Journal of Experimental Psychology:
Learning, Memory, and Cognition
2018, Vol. 44, No. 3, 421–439

0278-7393/18/$12.00 http://dx.doi.org/10.1037/xlm0000442

Individual Variability in the Semantic Processing of English Compound Words

Daniel Schmidtke
McMaster University

Julie A. Van Dyke
Haskins Laboratories, New Haven, Connecticut

Victor Kuperman
McMaster University

1883

Semantic transparency effects during compound word recognition provide critical insight into the organization of semantic knowledge and the nature of semantic processing. The past 25 years of psycholinguistic research on compound semantic transparency has produced discrepant effects, leaving the existence and nature of its influence unresolved. In the present study, we examined the influence of semantic transparency and individual reading experience on eye-movement behavior during sentence reading. Eye-movement data were collected from 138 non–college-bound 16- to 26-year-old speakers of English in a sentence-reading task representing a total of 455 different compound words. Measures of individual differences in reading experience were collected from the same participants and consisted of standardized assessments of exposure to printed materials, vocabulary size, and word recognition skill. Statistical analyses revealed facilitatory effects of both Modifier-Compound and Head-Compound transparency throughout the eye-movement record. Moreover, the study reports interactions between Head-Compound transparency and measures of reading experience. Readers with a small amount exposure to printed materials and a limited vocabulary size exhibited slower processing in late eye-movement measures when reading highly transparent compounds relative to opaque compounds. The opposite effect was observed for readers with a relatively large amount of exposure to printed materials and a relatively larger vocabulary size, such that highly transparent compounds facilitated lexical processing. To account for the results, the authors posit a trade-off between 2 cognitive mechanisms, which is modulated by individual reading experience; that is, the benefit of semantic coactivation of closely related concepts, and the cost of discriminating between those concepts.

Keywords: morphology, compound word recognition, eye-tracking, individual differences, naive discriminative learning

Supplemental materials: http://dx.doi.org/10.1037/xlm0000442.supp

How do human minds process complex meanings? How might both the properties of individual minds and those of semantic entities affect this processing? A common approach to addressing these questions is through the study of the visual recognition of compound words. The orthography of a compound word, by definition, encodes not only the meaning of the whole word itself

This article was published Online First September 21, 2017.

Daniel Schmidtke, Department of Linguistics and Language, McMaster University; Julie A. Van Dyke, Haskins Laboratories, New Haven, Connecticut; Victor Kuperman, Department of Linguistics and Language, McMaster University.

This work was supported by the Ontario Trillium Award and a Graduate fellowship awarded by the Lewis & Ruth Sherman Centre for Digital Scholarship (McMaster University) to Daniel Schmidtke. Julie A. Van Dyke's contribution was supported by the following NIH grants to Haskins Laboratories: R01 HD-073288 (Julie A. Van Dyke, PI), and P01 HD-01994 (Jay G. Rueckl, PI). Victor Kuperman's contribution was partially supported by the Canadian Social Sciences and Humanities Research Council Insight Development (430-2012-0488), and the Canadian Natural Sciences and Engineering Research Council (402395-2012), and the Early Researcher Award from the Ontario Research Fund. The authors are grateful to Morgan Bontrager for assistance with data collection and Kazunaga Matsuki for his contributions to statistical analyses. For technical support, we thank the Sherman Centre for Digital Scholarship at McMaster University. Thanks are also due to the attendees of the Twenty-First Annual Meeting of the Society for the Scientific Study of Reading, Santa Fe, New Mexico, where this work was presented in July 2014; the attendees of the Eighteenth Annual European Conference on Eye Movements, Vienna, Austria, where this work was presented in August 2015; attendees of the Third American International Morphology Meeting, which was held in Amherst, Massachusetts, in October 2015; and also to members of the Quantitative Linguistics research unit at the University of Tübingen, Tübingen, Germany, where this work was presented in April 2016. We also wish to thank Harald Baayen, Thomas Spalding and Christina Gagné for commenting upon earlier versions of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the U.S. National Institutes of Health or the Canadian Government.

Correspondence concerning this article should be addressed to Daniel Schmidtke, who is now at Department of Psychology, Biological Sciences Building, University of Alberta, Edmonton, Alberta, Canada T6G 2R3. E-mail: schmidtk@ualberta.ca

(e.g., *bootlace*), but also both of the meanings of its constituent morphemes (e.g., *boot* and *lace*). The processing consequences of the interaction between the many meanings contained in the orthography of complex words has been examined in many psycholinguistic studies (for recent reviews of complex word processing, including derived words, see Amenta & Crepaldi, 2012; Diependaele, Grainger, & Sandra, 2012; Hyönä, 2012; Libben, 2014). This article investigates the role of compound semantics on the visual recognition of compound words by focusing on *semantic transparency*, a measure of the semantic similarity between the meaning of the full compound and the meanings of either of the compound's constituents (e.g., the meanings of *boot*, *lace*, and *bootlace*; for reviews of other semantic variables pertinent to compound processing, see Juhasz, Lai, & Woodcock, 2015; Kuperman, 2013).

The aim of this study is to contribute to the conceptualization of the cognitive processes associated with semantic transparency and to the empirical base for this line of research. The emphasis of the latter aim is to provide further insight into the influence of the largely neglected role of individual differences in reading skill on morphological processing (but see Andrews & Lo, 2013; Falkauskas & Kuperman, 2015; Kuperman & Van Dyke, 2011). In what follows, we briefly review theoretical and empirical work on the semantic transparency effect during compound word processing and then motivate the theoretical and methodological approach of this study.

Theoretical and Experimental Work on the Role of Semantic Transparency

Current theoretical accounts of morphological processing generally agree on the expected *facilitatory* role of semantic transparency. Several accounts argue that the degree of similarity between multiple meanings encoded in transparent compounds triggers a "semantic flooding effect" (Libben, 1998, p. 43; see also Libben, Gibson, Yoon, & Sandra, 2003; Zwitserlood, 1994). This approach assumes that connections among a network of semantic representations facilitate recognition when those semantic representations are strongly associated (i.e., for the case of transparent compounds, such as *bootlace*), but inhibit processing when the semantic representations are weakly associated (i.e., for the case of opaque compounds, such as *bootleg*, which can mean "an illegal musical recording"). A family of analogous accounts are also available in the literature; for example, the "conjunctive activation" (El-Bialy, Gagné, & Spalding, 2013) and "semantic resonance" hypotheses (De Jong, 2002). Each of these hypotheses outlines a distinct processing sequence and processing architecture, yet all are united in their prediction that transparent compounds are recognized faster than opaque compounds. As a consequence, we remain broad in our construal of these effects and henceforth use the term *semantic boost* when referring to the facilitatory (i.e., "speeding up") effect originating from the close semantic relationship between the meanings of the full compound form and its embedded constituents.

Over the last 25 years, experimental work on the role of semantic transparency during compound word recognition has produced inconsistent effects. Many experiments that have found support for the facilitatory effect of semantic transparency have involved the use of visual lexical decision or priming paradigms. For example, relative to semantically unrelated control primes, Sandra (1990) reported shorter reaction times (RTs) during lexical decision when priming morphological constituents of semantically transparent compounds (e.g., *death* used as a prime for *birth* in *birthday*) relative to opaque compounds (e.g., *moon* used as a prime for *sun* in *Sunday*). In another seminal priming study in Dutch, Zwitserlood (1994) used compounds as primes and their constituents as target words. Zwitserlood (1994) found facilitatory effects on response latencies to both constituents when they were primed by fully transparent or partially transparent compounds (for similar results, see El-Bialy et al., 2013, and Libben et al., 2003).

In addition, Ji, Gagné, and Spalding (2011) revealed that greater semantic transparency sped up recognition under conditions in which decomposition was forced, such as when the compound was presented in a spaced format (e.g., *bull dog* instead of *bulldog*), and when constituents were presented in different colors. However, in a masked priming experiment, Fiorentino and Fund-Reznicek (2009) found an equivalent priming effect for transparent and opaque compounds in an experiment in which compound words served as primes for target words that were either their modifier or head constituents.

Moreover, Pham and Baayen (2013) found a nonlinear effect of the semantic similarity between the modifier and head of compound words (e.g., the degree of similarity between *shoe* and *horn* in the compound *shoehorn*). They found that both extremely transparent and opaque compounds inhibited processing, but that those at neither extreme of the transparency continuum were recognized the fastest. More recently, Juhasz et al. (2015) reported facilitatory effects of transparency on unprimed lexical decision latencies. Excluding Fiorentino and Fund-Reznicek's (2009) and Pham and Baayen's (2013) studies, lexical decision studies demonstrate that recognition of the compound is aided by a close relationship between the meanings of the compound and its constituents.

In eye-tracking studies, the findings are mixed. Two eye-tracking experiments (Juhasz, 2007; Underwood, Petley, & Clews, 1990) both reported shorter gaze durations on semantically transparent English compounds relative to semantically opaque compounds. Also, consistent with the findings of Ji et al. (2011) in lexical decision, Frisson et al. (2008; Experiment 2) reported a reliable effect of semantic transparency on eye movements when compounds were presented in a spaced format, providing further evidence that transparency may have a stronger effect in conditions in which decomposition is "forced." Finally, Marelli and Luzzatti (2012) reported interactive effects of semantic transparency, constituent frequency, and headedness during the processing of Italian compounds in lexical decision and eye-tracking data. Marelli and Luzzatti gauged semantic transparency using human judgments and reported that semantic transparency interacted with the headedness, frequency, and transparency of the compound word. The results of Marelli and Luzzatti (2012) are concomitant with evidence in support of the semantic boosting effect of semantic transparency (e.g., Juhasz, 2007; Sandra, 1990; Underwood et al., 1990; Zwitserlood, 1994) and additionally indicate that this effect may also be contingent on the constituent frequency and headedness of the compound word under visual inspection.

Finally, a handful of eye-movement experiments have failed to find effects of semantic transparency. For example, Juhasz (in press); Frisson, Niswander-Klement, and Pollatsek (2008; Experiment 1); and Pollatsek and Hyönä (2005) each reported no reliable influence of semantic transparency on the eye-movement patterns of reading English and Finnish compound words. In sum, lexical decision and eye-movement evidence suggests that, when present, semantic transparency facilitates compound word processing, but the overall picture remains inconclusive.

What Could Underlie the Inconsistent Effect of Semantic Transparency?

Based on the results from the past decades of research, the emergent picture of the role of semantic transparency in visual compound processing is far from clear. An obvious source of the discrepancy in research results could be traced to differences in tasks and behavioral measures (see Baayen, 2014, for a discussion of lexical decision vs. eye-tracking methodologies in morphological processing research).

In addition, the body of mixed results could stem from differences in how the theoretical construct of semantic transparency is conceptualized and operationalized. One common definition of transparency concentrates on semantic *relatedness*; that is, it taps into a degree of semantic overlap between the compound and its constituents and gauges how strong the semantic contribution of a constituent is within the meaning of a compound. Some studies (e.g., El-Bialy et al., 2013) operationalize this notion of relatedness as sets of categorical contrasts between fully opaque (OO; e.g., *blackmail*), partially opaque (OT or TO; e.g., *turtledove* or *footprint*), and fully transparent (TT; e.g., *bathtub*) semantic relationships between the meanings of compounds and their constituents (e.g., Zwitserlood, 1994). In some studies, human ratings (e.g., a 4-point scale; Libben et al., 2003) are converted into the aforementioned type of categorical variable (OO, OT, TO, TT) and are then entered into statistical analyses as primary predictors, but in other studies the same kind of human ratings are not converted to categorical predictors and are included in statistical models as continuous predictors (e.g., Juhasz et al., 2015; Marelli & Luzzatti, 2012).

A somewhat different definition of semantic transparency zooms in on the *similarity* between meanings of constituents and compounds. Similarity is most often assessed computationally by using Distributional Semantic Modeling (DSM; Turney & Pantel, 2010). Distributed Semantic Models are a collection of techniques that use textual co-occurrence-based vector representations to estimate semantic distance between words (see Marelli, Dinu, Zamparelli, & Baroni, 2015, for an overview of its application to compound transparency). Put simplistically, words are deemed similar in meaning if they tend to occur in similar contexts and thus, are more substitutable. One particular method of deriving word similarity is Latent Semantic Analysis (LSA; Landauer & Dumais, 1997), which has been previously used as a measure of semantic transparency in many morphological processing studies (e.g., Feldman, O'Connor, & Moscoso del Prado Martín, 2009; Fruchter & Marantz, 2015; Gagné & Spalding, 2009; Kuperman & Bertram, 2013; Moscoso del Prado Martín, Deutsch, Frost, Schreuder, De Jong, & Baayen, 2005; Pham & Baayen, 2013; Rastle, Davis, Marslen-Wilson, & Tyler, 2000; Rastle, Davis, & New, 2004; Wang, Hsu, Tien, & Pomplun, 2012). The primary advantage of LSA (and indeed other DSM measures) is that meaning relations are defined as an emergent property of the distributional information available for words in a corpus, and thus do not rely on subjective human judgment.

There is also a possibility to collect human subjective ratings of semantic similarity between compounds and their constituents: This article realizes that possibility. We argue that semantic relatedness and similarity are not orthogonal to one another: It stands to reason that a constituent that does not contribute much to the meaning of the compound (e.g., *bird* in *jailbird*) will not often occur in the same contexts as the compound. A recent study by Gagné, Spalding, and Nisbet (2016) demonstrated that measures of similarity and of relatedness are not identical either. Gagné et al. showed that human ratings of relatedness between the meaning of the modifier and the meaning of the compound are predicted by LSA values estimated for the semantic similarity between (a) the modifier and head constituent, (b) the modifier and the compound, and (c) the head and the compound. In contrast, human ratings of relatedness between the head constituent and the compound were only associated with relationship (c). Furthermore, Gagné et al. found that whereas human ratings of semantic relatedness could discriminate between discrete categories of semantic transparency (i.e., categories classified by expert linguists), LSA scores were not able to do so as accurately. Most notably, and relevant to the current article, discriminant analysis of human ratings was able to differentiate between fully opaque (OO) compounds (e.g., *buttercup*) and pseudocompounds (e.g., *kitten*), yet LSA estimates of similarity treated these categories as equivalent. As a result, Gagné et al. argue that linguistic classification, human ratings of relatedness, and LSA estimates of similarity each tap into a different facet of a theoretical concept of transparency.

A full-scale examination of the validity of each kind of transparency measure is beyond the scope of the current study. For the purposes of the present experiment, we concentrate on semantic similarity between compounds and their constituents. This choice is motivated by the central role of substitutability of words in context as a driving factor of word learning and concept discrimination (see next section). Moreover, in the current article we opted for a continuous measure of semantic similarity based on human ratings, rather than computational measures. By using human raters, we ensure that all compounds across the entire range of the transparency continuum are less likely to be implicitly treated as words belonging to two different morphological classes (see critique of LSA in Gagné et al., 2016). Moreover, the adoption of human ratings makes our results more comparable with previous studies on compound transparency (e.g., Fiorentino & Fund-Reznicek, 2009; Frisson et al., 2008; Juhasz et al., 2015; Libben et al., 2003; Zwitserlood, 1994).

The Naive Discriminative Reader Model of Word Recognition

We believe that the methodological discrepancies outlined previously do not exhaust the range of reasons for the mixed evidence regarding semantic transparency effects. At a theoretical level, the only predicted behavioral outcome associated with semantic transparency so far is the facilitated recognition of highly transparent compounds. However, a crucial limitation of this theoretical standpoint is exposed when one considers the cognitive effects of compound semantics from a developmental perspective, within the framework of the *Naive Discriminative Reader* (NDR) model of morphological processing (Baayen, Milin, & Đurđević, Hendrix, & Marelli, 2011). The NDR model of word processing is a framework in which orthographic representations of letter unigrams and letter bigrams are mapped directly onto semantic representations. This is a "two-layer" word recognition model that performs without the intervention of (orthographic) form representations of morphemes, whole words, or phrases. To take an example of a simplex (monomorphemic) word such as *rain*, naive discriminative reading first involves the recognition of incoming orthographic cues via *n*-grams (e.g., the bigrams *ra*, *ai*, *in*). During learning, these cues are associated with events, states, and entities in the world (e.g., "wetness," "puddles," and "umbrellas"; see Ramscar, Yarlett, Dye, Denny, & Thorpe, 2010). In this example, the ease of accessing the meaning of the word *rain* is driven by the strength of the activation between the memory trace of these events and the linguistic cue. Thus, reaching the point at which the meaning "rain" is learned is contingent on *discriminating* between the rest of the meanings that are cued by orthographic strings that are consistent with the bigrams *ra*, *ai*, *in*. In summary, the success of learning a meaning associated with a word is ultimately driven by the amount of uncertainty that is attributed to the association between the linguistic event and the events and entities which it denotes.

However, the learning and reading of an unspaced compound word intensifies the challenge of associating orthographic strings with meanings. That is, the orthographic forms of a compound's constituents (and also other letter strings which are embedded within compounds, see Baayen, Wurm, & Aycock, 2007; Bowers, Davis, & Hanley, 2005) provide *multiple* symbolic cues, which in turn activate a symbolic layer consisting of many meanings. During reading, resolution of the compound meaning is achieved in part by successfully discriminating between all the distinct meanings that are cued by the whole word string and substrings (constituents and/or pseudoconstituents), which are embedded in the orthography of the compound (e.g., *eye*, *lash*, *ash*, and *eyelash*). Of particular relevance to the current study, the NDR model predicts that discrimination is more effortful when meanings are more similar to one another, as is the case in relatively transparent compounds.

A crucial aspect of the discriminative learning hypothesis (cf. Ramscar & Baayen, 2013) is the way in which similarity between word meanings is defined. Put simply, the degree of semantic similarity between events and entities in the world is a function of how difficult it is to discriminate between them. From this perspective, discrimination difficulty is driven by a word's co-occurrence patterns. The more a word occurs in similar environments as other words across a diverse range of contexts, that is, the more a word is "substitutable" with other words across a range of lexical environments, the more effortful it is to learn to correctly discriminate between the event or entity that the word denotes and other events that are denoted by words in the same lexical environment. As argued previously, this theoretical premise motivates our focus on semantic similarity between morphemes in a compound, and provides a concept of semantic transparency that can connect with theories of human learning.

The hypothesized processing cost associated with discriminating between similar meanings can be illustrated with the transparent compound *rainstorm*. For this compound, the embedded orthographic strings *rain* and *storm* cue meanings that are both strongly related to the meaning(s) of *rainstorm* (e.g., "a storm with heavy rain") and also to each other. Under the semantic boost account, this compound would benefit from the strong semantic priming of closely related (and substitutable) meanings, leading to a speed-up in recognition. However, it is precisely this kind of compound that would create an effect opposite to semantic facilitation: *a discrimination challenge*. It is the concept of "rainstorm," and not "rain" or "storm" that is the target meaning. Singling out the target meaning is made difficult by the fact that all three meanings tend to occur and be learned in very similar contexts.

On the other hand, for an opaque compound, such as *hogwash*, the embedded orthographic strings *hog* and *wash* cue meanings that are unrelated to the meaning of the compound under visual inspection. That is, the constituents *hog*, *wash*, and *hogwash* occur in different contexts (i.e., are less substitutable). Indeed, this makes the meanings less likely to create a semantic boost, but alleviates the cognitively effortful task of discriminating between meanings. To sum up, the two accounts appear to predict opposite behaviors: the more transparent the compound, the faster its recognition would take place under the "semantic boost" account and the slower it will be under the "discrimination cost" account propounded in the (NDR) model.

It is important that the NDR model makes explicit predictions about word learning and language experience. A core principle of the model is that the strength of the cues for a particular meaning is established through a mechanism of competitive error-driven learning. Following from this insight is the notion that as language experience accumulates, the strengths of associations between the cues and the linguistic events adjust accordingly. The model's prediction then is that the ability to discriminate between semantic entities improves as the reader gains more reading experience.

In sum, there are three insights that we derive from the NDR perspective: a possibility of the discrimination challenge (an inhibitory role of the semantic similarity between compounds and their constituents); similarity of meanings (substitutability) as a relevant aspect of the semantic relationship within a compound; and the predicted importance of language experience for semantic effects. In the current experiment, we investigate the impact of language experience on the semantic processing of compound words. To this end, we investigate several related indices of the accumulation of reading experience; that is, individual exposure to print, vocabulary size, and word recognition skill. In what follows, we further motivate these individual differences and outline our hypotheses.

The Current Study

The focal role of language experience as outlined in the NDR model shapes this investigation in two ways. First, individual language experience has consequences for the predictions that stem from the two theoretical accounts described above, the "semantic boost" account and the NDR model. We illustrate these predictions in Figure 1, which presents five hypothetical but theoretically plausible profiles for the interaction effect of reading experience and semantic transparency on recognition effort during compound reading. In each panel, the solid regression line illustrates the hypothetical effect of transparency for individuals who represent the lowest level (here, the lowest quintile) of language experience. The dashed, dotted, dotdash, and longdash regression

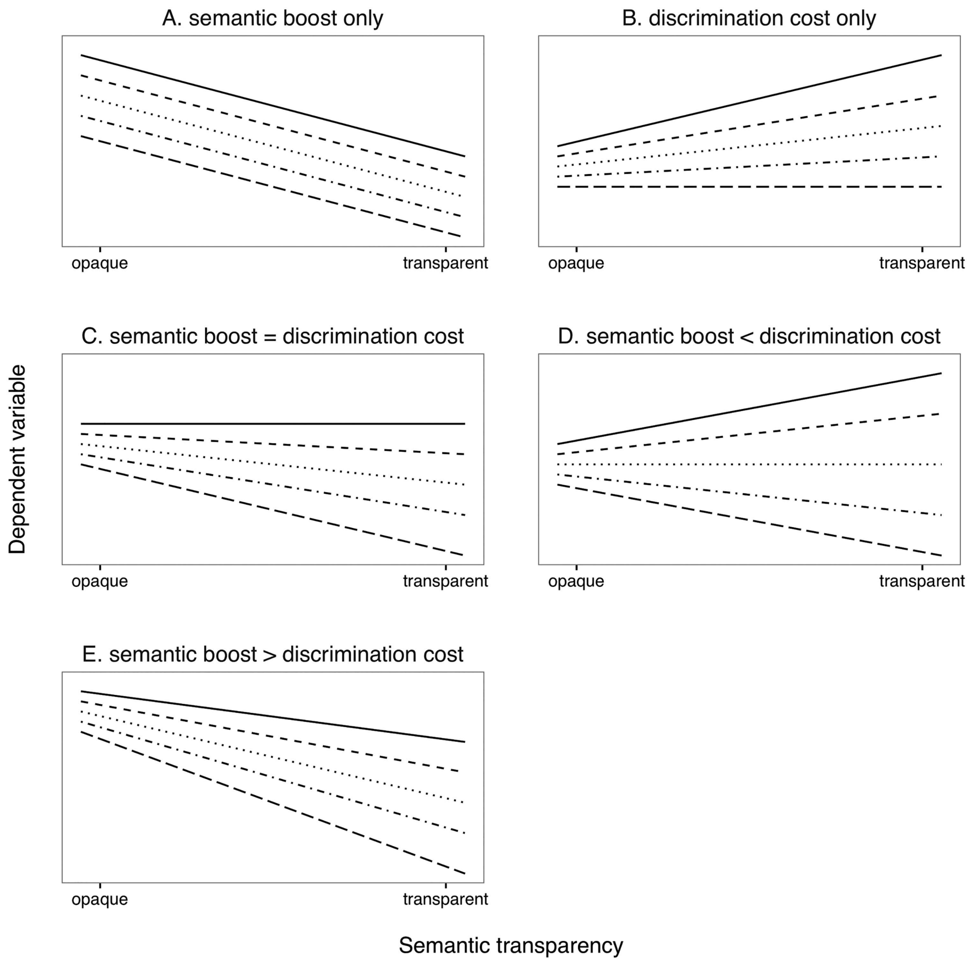

Figure 1. Hypothetical processing patterns for the interactive effect of semantic transparency (*x*-axis) and individual reading experience on compound word recognition effort (*y*-axis). Individual panels represent behavioral outcomes predicted under the semantic boost account (Panel A), the *Naive Discriminative Reader* (NDR) theory (B), or the potential trade-off between the semantic boost account and the NDR hypothesis (C–E). The lines plot the predicted effect of semantic transparency across the range of individual language experience, where quintiles of a metric of language skill are shown as the solid line (lowest quintile), and the dashed, dotted, dotdash and longdash (top quintile) lines, respectively.

lines each represent stepwise increases in language experience, respectively, such that the longdash regression slope visualizes a hypothetical effect of transparency for those individuals with the greatest amount of language experience.

Panel A visualizes a scenario in which there is an overall facilitatory semantic boost effect of transparency on processing effort for all levels of reading experience. Moreover, this panel reflects a hypothesized main effect of language experience, such that less experienced readers consistently show a greater effort of compound recognition. Because no account that we are aware of proposes an interaction between facilitation via semantic boosting and individual exposure to language, we do not incorporate such an interaction, leaving the mock-up regression lines parallel. Second, Panel B engenders the predictions of the NDR model, where more transparent compounds would give rise to a greater discrimination effort, especially for individuals with impoverished language experience, that is, individuals less likely to establish sufficiently strong associations between contextual cues and meanings. We visualize this result as a transparency by experience interaction effect, whereby less experienced individuals show inflated response times, but with a stronger inhibitory effect of transparency. In this scenario, the most experienced readers progressively accumulate enough exposure to language and vocabulary to be able to fully discriminate between meanings and thus the effect of greater transparency gradually weakens.

Of course, the semantic boost and NDR account need not be mutually exclusive as is depicted in Panels A and B. The remaining panels illustrate the case where both cognitive mechanisms are simultaneously present in the reading population. The behavioral outcome of the trade-off between the two conflicting processes is contingent on the relative contribution of both processes during compound recognition. Each of the potential trade-off outcomes play out in Panels C–E; that is, a pattern where the benefit of the semantic boost is equal to the cost of discrimination (Panel C), a pattern where the benefit of the semantic boost is weaker than the discrimination cost (Panel D), and a pattern where the benefit of the semantic boost outweighs the discrimination cost (Panel E). It is worth noting that a negative slope can only be found in the case of semantic facilitation, because discrimination does not predict any benefit from semantic similarity. Conversely, evidence of inhibition (i.e., a positive slope) can only point to the discrimination effort posited by the NDR account, but is not predicted under the semantic boost account. By comparing this theoretical space of possibilities to the patterns present in experimental data, we will be able to determine whether the trade-off is in place, and additionally, gauge the extent of the contribution of each process.

The second way in which language experience shapes this investigation is as follows. Any of the hypothesized interactive effects in Panels C to E of Figure 1 are most likely to be uncovered when there is a broad range of reading experience in the sample of the population under consideration. An ideal population to afford such experiential variability would be composed of individuals widely different in their educational backgrounds and reading habits, something rarely found in relatively homogeneous, highly proficient convenience pools of undergraduate student readers. To this end, in this study we consider the reading behavior of a "community" population of noncollege bound students. We do not argue that the cognitive mechanisms hypothesized above would not generalize to a student population. Instead we contend that they will be more likely detected within the community sample, where there lies a greater variability in reading skill. In the Method section, we compare a community sample against undergraduate students on their performance of one set of indices of reading experience and indeed observe a greater variability in the former sample.

In what follows, we investigate the effects of compound semantic transparency and individual reading experience, with the aim of adjudicating between each of the processing predictions outlined above (Figure 1). We collected eye-tracking data from 138 participants from the local community, who represent a noncollege bound population of readers. Participants read a total of 455 unique English compounds that were each embedded within a sentence context and varied in semantic transparency. In addition to partaking in an eye-tracking experiment, participants were administered a selection of tests which measure individual reading experience and other verbal skills.

With data from this experiment, the current study introduces two analyses which shed light on the role of semantic transparency on the reading of English compound words. Analysis 1 is a validation of prior mixed results, where we first focus on the main effects of semantic transparency in the eye-movement record. Based on previous results, only a main facilitatory effect of transparency is expected, if any main effect of transparency should arise in the first place. Analysis 2 puts to the test the predictions outlined in the current section, which each hypothesize that reading experience will in some way affect the processing of semantic transparency. This analysis therefore involves the interaction between semantic transparency and tests of individual reading experience.

Method

Participants

One hundred thirty-eight participants (74 female; 64 male) were recruited in New Haven, Connecticut, within an age range of 16–26 ($M = 21$, $SD = 2.29$). Participants were paid $15/hr and recruited from the local community in a number of ways, including presentations at adult education centers; advertisements in local newspapers; posters/flyers placed on adult school and community college campuses, public transportation hubs, local retail and laundry facilities; and from referrals from past and current study participants. All participants were noncollege-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. This study was approved by the Yale University Institutional Review Board.

Our choice of a community sample was because of an expected greater variation in their reading skill compared with an undergraduate sample. We tested this expectation by analyzing performance in the Author Recognition Test and Magazine Recognition Tests of individual exposure to print (ART and MRT; Acheson, Wells, & MacDonald, 2008, see also Moore & Gordon, 2015, for a recent review and the following sections for a full description of the tests). The distributions of ART and MRT test scores for the current study's 138 community participants were compared with that of the same scores acquired from an unrelated study of 173 students enrolled at McMaster University, Canada.

As is visualized in the density plots in Figure 2, the dispersion of scores in both the ART and MRT is greater among participants drawn from the community sample (ART; $SD = 10.44$, MRT; $SD = 7.81$) compared with those drawn from the convenience sample (ART; $SD = 7.3$, MRT; $SD = 4.64$). A larger variance of scores in the community sample was confirmed by F tests for both ART, $F = 2.04$, $p < .001$, and the MRT results, $F = 2.84$, $p < .001$.[1] In view of the demonstrably broader variability in both indices of exposure to print, we opted for testing our hypotheses on the role of reading experience within a population of noncollege bound readers. In studying an underresearched population, this study also aims to supplement the growing body of work which addresses the oft-neglected role of individual variability in eye-tracking studies (see reviews in Andrews, 2012; Radach & Kennedy, 2013; Rayner, Pollatsek, Ashby, & Clifton, 2012).

Apparatus

Eye movements were recorded with an EyeLink 1000 eyetracker, manufactured by SR Research Ltd. (Kanata, Ontario, Canada). The eyetracker is an infrared video-based tracking system combined with hyperacuity image processing. The eye-movement camera and a conjoined infrared illuminator was mounted on a desktop beneath the stimulus display. The recording was monocular (right eye). The camera sampled pupil location and pupil size at a rate of 1,000 Hz. A chin support and forehead rest was used to stabilize participants' gross head movements. The stimuli were presented on a 10.75-in × 13.25-in screen, with a refresh rate of 60 Hz. The average gaze position error of the

[1] As a side note, we discovered—much to our chagrin—that the community population exhibited a greater average amount of exposure to printed materials (ART; $M = 12.74$, MRT; $M = 10.22$) than did the sample of undergraduate students enrolled in the linguistics courses at McMaster University (ART; $M = 10.1$, MRT; $M = 6.33$). The independent Mann–Whitney U test confirmed that these differences were reliable for the ART ($W = 13533$, $p = .04$) and the MRT ($W = 15646$, $p < 0.0001$).

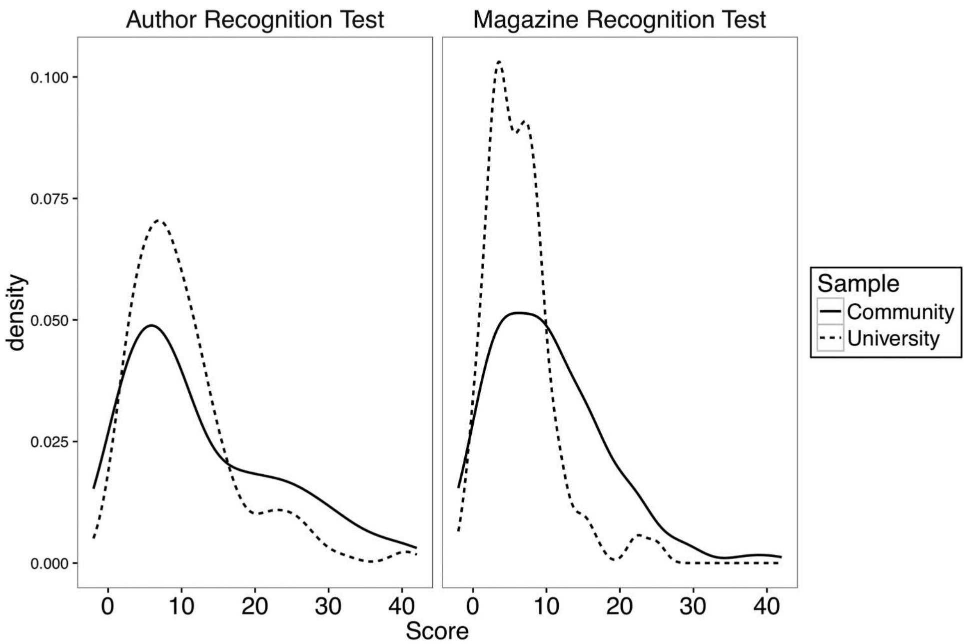

Figure 2. Density plots of the distribution of results of the Author (left) and Magazine (right) Recognition Tests split by two samples; a university student population ($n = 173$) and a noncollege bound adult population ($n = 138$).

EyeLink 1000 is <.05°, while its resolution is .01° (root mean square error), with a microsaccade resolution of .01°.

Materials

A total of 455 unique English concatenated noun-noun compounds (e.g., *applesauce*) were generated and were each embedded within a single sentence context (e.g., *There is nothing better than applesauce that you have made yourself.*). The sentences were split across 12 different stimulus lists: Each list consisted of 112 stimuli including 55 or 56 sentences with an unrelated syntactic manipulation that tests for syntactic interference effects. The number of participants assigned to each stimuli list was not identical, so compounds may have been seen by a differing numbers of participants (the full list of experimental stimuli with sample sizes for each compound are reported in the supplemental materials, which are available online).

The sentence context preceding each compound was neutral, and each compound 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. All sentences were displayed in Arial 14-point font. The height for the font was .5 cm and was viewed at a distance of 60 cm, where one character space subtended approximately .477° of visual angle. To track the performance of participants, comprehension questions would appear after certain trials, which required a "yes" or "no" response. The number of comprehension questions to critical trials per each list ranged from 27 to 32 (24%–29% of trials per list). The order of sentences was pseudorandomized for each participant. Each experimental session began with a series of eight practice trials and four comprehension questions.

Procedure

Prior to presentation of the stimuli, the eye-tracker was calibrated using a series of nine fixed targets distributed around the display, followed by a 9-point accuracy validation. Calibration was monitored throughout the experiment and was repeated after any breaks or whenever the experimenter judged necessary. Participants were instructed to read each sentence silently for comprehension and were told that they would be required to answer a comprehension question after some trials. Participants were told that they could take a break at any point during the experiment.

Each trial began with a screen containing a fixation point in the middle left of the screen. While fixating on this point, the experimenter pressed a button which would display a sentence on the screen. Participants were limited to 10 s to complete the reading of the stimulus sentence. After they had read the sentence, participants were instructed to look at a dark rectangle at the bottom right corner of the screen, which triggered the comprehension question to appear. The question appeared in the center of the screen; two possible answers appeared three lines below, one to the left of center and one to the right of center. Participants indicated their answer by pressing the associated button on a keyboard; for example, if the answer they chose appeared to the left of center, they were told to press a corresponding key. The position of the correct answer was counterbalanced throughout the experiment. If participants had not signaled that they had completed reading the sentence within the 10-s limit, the computer moved onto the comprehension question automatically. Participants were told to make their best guess at the comprehension question if they were unsure of the answer. If they had not answered within a 10-s limit, the computer moved onto the next item. The proportion of correct responses to the comprehension questions was 83%.

The duration of the eye-tracking experiment, consisting of one stimulus list, was 30 min. Participants also completed a battery of over 20 psychometric skill tests, whose duration did not exceed 5 hr (across separate sessions). In this article, we only discuss those tests that are theoretically motivated as relevant for an inquiry into semantic access during compound word recognition, see below Assessments of Individual Reading Experience section.

Response Variables

We analyzed the following durational eye-movement measures: single fixation duration (SFD; duration of fixation in cases where one fixation is made on a given word before gaze leaves that word for the first time); first-of-many fixation duration (FOM; duration of FOM fixation when there are more than one fixations on a given word before gaze leaves that word for the first time); gaze duration (GZD; summed duration of fixations on a given word before gaze leaves the word for the first time); go-past time (GPT; the sum of all fixation durations starting the moment the eyes first land on the target word until they make a rightward saccade past the target word), and the total viewing time (TVT; the summed duration of all fixations on a word). We also analyzed the probability of refixation (RefP; the likelihood of fixating on the word more than one time in the first reading pass), and regression-in probability (RegP; the likelihood of a regression to the compound after the first reading pass has been completed).

These measures are associated with different stages of the time course of processing a target word (Liversedge & Findlay, 2000; Rayner, 1998). FOM reflects the reader's first encounter with the target word. We use FOM instead of a more common first fixation duration measure (i.e., duration of any first fixation on the word regardless of the number of subsequent fixations on that word) because the first fixation duration conflates SFD and FOM. The first fixation duration measure does not allow for an independent examination of two cognitively different processes: one in which grapheme-to-phoneme conversion, lexical access to all morphemic meanings, and integration of the compound meaning with the sentence representation takes place within the timeframe of one single fixation on the word; and another in which the first fixation is a timeframe for only initial formal stages of word processing, and is a predecessor of semantic processing that takes place in subsequent fixations. Thus, whereas FOM has the ability to separate early from late processes, SFD subsumes all processes—early and late—associated with compound word recognition.[2]

RefP and GZD reflect further cognitive processing within the first-pass of reading. GPT incorporates any time spent reading earlier parts of the sentence before proceeding past the target word to inspect new parts of the sentence, while RegP indicates whether a regressive look-back has occurred. These measures therefore reflect higher order processes such as semantic integration and ambiguity resolution in the context preceding the target word. Finally, the cumulative measure of TVT reflects the overall amount of effort of word processing and likely shows the ease of the word's integration in the semantic representation of the entire sentence. All durational eye-movement measures were log-transformed to reduce the skewness of their distributions. Part A of Table 1 provides a list of the dependent variables including their ranges, standard deviations and means of their original and transformed values.

Predictor Variables

Our primary interest is in the combined influence of the participant-level variables (i.e., individual differences in exposure to printed materials) and word-level variables on the eye-movement record.

Lexical characteristics.

Measures of semantic transparency. The critical lexical variable of interest was semantic transparency. While different definitions and operationalizations of transparency are possible (see the section What could underlie the inconsistent effect of semantic transparency?), we focused on gauging semantic similarity of the modifier to the whole compound word (Modifier-Compound transparency; e.g., *rain-rainstorm*), and of the head to the whole compound (Head-Compound transparency; e.g., *storm* and *rainstorm*), and the simultaneous influence of these similarity measures on eye-movement behavior. The theoretical motivation for operationalizing semantic transparency as similarity (or substitutability) between morphemes originates from the Naive Discriminative Learning framework, whereas our choice of human ratings of similarity over computational ones is motivated by a recent demonstration that computational measures do not capture some distinctions as language users: see the earlier section entitled What could underlie the inconsistent effect of semantic transparency?, and Gagné et al. (2016). Moreover, our consideration of the Modifier-Compound and Head-Compound semantic relationships allows for comparability with previous studies using human ratings (e.g., Frisson et al., 2008). We acknowledge that there are other subtly distinct operationalizations of semantic relations between compounds and their morphemes (e.g., Ji et al., 2011; Juhasz et al., 2015; Libben et al., 2003) and relegate a comprehensive examination of their respective influence on word recognition behavior to future research.

In a separate study, we collected human-rated estimates of the Modifier-Compound and Head-Compound similarity for all compounds that were presented as stimuli in the eye-movement data set (see online supplemental materials S1 for procedure). To give a few examples, the compound *earthworm* showed a low Modifier-Compound and a high Head-Compound similarity: mean Modifier-Compound rating = 2.57; mean Head-Compound rating = 6.17. Conversely, the compound *eyeball* showed a high Modifier-Compound similarity and a low Head-Compound similarity: Mean Modifier-Compound rating = 5.92; mean Head-Compound rating = 2.73. We provide further compound examples, where ratings for the Modifier-Compound and Head-Compound Similarities were comparably high, such as *thunderstorm*: mean Modifier-Compound rating = 5.92; mean Head-Compound rating = 5.81, and also where ratings for the Modifier-Compound and Head-Compound Similarities were comparably low, such as *crowbar*: mean Modifier-Compound rating = 2.17; mean Head-Compound rating = 2.71. Overall, the correlation between Modifier-Compound and Head-Compound similarity is weak (r = .14).

[2] In our sample, first-fixation duration patterns with FOM: it shows a main effect of Modifier-Compound transparency and a weak unreliable effect of Head-Compound transparency. Our differentiation of SFD and FOM enabled us to pin down FOM as part of a behavioral pattern responsible for the effect.

Table 1
Summary of Eye Movement Measures as Response Variables (A) and Continuous Lexical Predictor Variables (B), Including Labels Used in Statistical Models

Variable	Original				Transformed		
	Range	*M*	*SD*	*n* trials	Range	*M*	*SD*
A. Response variables							
Single fixation duration (SFD)	127:1371	271	116	1,623	4.84:7.22	5.53	.36
First-of-many fixation duration (FOM)	53:1154	246	108	4,101	3.97:7.05	5.43	.39
Refixation probability (RefP)	0:1	.39	.49	5,724			
Gaze duration (GZD)	55:1711	374	233	5,724	4.01:7.44	5.77	.55
Go-past time (GPT)	55:1757	436	260	5,724	4.01:7.47	5.92	.55
Regression-in probability (RegP)	0:1	.24	.43	5,724			
Total viewing time (TVT)	127:1757	573	328	5,724	4.84:7.47	6.19	.59
B. Lexical predictor variables							
Modifier-Compound Transparency Rating	1.92:6.77	4.42	1.01		−2.49:2.33	0	1
Head-Compound Transparency Rating	1.4:6.79	4.39	.96		−3.1:2.49	0	1
Compound frequency	0:4958	133	432		−2.49:3.37	0	1
Modifier frequency	18:102467	6709	12492		−3.43:2.55	0	1
Head frequency	8:204428	13315	27039		−3.1:2.25	0	1
Modifier family size	0:54	11	11		−2.34:2.07	0	1
Head family size	0:155	18	27		−2.31:2.57	0	1
Compound length	6:13	8.5	1.22		−2.06:3.7	0	1

Note. Mean values, *SD*s, and ranges are reported for raw values and transformed variables.

The full list of 455 compound stimuli with the corresponding human transparency ratings and sentence contexts is available as an online supplemental data file.

Lexical control measures. We also considered the frequency of occurrence of the compound, the stand-alone frequency of occurrence of the modifier constituent, and the stand-alone frequency of the head constituent. These frequency statistics were extracted from the 50-million token SUBTLEX US (Brysbaert & New, 2009). We included the 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 additional predictors: these measures were obtained from the CELEX database (Baayen, Piepenbrock, & Van Rijn, 1995). We also computed the length (in characters) of all compound stimuli.

Additional norming studies were conducted to collect judgments about the plausibility and predictability of the critical compounds within their sentence contexts, see online supplemental materials S2 for procedure. Plausibility did not show any effect on eye-movement behavior, and predictability did not show substantial variability (with virtually all compounds and their morphemes having a zero predictability in their contexts). These variables are not discussed further.

Distributional characteristics of all lexical predictor variables are reported in Part B of Table 1. Correlations between all independent lexical predictors are presented in Table 2.

Assessments of individual reading experience. Educational psychology and psycholinguistics has proposed a number of assessments that tap into the multifaceted concept of reading experience (see Moore & Gordon, 2015; Stanovich & West, 1989). Here we consider three of the most commonly used measures; that is, amount of exposure to printed materials, vocabulary size and word recognition skill.

To estimate individual exposure to printed materials, we administered an updated version of Stanovich and West's (1989) ART and MRT; Acheson et al., 2008; see online supplemental materials S3 for procedure. Because the ART and MRT target the same individual characteristics, we summed the ART and MRT test scores to create a composite score of print exposure (henceforth referred to as Print Exposure). As is visualized in Figure 2, the distribution of ART and MRT is strongly right skewed, so to

Table 2
Correlation Matrix of All Lexical Predictor Variables

Variable	1	2	3	4	5	6	7
1. Modifier-Compound Transparency Rating							
2. Head-Compound Transparency Rating	.14**						
3. Compound frequency	.05	.01					
4. Compound length	.15**	.20***	−.06				
5. Modifier frequency	.00	.03	.10*	.06			
6. Head frequency	−.10*	.15**	.06	−.13**	.01		
7. Modifier family size	.10*	−.04	.00	−.23***	.25***	−.01	
8. Head family size	−.07	.09	.02	−.13**	−.03	.56***	.01

* Correlation is significant at the.05 level. ** Correlation is significant at the .01 level. *** Correlation is significant at the .001 level.

combat this we applied the square root transformation to normalize the distribution. A Shapiro-Wilk test indicated that the transformed Print Exposure score is normally distributed ($W = 0.9859$, $p = .1692$). The Print Exposure measure was scaled to ensure comparability with other variables in the regression analyses. The correlation between raw scores in ART and MRT tests was high ($r = .74$). Both ART and MRT correlated with Print Exposure at $r = .93$.

We measured individual vocabulary size using the Peabody Picture Vocabulary Test 4th Edition (PPVT; Dunn & Dunn, 1981) and the Vocabulary subtest of the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999)–2nd Edition. As both the PPVT and WASI vocabulary tests measure the same skill, we created composite Vocabulary Size composite test score (henceforth referred to as Vocabulary Size) by scaling and summing their values. To measure fluid word recognition skill, we administered the Sight Word Efficiency subtest of the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999). The scaled accuracy of this test score was used in analyses and is henceforth referred to as Word Recognition Skill. Both Vocabulary Size and Word Recognition Skill were normally distributed.

Correlations between variables were high in our sample (Print Exposure and Vocabulary Size, $r = .7$; Print Exposure and Word Recognition Skill, $r = .6$; Vocabulary Size and Word Recognition Skill, $r = .52$). As reviewed in Moore and Gordon (2015), exposure to printed materials (as measured by the ART) often shows moderate to strong correlations with other received indices of reading experience, for example, vocabulary size and word recognition skill. This is expected, given a reciprocal "causal spiral" relationship between the amount of reading, the rate of acquisition in one's vocabulary, and the ease of word recognition as a foundational skill for reading comprehension (e.g., Stanovich, 1986; Stanovich & Cunningham, 1992). This causal spiral implies that as print exposure intensifies "individuals acquire increasingly more word meanings and word forms from books, which further facilitates their reading development and their willingness to read for pleasure" (Mol & Bus, 2011, p. 289). Because the exact nature of the relationship between these two variables is not the theoretical focus of the current study, we argue that either measure can be entered into an analysis as more (exposure to print) or less (vocabulary knowledge and word recognition) generic proxies of reading experience.

Distributional characteristics of the ART, MRT, and Print Exposure (the composite of ART and MRT), as well as the Vocabulary Size composite and the Word Recognition Skill tests, are reported in Table 3. See online supplemental materials S3 for the procedure of each test. Table 3 confirms the notion of broad variability in our sample. For example, the age equivalent of the WASI vocabulary score ranged from 9 to 20 years of age, with 17.5 as a mean (the actual mean age of our population sample was 21), and word recognition skill from 7 to 18, with 15.5 as a mean.

Statistical Considerations

Our interest in studying this behavioral data set lies in tracing the impact of semantic transparency and the way it is modulated by individual differences in reading experience. In Analysis 1, we fit a separate regression model to each dependent variable with both estimations of Modifier-Compound and Head-Compound transparency included within the same model as critical main effects. In Analysis 2, each model contained both measures of semantic transparency, each separately interacting with indices of reading experience.

Linear mixed-effects multiple regression models were used for this study, with Gaussian (for continuous response variables) or binomial (for binary response variables) underlying distributions (Baayen, 2008; Baayen, Davidson, & Bates, 2008; Jaeger, 2008; Pinheiro & Bates, 2000). Across all models, we used restricted maximum likelihood (REML) estimations. This procedure ensures that nuisance parameters are restrained by producing unbiased estimates of variance and covariance parameters. Each model contained by-compound and by-participant random intercepts, as well as separate by-participant random slopes of both semantic similarity measures defined in the fixed effects part of the model. Because the performance of each participant in each assessment of reading experience is represented by

Table 3
Summary of Individual Differences Tests and Measures

	Raw scores									Equivalent scores			
Skill test	*M*	*SD*	Minimum	10%	30%	50%	70%	90%	Maximum	*M*	*SD*	Minimum	Maximum
1. Print exposure[a]	0	1	−2.48	−1.27	−.55	−.11	.57	1.3	2.45	—	—	—	—
i. Author Recognition Test	12.74	10.44	−1	2	5	9	18	28.3	42	—	—	—	—
ii. Magazine Recognition Test	10.22	7.81	−2	1.7	5	9	13	21.3	41	—	—	—	—
2. Vocabulary size[a]	0	1	−2.38	−1.64	−.26	.08	.55	1.26	1.74	—	—	—	—
i. Vocabulary (WASI)[b]	39	6.33	24	30	37	39	42	47.3	52	17:6	3:6	9:2	20
ii. Vocabulary (PPVT)[cd]	200.83	18.57	85	19.4	200	205	210	218	228	10.97	2.52	<k.0	>12.5
3. Word Recognition Skill	0	1	−5	−.95	−.27	.09	.76	1	1	—	—	—	—
i. Sight Word Efficiency (TOWRE)[b]	91.35	12.67	25	79.3	87.9	92.5	101	104	104	15:6	2:4	6:8	17:8

Note. WASI = Wechsler Abbreviated Scale of Intelligence; PPVT = Peabody Picture Vocabulary Test 4th Edition; TOWRE = Test of Word Reading Efficiency. Numbers 1:3 label the transformed and/or scaled values for individual differences measures included in statistical analyses. Raw values of tests from which scores composed are nested beneath each measure. Mean values, *SD*s, and quartiles are reported. Where available, mean, *SD*s, and ranges are reported for percentile rank, grade, or age equivalent scores.
[a] Composite score. [b] Age (year:month) is used here as the standardized equivalent score. [c] Grade equivalent of "second year of graduate school" is indicated as 18. [d] Grade levels prior to first grade are indicated as k [month], where [month] indicates months into the kindergarten year.

one value (and hence offers no variability) we did not model by-participant random slopes of individual-differences tests. Furthermore, across all analyses, we refitted models after removing outliers from all data sets by excluding absolute standardized residuals exceeding 3 standard deviations. We used log-transformed values of frequency-based measures to account for distributional skewness. All lexical predictors and scores of individual differences tests were scaled in order to ensure interpretable comparison of effects across predictor variables of varying scales. We also included trial number as a fixed effect across all statistical models. We used the lme4 (Bates, Maechler, Bolker, & Walker, 2013) package in the R statistical computing software program (R Core Team, 2014) to compute statistical models. A possible nonlinear nature of the critical main effects and nonplanar nature of critical interactions was explored using generalized additive multiple regression models. No effect showed an appreciable deviation from a linear/planar functional form.

The simultaneous consideration of multiple models is expected to lead to inflated Type I error rates. A recent article by von der Malsburg and Angele (2017) suggests two kinds of family wise corrections for multiple comparisons: we followed both in our analyses. We used seven dependent variables in Analyses 1 and 2 (see the list in Part A of Table 1), and defined the resulting set of seven models produced by each analysis as a "family." The first rule-of-thumb correction criterion relies on a statistical fact that the probability of discovering two out of seven effects with p values $<.05$ is estimated at .044 (under the simplifying assumption that the dependent variables are independent from each other). Under this family wise correction criterion, critical main effects (Analysis 1) or interactions (Analysis 2) were determined as reliable only if there were two or more effects (with p value <0.05) within the family of dependent measures. For instance, the measure of Modifier-Head transparency was interpreted as a reliable predictor of compound processing only if it had a main effect on two or more dependent eye-movement variables.

An alternative, second type of correction was the false discovery rate method (Benjamini & Hochberg, 1995), which we also separately applied to each family of models. This method keeps under control the proportion of rejected null hypotheses in the entire set of available null hypotheses through p value adjustment: Other common methods (e.g., the Bonferroni correction investigated in von der Malsburg & Angele, 2015) have been shown to be overly conservative. In other words, for this test, an effect of a transparency measure on compound recognition was deemed reliable if its p value was below the nominal .05 threshold after the false discovery rate correction. The family wise probability and false discovery tests were applied independently of one another. If either test criterion was satisfied, we considered an effect to qualify as a true rejection of the null hypothesis. Because not all significant effects that we report passed the false discovery rate p value adjustment (i.e., some models satisfied just the family wise correction criterion), we henceforth do not report adjusted p values in our reporting of results.

Results and Discussion

The initial raw data set consisted of 6,582 trials. We removed trials in which the eye-tracking signal was lost, the target word was skipped (485 trials), was fixated on for more than 6 times, was fixated on for the first time after gaze proceeded past the target word, or was fixated on for less the 50 ms. We also removed outlier responses by eliminating the top and bottom 1% of the total fixation time distribution per each participant. All together, these data cleaning procedures led to a loss of 858 (13%) trials from the initial raw data set. The resulting final data set comprised of 5,724 valid trials.

For presentational purposes, we report our results as two analyses. Analysis 1 tests main effects of compound transparency on the eye-movement record and does not take into account individual differences between participants. Analysis 2 conducts a confirmatory analysis of the hypothesized interaction between semantic transparency and individual differences in reading experience.

Analysis 1: Main Effects of Semantic Transparency

We fitted a model for each of the seven eye-movement measures (listed in Part A of Table 1) as a response variable, with both transparency (Modifier-Compound and Head-Compound) scores included together as fixed predictor variables in each model. Along with each of the transparency measures, we also included the set of predictors listed in Part B of Table 1 as fixed effects. For number of trials per dependent variable (EM), see Part A of Table 1. Random effects were defined as stated in Statistical Considerations. In what follows, besides inferential statistics we reported effect sizes which we define as the amount of change in the dependent variable estimated for the contrast between the 10th and the 90th percentile of the critical semantic transparency variable: Where necessary, values of the dependent variables were back-transformed to milliseconds or percentages.

Modifier-compound transparency. We found significant main effects of Modifier-Compound semantic transparency on six out of seven different eye-movement measures. Two of these eye-movement measures also demonstrated a significant main effect of Head-Compound semantic transparency. These effects had a p value $<.05$ after the application of the family wise probability correction or false discovery rate correction criterion. Table 4 provides a summary of the models in which a main effect of semantic transparency was reliably present.

There was a significant main effect of Modifier-Compound Transparency Rating on first-of-many fixation duration [$\hat{\beta} = -0.01$; $SE = 0.01$; $p = .046$, effect size (defined above) = 8 ms], gaze duration [$\hat{\beta} = -0.02$; $SE = 0.01$; $p = .008$, effect size = 19 ms], go-past time [$\hat{\beta} = -0.02$; $SE = 0.01$; $p = .006$, effect size = 24 ms], and total viewing time [$\hat{\beta} = -0.04$; $SE = 0.01$; $p = .001$, effect size = 53 ms] in such a way that greater Modifier-Compound transparency led to faster reading times. Moreover, Modifier-Compound Transparency also affected refixation probability [$\hat{\beta} = -0.07$; $SE = 0.03$; $p = .037$, effect size = 0.11] and regression-in probability [$\hat{\beta} = -0.13$; $SE = 0.05$; $p = .009$, effect size = 0.09] such that greater Modifier-Compound transparency reduced the likelihood of fixating on the compound more than one time during the first reading pass and reduced the probability of regression to the compound after the first pass of reading had been completed.

The effects for Modifier-Compound Transparency demonstrates that the greater semantic similarity between the modifier and the compound word (e.g., *rain* and *rainstorm*), the smaller the amount of processing effort is required to read the compound (i.e., shorter reading times and fewer refixations and regressions). The results show that this effect has an influence on processing effort from the initial stages of reading a word (first-of-many fixation duration) and continues (gaining larger effect sizes) throughout the cumu-

Table 4
Summary of Main Effects for Semantic Transparency

Eye movement measure	Modifier-compound transparency	Head-compound transparency
Single fixation duration	*ns*	*ns*
First-of-many fixation duration	$\hat{\beta} = -.01$, $SE = .01$, $p = .046$	*ns*
Refixation probability	$\hat{\beta} = -.07$, $SE = .03$, $p = .037$	$\hat{\beta} = -.09$, $SE = .04$, $p = .013$
Gaze duration	$\hat{\beta} = -.02$, $SE = .01$, $p = .008$	$\hat{\beta} = -.02$, $SE = .01$, $p = .029$
Go-past time	$\hat{\beta} = -.02$, $SE = .01$, $p = .006$	*ns*
Regression-in probability	$\hat{\beta} = -.13$, $SE = .05$, $p = .009$	*ns*
Total viewing time	$\hat{\beta} = -.04$, $SE = .01$, $p = .001$	*ns*

Note. *ns* = nonsignificant. Included are only the models which passed either the family-wise probability or the false discovery rate *p* value correction. Reported are regression coefficients and *SE*s, as well as false discovery-rate corrected *p* values.

lative durational measures of eye-movement fixations (gaze duration and total viewing time). The inflation in reading times is also reflected in the effects of probability of refixation on the compound within the first pass of reading and regression to the compound after the first pass of reading.

Head-compound transparency. With respect to Head-Compound semantic transparency, we found that Head-Compound Transparency Rating exerted a main effect on gaze duration [$\hat{\beta} = -0.02$; $SE = 0.01$; $p = .029$, effect size = 17 ms] and refixation probability [$\hat{\beta} = -0.09$; $SE = 0.04$; $p = .013$, effect size = 0.14]. These effects show that greater transparency of the Head-Compound transparency relation (e.g., *storm* and *rainstorm*) reduces the likelihood of refixating on a compound word during the first pass of reading and leads to faster overall reading times during the first pass of reading.

Tables S5.1–6 in the online supplemental materials provide full specifications of each model in Analysis 1. To ensure that the trends that were estimated as reliable in regression models do in fact reflect the structure of the raw data, and for interpretability of the models' results, in Supplementary materials S4, we report plots displaying raw trends in the data. Trends in the raw data confirm the critical findings that result from the regression models.

The procedure of Analysis 1 is compatible with prior lexical decision (Juhasz et al., 2015; Sandra, 1990) and eye-movement (Juhasz, 2007; Underwood et al., 1990) studies of semantic transparency in that it estimates the critical effect on eye movements without regard to individual differences between participants. These earlier studies reported a reduced cognitive effort associated with greater semantic transparency when recognizing compound words. The results of this analysis provide corroborating evidence in favor of semantic transparency as a codeterminant of compound processing, as indicated by main effects of human ratings of Modifier-Head and Head-Compound transparency on eye-movement behavior. We therefore conclude that Analysis 1 is inconsistent with the null effects reported in Frisson et al. (2008); Juhasz (in press); and Pollatsek and Hyönä (2005).

As reviewed earlier, Marelli and Luzzatti (2012) demonstrated that semantic transparency interacts with constituent frequency and headedness during the processing of Italian compounds. We investigated this possibility in our own data set by interacting each transparency measure with the modifier and head constituent frequencies of a compound word. Both transparency measures were entered as a multiplicative interaction with both modifier and head constituent frequencies (and with compound frequency, for completeness) in separate models for each dependent variable. After the application of either the family wise probability criterion or the false discovery rate correction for multiple comparisons, we found that neither of the transparency measures entered into a statistically significant interaction with either of the frequency measures. We thus failed to replicate in English the findings of Marelli and Luzzatti's (2012) investigation of Italian nominal compounds.

The analysis to which we now turn goes beyond the aggregation of individual reading performance. In the next analysis, we examine the effect of semantic transparency as modulated by reading experience.

Analysis 2: Interactions of Semantic Transparency With Reading Experience

We hypothesized in the introduction to the article that reading experience plays an important role in compound processing, and especially in the individual ability to use orthographic cues toward discriminating between subtle differences in meaning. The chosen diagnostic for this role was whether the scores of individual Print Exposure, Vocabulary Size, or Word Recognition Skill entered into statistically significant interactions with either Modifier-Compound or Head-Compound semantic transparency measures.

We illustrate our analytical procedure by example of the interaction of semantic transparency with Print Exposure. For each eye-movement variable listed in Part A of Table 1, we fitted a model which included an interaction of Print Exposure with (i) Modifier-Compound semantic transparency and, (ii) Head-Compound semantic transparency. As in Analysis 1, this procedure resulted in seven different models (one for each eye-movement measure). Along with each transparency measure by Print Exposure interaction, we also included in each model the set of predictors listed in Part B of Table 1 as fixed effects in our models. For number of trials per dependent variable, see Part A of Table 1. Random effects were included as described in Statistical considerations. We report effect sizes as the predicted amount of change in the dependent variable estimated for the contrast between the 10th and the 90th percentile of the critical semantic transparency variable for participants with a low and a high value of Print Exposure. The values were chosen to represent extremes (the 10th percentile and the 90th percentile) of the Print Exposure range. Where necessary, dependent variables were back-transformed into milliseconds or percentages.

Interactions of semantic transparency with print exposure. For Print Exposure, two significant interaction effects were detected after the application of the false discovery rate criterion (we report adjusted *p* values for models that pass this criterion). Head-Compound Semantic Transparency entered into a significant interaction with the Print Exposure on regression-in probability [$\hat{\beta} = -0.09$; $SE = 0.04$; $p = .01$] and on total viewing time [$\hat{\beta} = -0.02$; $SE = 0.01$; $p = .02$]. Tables S5.7–8 in the online supplemental materials provide a full specification of these models.

Figure 3 depicts the partial effects of Head-Compound semantic transparency on regression-in probability and total viewing time, broken down by 10th, 30th, 50th, 70th, and 90th percentiles of Print Exposure. The critical interactive effect of Head-Compound transparency and Print Exposure on regressive saccades demonstrates that readers with greater exposure to printed materials were less likely to execute a regressive saccade onto a compound word after the first pass of reading that is semantically transparent i.e. when the individual meaning of the right constituent is similar to the compound word in which it is embedded (e.g., *storm* in *rainstorm*). Specifically, individuals in the 90th percentile of the distribution of Print Exposure (cf. the negative longdash slope denoting the highest Print Exposure value in Figure 3), the effect size range (see above for definition) indicates an overall 8.34% decrease in the execution of regressive saccades. However, the opposite outcomes were true for readers with relatively less exposure to printed materials. The interaction pattern indicates that the facilitatory effect of transparency progressively weakens as exposure to printed materials decreases in the population sample. For participants with relatively low Print Exposure scores (i.e., those in the bottom 10th percentile), compounds with higher Head-Compound transparency elicited a higher probability of performing a regressive saccade as compared to opaque compounds. Specifically, the effect size range indicates that for these individuals (cf. the solid positive slope denoting the lowest Print Exposure value in Figure 3), there was an overall 7.86% increase in regressive saccades on the compound.

The interaction between Head-Compound transparency and Print Exposure in total viewing time demonstrates a similar pattern in that the difference in total viewing time between readers with relatively small and large quantities of exposure to printed materials was maximal for highly transparent compounds. Here the interaction effect indicates that readers with more exposure to printed materials were facilitated by semantic transparency (i.e., shorter reading durations). The processing advantage associated with greater transparency was restricted to individuals in the top 50% of the Print Exposure range (the dotted, dotdash and longdash lines in Figure 3 denoting the 50th, 70th and 90th percentiles of the Print Exposure range) and the facilitation effect becomes progressively stronger with increases in the sample population's exposure to printed materials. The effect size for individuals in the 90th percentile of the Print Exposure range demonstrated a total recognition boost of 27 ms. However, as is visualized in Figure 3 and shown in Table 3, this facilitative effect progressively attenuates

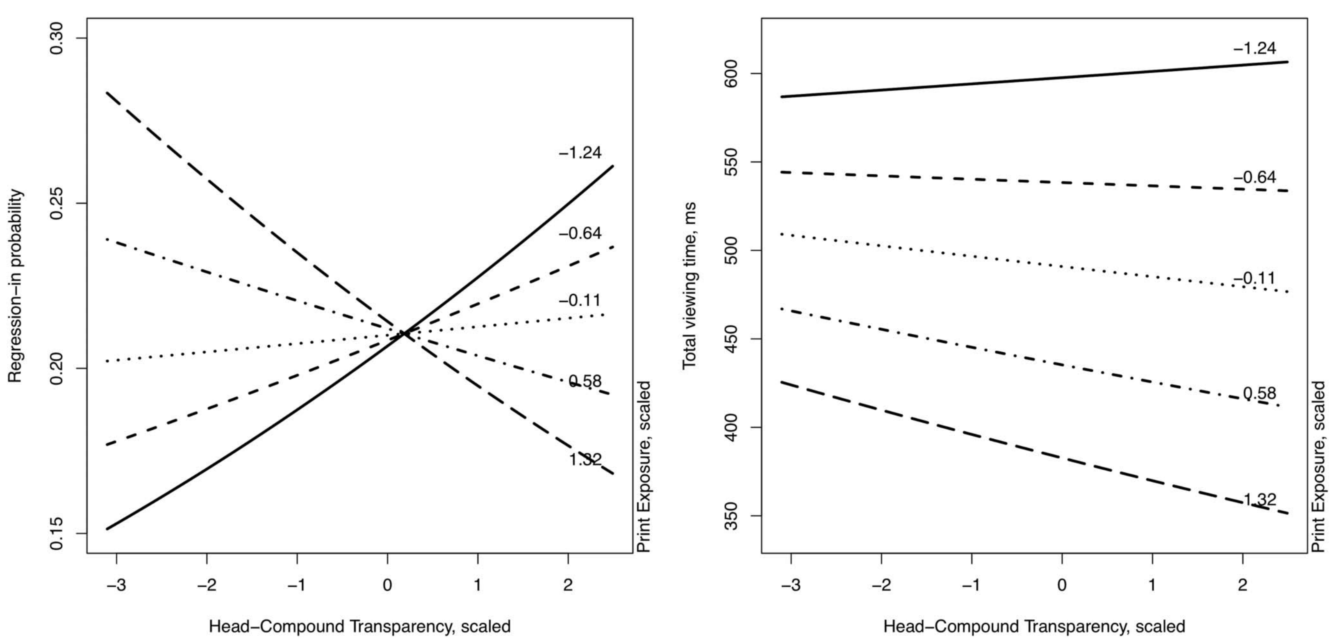

Figure 3. Interaction effects between semantic transparency and reading experience. Left panel: partial effects of Head-Compound Semantic Transparency Rating (scaled) modulated by Print Exposure, on regression-in probability. Right panel: partial effects of Head-Compound Semantic Transparency Rating (scaled) modulated by Print Exposure on total viewing time. Scaled values of the Print Exposure are provided in the right margin and report, where available, the 10th (solid line), 30th (dashed line), 50th (dotted line), 70th (dotdash line), and 90th (longdash line) percentiles of each test score.

only until it is "levelled out" among individuals in the 30th percentile of the range of Print Exposure scores (cf. the dashed lines), and constitutes an inhibitory effect for readers in the bottom 10th percentile of Print Exposure scores (cf. the solid line). For these readers, the inhibitory effect size demonstrated a total size of 26 ms. In sum, for both regression-in probability and total viewing time, both extremes of the Print Exposure score range demonstrated counterdirected, but near equivalent effect size ranges. The plots of raw trends provided in Supplementary materials S4 support the regression models' predictions and aid interpretability of interactions.

Interactions of semantic transparency with vocabulary size and word recognition skill. We repeated our analytical procedure with Vocabulary Size and, separately, with Word Recognition Skill, as two further proxies of individual reading experience. Two significant interactions of Head-Compound transparency by Vocabulary Size were found, and were qualitatively identical to the interactions between Head-Compound transparency and Print Exposure. These were identified as statistically reliable (after the correction for multiple comparisons) in models fitted to the same dependent eye-movement measures (regression-in probability and total viewing time). Moreover, the interactions showed the same fanning pattern, in which participants with large vocabularies were facilitated by increased similarity between the head and compound while those with smaller vocabularies were impeded by it. Tables S5.9–10 report statistical models for these effects. Individual performance in the TOWRE word identification test did not interact reliably with either measure of semantic transparency under consideration. As in Analysis 1, we tested three-way interactions between transparency, compound frequency, and each individual differences measure. None reached significance in our data.

Given the strong correlation between Print Exposure and Vocabulary size scores, we cannot establish in the present cohort whether and which one of these tests primarily drives the interactive effect of transparency on reading behavior. We propose, however, that this distinction is immaterial for our hypotheses, because both the amount of exposure to printed materials and the estimated number of words in one's mental lexicon are theoretically and empirically intertwined facets of an individual's reading experience, habits and behavior (see reviews in Mol & Bus, 2011; Moore & Gordon, 2015). Both facets demonstrably correlate with one's ability to discriminate between (often subtle) differences in meanings associated with compounds and their morphemes and thus impact the ability to overcome the 'discrimination challenge' posed by transparent compounds.

We confirm the hypothesized functional relationship between individual reading experience and the effect of semantic transparency on eye movements. For interactions between transparency and reading experience on regression-in probability, the qualitative pattern of the interaction accords with the hypothesized processing trade-off pattern illustrated in Pattern D of Figure 1. This pattern indicates the existence of joint effects of semantic facilitation ("semantic boosting") and meaning discrimination during the processing of transparent compounds. For individuals with greatest reading experience, the effect of semantic facilitation is unimpeded by the inhibitory force of semantic discrimination. However, with less reading experience the processing benefit of greater semantic transparency is gradually counteracted by the processing cost of semantic discrimination. Thus, while the greater discrimination cost counteracts the benefit from the semantic boost in less experienced individuals, that benefit is still present in more experienced readers. The implications of these two trade-off patterns and their apparent association with separate semantic transparency measures will be addressed in the General Discussion.

In summary, our results are compatible with the prediction that reading experience, as gauged by Print Exposure and Vocabulary Size, modulates the processing of semantic transparency. On the whole, these patterns suggest a trade-off between the effect of conjunctive semantic activation and the additive effect of meaning discrimination during the processing of transparent compounds.

General Discussion

This article addresses the question of how human minds are able to handle many concurrently activated meanings when there is variability in the degree to which those meanings are similar. Historically, the effect of semantic transparency (defined as similarity between meanings of compounds and their constituent morphemes) has produced mixed results during compound word recognition in lexical decision and sentence reading tasks. At the outset, we argued that the inconsistency in findings may be because of at least two factors. One is the incompleteness of theoretical approaches to semantic processing in complex words. A dominant theoretical family of models that we label here "semantic boosting" invariably predicts a facilitatory effect of high semantic overlap between constituents and the whole word that one finds in relatively transparent compounds. Under this account, a processing advantage for transparent compounds arises from an easier composition of the compound meaning via the related meanings of its constituents, or the conjunctive activation of strongly associated semantic representations within a lexical network.

Crucially, the aforementioned family of accounts overlooks the dimension of discriminability, that is, the effort that is required to dissociate a target meaning of a compound from all other meanings (including those of its embedded constituents) that are activated by largely overlapping orthographic cues. The NDR model (Baayen et al., 2011) formalizes the process of discrimination as follows. In the case of compound word recognition, the reader is confronted with a meaningful orthographic string which is composed of at least two substrings that each denote distinct meanings of their own. For example, when reading a highly transparent compound word such as *rainstorm*, the reader must be able to discriminate between the network of events and entities that have been previously associated with the words *rain*, *storm* and *rainstorm*. These conditions make it difficult for the cognitive system to minimize the uncertainty between the outcomes associated with each word in the above pairs of relations and extract the true meaning outcome, that is, the event or entity associated with "a storm with heavy rain." At the opposite end of the transparency spectrum, opaque compounds, such as *hogwash*, are predicted to impose relatively little meaning discrimination cost (and no semantic boost). This is because the constituents *hog* and *wash* are not learned in similar lexical contexts as the whole word and so do not activate similar meanings.

The "semantic boost" account and the NDR model make opposing predictions as to the direction of the semantic transparency effect. Under the semantic boost account, a compound that encodes semantically similar meanings is expected to elicit an increased coactivation of associated lexical representations, leading to an advantage in processing. However, under the NDR account, the same compound would trigger an increased cognitive load because of the requirement of discriminating between closely related meanings. In Figure 1, we formulated the entire space of theoretical possibilities for the outcome that could derive from either one account exclusively, or from their superposition.

In our theorizing, we also addressed a second factor that might give rise to the inconsistent findings in the literature, namely, the largely overlooked role of individual variability in reading experience and skill in modulating semantic processing of complex words. In line with the learning principles that underlie NDR, we argued that tests serving as proxies of one's reading experience influence cognitive behavior in such a way that more experienced readers will be more proficient in resolving the discriminability challenge of transparent compound word processing. We showed that the effects of written language experience are most likely to be demonstrated in the community population that we examined, where exposure to printed materials is heterogeneous, compared with a convenience participant pool of university students (Figure 2).

Based on the reported results, we outline three empirical contributions that this article offers and then address each of the findings in turn:

1. Semantic transparency affects compound recognition, and does so consistently and early in the eye-movement record.

2. Individual reading experience modulates the effect of semantic transparency during the later stages of visual identification of compound words (e.g., regressions to the compound and in total viewing time). For the Head-Compound semantic transparency relation, exposure to written language and vocabulary size affect one's ability to discriminate between meanings during compound word reading. This finding demonstrates that semantic facilitation is present for all readers, while for the most inexperienced readers such a facilitative boost is offset by a meaning discrimination cost. This processing cost and its link with inexperience with written language is predicted by the Naive Discriminative Reader account of word recognition (Baayen et al., 2011).

3. The Head-Compound semantic transparency relation (e.g., the meaning similarity between *tub* and *bathtub*), and the Modifier-Compound relation (e.g., the similarity between *bath* and *bathtub*) are differentially sensitive to effects of reading experience during compound word reading.

Semantic Transparency and the time-course of Its Effect During Compound Reading

First, as is reported in Analysis 1, we found main effects of semantic transparency for both the Modifier-Compound relation (e.g., the semantic similarity between *spear* and *spearmint*) and the Head-Compound relation (e.g., the semantic similarity between *mint* and *spearmint*). The effects demonstrate a processing advantage when reading compounds for which semantic transparency is relatively high. These effects are compatible with previous studies that have also found main effects of semantic transparency on visual word recognition (e.g., Ji et al., 2011; Juhasz, 2007; Juhasz et al., 2015; Libben et al., 2003; Zwitserlood, 1994). We align our interpretation of effects with studies which propose that the semantic processing of compound words during visual word recognition is facilitated by the conjunctive activation (i.e., "semantic boosting") of similar whole-word and morpheme meanings (e.g., El-Bialy et al., 2013; Libben, 1998).

Our findings also pertain to the question of the time-course of semantic access during visual processing of complex word forms. The earliest durational effect occurs for the Modifier-Compound semantic transparency relation on first-of-many fixation duration. Given that the mean first-of-many fixation duration is 228 ms and that 50% of first-of-many fixations fall below 221 ms (min = 53 ms, max = 1,154 ms), it is possible to conclude that semantic effects of compound word recognition appear early in the time-line of compound processing, and do so simultaneously with the processing of their orthographic forms. This claim is supported by the average time-point estimates of semantic effects on first-of-many-fixation durations in previous studies on complex word processing (of words in context and in isolation), for example, Amenta, Marelli, and Crepaldi (2015), 243 ms; Marelli and Luzzatti (2012), 231 ms; Marelli, Amenta, Morone, and Crepaldi (2013; Experiment 1), 256 ms. In summary, the effect of transparency on first-of-many fixation duration indicates that during the initial fixation on the compound word, the cognitive system is already beginning to process the meaning similarity of the the compound and its modifier.

We also found effects of Modifier-Head and Head-Compound semantic transparency on refixation probability (71.6% of trials elicited more than one fixation on the compound target in the first pass of reading). Refixation probability is an eye-movement measure which quantifies a decision that takes place during the first fixation on a word, and is argued by Dimigen, Sommer, Hohlfeld, Jacobs, and Kliegl (2011) to be a decision that is made before the refixating saccade is planned and enters a nonlabile execution stage, that is, 80 ms before the fixation ends (Becker, 1991; Findlay & Harris, 1984). This confines the temporal locus of the transparency effect to an average of $228 - 80 = 148$ ms. This is remarkable given that the first fixation on compounds is often insufficient to provide a full foveal view of the word because of visual acuity constraints (see, e.g., Bertram, 2011; Bertram & Hyönä, 2003; Hyönä, 2012): compound length in this experiment ranged from 6 to 13 characters. This finding indicates that processing of semantic transparency is initiated prior to the visual inspection of the entire compound, a processing phase which must logically precede decomposition of the compound into its constituent morphemes (for prior discussion, see Kuperman, Bertram, & Baayen, 2008; Kuperman, Schreuder, Bertram, & Baayen, 2009).

In sum, the early effects of semantic transparency reported in the present article run counter to the influential account of semantics-blind obligatory morphological decomposition (e.g., Rastle & Davis, 2008; Taft, 2004) and support accounts that advocate parallel and simultaneous access to both form *and* meaning of the entire complex word and its morphemes (among others, Baayen, Dijkstra, & Schreuder, 1997; Kuperman et al., 2008).

Individual Reading Experience Modulates the Semantic Transparency Effect

The second insight from the data is that the effect of Head-Compound semantic transparency was modulated by individual exposure to printed materials and vocabulary size and that these interactions were present in later eye-movement measures (regression probability and total viewing time). Both measures of individual differences (Print Exposure and Vocabulary Size) produced a qualitatively identical pattern of a gradient processing outcome. For individuals scoring in the lowest 10th percentile of the Print Exposure assessment, greater semantic transparency between the head and the compound (e.g., *tub* and *bathtub*) served to (a) increase the probability of regressive saccades and (b) slow down viewing times. However, the same lexical property of high semantic transparency yielded a reverse effect in more experienced readers, such that as print exposure gradually increases, greater semantic transparency of compounds was associated with (a) a decrease in the number of regressive saccades and (b) shorter total viewing time fixation durations. Thus, readers with impoverished reading experience exhibited inhibited processing (with equivalent processing for those in the 30th percentile) when reading highly transparent compounds compared with opaque compounds. For readers with relatively large amounts of reading experience, greater Head-Compound transparency facilitated processing throughout. As argued earlier, Print Exposure and Vocabulary Size stand in a reciprocal relationship in the development of reading skill, and are both theoretically and empirically intertwined (correlation $r = .7$). In this article, we interpret both as proxies for different facets of reading experience (see Stanovich & West, 1989; Mol & Bus, 2011; Moore & Gordon, 2015) and—because our hypothesis is couched in terms of generic reading experience—do not pursue a question of whether one facet is primary or dominant compared with another.

As we argue in the introduction, neither the "semantic boost" account (e.g., Libben, 2005) nor the discrimination learning account (Baayen et al., 2011) can give rise to the observed pattern on their own, as the former model imposes no penalty for handling more or less overlapping meanings (regardless of individual exposure to language), while the latter only predicts a bigger or smaller cost (and no benefit) for resolving the discrimination challenge. Rather, the pattern of results can only be explained by the simultaneous operation of *both* processing mechanisms. Specifically, the trade-off patterns that we observe in Figure 3 match the theoretical prediction illustrated in Panel D of Figure 1. Under this scenario, the pattern observed for transparent compounds can only be true if, for readers with impoverished language skill, the facilitatory boost caused by conjunctive semantic activation is weaker than the inflation in processing time associated with meaning discrimination.

The above point begs the question of how exactly the cost of conceptual discrimination might trade-off with facilitation from conjunctive semantic activation. To revert back to the naive discriminative learning perspective, it is proposed that the strengths of the associations between the mapping of form and meaning that learners acquire from experience are in constant flux. Thus, we argue that our results support the notion that the uncertainty brought about by discriminating between two similar meanings is subject to modification within the language user, and that these modifications are driven by the accumulation of language experience over time (MacDonald, 2013; Ramscar, Hendrix, Shaoul, Milin, & Baayen, 2014). This interpretation is supported by the phenomenon that we have observed empirically: both Print Exposure and Vocabulary Size are reasonable indices of language experience, which is a central component of the NDR account.

Notably, the restriction of interactive effects to regression probability and total viewing time suggests that the modulation of the transparency effect by experience happens during later stages of compound processing, for example, its integration with the semantic representation of the sentence. We speculate that later phases of meaning integration are the most likely time for the challenge of semantic discrimination to bear importance for compound word processing and for making subtler distinctions between the compound meaning and the meaning of its head (see further discussion below on the role of the morphological head of English compound words).

An effect of inhibition during the processing of closely related concepts is not new to the literature. For example, studies employing the visual world eye-tracking paradigm have demonstrated that listeners fixate on images of semantically associated objects more than they do for semantically unrelated entities when simultaneously presented with a spoken target word (Huettig & Altmann, 2005; Huettig & McQueen, 2007; Huettig, Quinlan, McDonald, & Altmann, 2006; Mirman & Magnuson, 2009). In particular, Mirman and Magnuson (2009) reported that refixation likelihood was increased among presentations of words defined as near semantic neighbors compared to words that are distant semantic neighbors (semantic neighborhood density calculated over counts of semantic feature ratings of words). In addition, in a lexical decision study, Mirman and Magnuson (2008) reported slowed processing of words with near neighbors and speeded processing to words with distant semantic neighbors. Although methodologically distinct from the current study, these studies report results that are analogous to the effects we observe for less experienced readers when reading highly transparent compounds.

In summary, the present results demonstrate evidence for the role of individual differences in reading experience on the processing of semantic transparency during compound word reading. We also argue that the presented interactive effects may not be detected in a homogeneous population of readers in which the range of reading experience is narrow. In the present experiment, as is visualized in Figure 2, we explored the influence of reading experience in a population of readers from the local community, where there is a broad range of reading experience. We therefore conclude that reading experience is an important component factor of individual reading skill which modulates the effect of semantic transparency during reading, and that differences in strategies of semantic processing of compound words may be more easily uncovered when tested in a heterogeneous sample of the reading population.

Differential Effects of Semantic Transparency

Third, we observed that the exact nature of the trade-off between the cost of conceptual discrimination and facilitation from conjunctive semantic activation is contingent on the particular type of transparency relation. We observed that a higher degree of semantic similarity between the modifier and the compound facilitated the visual recognition of compound words for *all* individuals across the entire continuum of reading experience. That is, we observed main effects of Modifier-Compound semantic transparency on six out of seven eye-movement measures. By the same token, all readers, regardless of experience, showed facilitation because of increased Head-Compound similarity in two first-pass eye-movement measures (refixation probability and gaze duration).

Contrary to the proliferation of main effects for Modifier-Compound transparency, individual variability in reading experience only modulated compound recognition effort in relation to Head-Compound transparency, and only in late behavioral measures (regression probability and total viewing time; see our interpretation of the late effects). That is, less experienced readers were only sensitive to the inhibitory force of the discrimination challenge—along with the more ubiquitous facilitatory force of a semantic boost—when processing compounds with a high semantic similarity between the head of the compound and the whole compound word.

The specificity of the interaction may point to an important, distinct function of the compound head during semantic processing. From a structural perspective, the head of an English compound is the conceptual kernel of the whole word. We speculate that the strength of the semantic similarity between the concept(s) denoted by the whole compound word and that of the same compound's head is likely to pose a more acute meaning discrimination problem during reading and elicit differential processing efforts in individuals varying in their ability to discriminate between meanings. To sum up, our results do not only enable us to posit the simultaneous operation of two cognitive mechanisms involved in the processing of semantic transparency—a semantic boost and a discrimination challenge, but also identify their relative strengths across the range of individual variability, time-course of compound processing and type of transparency measure.

Future investigations are required to further track the influence of reading experience on the reader's sensitivity to compound transparency. For example, a future inquiry may be able to incorporate transparency measured via DSM techniques that are computed over texts which are more characteristic of the lexicons of individuals with varying levels of reading experience. This approach will shed more light on the role of reading experience and compound word processing, and will also expand upon the work by Gagné et al. (2016) on the conceptualization (e.g., substitutability vs. meaning relatedness) and operationalization of compound semantic transparency. In addition, this research has used reading experience as a proxy of meaning discrimination skill. Even though the NDR model predicts that discrimination improves with reading experience, we did not independently nor directly test individual skill differences in semantic discrimination skill. Future work will need to examine this individual skill more finely and explore its relationship with reading experience. Finally, the structure of the English compound is such that the grammatical head of a noun-noun compound word is also systematically the rightmost constituent in English. Therefore, the present experiment is not able to disentangle headedness effects from positional effects (but see Marelli & Luzzatti, 2012). Thus, experimentation in other languages is necessary to establish whether the specific sensitivity to the discrimination challenge is tied to the position or the function (head) of a compound's constituent.

In summary, the present study reveals that the particular nature of processing an entanglement of semantic representations during compound word recognition can be predicted and interpreted in the light of a detailed understanding of mechanisms underlying both word learning, morphological processing and individual variation in reading skill.

References

Acheson, D. J., Wells, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods, 40,* 278–289.

Amenta, S., & Crepaldi, D. (2012). Morphological processing as we know it: An analytical review of morphological effects in visual word identification. *Frontiers in Psychology, 3,* 232.

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,* 1587–1596.

Andrews, S. (2012). Individual differences in skilled visual word recognition and reading. In J. Adelman (Ed.), *Visual word recognition Vol. 2: Meaning and context, individuals and development* (pp. 151–172). Sussex, United Kingdom: Psychology Press.

Andrews, S., & Lo, S. (2013). Is morphological priming stronger for transparent than opaque words? It depends on individual differences in spelling and vocabulary. *Journal of Memory and Language, 68,* 279–296.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* New York, NY: Cambridge University Press.

Baayen, R. H. (2014). Experimental and psycholinguistic approaches to studying derivation. In R. Lieber & P. Stekauer (Eds.), *Handbook of derivational morphology* (pp. 95–117). New York, NY: Oxford University Press.

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

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language, 37,* 94–117.

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review, 118,* 438–481.

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1995). *The CELEX database* [CD-ROM]. Nijmegen, the Netherlands: Center for Lexical Information, Max Planck Institute for Psycholinguistics.

Baayen, R. H., Wurm, L. H., & Aycock, J. (2007). Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. *The Mental Lexicon, 2,* 419–463.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). lme4: Linear mixed-effects models using Eigen and S4 (R package version 1.0–4) [Computer software]. Retrieved June 1, 2013 from https://cran.r-project.org/web/packages/lme4/index.html

Becker, W. (1991). Saccades. *Eye Movements, 8,* 95–137.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57,* 289–300.

Bertram, R. (2011). Eye movements and morphological processing in reading. *The Mental Lexicon, 6,* 83–109.

Bertram, R., & Hyönä, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of Memory and Language, 48,* 615–634.

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

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

De Jong, N. H. (2002). *Morphological families in the mental lexicon* (Unpublished doctoral dissertation). University of Nijmegen, Nijmegen, the Netherlands.

Diependaele, K., Grainger, J., & Sandra, D. (2012). Derivational morphology and skilled reading: An empirical overview. In M. Spivey, K. McRae, & M. Joanisse (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 311–332). New York, NY: Cambridge University Press.

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,* 552–572.

Dunn, L. M., & Dunn, L. M. (1981). *Manual for Peabody Picture Vocabulary Test–Revised: PPVT-R*. Circle Pines: MI: American Guidance Service.

El-Bialy, R., Gagné, C. L., & Spalding, T. L. (2013). Processing of English compounds is sensitive to the constituents' semantic transparency. *The Mental Lexicon, 8,* 75–95.

Falkauskas, K., & Kuperman, V. (2015). When experience meets language statistics: Individual variability in processing English compound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41,* 1607–1627.

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,* 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., & Fund-Reznicek, E. (2009). Masked morphological priming of compound constituents. *The Mental Lexicon, 4,* 159–193.

Frisson, S., Niswander-Klement, E., & Pollatsek, A. (2008). The role of semantic transparency in the processing of English compound words. *British Journal of Psychology, 99,* 87–107.

Fruchter, J., & Marantz, A. (2015). Decomposition, lookup, and recombination: MEG evidence for the full decomposition model of complex visual word recognition. *Brain & Language, 143,* 81–96.

Gagné, C. L., & Spalding, T. L. (2009). Constituent integration during the processing of compound words: Does it involve the use of relational structures? *Journal of Memory and Language, 60,* 20–35.

Gagné, C. L., Spalding, T. L., & Nisbet, K. A. (2016). Processing English compounds: Investigating semantic transparency. *SKASE Journal of Theoretical Linguistics, 13,* 2–22.

Huettig, F., & Altmann, G. T. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition, 96,* B23–B32.

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language, 57,* 460–482.

Huettig, F., Quinlan, P. T., McDonald, S. A., & Altmann, G. T. (2006). Models of high-dimensional semantic space predict language-mediated eye movements in the visual world. *Acta Psychologica, 121,* 65–80.

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

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434–446.

Ji, H., Gagné, C. L., & Spalding, T. L. (2011). Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory and Language, 65,* 406–430.

Juhasz, B. J. (2007). The influence of semantic transparency on eye movements during English compound word recognition. In R. P. van Gompel, M. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 373–389). Amsterdam, the Netherlands: Elsevier.

Juhasz, B. J. (in press). Experience with compound words influences their processing: An eye movement investigation with English compound words. *The Quarterly Journal of Experimental Psychology*.

Juhasz, B. J., Lai, Y.-H., & Woodcock, M. L. (2015). A database of 629 English compound words: Ratings of familiarity, lexeme meaning dominance, semantic transparency, age of acquisition, imageability, and sensory experience. *Behavior Research Methods, 47,* 1004–1019.

Kuperman, V. (2013). Accentuate the positive: Semantic access in English compounds. *Frontiers in Psychology, 4,* 1–10. http://dx.doi.org/10.3389/fpsyg.2013.00203

Kuperman, V., & Bertram, R. (2013). Moving spaces: Spelling alternation in English noun-noun compounds. *Language and Cognitive Processes, 28,* 939–966.

Kuperman, V., Bertram, R., & Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes, 23,* 1089–1132.

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,* 876–895.

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

Kuperman, V., & Van Dyke, J. A. (2011). Individual differences in visual comprehension of morphological complexity. In *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 1643–1648). Austin, TX: Cognitive Science Society.

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,* 211.

Libben, G. (1998). Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment. *Brain & Language, 61,* 30–44.

Libben, G. (2005). Everything is psycholinguistics: Material and methodological considerations in the study of compound processing. *Canadian Journal of Linguistics/Revue Canadienne De Linguistique, 50*(1–4), 267–283.

Libben, G. (2014). The nature of compounds: A psychocentric perspective. *Cognitive Neuropsychology, 31,* 8–25.

Libben, G., Gibson, M., Yoon, Y. B., & Sandra, D. (2003). Compound fracture: The role of semantic transparency and morphological headedness. *Brain & Language, 84,* 50–64.

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences, 4,* 6–14.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology, 4,* 226.

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,* 534–541.

Marelli, M., Dinu, G., Zamparelli, R., & Baroni, M. (2015). Picking buttercups and eating butter cups: Spelling alternations, semantic relatedness, and their consequences for compound processing. *Applied Psycholinguistics, 36,* 1421–1439.

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,* 644–664.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods, 44,* 1–23.

Mirman, D., & Magnuson, J. S. (2008). Attractor dynamics and semantic neighborhood density: Processing is slowed by near neighbors and speeded by distant neighbors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 65–79.

Mirman, D., & Magnuson, J. S. (2009). Dynamics of activation of semantically similar concepts during spoken word recognition. *Memory & Cognition, 37,* 1026–1039.

Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin, 137,* 267.

Moscoso del Prado Martín, F., Deutsch, A., Frost, R., Schreuder, R., De Jong, N. H., & Baayen, R. H. (2005). Changing places: A cross-language perspective on frequency and family size in Dutch and Hebrew. *Journal of Memory and Language, 53,* 496–512.

Moore, M., & Gordon, P. C. (2015). Reading ability and print exposure: Item response theory analysis of the author recognition test. *Behavior Research Methods, 47,* 1095–1109.

Pham, H., & Baayen, R. H. (2013). Semantic relations and compound transparency: A regression study in CARIN theory. *Psihologija, 46,* 455–478.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. New York, NY: Springer.

Pollatsek, A., & Hyönä, J. (2005). The role of semantic transparency in the processing of finnish compound words. *Language and Cognitive Processes, 20,* 261–290.

Radach, R., & Kennedy, A. (2013). Eye movements in reading: Some theoretical context. *The Quarterly Journal of Experimental Psychology, 66,* 429–452.

Ramscar, M., & Baayen, H. (2013). Production, comprehension, and synthesis: A communicative perspective on language. *Frontiers in Psychology, 4,* 233. http://dx.doi.org/10.3389/fpsyg.2013.00233

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science, 6,* 5–42.

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science, 34,* 909–957.

Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes, 23*(7–8), 942–971.

Rastle, K., Davis, M. H., Marslen-Wilson, W. D., & Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes, 15,* 507–537.

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

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

Rayner, K., Pollatsek, A., Ashby, J., & Clifton Jr., C. (2012). *Psychology of reading*. New York, NY: Psychology Press.

R Core Team. (2014). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org/

Sandra, D. (1990). On the representation and processing of compound words: Automatic access to constituent morphemes does not occur. *The Quarterly Journal of Experimental Psychology, 42,* 529–567.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21,* 360–407.

Stanovich, K. E., & Cunningham, A. E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory & Cognition, 20,* 51–68.

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly, 24,* 402–433.

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

Torgesen, J. K., Wagner, R., & Rashotte, C. (1999). *Towre–2 test of word reading efficiency*. Austin, TX: Pro-Ed.

Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research, 37,* 141–188.

Underwood, G., Petley, K., & Clews, S. (1990). Searching for information during sentence comprehension. In R. Groner, G. d'Ydewalle, & R. Parham (Eds.), *From eye to mind: Information acquisition in perception* (pp. 191–203). Amsterdam, the Netherlands: Elsevier Science.

von der Malsburg, T., & Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of Memory and Language, 94,* 119–133.

Wang, H.-C., Hsu, L.-C., Tien, Y.-M., & Pomplun, M. (2012). Estimating semantic transparency of constituents of English compounds and two-character Chinese words using Latent Semantic Analysis. In *Proceedings of CogSci* (pp. 2499–2504). Austin, TX: Cognitive Science Society.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI)*. San Antonio, TX: Psychological Corporation.

Zwitserlood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes, 9,* 341–368.

Received February 22, 2017
Revision received May 8, 2017
Accepted May 15, 2017 ■