-

movements and either EEG or MEG signals (Dimigen et al., 2011; Kretzschmar et al., 2015), and relegate such studies of complex word processing to future research (for other limitations and future directions see Schmidtke et al., 2017). We have also demonstrated the utility of survival analysis for characterizing a relative order of formal, morphological and semantic effects during compound recognition and adjudicating between theories critically dependent on the knowledge of that relative order. It is our hope that neurolinguistic research will direct its attention to revisiting the methodological or analytical premises of its current practice, and find reasons for the currently irreconcilable clash of two bodies of knowledge. Ideally, no further brain imaging data report would be able to gloss over the firm temporal constraints posited by the current wealth of available behavioral findings (Krakauer, Ghazanfar, Gomez-Marín, MacIver, & Poeppel, 2017). To continue doing so is to allow for this paradox of “brainless behavior” to perpetuate.

Acknowledgements

Daniel Schmidtke, McMaster English Language Development Diploma (MELD) programme, Department of Linguistics and Languages, McMaster University; Victor Kuperman, Department of Linguistics and Languages, McMaster University. Daniel Schmidtke's contribution was partially completed during his Ph.D. studies, which was supported by the Ontario Trillium Award and a Graduate fellowship awarded by the Lewis & Ruth Sherman Centre for Digital Scholarship (McMaster University). The remainder of Daniel Schmidtke's contribution was supported by a Post-doctoral fellowship appointment at McMaster English Language Development Diploma (MELD) Programme, Faculty of Humanities, McMaster University, Canada. Victor Kuperman's contribution was partially supported by the Canadian NSERC Discovery grant RGPIN/402395-2012 415 (Kuperman, PI), the Ontario Early Researcher award (Kuperman, PI), the Canada Research Chair (Tier 2; Kuperman, PI), the SSHRC Partnership Training Grant 895-2016-1008 (Libben, PI), and the CFI Leaders Opportunity Fund (Kuperman, PI).

We are thankful to Noor Al-Zanoon, Emma Bridgwater, Kaitlin Falkauskas, Irena Grusecki, Brooke Osborne, Sadaf Rahmanian, Katrina Reyes, Aaron So, and Chloe Sukkau for eye-movement data collection. We are also grateful to Jon Andoni Duñabeitia, Christos Pliatsikas, Alina Leminen and Eva Smolka for references to a compilation of neuroimaging research papers and presentation materials of speakers who presented neuroimaging research on morphological processing at the 10th International Morphological Processing Conference, Trieste, Italy in June 2017. Finally, we also wish to thank audience members of the aforementioned conference for their valuable feedback on the present research paper.

Appendix

Appendix A – Correlation matrix of lexical variables and first fixation durations in study 1.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1. Whole freq.													
2. Left freq.	.17***												
3. Right freq.	-.08***	.18***											
4. Whole valence	-.10***	-.10***	-.05*										
5. Left valence	-.15***	.06**	-.17***	.44***									
6. Right valence	-.08***	-.18***	.21***	.42***	.17***								
7. LW similarity	-.10***	-.16***	-.07**	.23***	.14***	.17***							
8. RW similarity	-.18***	-.05*	-.04	.01	-.16***	-.16***	.06**						
9. Left fam. size	-.08***	.43***	-.10***	.07**	.24***	-.16***	-.18***	-.13***					
10. Right fam. size	-.06**	.10***	.78***	.03	.00	.24***	-.08***	-.11***	-.14***				
11. Length	-.14***	-.01	-.03	.07**	.11***	.05*	.08***	.06*	-.15***	-.06*			
12. Orth. density	-.10***	-.14***	-.20***	.03	.00	.05*	.07**	.01	-.23***	-.23***	.83***		
13. TPB	.13***	.02	-.16***	-.03	-.01	-.07**	-.08***	.11***	.08***	-.03	.07**	-.01	
14. First fixation duration	-.08***	.00	.02	-.02	.02	.00	-.04	.03	.01	.00	-.03	-.04	.00

The lower triangle provides Spearman correlation coefficients. ***Correlation is significant at the .001 level. **Correlation is significant at the .01 level. *Correlation is significant at the .05 level.

Appendix B – Correlation matrix of lexical variables and first fixation durations in study 2.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1. Whole freq.													
2. Left freq.	.08***												
3. Right freq.	.10***	.06*											
4. Whole valence	.20***	-.07***	.28***										
5. Left valence	.08***	.17***	.14***	.47***									
6. Right valence	.15***	-.02	.07**	.27***	.09***								
7. LW similarity	-.34***	-.09***	-.11***	-.23***	-.07**	-.03							
8. RW similarity	-.27***	-.15***	.04*	-.10***	-.20***	-.31***	.09***						
9. Left fam. size	.09***	.57***	-.04	.01	.30***	.01	.07**	-.07***					
10. Right fam. size	.05*	-.02	.65***	.01	-.04	.00	-.03	.13***	-.04				
11. Length	.05*	.15***	-.11***	-.08***	-.03	.25***	-.14***	-.08***	-.10***	-.08***			
12. Orth. density	-.01	-.06**	-.20***	-.07**	-.15***	.20***	-.17***	-.08***	-.34***	-.17***	.74***		
13. TPB	-.01	-.04	-.07**	-.02	-.17***	-.12***	-.07**	-.01	.03	-.11***	-.25***	-.35***	
14. First fixation duration	-.11***	-.02	-.08***	-.04	-.03	-.01	.06**	.00	-.04	-.07**	.05*	.00	.02

The lower triangle provides Spearman correlation coefficients. ***Correlation is significant at the .001 level. **Correlation is significant at the .01 level. *Correlation is significant at the .05 level.

Appendix C – Correlation matrix of lexical variables and first fixation durations in study 3.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1. Whole freq.													
2. Left freq.	.16***												
3. Right freq.	.24***	.17***											
4. Whole valence	.15***	.14***	.06***										
5. Left valence	-.02	.18***	-.10***	.48***									
6. Right valence	.14***	.03	.19***	.25***	.05**								
7. LW similarity	-.04*	.04*	-.05**	-.18***	-.23***	-.03							
8. RW similarity	-.10***	.09***	.21***	-.05**	.00	-.15***	.19***						
9. Left fam. size	.02	.58***	-.02	.10***	.23***	-.02	.01	.07***					
10. Right fam. size	.13***	-.06***	.62***	-.08***	-.12***	.02	-.03	.11***	-.19***				
11. Length	.00	-.07***	-.08***	.10***	.03	.11***	-.04*	.01	-.20***	-.05**			
12. Orth. density	-.09***	-.09***	-.26***	.08***	.00	.09***	-.02	-.11***	-.22***	-.30***	.76***		
13. TPB	-.06***	.01	-.14***	.03	.05**	.01	-.03	-.04*	.14***	-.17***	.05**	.03	
14. First fixation duration	-.06***	-.04*	.01	-.04*	-.04*	.00	.01	.03	.00	.03	-.02	.00	.02

The lower triangle provides Spearman correlation coefficients. ***Correlation is significant at the .001 level. **Correlation is significant at the .01 level. *Correlation is significant at the .05 level.

Appendix D – Correlation matrix of lexical variables and first fixation durations in study 4.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1. Whole freq.													
2. Left freq.	.19***												
3. Right freq.	-.10***	.18***											
4. Whole valence	-.10***	-.11***	-.04*										
5. Left valence	-.13***	.10***	-.15***	.42***									
6. Right valence	-.09***	-.18***	.19***	.42***	.17***								
7. LW similarity	-.07***	-.14***	-.07**	.21***	.14***	.14***							
8. RW similarity	-.18***	-.08***	-.05*	.00	-.17***	-.18***	.07***						
9. Left fam. size	-.04	.43***	-.07***	.07**	.22***	-.16***	-.15***	-.17***					
10. Right fam. size	-.10***	.08***	.76***	.05*	.01	.23***	-.09***	-.12***	-.15***				
11. Length	-.18***	.01	-.05*	.04*	.12***	.02	.12***	.11***	-.16***	-.08***			
12. Orth. density	-.12***	-.12***	-.21***	.02	.01	.05*	.10***	.06**	-.23***	-.24***	.83***		
13. TPB	.13***	.03	-.16***	.00	-.01	-.09***	-.08***	.13***	.11***	-.04	.06**	-.01	
14. First fixation duration	-.12***	.00	.05**	-.04*	-.05*	.03	-.07**	.04	.03	.01	-.02	-.03	.03

The lower triangle provides Spearman correlation coefficients. ***Correlation is significant at the .001 level. **Correlation is significant at the .01 level. *Correlation is significant at the .05 level.

Appendix E – Correlation matrix of lexical variables and first fixation durations in study 5.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1. Whole freq.													
2. Left freq.	.12***												
3. Right freq.	.22***	.24***											
4. Whole valence	.20***	.13***	.04*										
5. Left valence	.10***	.31***	.17***	.50***									
6. Right valence	-.09***	.08***	.11***	.20***	.16***								
7. LW similarity	-.08***	.10***	-.09***	-.17***	-.25***	-.09***							
8. RW similarity	-.08***	.07***	.11***	.07***	-.07***	-.15***	.34***						
9. Left fam. size	-.01	.17***	-.05***	.10***	.11***	-.17***	-.12***	.00					
10. Right fam. size	-.03*	-.05**	.46***	.00	-.03	.14***	.01	.14***	-.02				
11. Length	-.23***	-.15***	-.14***	.02	-.08***	.01	-.04*	.03	-.06***	.01			
12. Orth. density	-.20***	-.23***	-.26***	-.04*	-.10***	-.03	-.14***	-.10***	-.17***	-.29***	.72***		
13. TPB	.03	-.01	-.09***	.06***	-.05***	-.13***	-.09***	.04*	.05**	-.08***	.02	-.03*	
14. First fixation duration	-.05***	-.02	-.01	-.06***	-.01	-.02	.01	-.01	.02	.05**	.04**	.02	.01

The lower triangle provides Spearman correlation coefficients. ***Correlation is significant at the .001 level. **Correlation is significant at the .01 level. *Correlation is significant at the .05 level.

REFERENCES

- Amenta, S., Marelli, M., & Crepaldi, D. (2015). The fruitless effort of growing a fruitless tree: Early morpho-orthographic and morpho-semantic effects in sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(5), 1587–1596.
- Arcara, G., Marelli, M., Buodo, G., & Mondini, S. (2014). Compound headedness in the mental lexicon: An event-related potential study. *Cognitive Neuropsychology*, 31(1–2), 164–183.
- Assadollahi, R., & Pulvermüller, F. (2003). Early influences of word length and frequency: A group study using MEG. *Neuroreport*, 14(8), 1183–1187.
- Baayen, R. H., Milin, P., Đurevic, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, 118(3), 438–481.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (release 2)*. Distributed by the Linguistic Data Consortium. University of Pennsylvania.
- Balota, D., Cortese, M., Hutchison, K., Neely, J., Nelson, D., Simpson, G., et al. (2002). *The English lexicon project: A web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords*. Washington University. Online <http://lexicon.wustl.edu>.
- Becker, W. (1991). Saccades. *Eye Movements*, 8, 95–137.
- Brooks, T. L., & Cid de Garcia, D. (2015). Evidence for morphological composition in compound words using meg. *Frontiers in Human Neuroscience*, 9, 215.
- Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977–990.
- Burani, C., & Caramazza, A. (1987). Representation and processing of derived words. *Language and Cognitive Processes*, 2(3–4), 217–227.
- Cavalli, E., Colé, P., Badier, J.-M., Zielinski, C., Chanoine, V., & Ziegler, J. C. (2016). Spatiotemporal dynamics of morphological processing in visual word recognition. *Journal of Cognitive Neuroscience*, 28(8), 1228–1242.
- Cavalli, E., Colé, P., Pattamadilok, C., Badier, J.-M., Zielinski, C., Chanoine, V., et al. (2017). Spatiotemporal reorganization of the reading network in adult dyslexia. *Cortex*, 92, 204–221.
- Dambacher, M., & Kliegl, R. (2007). Synchronizing timelines: Relations between fixation durations and N400 amplitudes during sentence reading. *Brain Research*, 1155, 147–162.

- Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research*, 1084(1), 89–103.
- Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: Analyses and review. *Journal of Experimental Psychology: General*, 140(4), 552–572.
- El Yagoubi, R., Chiarelli, V., Mondini, S., Perrone, G., Danieli, M., & Semenza, C. (2008). Neural correlates of Italian nominal compounds and potential impact of headedness effect: An ERP study. *Cognitive Neuropsychology*, 25(4), 559–581.
- Feldman, L. B., Milin, P., Cho, K. W., del Prado Martín, F. M., & O'Connor, P. A. (2015). Must analysis of meaning follow analysis of form? A time course analysis. *Frontiers in Human Neuroscience*, 9. Retrieved from: <https://doi.org/10.3389/fnhum.2015.00111>.
- Feldman, L. B., O'Connor, P. A., & Moscoso del Prado Martín, F. (2009). Early morphological processing is morphosemantic and not simply morpho-orthographic: A violation of form-then-meaning accounts of word recognition. *Psychonomic Bulletin & Review*, 16(4), 684–691.
- Findlay, J. M., & Harris, L. R. (1984). Small saccades to double-stepped targets moving in two dimensions. *Advances in Psychology*, 22, 71–78.
- Fiorentino, R., Naito-Billen, Y., Bost, J., & Fund-Reznicek, E. (2014). Electrophysiological evidence for the morpheme-based combinatoric processing of English compounds. *Cognitive Neuropsychology*, 31(1–2), 123–146.
- Fiorentino, R., & Poeppel, D. (2007). Compound words and structure in the lexicon. *Language and Cognitive Processes*, 22(7), 953–1000.
- Grainger, J., & Ziegler, J. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology*, 2, 54.
- Hauk, O., Davis, M. H., Ford, M., Pulvermüller, F., & Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage*, 30(4), 1383–1400.
- van Heuven, W. J., 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.
- Jared, D., Jouravlev, O., & Joanisse, M. F. (2017). The effect of semantic transparency on the processing of morphologically derived words: Evidence from decision latencies and event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(3), 422–450.
- Juhasz, B. J. (2016). Experience with compound words influences their processing: An eye movement investigation with English compound words. *The Quarterly Journal of Experimental Psychology*, 1–10.
- Juhasz, B. J., & Berkowitz, R. N. (2011). Effects of morphological families on English compound word recognition: A multitask investigation. *Language and Cognitive Processes*, 26(4–6), 653–682.
- Keuleers, E. (2013). *vwr: Useful functions for visual word recognition research*. R package version 0.3.0. Retrieved from: <http://CRAN.Rproject.org/package=vwr>.
- Keuper, K., Zwitserlood, P., Rehbein, M. A., Eden, A. S., Laeger, I., Junghöfer, M., et al. (2013). Early prefrontal brain responses to the hedonic quality of emotional words—a simultaneous EEG and MEG study. *PLoS One*, 8(8).
- Koester, Dirk, Gunter, Thomas C., & Wagner, Susanne (2007). The morphosyntactic decomposition and semantic composition of German compound words investigated by ERPs. *Brain and Language*, 102(1), 64–79.
- Koester, Dirk, Gunter, Th C., Wagner, Susanne, & Friederici, Angela D. (2004). Morphosyntax, prosody, and linking elements: The auditory processing of German nominal compounds. *Journal of Cognitive Neuroscience*, 16(9), 1647–1668.
- Koester, Dirk, Holle, Henning, & Gunter, Thomas C. (2009). Electrophysiological evidence for incremental lexical-semantic integration in auditory compound comprehension. *Neuropsychologia*, 47(8–9), 1854–1864.
- Krakauer, J. W., Ghazanfar, A. A., Gomez-Marín, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience needs behavior: Correcting a reductionist bias. *Neuron*, 93(3), 480–490.
- Kretzschmar, F., Schleewsky, M., & Staub, A. (2015). Dissociating word frequency and predictability effects in reading: Evidence from coregistration of eye movements and EEG. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(6), 1648–1662.
- Krott, A., Baayen, R. H., & Hagoort, P. (2006). The nature of anterior negativities caused by misapplications of morphological rules. *Journal of Cognitive Neuroscience*, 18(10), 1616–1630.
- Kuperman, V. (2013). Accentuate the positive: Semantic access in English compounds. *Frontiers in Psychology*, 4, 203.
- Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance*, 35(3), 876–895.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 211–240.
- MacGregor, L. J., & Shtyrov, Y. (2013). Multiple routes for compound word processing in the brain: Evidence from EEG. *Brain and Language*, 126(2), 217–229.
- Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, 92, 57–78.
- Marelli, M., Amenta, S., Morone, E. A., & Crepaldi, D. (2013). Meaning is in the beholder's eye: Morpho-semantic effects in masked priming. *Psychonomic Bulletin & Review*, 20(3), 534–541.
- Marelli, M., & Luzzatti, C. (2012). Frequency effects in the processing of Italian nominal compounds: Modulation of headedness and semantic transparency. *Journal of Memory and Language*, 66(4), 644–664.
- Penolazzi, B., Hauk, O., & Pulvermüller, F. (2007). Early semantic context integration and lexical access as revealed by event-related brain potentials. *Biological Psychology*, 74(3), 374–388.
- Pulvermüller, F., Shtyrov, Y., & Hauk, O. (2009). Understanding in an instant: Neurophysiological evidence for mechanistic language circuits in the brain. *Brain and Language*, 110(2), 81–94.
- Pylkkänen, L., Feintuch, S., Hopkins, E., & Marantz, A. (2004). Neural correlates of the effects of morphological family frequency and family size: An MEG study. *Cognition*, 91(3), B35–B45.
- R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org/>.
- Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, 23(7–8), 942–971.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422.

- Rayner, K., & Clifton, C. (2009). Language processing in reading and speech perception is fast and incremental: Implications for event-related potential research. *Biological Psychology*, 80(1), 4–9.
- Reichle, E. D., Tokowicz, N., Liu, Y., & Perfetti, C. A. (2011). Testing an assumption of the E-Z Reader model of eye-movement control during reading: Using event-related potentials to examine the familiarity check. *Psychophysiology*, 48(7), 993–1003.
- Reingold, E. M., Reichle, E. D., Glaholt, M. G., & Sheridan, H. (2012). Direct lexical control of eye movements in reading: Evidence from a survival analysis of fixation durations. *Cognitive Psychology*, 65(2), 177–206.
- Reingold, E., & Sheridan, H. (2014). Estimating the divergence point: A novel distributional analysis procedure for determining the onset of the influence of experimental variables. *Frontiers in Psychology*, 5, 1432–1432.
- Reingold, E. M., & Sheridan, H. (2018). On using distributional analysis techniques for determining the onset of the influence of experimental variables. *Quarterly Journal of Experimental Psychology*, 71(1), 260–271.
- Schmidtke, D., Matsuki, K., & Kuperman, V. (2017). Surviving blind decomposition: A distributional analysis of the time-course of complex word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(11), 1793–1820.
- Schmidtke, D., Van Dyke, J. A., & Kuperman, V. (2018). Individual variability in the semantic processing of English compound words. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 44(3), 421–439.
- Schreuder, R., & Baayen, H. (1995). Analysis of response time distributions. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 131–154). Mahwah, NJ: Lawrence Erlbaum Associates.
- Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: Eye movements and event-related potentials. *Trends in Cognitive Sciences*, 7(11), 489–493.
- Sereno, S. C., Rayner, K., & Posner, M. I. (1998). Establishing a time-line of word recognition: Evidence from eye movements and event-related potentials. *Neuroreport*, 9(10), 2195–2200.
- Shtyrov, Y., & MacGregor, L. J. (2016). Near-instant automatic access to visually presented words in the human neocortex: Neuromagnetic evidence. *Scientific Reports*, 6, 26558.
- Solomyak, O., & Marantz, A. (2010). Evidence for early morphological decomposition in visual word recognition. *Journal of Cognitive Neuroscience*, 22(9), 2042–2057.
- Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology Section A*, 57(4), 745–765.
- Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14(6), 638–647.
- Vergara-Martínez, M., Duñabeitia, J. A., Laka, I., & Carreiras, M. (2009). Erp correlates of inhibitory and facilitative effects of constituent frequency in compound word reading. *Brain Research*, 1257, 53–64.
- Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45(4), 1191–1207.
- Zandt, T. V. (2002). Analysis of response time distributions. In J. Wixted (Ed.), *Stevens' handbook of experimental psychology* (pp. 461–516). Hillsdale, NJ: Wiley.