SEMANTIC PROCESSING OF MORPHOLOGICALLY COMPLEX WORDS

SEMANTIC PROCESSING OF MORPHOLOGICALLY COMPLEX WORDS: EXPERIMENTAL STUDIES IN VISUAL WORD RECOGNITION

BY
DANIEL SCHMIDTKE, M.Sc., B.A. (Hons)

A THESIS
SUBMITTED TO THE DEPARTMENT OF LINGUISTICS & LANGUAGES
AND THE SCHOOL OF GRADUATE STUDIES
OF MCMASTER UNIVERSITY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

McMaster University

Doctor of Philosophy (2016)
Linguistics & Languages

McMaster University
Hamilton, Ontario, Canada

TITLE:	Semantic Processing of Morphologically Complex Words: Experimental Studies in Visual Word Recognition
AUTHOR:	Daniel Schmidtke M.Sc., University of Edinburgh B.A. (Hons), York St. John University
SUPERVISOR:	Dr. Victor Kuperman
NUMBER OF PAGES:	vi, 210

To my parents,

And for Oma,

Acknowledgements

This thesis would not have been possible without the efforts of many kind-hearted souls. The faults are still all mine, but so many parts of this thesis have been intentionally or unintentionally developed, inspired, or sparked by clever ideas or practical assistance from folk that I feel I simply must mention by name. I would like to take a moment to express my gratitude towards these people, who I hope will continue to be a part of my academic and personal life for many years to come.

First and foremost I want to thank my supervisor Victor Kuperman. Victor took me on as his Ph.D. student without us having met face to face. Because of this, I am immensely appreciative of how my ~5500 km relocation to Canada was probably as much a leap of faith for him as it was for me. Over the years, Victor's guidance, teaching and creativity has been an immense source of inspiration. Right from the beginning he devoted many hours teaching me many of the practicalities of the psycholinguistic trade. I will never forget the countless hours we spent together in his office as he taught me the nuts and bolts of eye-movement research, psycholinguistic theory, experimental design, statistics, data analysis, coding, and many other valuable lessons.

I appreciate all of Victor's contributions of time, ideas, and funding to make my Ph.D. experience productive and stimulating. There were countless occasions (throughout all times of day and across many timezones) where Victor would abandon another project to share in the delight of a new finding, challenge my thinking, or help me think through a new idea. There was never a period of intellectual inertia during my time as a student, and I am grateful to Victor for allowing me the freedom to sate my desire to carry out many other projects that did not make it into this thesis. Most of all, the joy and enthusiasm he has for his research was contagious and motivational for me, even during the tough times when I doubted my own abilities. Thank you, Victor, for this life-changing opportunity, and for inspiring me to squeeze out as many of the juices of Ph.D. life as I could.

Apart from his academic guidance, I would also like to thank Victor for his personal support and friendship. Whether it was taking a long drive through the US to a conference, calming my pre-talk nerves, or playing music and singing together, I will always be grateful for guiding me through the past four years with so much bonhomie, kind-heartedness, and good humour. I also wish to express my thanks to Katya, Boris, and Joseph for allowing me to take up generous portions of Victor's time, and for providing great entertainment and delicious food on so many occasions.

I would also like to thank my supervisory committee members, Anna Moro, Lili Service, and Julie Van Dyke for their guidance and constructive feedback. I very much appreciate Anna and Lili for treating me as a colleague, I suspect, long before I had produced anything to warrant such a designation. It is thanks to Julie and her lab at Haskins Laboratories that I had such a rich and intriguing dataset to work with from the beginning. I am also thankful to Julie for her careful editing of my manuscripts and for enhancing my work with fresh insights.

I wish to thank past and present members of the Reading Lab who contributed so much to my graduate experience. More specifically, I want to thank Gina Henry for providing me with huge amounts of guidance and help when I first started in the lab. Thank you also to my brother in arms, Bryor Snefjella. Bryor, I feel privileged to have had you as my office mate for the past four years. Thank you for helping me through so many problems (both academic and personal), for allowing so much time for stimulating discussion, sharing in my love for self-deprecating humour and for the borderline psychotic angry rants. Your company, friendship and penetrating insight made this past four years bearable.

Thank you also to Kazunaga Matsuki (not least, may I add, because of his deft abilities in ploughing through unsuspecting ungulates on the Hawaiian freeway), Amy-Beth Warriner, Connie Imbault and Kaitlin Falkauskas for bravely sticking through and providing lots of constructive feedback on my practice conference talks. Of course, I would not have had any experiments to run or data to analyze if it were not for the small army of research assistants (past and present) in the Reading Lab who were always ready and willing to help out with my projects. I would like to express my gratitude to Noor Al-Zanoon, Emma Bridgwater, Irena Grusecki, Brooke Osborne, Sadaf Rahmanian, Katrina Reyes, Aaron So, and Chloe Sukkau. Thank you all!

During the past four years I have been fortunate enough to collaborate with many incredible and inspiring researchers. I am thankful to Christina Gagné and Tom Spalding for the collaboration which materialized as a chapter of this thesis. Thank you, Tom and Christina, for bringing so much theoretical expertise to our projects, and for your willingness to always share with me your time, data, and deep insight into the topic of conceptual structure of compound words. I can't wait to start my post-doctoral research programme with you and your lab in Edmonton.

I am grateful also to Harald Baayen who sponsored a three week stay at his Quantitative Linguistics research unit at the University of Tübingen. My visit to Harald's lab came with only a few months left before I needed to submit this thesis, yet my time working with my 'academic grandfather' pushed this work and future projects forward in so many promising ways. Thank you, Harald, I learned so much from our discussions and intense data analysis sessions. It was a privilege and an honour to be able to work alongside someone with such vast amounts of knowledge on the psycholinguistic field and seemingly endless supplies of energy and enthusiasm. Whilst at Tübingen, I also enjoyed many stimulating and fruitful discussions on survival analysis methods with Tino Sering, Petar Milin, and Peter Hendrix. I would also like to thank Petar and Tineke Baayen-Oudshoorn especially, for being such great hosts during my stay.

I also wish to thank John Connolly and the Language, Memory and Brain Lab who I was lucky enough to work and collaborate with during my Ph.D. time. John, you took a huge risk in allowing me to run with an ERP experiment idea I had, and for that I am hugely grateful. Thank you for showing me the ropes, I owe a lot of my knowledge and skills in designing and running an ERP experiment to you. Thank you also to Lili Service, Richard Mah, Rober Boshra, and the rest of the lab for all their hands-on training and help with running my experiments. Richard, thank you for devoting so much time in showing me how to implement my experiment in Presentation, for the debugging, and of course for prepping the data so I could 'model the wiggles'.

Nanci Cole and Bianca James provided indispensable administrative assistance on too many occasions to list (many of which required last minute help with immigration documents and expense reports!). I would like to thank Nanci especially for her genuinely caring attitude and for steering me through the annual administrative gauntlet of hell.

On a practical level, I would not have been able to carry out my Ph.D. without the financial support of the Ontario Trillium Foundation. Thank you, your support gave me the financial stability I needed during my studies. The Department of Linguistics and Languages here at McMaster University has also provided a wonderfully supportive environment, both morally and financially, in which to carry out my research. Thank you for providing me with the opportunity to assist in teaching and for allowing me to travel to so many exotic places to give conference talks and poster presentations. I would also like to thank Dale Askey and his team at McMaster's Lewis and Ruth Sherman Centre for Digital Scholarship for providing me with financial support and a wonderful office space. Thank you also for supplying the technical and logistical infrastructure for my projects.

Right from the beginning of my Canadian adventure I was able to forge so many new friendships, many of which will last for many years to come. I wish to extend a heartfelt thank you to all the friends in the Department of Linguistics and Languages, across campus, and in the Hamilton community who have carried me through this journey with so much passion, excitement and steadfast reliability. Thank you to Jitka, Rachael, Richard, Rober, Kyle, Heather, Bryor, Connie, Amanda, Cassandra, Bryan, Edalat, Mary (and Woodstock), Chris, Rob, and Kandice. I am also fortunate enough to have a loyal group of friends back home in Newcastle who, when I took my annual visits back home, always managed to keep me sane and show me what life really means. Thanks especially to Matthew, Andy, and Alice. I owe you.

I am grateful to my family for their unconditional love, motivation and support. Mam, dad, thank you for always being there for me and allowing me to be me. I also want to thank my brother Liam, for the time we spend together when I come back home. It means a lot to me.

Last but not least, I wish to thank Diane, for nurturing me with so much love and unwavering support, and for listening patiently when I was at my most doubtful and most obsessed. I could not have done this without you at my side.

Finally, I would like to dedicate this thesis to my Oma Inge, who left us just as I reached Canada. I wish she could still be here to see me finish.

Abstract

This thesis examines the semantic processing of morphologically complex words during visual word recognition. In a series of three experiments this thesis addresses (i) the role semantic transparency during compound word reading, (ii) the nature of the conceptual structure of compound words and its effect on visual word recognition, and (iii) the time-course of semantic access during the visual comprehension of derived words.

Chapter 2 documents evidence that the outcome of the compound semantic transparency effect is dependent upon the amount of language experience of the reader. We report that high compound transparency inhibits less experienced readers during naturalistic reading, yet facilitates processing among relatively more experienced readers. This study is the first to demonstrate that semantic processing of compound words is driven by individual reading skill.

The study reported in Chapter 3 tests the hypothesis that the conceptual representation of a compound is based on a relational structure linking the compound's constituents. Across two lexical decision datasets, Chapter 3 reports that greater entropy (i.e., increased competition) among a set of conceptual relations associated with a compound gives rise to longer lexical decision latencies. This finding indicates that the same compound word form is associated with many potential relational meanings, and that these meanings compete for selection during visual word recognition.

Chapter 4 concerns the time-course of lexical-semantic access during derived word recognition. Existing accounts of derived word recognition widely disagree about whether access to conceptual information is granted prior to morphological decomposition. We report evidence which shows that the semantics of derived words and their stems are accessed in concert with morphological sources of information. These results challenge theoretical accounts that advocate strictly serial access to (morpho-orthographic then morpho-semantic) lexical cues.

Overall, the empirical evidence presented in this thesis suggests that morphological processing involves rapid and concurrent access to many sources of conceptual information. These findings align with a view of complex word processing in which the cognitive system utilizes as many cues as possible in order to maximize the opportunity of obtaining the meaning of the word.

Contents

Declaration of academic achievements

This is a ‘sandwich’ thesis. One of the empirical chapters has been published in a peer reviewed journal. The remaining two empirical chapters have been submitted to journals and are currently under review. The following outlines the status of each chapter and the author contributions of each manuscript.

Chapter 2

This chapter has been submitted to The Quarterly Journal of Experimental Psychology as **Schmidtke, D**, Kuperman, V., & Van Dyke, J.A. Individual variability in the semantic processing of English compound words.

DS and VK came up with the concept of the study. DS and VK conducted the statistical analyses. DS was the primary writer of the manuscript and VK helped edit. JAVD provided the eye-movement and individual differences data, and also edited the manuscript.

Chapter 3

This is a reprint of **Schmidtke, D.**, Kuperman V., Gagné C., & Spalding, T. (2016). Competition between conceptual relations affects compound recognition: the role of entropy. *Psychonomic Bulletin & Review 23*(2), 556-570.

The concept for this study came from CG, VK, DS, and TS. CG and TS collected the conceptual relations data from participants at the University of Alberta. DS performed the statistical analysis under the supervision of VK. CG and TS also contributed writing to the introduction of the paper. I also wrote parts of the Introduction, the Methods, results and Discussion, and General Discussion. VK, CG and TS helped edit.

Chapter 4

This chapter has been submitted to The Journal of Experimental Psychology: Human Perception & Performance as **Schmidtke, D.**, Matsuki, K, & Kuperman. V. Surviving blind decomposition: a distributional analysis of the time-course of complex word recognition.

The initial idea of adopting the survival analysis for the study of derived word recognition came from VK and DS. The design of the experiment was further refined with help from further discussion with DS and VK. DS and VK collected and computed the lexical data, merged it with the lexical decision data, and performed the analysis. KM converted the survival analysis script from MATLAB to R. DS was the primary writer. Parts of the introduction were written by VK. Both VK and KM helped edit.

Additional achievements

In addition to the chapters presented in this thesis, I published an extra paper that could not be included in this thesis for thematic reasons. This paper appears as **Schmidtke, D.** and Kuperman V. (2016). Mass counts in World Englishes: A corpus linguistic study of noun countability in non-native varieties of English. *Corpus Linguistics and Linguistic Theory.*

List of Tables

List of Figures

1 Introduction

The challenge of reading and successfully understanding a word is typically taken for granted by users of a language. In a single day, the same word may be read hundreds of times by one person, but may be introduced to another reader for the first time. The challenge of reading is further complicated given that new words enter into a language on a daily basis, and some words may only turn up in specific contexts, like a novel, scientific article, or a poem. The challenge of reading gains an extra degree of complexity when the reader is confronted with a word that is assembled out of other words. For example, the word *catnap* is composed of the individual words *cat* and *nap*. A word such as this is morphologically complex, that is, it is composed of two or more meaningful linguistic units, or morphemes (e.g., *ant-eater*, *fire-man*, or *mind-ful*).

To the lay person, reading a morphologically complex word may appear as a simple and effortless process, but to the psycholinguist, the apparent automaticity and seamlessness of complex word recognition presents a myriad of questions about the cognitive processes involved in language comprehension. How does the language system reach the intended meaning of the complex word during reading? Does the human mind achieve this by exploiting the individual meanings that are embedded in a complex word? The question of whether complex word meanings are stored in memory as one unit, or are computed from the individual meanings of their morphemes is central to the field of psycholinguistics. Additionally, if the meanings of morphemes do play a role in complex word recognition, what exactly is this role and how can it be characterized? This thesis is an attempt to extend our current knowledge of the answers to these questions.

Though much experimental work has already been carried out on the semantic processing of complex words during visual word recognition, this thesis aims to address a number of contentious or unresolved issues that either require further exploration, or demand further empirical scrutiny. These issues fall into three inter-related subtopics, and are distilled into the following research questions:

1. Does the retrieval of complex word meanings from memory depend on reading skill?
2. How does the conceptual structure of a compound word affect its recognition?
3. How does access to the meaning(s) of complex words unfold during the time-course of complex word recognition?

This thesis addresses specific gaps in the current understanding of the answers to these questions via a series of experiments on compound and derived word reading. The remainder of this chapter provides an introduction to each of these experiments in turn before providing an outline of the thesis.

1. Does the retrieval of complex word meanings from memory depend on reading skill?

An important question in psycholinguistics is the extent to which accessing the meaning of a morphologically complex word also involves the meanings of its constituents (Libben, 2007; Marslen-Wilson, Tyler, Waksler, & Older, 1994). An often-used diagnostic of access to the meanings of morphemes during complex word reading is *semantic transparency*, and much of the experimental work on this semantic property has been carried out on compound words. As Zwitserlood (1994) notes, there is a great deal of variation in the transparency of compounds relative to their constituents, and so the study of compound words provides an ideal test-case with which to avoid confounding effects of form and meaning. For example, consider that the constituents *fire*- and -*guard* are both related in meaning to the compound word *fireguard*, 'a person who watches for the outbreak of fire'. This is an exemplification of a semantically *transparent* compound, as either the modifier or the head of the compound are within the same semantic field as the whole compound word. On other hand, the meaning of the compound *firedamp*, 'a combustible mine gas' is not related to the meanings of either *fire*- or -*damp*. Compounds that have either one or two constituents that are semantically dissimilar to the meaning of the whole word are described as semantically *opaque*.

The past 25 years of psycholinguistic research on compound transparency has yielded many results which imply that semantic transparency has an effect on visual recognition of compound words. Many of these results come from lexical decision experiments, in which participants are required to decide upon the lexicality of a word and the time taken to make this decision is measured (see e.g., Ji, Gagné, & Spalding, 2011; Juhasz, Lai, & Woodcock, 2015; Jarema, Busson, Nikolova, Tsapkini, & Libben, 1999; Libben, 1998; Libben et al., 2003; Monsell, 1985; Sandra, 1990; Zwitserlood, 1994). There is also a growing number of results from eye-tracking studies, in which eye-movements to compound words during naturalistic reading are used as a dependent measure (e.g., Juhasz, 2007; Marelli & Luzzatti, 2012; Sandra, 1990; Underwood, Petley, & Clews, 1990). Collectively, these studies demonstrate that more transparent compounds (e.g., *bootlace*, *fireguard*, and *rainstorm*) are processed faster than opaque ones (e.g., *bootleg*, *firedamp*, and *brainstorm*). This facilitatory effect is assumed to be driven by the 'semantic boost' which emerges from the excitatory connections that exist between related concepts in the mental lexicon (El-Bialy, Gagné, & Spalding, 2013). However, compound transparency effects have not been consistent across all studies. For example, Pollatsek and Hyönä (2005) and Frisson, Niswander-Klement, and Pollatsek (2008) did not find significant effects in the eye-movement record.

Given the discrepant results in the literature, the overall picture of the role of semantic transparency in printed word recognition is far from clear. There could be many underlying reasons for these discrepant results, from the choice of experimental method (as discussed in Baayen, 2014), to the operationalization of semantic transparency itself (see Bell and Shäfer, 2016). In Chapter 2 of this thesis, we bring to light the cognitive underpinning of the semantic transparency effect, which we reveal is theoretically underspecified and may be a contributing factor of the inconsistency in past results.

The eye-movement study reported in Chapter 2 broadens the empirical and theoretical foundation of the semantic transparency effect by moving from examining its effect at the aggregate level (i.e., the average effect among a population) to its effect as conditioned by individual variability in reading skill. Moving from the aggregate level to the individual level raises two important questions. First, it is not clear what cognitive skills are pertinent to the processing of compound transparency. As well as supplementing current theory of morphological processing, isolating these skill-sets can also shed light on the organization of semantic knowledge and on the nature of semantic processing in general. Second, if semantic transparency is contingent on continuous developmental changes in the lexicon, it is not yet known what the cognitive outcome of these changes might be.

Chapter 2 documents evidence that the compound semantic transparency effect in English is indeed co-influenced by reading skill. The key finding is that relatively less experienced readers are *inhibited* by greater transparency during compound word processing, whereas relatively experienced readers are *facilitated* by greater transparency during compound word recognition. Furthermore, the results suggest that these counter-directed effects are driven by the *language experience* of the individual, as gauged by measures of vocabulary knowledge and experience with printed materials.

2. How does the conceptual structure of a compound word affect its recognition?

Relational sources of lexical semantic information are fundamental to the organization of conceptual structure in the mind (for an overview see Medin, 1989). As we demonstrate in Chapter 2, semantic transparency is a relational property of compound words that captures the consequences of the concurrent activation of multiple meanings. However, earlier work by Levi (1978), Warren (1978) and Shoben (1991) has called attention to another crucial relational property of compound word semantics, namely *conceptual relations*. The topic of conceptual relations, and their relevance for compound processing is addressed in Chapter 3 of this thesis.

Conceptual relations capture the way in which compound words encode implicit relations between constituents. To illustrate, the meaning of the compound *firewood* is more than just than an expression of the meaning of 'fire' that is in someway related to the meaning of 'wood'. Instead, there is a particular connection between combination of constituents, such that *firewood* can be paraphrased as "wood *for* fire". The same relation does not apply to other compounds, for example *firepit* is a compression of the phrase "a pit that *has* fire", and *fireball* is an abbreviation of "a ball *made of* fire".

From a purely conceptual semantic perspective, the combinatorial relations exemplified above allow new meanings to be computed from existing concepts. However, though relational structures were prominent in earlier linguistic theories (Downing, 1977; Gleitman & Gleitman, 1970; Kay & Zimmer, 1976; Lees, 1960; Levi, 1978; Warren, 1978), such structures are not often discussed in current theories of complex word recognition. Based on these insights from earlier work (e.g., Levi, 1978; Shoben 1991), researchers have conceived of 16 core conceptual combinations (provided in Table 1.1).

Table 1.1: Semantic relations coding, adopted from Shoben (1991). H: head; M: modifier.

Conceptual relation	Compound example	Conceptual relation	Compound example
H ABOUT M	*newsflash*	M HAS H	*doorframe*
H BY M	*handclap*	H LOCATION IS M	*farmyard*
H CAUSES M	*joyride*	M LOCATION IS H	*neckline*
H CAUSED BY M	*sunbeam*	H MADE OF M	*snowman*
H DERIVED FROM M	*seafood*	H MAKES M	*flourmill*
H DURING M	*nightlife*	H IS M	*girlfriend*
H FOR M	*mealtime*	H USES M	*steamboat*
H HAS M	*bookshop*	H USED BY M	*witchcraft*

Interestingly, the conceptual relations listed in Table 1.1 are not just based purely on abstract linguistic theorizing. Previous psycholinguistic work has revealed that conceptual relations such as those in Table 1.1 do play a role in the recognition of compound words and thus provide a window into the organization of the conceptual system (Fiorentino & Poeppel 2007; Gagné & Shoben, 1997; Gagné & Spalding, 2004; Gagné & Spalding 2009; Ji et al. 2011). These studies imply that conceptual relations aid the integration of constituent semantic representations, and contribute to the interpretation of compound meanings.

The claim that conceptual relations are integral to compound word recognition is formalized in the relational-interpretation-competitive-evaluation (RICE) theory of modifier-noun compound processing (Spalding, Gagné, Mullaly & Ji, 2010). A critical proposal of the RICE theory is that the process of selecting a relation for a compound is competitive and that this competition unfolds in three stages. Based on Spalding et. al. (2010), these stages can be summarized as follows:

Stage 1. The conceptual relations associated with the modifier compete to be selected as a potential interpretation.

Stage 2. The possible conceptual relations linking the modifier *and* the head are concurrently activated. The strongest competitor is selected as the interpretation of the compound.

Stage 3. The selected interpretation can be elaborated in order to derive fuller meanings of the compounds.

The hypothesis outlined in Stage 1 has been confirmed by Gagné and Shoben (1997) and Pham & Baayen (2013). These findings demonstrate inhibited lexical processing when there is greater competition among relations for a specific modifier. For example, some modifiers are used with a range of many different relations (e.g., *fireball* - H MADE OF M, *fireman* - H FOR M, *firepit* - H HAS M, *firebomb* - H CAUSES M) and some are strongly associated with just a few (e.g., *battleground* - H FOR M, *battlefleet* - H FOR M, *battlefront* - H FOR M, *battlecry* - H DURING M).

Whereas Stage 1 of the RICE theory of compound recognition involves selecting the relation that is associated with a specific *modifier*, Stage 2 involves selecting a conceptual relation that fits a specific *compound*. More specifically, Stage 2 makes the critical prediction that the uncertainty inherent in choosing a suitable relation for a compound corresponds with the time it takes to process the compound word itself. This hypothesis is driven by the observation that some compounds can not be easily associated with one specific conceptual relation. For example, it is clear that the compound *heatstroke* has a very restricted range of possible conceptual relations:

the H caused by M relation has a high semantic plausibility. On the other hand, the range of potential conceptual relations is considerably greater for a compound such as *homeland*, e.g., H for M, H has M, H is M, M located H, H located M or M made of H.

So far, the competition between relations associated with a compound has been quantified using data from the possible relations task. This task uses data elicited from human participants to generate a distribution of potential relations for a compound word. In the task, participants are presented with a series of compounds, and for each compound select which conceptual relation they think is the most appropriate. This data is then used to predict lexical decision reaction times. Chapter 3 of this thesis argues that previous measures that were derived from this data (e.g., Spalding and Gagné, 2014) do not validly quantify the assumed competition among active relation candidates during compound word recognition.

In order to remedy this shortcoming, Chapter 3 implements the information-theoretic measure of entropy (Shannon, 1948). Using Shannon entropy computed over the set of possible relations associated with a compound, Chapter 3 confirms Stage 2 of the RICE theory by reporting that greater entropy (i.e., increased competition) in a set of conceptual relations associated with a compound produces longer lexical decision latencies. This result (i) demonstrates that entropy can also be used as a concise specification and interpretation of the probabilistic factors that influence semantic aspects of compound word recognition and (ii) complements other applications of entropy as a measure of lexical competition during complex word processing (e.g., Moscoso del Prado Martín, Kostić, and Baayen, 2004).

3. How does access to meaning unfold during the time-course of complex word recognition?

An important question in psycholinguistics is the relationship between orthographic, morphological and semantic levels of representation. One approach to investigating the relationship between orthographic, morphological and semantic properties of complex words is to establish the points in time at which each of these properties become available during visual word recognition. An estimate of the time-points at which characteristics of complex words become engaged during word recognition can be used as a proxy of how complex word recognition unfolds in time. Thus, whereas Chapters 2 and 3 isolate and document the influence of two semantic variables as independent semantic factors of compound word recognition, Chapter 4 zooms-out and takes a macro-level perspective of the sequence of processes involved in complex word recognition. As we will elaborate upon below, current theories of complex word recognition disagree widely in their predictions about what properties of complex words are activated during lexical processing, and in what order these properties are activated. Chapter 4 aims to address this contentious issue using a relatively novel distributional method of estimating the time-course of complex word recognition.

A dominant family of accounts in morphological processing predicts that the morpho-orthographic units of a complex word are analyzed by the language processing system before access to the word's meaning is able to proceed (e.g., Rastle & Davis, 2008; Solomyak & Marantz, 2010; Taft & Forster, 1976; Taft, 2004). This semantics-blind obligatory decomposition account (or form-then-meaning account) claims that complex word reading unfolds in a strictly serialized and semantics-blind manner. This theory contends that upon visual presentation of a complex word (e.g., *plumber*), the word string is first decomposed into separate morphemes, i.e., *plumb* + *er*. Once the orthographic cues to the word's morpheme representations have been isolated, the morpheme representations are then extracted from the mental lexicon. At the final stage, semantic access occurs as component morphemes of the complex word are recombined and the whole-word meaning -'a person who plumbs'- is retrieved from the lexicon.

On the other hand, an alternative class of models predict that morphological activation is conditioned simultaneously by form and meaning characteristics of the complex word (e.g., Diependaele, Sandra, & Grainger, 2005; Libben, 2014; Moscoso del Prado Martín, 2007; Plaut & Gonnerman, 2000), or that the activation of complex word meaning is not aided by a morphological level of representation at all (Baayen, Milin, Đurđević, Hendrix, & Marelli, 2011). Under these accounts, once the orthographic form of the complex word has been accessed, the system immediately seeks to access any possible source of semantic information. This semantic information may be processed concurrently with the parsing of morphological structure, or concurrently with the availability of (even partial) orthographic information that does not require decomposition (as implemented in Baayen et al., 2011).

There have been many studies that have been taken as support of semantics-blind obligatory decomposition. Most of this data has come from masked morphological priming studies on derived words (e.g., Beyersmann, Ziegler, Castles, Coltheart, Kezilas, & Grainger, 2015; Rastle, Davis, & New, 2004). However, averaged lexical decision latencies are suboptimal for the task of characterizing word recognition as it unfolds in time. Specifically, a mean lexical decision is only informative about the final outcome of visual word recognition (i.e., how long it took the participant to provide a response) and is not able to provide insight into the stages of processing that lead to that outcome. However, this lack of temporal precision has been remedied by the use of paradigms that allow for an excellent temporal resolution of neural behaviour as it unfolds during visual word recognition, primarily magneto- and electroencephalography (MEG and EEG).

Recently, two MEG studies carried out by Solomyak and Marantz (2010) and Fruchter and Marantz (2015) found support for the semantics-blind obligatory decomposition account of complex word recognition. These studies found that neural activity associated with morpho-orthographic parsing temporally preceded the neural activity associated with semantic processing of derived words. That is, distributional properties of the stem and affix and other properties pertaining to the morpho-orthographic structure of derived words affected neural activity during the early stages of word processing. However, relatively later stages of lexical processing in the MEG record exhibited sensitivity to properties of the whole word and its semantics.

In Chapter 4 we return to the lexical decision methodology and put to the test the evidence that is provided by Solomyak and Marantz (2010) and Fruchter and Marantz (2015). Instead of avoiding the problem of deriving a time-course of word recognition from lexical decision data, we address it. Chapter 4 presents a first attempt at using the non-parametric distributional survival analysis technique (Reingold & Sheridan, 2014) to help adjudicate between form-then-meaning and form-and-meaning accounts of complex word recognition.

The main advantage of the survival technique documented in Chapter 4 is that it is able to provide an estimate of the earliest discernible point in time at which a lexical variable exerts an influence over chronometric data. Chapter 4 describes how this technique can be adopted to construct an approximation of the time-course of complex word recognition. We demonstrate the utility of this technique by successfully establishing a chronology of a series of lexical variable onset times. Our findings are not compatible with the semantics-blind obligatory decomposition model of complex word recognition.

Outline of the Thesis

Chapter 2 reports an eye-tracking study on the silent reading of compound words embedded within sentence contexts. The focus of Chapter 2 is on the combined role of semantic transparency and individual differences in reading skill during the reading of English concatenated noun-noun compound words. Chapter 3 concentrates on the role of conceptual relations in compound words during visual word recognition. Across two separate sets of behavioural data, Chapter 3 reports the effect of competition between conceptual relations on visual lexical decision latencies to English compound words. The study in Chapter 4 tackles the issue of the time-course of complex word processing. With visual lexical decision latencies to English and Dutch derived words (e.g., *walker* and *blijheid* 'happiness'), English pseudo-derived words (e.g., *trumpet*), and English monomorphemic control words (e.g., *ballot*), Chapter 4 adopts a distributional 'survival' analysis to address the time-course of complex word recognition. Chapter 5 summarizes the research presented in this thesis and outlines topics for further investigation.

References

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

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*(3), 438–481.

Bell, M. J., & Schäfer, M. (2016). Modelling semantic transparency. *Morphology*, *26*(2), 157–199.

Beyersmann, E., Ziegler, J. C., Castles, A., Coltheart, M., Kezilas, Y., & Grainger, J. (n.d.). Morpho-orthographic segmentation without semantics. *Psychonomic Bulletin & Review*, *23*(2), 533–539.

Diependaele, K., Duñabeitia, J. A., Morris, J., & Keuleers, E. (2011). Fast morphological effects in first and second language word recognition. *Journal of Memory and Language*, *64*(4), 344–358.

Diependaele, K., Sandra, D., & Grainger, J. (2005). Masked cross-modal morphological priming: Unravelling morpho-orthographic and morpho-semantic influences in early word recognition. *Language and Cognitive Processes*, *20*(1-2), 75–114.

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*(1), 75–95.

Fiorentino, R., & Fund-Reznicek, E. (2009). Masked morphological priming of compound constituents. *The Mental Lexicon*, *4*(2), 159–193.

Fiorentino, R., & Poeppel, D. (2007). Compound words and structure in the lexicon. *Language and Cognitive processes*, *22*(7), 953–1000.

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*(1), 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., & Shoben, E. J. (1997). Influence of thematic relations on the comprehension of modifier–noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 71.

Gagné, C. L., & Spalding, T. L. (2004). Effect of relation availability on the interpretation and access of familiar noun–noun compounds. *Brain & Language*, *90*(1), 478–486.

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*(1), 20–35.

Gleitman, L. R., & Gleitman, H. (1970). *Phrase and paraphrase: Some innovative uses of language.* New York: Norton.

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*(4), 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: Elsevier.

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*(4), 1004–1019.

Kay, P., & Zimmer, K. (1976). On the semantics of compounds and genitives in English. In *Sixth California Linguistics Association Proceedings* (pp. 29–35).

Lees, R. B. (1966). *The grammar of English nominalizations.* (Unpublished doctoral dissertation).

Levi, J. N. (1978). *The syntax and semantics of complex nominals.* Academic Press, New York.

Libben, G. (1998). Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment. *Brain & Language*, *61*(1), 30–44.

Libben, G. (2007). Reading complex morphological structures. In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing.* (pp. 192–215). Hove: Psychology Press.

Libben, G. (2014). The nature of compounds: A psychocentric perspective. *Cognitive Neuropsychology*, *31*(1-2), 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*(1), 50–64.

Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the english mental lexicon. *Psychological Review*, *101*(1), 3–33.

Medin, D. L. (1989). Concepts and conceptual structure. *American psychologist*, *44*(12), 1469–1481.

Monsell, S. (1985). Repetition and the lexicon. *Progress in the psychology of language*, *2*, 147–195.

Moscoso del Prado Martín, F. (2007). Co-occurrence and the effect of inflectional paradigms. *Lingue e linguaggio*, *6*(2), 247–263.

Moscoso del Prado Martín, F., Kostić, A., & Baayen, R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition*, *94*(1), 1–18.

Pham, H., & Baayen, R. H. (2013). Semantic relations and compound transparency: A regression study in CARIN theory. *Psihologija*, *46*(4), 455–478.

Plaut, D. C., & Gonnerman, L. M. (2000). Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes*, *15*(4-5), 445–485.

Pollatsek, A., & Hyönä, J. (2005). The role of semantic transparency in the processing of Finnish compound words. *Language and Cognitive Processes*, *20*(1-2), 261–290.

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., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, *11*(6), 1090–1098.

Reingold, E. M., & Sheridan, H. (2014). Estimating the divergence point: a novel distributional analysis procedure for determining the onset of the influence of experimental variables. *Frontiers in Psychology*, *5*. Retrieved from `http://dx.doi.org/10.3389/fpsyg.2014.01432`

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*(3), 529–567.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell Technical Journal*, *27*(4), 379–423.

Shoben, E. J. (1991). Predicating and nonpredicating combinations. In P. J. Schwanenflugel (Ed.), *The psychology of word meanings.* Hillsade, NJ: Lawrence Erlbaum.

Solomyak, O., & Marantz, A. (2010). Evidence for early morphological decomposition in visual word recognition. *Journal of Cognitive Neuroscience*, *22*(9), 2042–2057.

Spalding, T. L., & Gagné, C. L. (2014). Relational diversity affects ease of processing even for opaque English compounds. *The Mental Lexicon*, *9*(1), 48–66.

Spalding, T. L., Gagné, C. L., Mullaly, A. C., & Ji, H. (2010). Relation-based interpretation of noun-noun phrases: A new theoretical approach. *Linguistische Berichte Sonderheft*, *17*, 283–315.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology Section A*, *57*(4), 745–765.

Taft, M., & Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior*, *15*(6), 607–620.

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

Warren, B. (1978). Semantic patterns of noun-noun compounds. *Acta Universitatis Gothoburgensis. Gothenburg Studies in English Goteborg*, *41*, 1–266.

Zwitserlood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes*, *9*(3), 341–368.

2 Individual variability in the semantic processing of English compound words

This chapter has been submitted to the Quarterly Journal of Experimental Psychology as Daniel Schmidtke, Victor Kuperman, and Julie A. Van Dyke. Individual variability in the semantic processing of English compound words.

Abstract

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 skills on eye-movement behaviour during sentence reading. Eye-movement data were collected from 100 non-college-bound 16-26 year-old speakers of English in a sentence reading task representing a total of 394 different compound words. Individual differences data were collected from the same participants using multiple offline verbal and cognitive skill assessment batteries. Statistical analyses revealed a qualitatively consistent interaction pattern between individual differences measures and semantic transparency across skill tests that measure experience with printed material. Readers with impoverished reading experience and smaller vocabulary size exhibited slower processing when reading highly transparent compounds relative to opaque compounds. The opposite effect was observed for readers with greater reading experience and vocabulary size, such that highly transparent compounds facilitated lexical processing. To account for the results, the authors posit a trade-off between two cognitive mechanisms, i.e., the benefit of semantic co-activation of closely related concepts and the cost of discriminating between those concepts, which is influenced by individual reading experience.

2.1 Introduction

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 compound words, by definition, encodes not only meaning of the whole word itself (e.g., *bootlace*), but also the multiple distinct meanings of its constituent morphemes (e.g., *boot* and *lace*). The processing consequences of the interaction between the meanings contained in 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 paper 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 (for reviews of other semantic variables pertinent to compound processing, see Juhasz, Lai, & Woodcock, 2015; Kuperman, 2013). For example, the meaning of *bootleg*, 'an illegal music recording', is relatively opaque as it is unrelated to meanings of its constituents *boot* and *leg*, while the meaning of *bootlace* is relatively transparent.

Even though current theoretical accounts of morphological processing are at odds about the time at which the semantics of a complex word is accessed relative to the access of its orthographic form (for discussion see Baayen, 2014), they generally agree on the expected *facilitatory* role of semantic transparency. First, compound words with meanings that are closely associated to the meanings of their embedded morphemes are argued to allow for easier *meaning computation* (El-Bialy, Gagné, & Spalding, 2013). Under the meaning computation account, compound recognition is assisted by the automatic activation and interpretation of a compound's possible relational interpretations (e.g., a *snowball* is a *ball* 'made of' *snow*; Gagné & Spalding, 2004, 2006; Gagné, Spalding, Figueredo, & Mullaly, 2009; Schmidtke, Kuperman, Gagné, & Spalding, 2015). In opaque compounds but not in transparent ones, the meaning computed via a relational link clashes with its established meaning, and consequently impedes the recognition of the compound. For example, an opaque compound such as *hogwash* does not have an interpretable relational link that can plausibly combine the constituents of the compound. Therefore, the language processing system's automatic attempt to find such a conceptual relation slows down the recognition of the word.

In the same vein, several accounts argue that the similarity of multiple meanings encoded in transparent compounds triggers a "semantic flooding effect" (Libben, 1998, pp. 43; see also Libben, Gibson, Yoon, Bom, & 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), but inhibit processing when the semantic representations are weakly associated (i.e., for the case of opaque compounds). A family of analogous accounts are available in the literature, e.g., "conjunctive activation" (El-Bialy et al., 2013), "semantic flooding" (Libben, 1998), and "semantic resonance" hypotheses (De Jong, 2002). Each of these hypotheses outlines a distinct processing sequence and processing architecture. However, of importance to the current study, all hypotheses are united in their prediction that transparent compounds are recognized faster than opaque ones. Consequently, 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 of semantic transparency. The semantic boost account thus describes the presence of any facilitatory effect originating from the conjunctive activation of a close semantic relationship between the meanings of the full compound form and its embedded constituents (e.g., the semantic similarity between any pair of words formed by *bootlace*, *boot* and *lace*).

2.1.1 Evidence for the role of compound transparency

Experimental work on the role of semantic transparency during compound word recognition has produced inconsistent and discrepant effects reported in the psycholinguistic research literature over the last 25 years (e.g., Juhasz, 2007; Libben, 1998; Marelli & Luzzatti, 2012; Sandra, 1990; Underwood, Petley, & Clews, 1990; Zwitserlood, 1994). The aim of this study is to contribute to the conceptualization of the cognitive processes associated with semantic transparency and also to the empirical base for this line of research. The emphasis of the latter aim is on providing further insight into the influence of the largely neglected (but see Andrews & Lo, 2013; Kuperman & Van Dyke, 2011a, 2011b) role of individual differences in reading skill on compound processing. In what follows, we briefly review the prior empirical reports on semantic transparency during compound processing and then motivate the theoretical and methodological approach of this study.

A significant proportion of the experiments on semantic transparency has involved visual lexical decision and priming paradigms. For example, in Sandra (1990), semantic associates of constituents were used as primes for target compound words that were either semantically transparent or opaque (e.g., *death* used as a prime for *birth* in the semantically transparent compound *birthday*, and *moon* used as a prime for *sun* in the opaque compound *sunday*). Relative to semantically unrelated control primes, Sandra (1990) found shorter lexical decision latencies (RTs) to semantically transparent compounds compared to semantically opaque compounds. In another seminal study, Zwitserlood (1994) used Dutch compounds as primes and either the left or right constituent of the compound as the target word. Zwitserlood (1994) found facilitatory effects on response latencies for both fully transparent and partially transparent compounds as compared to fully opaque compounds.

The above results are supported by Libben et al. (2003), who found an effect of transparency when the whole word is similar to the semantics of the head constituent, such that compounds are recognized faster when the head is transparently related to the whole word. Moreover, the recent findings of Ji, Gagné, and Spalding (2011) also revealed a facilitatory effect of semantic transparency in a series of visual lexical decision experiments. Ji et al. (2011) forced the decomposition of compounds in a number of ways, such as presenting the compounds in a spaced format (e.g., *bull dog* instead of *bulldog*), and presenting constituents in different colours. They found effects of transparency arise under conditions that forced morphological decomposition, such that more transparent compounds were recognized faster than opaque compounds and monomorphemic words. In a series of related experiments, El-Bialy et al. (2013) found semantic priming for compounds composed of two transparent constituents (TT; e.g., *carwash*) but not for compounds with a transparent head and an opaque modifier (OT; e.g., *strawberry*) or compounds with a transparent modifier and an opaque head (TO; e.g., *jailbird*). They also found semantic priming for fully opaque compounds (OO; e.g., *hogwash*) relative to TO compounds. The results of this experiment demonstrate that the effect of transparency occurs only when the meanings of *both* constituents relate to the whole compound word.

Additionally, Juhasz et al. (2015) recently reported facilitatory effects of transparency for 629 compounds on reaction times present in the English Lexicon Project (ELP). 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 come with inhibited processing, but that those that were at neither extreme of the transparency continuum were recognized the fastest. To summarize, this collection of results demonstrates that semantic transparency exerts an influence during compound word recognition. Excluding Pham and Baayen's (2013) study, these results indicate that constituents aid the comprehension of compound words only when the meanings of constituents are related to the meaning of the compound word.

Despite the large body of evidence suggesting that semantic transparency boosts compound word recognition, a handful of experiments have not yielded such facilitatory effects. In a masked priming experiment by Fiorentino and Fund-Reznicek (2009) in which the prime was the compound word, and the target was either the modifier or head of the compound, they found that both transparent *and* opaque compounds significantly primed their constituents. These results demonstrate an equivalent priming effect for transparent and opaque compounds.

Findings in the eye-movement literature are also mixed. In convergence with the facilitatory effects of transparency observed in several lexical decision studies, two eye-tracking experiments (Underwood, Petley, & Clews, 1990; Juhasz, 2007) both reported shorter gaze durations on semantically transparent English compounds relative to semantically opaque compounds. More recently, Marelli and Luzzatti (2012) reported interactive effects of semantic transparency, constituent frequency and headedness during the processing of Italian compounds in lexical decision and in eye-tracking data. Marelli and Luzzatti (2012) gauged semantic transparency using human judgements and reported that semantic transparency interacted with the headedness, frequency and transparency of the compound word. The lexical decision and eye-tracking (total fixation times) results show that in head-final compounds, the facilitatory role of the frequency of the second constituent is attenuated as the degree of semantic transparency progressively decreases. The same pattern is reported, but only in lexical decision times and with a larger effect size, for interactions between the frequency of the modifier as the first constituent and semantic transparency. However, in head-initial compounds, higher frequency modifiers facilitated the processing of opaque compounds and inhibited the processing of transparent compounds. This pattern was confirmed for lexical decision RTs, and in gaze durations and total fixation durations in the eye-tracking component of the study.

The results of Marelli and Luzzatti's study are concomitant with the results reported by Juhasz (2007), Sandra (1990), Underwood et al. (1990) and Zwitserlood (1994) which collectively indicate that the relationship between meanings of constituents affects the ease of compound processing. Marelli and Luzzatti's results additionally indicate that the effect of semantic transparency is also contingent on the constituent frequency and headedness of the compound word under visual inspection. However, effects of semantic transparency have not been replicated in all eye-movement studies. Pollatsek and Hyönä (2005) and Frisson, Niswander-Klement, and Pollatsek (2008), both reported no reliable influence of semantic transparency on the eye-movement patterns of reading Finnish and English compound words respectively. Thus, although effects of transparency have some support in the eye-movement data reported above, the overall picture is still not conclusive.

2.1.2 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. The body of evidence suggests that the effect of semantic transparency is not always empirically substantiated, but when present, the effect indicates that increased transparency facilitates compound recognition. An obvious source of the discrepancy in research results could be traced to differences in tasks and behavioural measures (see Baayen, 2014 for a discussion of lexical decision vs. eye-tracking methodologies in morphological processing research). Additionally, the body of mixed results could also be attributed to differences in the operationalization of semantic transparency. To demonstrate, many studies use experimental designs with orthogonal contrasts between opaque and transparent compounds (e.g., Libben et al., 2003; Zwitserlood, 1994). As we reviewed earlier, these studies conceive of semantic transparency as sets of discrete 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 constituents and the compounds in which they embed. These measures are usually defined by experimentally obtained human ratings (Libben et al., 2003; Marelli & Luzzatti, 2012), and research suggests that the experimental outcome of these ratings are contingent upon the way in which participants are asked to evaluate the semantic relation between two meanings (Marelli, Dinu, Zamparelli, & Baroni; 2015).

In fact, Marelli et al. (2015) have argued that semantic similarity can be profitably measured when quantified computationally using Distributional Semantic Modelling (DSM; Turney & Pantel, 2010). Distributed Semantic Models are a collection of techniques which use textual co-occurrence-based vector representations to recreate semantic distance between words. The motivation behind using these tools is to use word co-occurrence statistics to simulate a network of human conceptual knowledge. One particular method of deriving word similarity is Latent Semantic Analysis (LSA; Landauer & Dumais, 1997), which has also 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 advantage of LSA (and indeed other DSM measures) is that it does not rely on subjective human judgment. Instead, meaning relations are defined as an emergent property of the distributional information available for words in a corpus. The present study takes into consideration the advantages that the LSA technique offers.

2.1.3 The Naive Discriminative Reader theory

We believe, however, that the methodological discrepancies outlined above 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, based on current sources of evidence, 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 firstly 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., 'cold', 'wet', 'puddle' and 'umbrella') (see Ramscar, Yarlett, Dye, Denny, & Thorpe, 2010). Accessing the meaning of the word *rain* originates from the strength of the activation between the memory trace of these events and the linguistic cue. Reaching the point at which the meaning 'rain' is learned is thus 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, learning and reading an unspaced compound word presents the processing system with a challenge that is unlike the processing of a simplex word such as *rain*. For the case of compound word processing, the idea is that the orthographic forms of the 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 of many meanings. The NDR model predicts that upon visual inspection of a compound word, the cognitive system resolves the meaning of a compound word by discriminating between all of the distinct meanings that are cued by the substrings (constituents and/or pseudo-constituents) present in the orthography of the compound (e.g., *eye*, *lash* and *ash* in *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, a discriminative perspective of semantic 'relatedness' conceives of the degree of semantic similarity between events and entities in the world as a function of how difficult it is to discriminate between them. From this perspective, the more a word occurs in similar environments as other words across a diverse range of contexts, i.e., 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. Importantly, LSA itself, as well as other DSM methods, provides a measure of the substitutability of words across a range of text documents, and thus is an appropriate technique with which to gauge the processing difficulty associated with highly indiscriminate meanings.

The hypothesized processing cost associated with discriminating between similar meanings can be illustrated with the transparent compound *rainstorm*. In 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 the closely related (and substitutable) meanings, leading to a recognition boost. However, it is precisely this kind of compound that would create an effect opposite to semantic facilitation: *a discrimination challenge*. It is 'rainstorm', and not 'rain' or 'storm' that is the target meaning (Baayen, Kuperman, Shaoul, Milin, Kliegl & Ramscar, unpublished). 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 *bootleg*, the embedded orthographic strings *boot* and *leg*, cue meanings that are unrelated to the meaning of the compound under visual inspection, i.e., 'an illegal music recording'. As operationalized by LSA, the constituents *boot*, *leg* and *bootleg* occur in different contexts (i.e., are relatively 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 *Naive Discriminative Reader* (NDR) model.

Importantly, 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, also known as *Naive Discriminative Learning*. Using a Rescorla-Wagner model of human learning (Rescorla & Wagner, 1972; Rescorla, 1988), Ramscar et al. (2010), and Ramscar, Dye, Popick, and O'Donnell-McCarthy (2011) demonstrated that word learning involves associating a linguistic event (e.g., a meaning) with cues in the world preceding its occurrence. In this way, NDL predicts that learning the meaning of a word proceeds only when multiple cues in the world compete with one another for a single linguistic expression. For example, learning the meaning of the word *apple* depends upon the probability of occurrence of the experiential cue of 'a round green fruit' relative to that of 'a long peelable yellow fruit' or 'a round orange fruit'. What follows from this insight is the critical notion that as language experience accumulates, the strengths of associations between the cues and the linguistic events adjust accordingly. The NDR model therefore makes explicit predictions about the way in which the semantic transparency and individual language experience may co-determine the processing of compound words: the ability to discriminate between semantic entities improves as the individual gains more reading experience.

2.1.4 The current study

The focal role of language experience in the NDR model shapes this investigation in two ways. First, the incorporation of the effect of 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 2.1, 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, lowest quintile) of language experience. The dashed, dotted, dotdash and longdash 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. Since no account that we are aware of proposes an interaction between facilitation via semantic boosting and individual reading skill, we do not incorporate such an interaction, leaving the mock-up regression lines parallel. Secondly, 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, i.e., individuals less likely to establish sufficiently strong associations between contextual cues and meanings. We visualize this result as a transparency by skill 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 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. To account for this, the remaining panels illustrate the case where both cognitive mechanisms are simultaneously present in the reading population. The behavioural 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: i.e., a

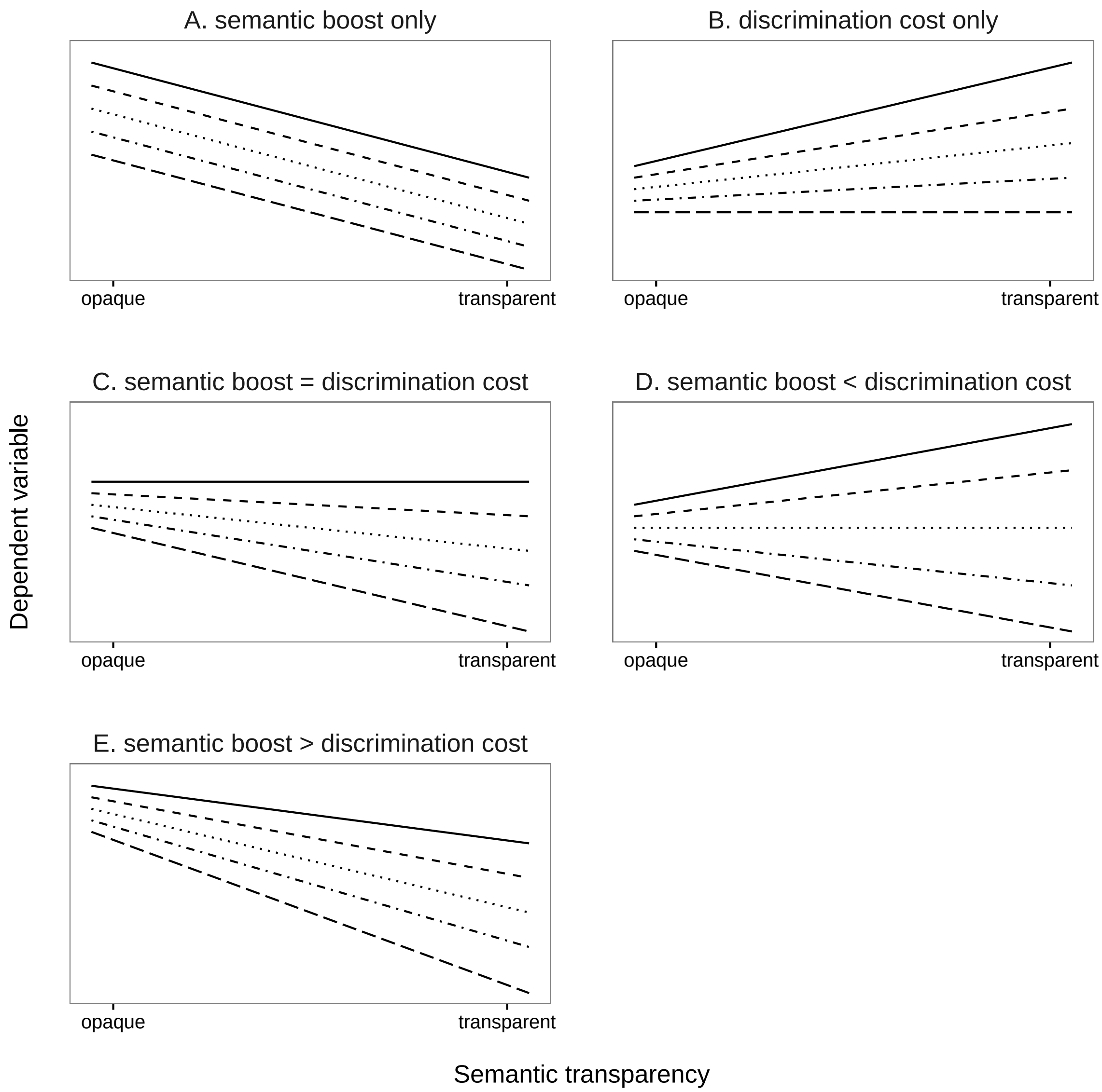

Figure 2.1: Hypothetical processing patterns for the interactive effect of semantic transparency (x axis) and individual reading experience on compound word recognition effort (y axis). Individual panels represent behavioural outcomes predicted under the semantic boost account (Panel A), the NDR theory (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.

pattern where the benefit of the semantic boost is equal to the discrimination cost (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, as 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, and 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 focal role of language experience in the NDR account shapes this investigation in a second way. Crucially, any of the hypothesized interactive effects in Panels C to E of Figure 2.1 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 behaviour of a "community" population of non-college bound students. We do not argue that the cognitive mechanisms hypothesized above would not be generalizable to a student population, just that they will be more likely to be detected within the community sample, where there lies a greater variability in reading skill. To verify this assumption, we compared the community population against undergraduate students on their performance of the Author Recognition Test and Magazine Recognition Test (ART and MRT; Acheson, Wells, & MacDonald, 2008, see Appendix A for full description of the test), which we take as our index of reading experience. The distributions of these test scores for the current study's 100 community participants were compared with that of the same scores acquired from an unrelated study of 135 McMaster University students.

As is visualized in the density plots in Figure 2.2, the dispersion of scores in both the Author and Magazine recognition tests was greater among participants drawn

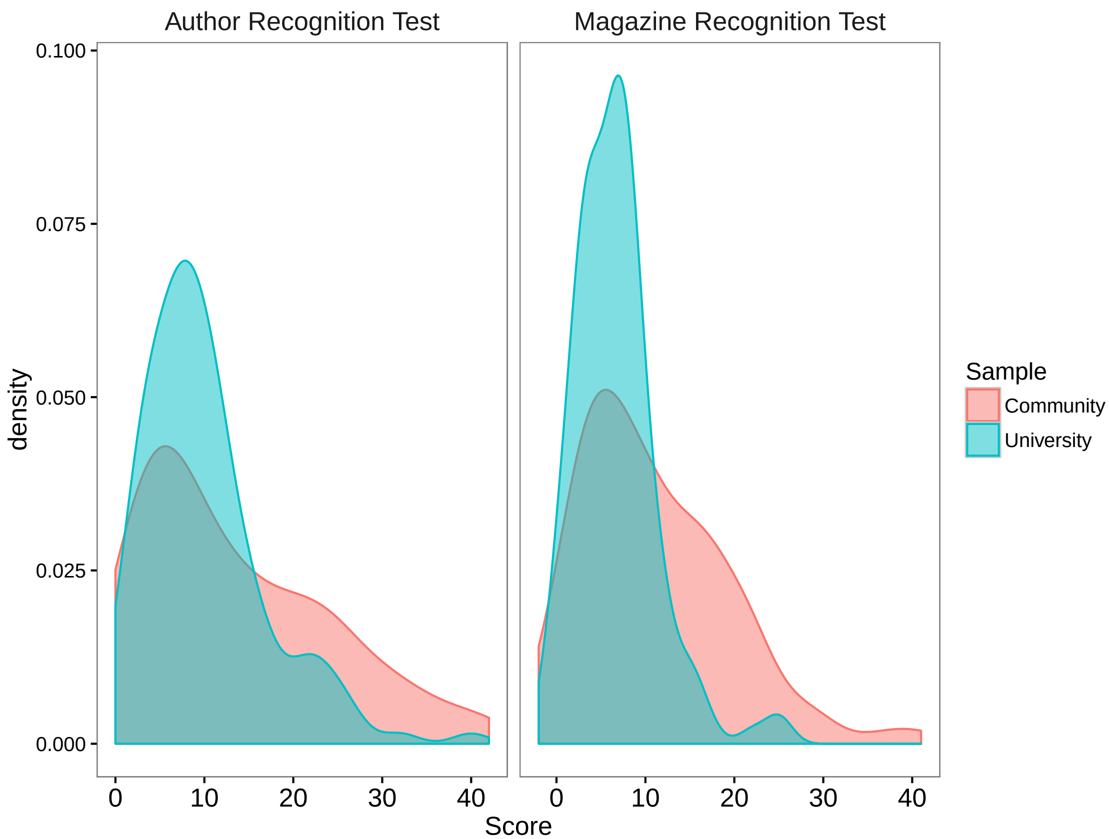

Figure 2.2: Density plots of the distribution of results of the Author (left) and Magazine (right) recognition tests split by two samples; a university student population ($n = 135$) and a non-college bound adult population ($n = 100$).

from the community sample (ART; $SD = 10.7$, MRT; $SD = 8.36$) as compared to those drawn from the convenience sample (ART; $SD = 6.87$, MRT; $SD = 4.54$). A larger variance of scores in the community sample was confirmed by F-tests for both the Author Recognition Test ($F = 2.43$, $p < .001$), and the Magazine Recognition Test results ($F = 3.39$, $p < .001$). As a side note, we discovered – much to our chagrin – the community population exhibited a higher average amount of experience with printed materials (ART; $M = 13.58$, MRT; $M = 11.05$) than did the sample of undergraduate students enrolled in the linguistics courses at McMaster University (ART; $M = 9.96$, MRT; $M = 6.64$). The independent Mann-Whitney U-test confirmed that these differences were reliable for the ART ($W = 7737.5$, $p = 0.04$) and the MRT ($W = 8839$, $p < 0.0001$). 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 against a population of non-college bound readers.

In studying an under-researched 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. A recent effort to investigate individual variability in reading performance in non-clinical adult populations has demonstrated that readers who perform poorly in tests designed to tap into a range of verbal and non-verbal skills also exhibit reading behaviours that contrast with readers who perform highly on skill tests (see reviews in Andrews, 2012; Radach & Kennedy, 2013; Rayner, Pollatsek, Ashby & Clifton, 2012). For example, less proficient readers have been shown to fixate on words for longer, make more regressions, skip fewer words, and be more likely to refixate on words. In addition, a number of eye-tracking studies have also shown that individual differences in reading comprehension and experience modulate sensitivity to the effects of distributional information present in the language. For instance, a series of experiments have demonstrated that less proficient readers are most affected by the influence of low frequency words, the frequency of embedded morphemes in derived words, the probabilistic bias for the presented format of compound words (e.g., spaced vs. concatenated), and predictability (cf. Ashby, Rayner, & Clifton, 2005; Falkauskas & Kuperman, 2015; Kuperman & Van Dyke, 2011a, 2011b, 2013; Whitford & Titone, 2014).

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 100 participants from the local community, who represent a non-college bound population of readers. Participants read a total of 394 unique English compounds that were each embedded within a sentence context and varied in semantic transparency (gauged by LSA). In addition to partaking in an eye-tracking experiment, participants were also administered a range of verbal and non-verbal skill tests, which together measure a variety of component skills of reading.

With data from this experiment, the current study introduces a series of three 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 2a puts to the test the predictions outlined in the Introduction, 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 an individual's exposure to print. Indeed, reading is a complex and multifaceted skill; reading proficiency is not only modulated by experience, but a full range of other components such as working memory, inhibitory/attentional control, phonological awareness, grapheme-to-phoneme decoding, vocabulary knowledge and general intellectual ability. To this end, Analysis 2b takes ones step further and examines whether any of the major components of reading skill are pertinent to the processing of compound transparency.

2.2 Methods

2.2.1 Participants

100 participants (53 female; 47 male) were recruited in New Haven CT, within an age range of 16-26 (M = 21.16, SD = 2.24). Participants were paid $15/hour 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 non-college bound individuals (formal level of education did not exceed the equivalent of high school level). All were native speakers of English, had normal or corrected to normal vision and none had a diagnosed reading or learning disability.

2.2.2 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 1000 Hz. A chin support and forehead rest was used to stabilize participants' gross head movements. The stimuli were presented on a 10.75-in x 13.25-in screen with a refresh rate of 60 Hz. The average gaze position error of the EyeLink 1000 is $< .05°$, while its resolution is $.01°$ (root mean square error), with a micro-saccade resolution of $.01°$.

2.2.3 Materials

A total of 394 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 filler sentences with an unrelated syntactic manipulation. 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 online supplementary materials).

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 pseudo-randomized for each participant. Each experimental session began with a series of 8 practice trials and 4 comprehension questions.

2.2.4 Procedure

Prior to presentation of the stimuli, the eye-tracker was calibrated using a series of 9 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 seconds 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 centre of the screen; two possible answers appeared three lines below, one to the left of centre and one to the right of centre. Participants indicated their answer by pressing the associated button on a keyboard; for example, if the answer they chose appeared to the left of centre, they were told to press a corresponding key. The position of the correct answer was counterbalanced throughout the experiment. If participants had not signalled that they had completed reading the sentence before the 10-second 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-second limit, the computer moved onto the next item.

Participants also completed a battery of skill tests (see Predictor variables below for a general summary, and Appendix A for a detailed description of the tests). The duration of the eye-tracking experiment, consisting of one stimulus list, was 30 minutes. The durations of the offline test battery did not exceed 5 hours (across separate sessions).

2.2.5 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 first-of-many fixation when there are more than one fixations on a given word before gaze leaves that word for the first time); gaze duration (GD; summed duration of fixations on a given word before gaze leaves the word for the first time) and the total viewing time (TVT; the summed duration of all fixations on a word). We also analyzed the probability of refixation (RP; the likelihood of fixating the word more than one time in the first reading pass), and first-pass regression rate (FRR; the likelihood of a regression to the compound after the first reading pass was completed).

All durational eye-movement measures were log-transformed to reduce the skewness of their distributions. Table 2.1 provides a list of dependent variables including the ranges, standard deviations and means of their original and transformed values.

2.2.6 Predictor variables

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

Lexical characteristics

Measures of semantic transparency: The critical lexical variable of interest was semantic transparency, which we considered alongside a range of lexical characteristics of compound words that have been demonstrated in prior research to affect compound processing. We focused on two measures of semantic transparency and their simultaneous influence on eye-movement behaviour. The first is the semantic similarity of the left constituent (modifier) to the whole compound word (Modifier-Compound; e.g., *flash* - *flashlight*), and the second is the semantic similarity of the right constituent (head) to the whole compound (Head-Compound; e.g., *light* and *flashlight*). The consideration of the Modifier-Compound and Head-Compound semantic transparency measures allows for comparability with numerous previous studies (e.g., Ji et al., 2011; Juhasz, 2007; Juhasz et al., 2015; Libben, 2003) that have operationalized transparency as the semantic overlap of the head and the modifier (compared either individually or both at the same time), with the whole compound word[1].

Several Distributional Semantics Modelling methods exist for operationalizing lexical substitutability as an important aspect of semantic similarity. We opted for the method that is most commonly used in studies of morphological processing, i.e., the Latent Semantic Analysis (LSA; Landauer & Dumais, 1997). For prior motivations of the choice, see also (e.g., Feldman et al., 2009; Fruchter & Marantz, 2015; Gagné & Spalding, 2009; Kuperman & Bertram, 2013; Moscoso del Prado Martín et al., 2005; Pham & Baayen, 2013; Rastle et al., 2000; Rastle et al., 2004). In LSA, each word is represented by a vector that contains co-occurrences of that word with a large number of high-frequency words in a text corpus. The degree of semantic similarity between the words that those vectors represent is estimated by the cosine of the angle between the vectors, which ranges from -1 to 1. Values closer to 1 imply a greater

[1]To our knowledge, there are two traditions of estimating compound transparency: (i) one evaluates how semantically similar each individual constituent is to the compound (see e.g., Juhasz et al., 2015), while (ii) another interrogates the compositionality of the compound (i.e., how semantically similar the constituents are to one another, see Marelli and Luzzatti, 2012). These approaches arguably tap into different dimensions of compound semantics. Our theorizing evokes the NDR account, for which substitutability of words in context is central. For this reason, we selected the LSA method, which directly gauges substitutability: this is a computational measure of a transparency relationship that is analogous to (i), and not to (ii).

semantic similarity/substitutability between the pair of words under comparison.

Currently available implementations of the LSA vary in the selection of corpora on which they are trained and, to a lesser extent, the parameters of the algorithm underlying the analysis. A comprehensive investigation of the most appropriate technique and training corpus is beyond the scope of our study. However, we compared two off-the-shelf LSA implementations in how closely related they are to human ratings of semantic transparency.

As a benchmark, we used human ratings of Modifier-Compound and Head-Compound semantic transparency, which were collected for 652 English noun-noun compounds (Baayen et al., unpublished). Ratings were obtained via the Amazon Mechanical Turk online crowdsourcing platform https://www.mturk.com from 295 USA-based native speakers of English. Participants were instructed to evaluate on a scale from 1 (completely unrelated) to 7 (extremely related) how close in meaning the pair of words was (e.g., *food* and *foodbank*, or *bank* and *foodbank*).

The term-to-term LSA scores for Modifier-Compound and Head-Compound relations in the target compounds were collected from http://zipf.ugent.be/snaut-english, with a default setting of 300 factors and a window of 6 words (Mandera, Keuleers, and Brysbaert, in press). These LSA scores (labeled here *SUBTLEX*) were calculated over word occurrences in the 51 million-token SUBTLEX corpus of US film and media subtitles (Brysbaert & New, 2009). Since greater values of the scores represent a greater semantic distance rather than a greater similarity, we inverted them (by multiplying by negative one) for comparability with other LSA scores, see below. Additionally, we collected LSA scores (labeled here *TASA*) for target compounds from http://lsa.colorado.edu, with a default setting of 300 factors. The training corpus was the "General Reading up to the first year of college" 17 million-token section of the Touchstone Applied Science Associates, Inc. (TASA) corpus (Zeno, 1995).

The overlapping set of compounds for which both human ratings and two LSA scores existed, both for Modifier-Compound and Head-Compound relations, consisted of 644 compounds. Correlations between average Modifier-Compound human ratings and respective *SUBTLEX* LSA scores was higher ($r = 0.545$) than that between human ratings and *TASA* LSA scores ($r = 0.508$). Similarly, the correlation between average Head-Compound human ratings with *SUBTLEX* LSA scores was higher ($r = 0.575$) than that with the *TASA* LSA counterpart ($r = 0.441$): all p-values below 0.001 and all r-to-z comparisons of correlations significant at the 0.01 threshold. In sum, correlations between the two computational and experimental estimates of semantic transparency showed a large effect size (defined as $|r| > 0.5$ by Cohen, 1992)[2]. However, LSA scores based on the *SUBTLEX* corpus demonstrated consistently and reliably stronger correlations with human judgments: in what follows we used this set of semantic similarity scores. As can be seen in Table B.1 of Appendix B, Modifier-Compound and Head Compound LSA are only weakly correlated ($r = 0.16$). This enabled us to include both measures in the same statistical model and estimate their independent effects on compound processing[3].

Control measures: We also considered the frequency of occurrence of the compound which was extracted from the 50-million token SUBTLEX US (Brysbaert & New, 2009) and compound length in characters. We also included the morphological family size of the left and right constituent (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). The stand-alone frequency of the left and right constituents nor the family size of the left and right constituents reached significance in the models in which semantic transparency also played a significant role: these variables are not discussed further.

Additional norming studies were conducted to collect judgments about the plausibility of a compound given the preceding sentence fragment on a scale from 1 (completely implausible) to 5 (extremely plausible). Mean plausibility scores per

[2]The upper bound for the correlation strength between these two comparisons is given by the reliability of human judgments. In these data, split-half reliability was not high and amounted to 0.68 for Modifier-Compound ratings and 0.67 for Head-Compound ratings. Thus when comparing the correlation coefficients between LSA estimates of transparency with the split-half reliability correlation coefficients of human ratings, the explanatory power of computational estimates is larger than the correlation coefficients alone suggest.

[3]For the sake of completeness, we considered the similarity between the modifier and the head (i.e., the compositional measure of semantic transparency mentioned earlier; e.g., *flash - light*). This variable was moderately correlated with the Modifier-Compound measure ($r = 0.31$) and the Head-Compound measure ($r = 0.4$). Moreover, this variable also elicited no independent effect on the eye-movement record when both measures were entered into regression models: we do not report the Modifier-Head LSA measure any further.

compound ranged from 2.43 to 4.91. We included plausibility scores for the compound word in the sentence and the position of the critical word in the sentence (the numeric position of the critical word from the beginning of the sentence) as control variables. We further conducted a cloze predictability study to ensure that the meanings of either the target compounds or their morphological constituents are not primed by the preceding sentence context. Participants were given the sentence fragment up to and excluding the target word and were instructed to type in their guess of the upcoming word. For the plausibility and cloze predictability tasks, the stimuli list was split into two and shown to different participants. Twenty native speakers of English from the McMaster undergraduate student pool contributed to each list in the plausibility study and to the cloze predictability task (80 students overall): none of the participants in either study was part of any other study reported here. Cloze predictability norms (defined as the ratio of correct guesses of the word to the total guesses) were equal to 0 for either whole compounds or their morphemes across all sentence contexts, and so this variable was not included in statistical models. Distributional characteristics of all lexical predictor variables are reported in Table 2.1. Correlations between all independent lexical predictors are presented in Table B.1 of Appendix B.

Individual differences assessments

We considered a battery of 21 verbal and cognitive skill tests which prior studies have hypothesized to represent major component skills of reading ability (e.g., Cain & Oakhill, 2012; Rayner et al., 2012; Share, 2008, Stanovich & West, 1989). These measures were collected from a battery of psychometric tests which included standardized measures of working memory, inhibitory/attentional control, phonological awareness, grapheme-to-phoneme decoding, reading comprehension, reading fluency, vocabulary knowledge, reading experience and general intellectual ability. In addition to these tests, we also considered participant age and years in education as individual differences measures. All 100 participants were administered the same test battery, however 4% of total number of individual differences scores were missing as some participants were not administered a complete test battery. We imputed the mean value of the given skill test for these missing cases. Finally, as there

Table 2.1: Summary of eye-movement measures as response variables (A) and continuous lexical predictor variables (B), including labels used in statistical models. Mean values, standard deviations and ranges are reported for raw values and transformed variables.

Variable	**Original**				**Transformed**		
A. *Response variables*	range	mean	sd	n trials	range	mean	sd
Single fixation duration (SFD)	56:1371	267	119	1288	4.03:7.22	5.59	4.78
First-of-many fixation duration (FOM)	53:1716	248	109	3478	3.97:7.45	5.44	0.39
Refixation probability (RP)	0:1	0.38	0.49	4766			
Gaze duration (GD)	56:2033	408	244	4766	4.03:7.62	5.85	0.57
First-pass regression rate (FRR)	0:1	0.27	0.45	4766			
Total viewing time (TVT)	56:2387	589	340	4766	4.03:7.78	6.21	0.61
B. *Lexical predictor variables*	**Original**				**Transformed**		
Modifier-Compound LSA Similarity	-0.99:0	-0.7	0.14		-2.19:5.14	0	1
Head-Compound LSA Similarity	-1.004:-0.36	-0.7	0.13		-2.31:2.66	0	1
Compound Frequency	1:4958	135	445		-2.1:3.43	0	1
Left constituent Frequency	18:102467	6517	12309		-3.34:2.55	0	1
Right constituent Frequency	12:204428	14039	27754		-2.82:2.18	0	1
Left constituent Family Size	0:54	12	11		-2.47:2.07	0	1
Right constituent Family Size	1:155	19	29		-1.59:2.52	0	1
Compound Length	6:13	8.32	1.13		-2.05:4.15	0	1
Compound Plausibility	2.43:4.91	3.67	0.54		-2.28:2.27	0	1

were many tests, some of which measure the same skill, we created composite skill test scores by scaling and summing the values of tests that measure similar skill sets. This process compressed the number of individual differences tests (including age and years in education) considered in the statistical analysis to a total 13 measures of verbal and cognitive skills.

Descriptive statistics of the individual differences measures (including those out of which composite scores are computed) are summarized in Table 2.2. As can easily be seen in Table 2.2, the average scores and age-equivalents on the individual skill tests are comparable with the scores of undergraduate samples, yet it is the minimum values in the distributions of these test scores that are drastically different from undergraduate samples. We provide a detailed description of the skill tests in Appendix A. A correlation matrix of all individual differences measures is provided in Table B.2 of Appendix B.

2.2.7 Statistical considerations

Our interest in studying this behavioural data set lies in tracing the impact of semantic transparency and the way it is modulated by individual differences in reading skill. For this reason, all our analyses fit a separate regression model to each dependent variable, and (in Analyses 2a and 2b) every unique combination of dependent variable and individual skill measure. Each model contained both critical measures of semantic transparency, Modifier-Compound and Head-Compound, either as independent predictors or in an interaction with skill measures.

Linear mixed-effects multiple regression models were used for this study, with Gaussian (for continuous predictors) or binomial (for binary predictors) 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. Since the performance of each participant in a given test is represented by one value (and hence offers no variability) we did not

Table 2.2: Summary of individual differences tests and measures. Numbers 1:13 label the individual differences measures included in statistical analyses (tests from which composite scores are composed are nested beneath each composite score). Mean values, standard deviations, and quartiles are reported for raw values. For interpretability, mean, standard deviations and ranges are reported for percentile rank, grade and age equivalent scores.

Domain	Skill Test	Raw scores							Equivalent scores				
		Mean	SD	Min	25%	50%	75%	Max	Mean	SD	Min	Max	Range (standardized)
Working Memory	1. Working Memory Composite	0	1.8	-5.1	-1.3	0.5	1.4	2.7	-	-	-	-	-2.9:1.5
	i. Operation Span Task	55.8	15	4	51.2	59.5	66	75	-	-	-	-	-3.4:1.3
	ii. Sentence Span Task	37.2	10.2	10	30.8	38.5	45	57	-	-	-	-	-2.7:1.9
Inhibitory control	2. Inhibitory Control Composite	0	2	-4.6	-1.2	-0.1	1.2	6.5	-	-	-	-	-2.3:3.2
	i. Stroop A	166.3	189.9	-478.3	21.9	185.2	305.5	600.4	-	-	-	-	-3.4:2.3
	ii. Stroop C	127.9	140.4	-291.5	35.7	124	216.5	472	-	-	-	-	-3:2.5
	iii. Stop Signal Task	236.7	71	92.5	200.8	226.6	266.1	568.3	-	-	-	-	-2:4.7
Phonological processing	3. Phonological Awareness Composite (CTOPP)[a]	101	13.1	69	93.5	101	112	127	52.9	27.26	2	97	-2.4:2
	i. Elision (CTOPP)[b]	30.1	4.3	15	29.8	31	33	34	8.13	2.61	1.4	9.7	-3.5:0.9
	ii. Word Blending (CTOPP)[b]	25.9	4.4	14	23	26	29	36	6.89	3.33	0.4	9.7	-2.7:2.3
	iii. Phonetic Isolation (CTOPP)[b]	25.9	2.8	18	24	26.5	28	30	4.35	2.78	1.4	9.7	-2.9:1.5
	4. Phonological Memory Composite (CTOPP)[a]	102.3	13.8	67	92	101.7	113	144	44.78	33.76	1	>99	-2.6:3
	i. Digit Span (CTOPP)[b]	21	3.1	12	19	21	23	27	7.85	2.49	0	9.7	-2.9:1.9
	ii. Non-word Repetition (CTOPP)[b]	18.9	2.7	12	17	19	21	24	6.53	3.15	0	9.7	-2.6:1.9
	5. Rapid Symbolic Naming Composite (CTOPP)[a]	101.4	12	49	95	101.4	107	128	51.35	27.27	<1	97	-4.4:2.2
	i. Rapid Digit Naming (CTOPP)[b]	11.6	2.5	7	10	11	13	27	8.7	1.65	1.4	9.7	-1.9:6.2
	ii. Rapid Letter Naming (CTOPP)[b]	12.4	2.1	8	11	12	14	21	8.74	1.48	2.7	9.7	-2:4
Word decoding	6. WJ Decoding Composite	0	1.9	-9.1	-0.4	0.6	1.2	1.7	-	-	-	-	-4.7:0.9
	i. Word Attack (WJ)[b]	27.8	4.9	0	27.9	29	30	32	12.28	4.65	0	18	-5.6:0.8
	ii. Letter-Word Identification (WJ)[b]	69.3	6.6	45	67	71	74	76	13.71	4.33	3.5	18	-3.7:1
Reading comprehension	7. Passage Comprehension (WJ)[b]	38.6	4.1	27	36	39	42	46	13.4	5.03	3.2	18	-2.8:1.8
Reading fluency	8. Reading Fluency (WJ)[b]	82.8	15.9	42	70.8	86.5	97	98	14.58	4.13	5.1	18	-2.6:1
Vocabulary knowledge	9. Vocabulary Knowledge Composite	0	1.9	-7.9	-1.1	0.3	1.2	2.8	-	-	-	-	-4.2:1.5
	i. Vocabulary (WASI)[d]	38.9	6.5	24	34	39	43.2	51	17.63	3.61	9.17	20	-2.3:1.9
	ii. Vocabulary (PPVT)[bc]	199.8	20.9	85	190.8	204.5	214.2	228	10.97	2.52	<k.0	>12.5	-5.5:1.3
Experience	10. Print Exposure Composite	0	1.9	-2.7	-1.5	-0.6	1.2	5.7	-	-	-	-	-1.4:3
	i. Author Recognition Test	13.6	10.7	0	5	11	21.2	42	-	-	-	-	-1.3:2.7
	ii. Magazine Recognition Test	11.1	8.4	-2	4	9	16	41	-	-	-	-	-1.6:3.6
	11. Age	21.2	2.2	16.1	19.6	21.3	22.8	26	-	-	-	-	-2.3:2.1
	12. Years in education	13.3	2.1	8	12	13	16	17	-	-	-	-	-2.6:1.8
Intellectual ability	13. Matrix Reasoning (WASI)[d]	20.9	4.1	7	18	21	24	28	16.28	4.94	6.17	20	-3.4:1.7

[a]Standardized percentile rank values are used here as the equivalent score.
[b]Grade equivalent of "second year of graduate school" is indicated as 18.
[c]Grade levels prior to first grade are indicated as k [month], where [month] indicates months into the kindergarten year.
[d]Age (year:month) is used here as the standardized equivalent score.

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 non-planar 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 independent models which were fitted 13 individual skill test scores across six different dependent variables (78 models in total) led to the issue of inflated Type I error rates. In the spirit of a recent paper by von der Malsburg & Angele (2015), we applied two kinds of family-wise multiple comparison corrections to the p-values in our proliferation of models. The first was the family-wise correction criterion. We used six dependent variables in our analyses (SFD, FOM, GD, TVT, RP, FRR), and defined as a family each set of the six models with the same individual differences measure included as a fixed effect. The probability of discovering two out of six effects with p-values < 0.05 is estimated at 0.033 (under the simplifying assumption that the dependent variables are independent from each other). Thus, we considered main effects or interactions that elicited two or more effects at the nominal significance level to pass this family-wise probability threshold: this is in line with the rule-of-thumb criterion recently tested by von der Malsberg and Angele (2015). The second type of correction method 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. These two false detection 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.

2.3 Results and discussion

The initial raw data set consisted of 5,397 trials. We removed trials in which the eye-tracking signal was lost, the target word was skipped (348 trials), was fixated on for the first time after gaze proceeded past the target word, was fixated on for more than 6 times 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 631 (11.7%) trials from the raw data set. The resulting final data set comprised of 4,766 valid trials.

For presentational purposes, we report our results as three 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 2a conducts a confirmatory analysis of the hypothesized interaction between semantic transparency and individual differences in reading experience. Finally, Analysis 2b is an exploratory study of the component skills of reading that are implicated in the processing of semantic transparency.

2.3.1 Analysis 1: main effects of semantic transparency

We fitted a model for each of the six eye-movement measures (listed in Part A of Table 2.1) as a response variable, with both LSA (Modifier-Compound and Head-Compound) scores included together as fixed predictor variables in each model. Along with all of the LSA measures, we also included the set of predictors listed in part B of Table 2.1 as fixed effects in our models. For number of trials per dependent variable (EM), see Part A of Table 2.1. Random effects were defined as stated in Statistical considerations. The resulting series of linear mixed-effect regression models followed this template structure,

$\text{EM}^i \sim$
LSA Modifier-Compound +
LSA Head-Compound +
Compound Frequency +
Left constituent Frequency +
Right constituent Frequency +
Left constituent Family Size +
Right constituent Family Size +
Compound Length +
Plausibility +
Word position +
Trial number +
random effects +
error,

where $i \in \{1 : 6\}$ is the index of the eye-movement measure.

No significant main effect (with a p-value < 0.05 after the application of the family-wise probability correction or false discovery rate correction criterion) was present for any of the LSA measures. Before the application of either correction criteria, one significant effect at the 0.05 threshold emerged for the Head-Compound Similarity in refixation probability, such that greater transparency reduced the probability of refixating on a word during the first pass of reading.

The procedure of this analysis 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. However, the results of this analysis did not provide any corroborating evidence in favour of semantic transparency as a co-determinant of compound processing, as no reliable main effects were observed out of either semantic transparency relation across the six eye-movement measures. We therefore conclude that Analysis 1 is consistent with the null effects reported in

Frisson et al. (2008) and Pollatsek and Hyönä (2005).

As reviewed earlier, Marelli & 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 LSA measure with the left and right constituent frequencies of a compound word. Both LSA measures were entered as a multiplicative interaction with both left and right frequencies 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 processing skill by examining the effect of semantic transparency as conditioned by level of exposure to printed material.

2.3.2 Analysis 2a: interactions of semantic transparency with reading experience

We hypothesized in the Introduction that reading experience plays an important role in compound processing, and especially in the individual ability to use orthographic cues towards discriminating between subtle differences in meaning. The chosen diagnostic for this role was whether the Print Exposure Composite score entered into statistically significant interactions with either of the LSA measures (see Appendix A for a full description of the Author Recognition Test and Magazine Recognition Test; Acheson et al., 2008).

Our effect of interest was the multiplicative interaction of semantic transparency (LSA) with individual exposure to print (Print Exposure Composite score). For each eye-movement variable listed in Part A of Table 2.1, we fitted a model which included an interaction of Print Exposure Composite scores with (i) LSA Modifier-Compound, and separately with, (ii) LSA Head-Compound. This resulted in 6 different models for each eye-movement measure. Along with each LSA by Print Exposure Composite interactions, we also included the set of predictors listed in part B of Table 2.1 as fixed effects in our models. For number of trials per dependent variable (EM), see Part A of Table 2.1. Random effects were included as described in Statistical considerations. The resulting series of linear mixed-effect regression models followed this template structure:

$\text{EM}^i \sim$
(LSA Modifier-Compound + LSA Head-Compound) * Print Exposure +
Compound Frequency +
Left constituent Frequency +
Right constituent Frequency +
Left constituent Family Size +
Right constituent Family Size +
Compound Length +
Plausibility +
Word position +
Trial number +
random effects +
error,

where $i \in \{1 : 6\}$ is the index of the eye-movement measure.

We present the results of statistical models after the application of either the family-wise probability criterion or the false discovery rate correction for multiple comparisons. An effect was detected after the application of the false discovery rate criterion (we henceforth report adjusted p-values for models that pass this criterion). Head-Compound LSA Similarity entered into a significant interaction with the Print Exposure Composite score on first-pass regression rate [$\hat{\beta}$ = -0.099; SE = 0.036; p = 0.032]. Table 2.3 provides a summary of the models in which the significant interaction between reading experience and estimates of semantic transparency was reliably present, while Table C.1 in Appendix C provides a full specification of this model.

The critical interactive effect demonstrates that readers with more experience of printed material were slightly less likely to execute a regressive saccade onto

Table 2.3: Summary of interactions between each LSA metric and scores of the Print Exposure Composite. Included are the significant models which passed either the family-wise probability or the false discovery rate p-value correction. Reported are regression coefficients and standard errors from models fitted to the entire dataset, as well as false discovery rate corrected p-values.

Eye-movement measure	Modifier-Compound x Print Exposure	Head-Compound x Print Exposure
Single fixation duration	ns	ns
First-of-many fixation duration	ns	ns
Refixation probability	ns	ns
Gaze duration	ns	ns
First-pass regression rate	ns	$\hat{\beta}$ = -0.099, SE = 0.036, p = 0.032
Total viewing time	ns	ns

compound words (after the first-pass of reading on that compound word), when the compound has a greater Head-Compound LSA Similarity value, 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*). However, the interaction pattern indicates that this facilitatory effect of transparency progressively reverses as reading experience decreases in the population sample. This reversal demonstrates that for participants with relatively low Print Exposure Composite scores, compounds with greater Head-Compound LSA Similarity scores instead elicited a higher probability of performing a regressive saccade as compared to compounds with low semantic transparency. Figure 2.3 depicts the partial effect of semantic transparency on first-pass regression rate, broken down by 10^{th}, 30^{th}, 50^{th}, 70^{th} and 90^{th} percentiles of the Print Exposure Composite score.

These results demonstrate that individual variability in exposure to print modulates the effect of semantic transparency in the predicted way of Pattern D of Figure 2.1, i.e., the benefit of the semantic boost is weaker than the discrimination cost. The processing benefit of greater semantic similarity between the head and the whole compound is counteracted by the cost of discrimination to a degree contingent on one's reading experience. The greater cost in less experienced individuals reverses the benefit from the semantic boost, while that benefit is still present in more experienced readers. Importantly, when averaged over all levels of reading experience, as in Analysis 1, the observed interactive "fan" pattern – with its counter-directed effects for extremes of the experiential continuum – collapses into a relatively flat line and appears as a null effect. The fact that an apparent null effect masks a reliable interaction, for this eye-movement measure, emphasizes the importance of considering individual differences in experimental research.

In summary, we confirm the hypothesized functional relationship between individual reading experience and the effect of semantic transparency on eye-movements. Moreover, the pattern of the interaction accords with the hypothesized processing trade-off pattern illustrated in Pattern D of Figure 2.1. This pattern suggests a joint effect of a semantic boost and meaning discrimination challenge during the processing of transparent compounds. This finding will be addressed in greater detail in the General Discussion. In the next analysis we expand the investigation to include, in

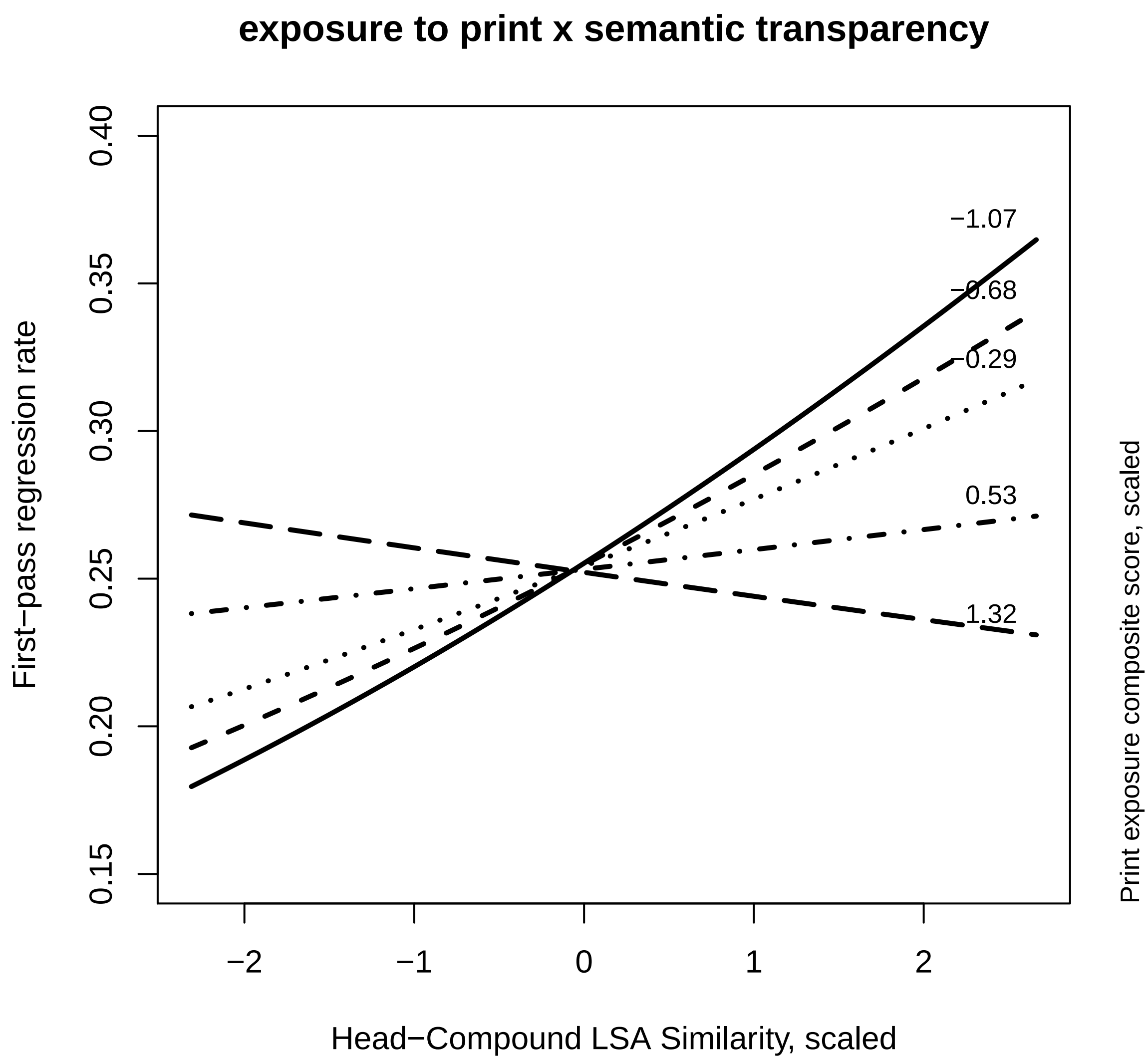

Figure 2.3: Interaction between Head-Compound LSA similarity (scaled) by Print Exposure Composite score (scaled) on first-pass regression rate. The lines plot the partial effect of Head-Compound LSA similarity for percentiles of the Print Exposure Composite score. Scaled values of the author recognition test scores are provided in the right margin and report the 10^{th} (solid line), 30^{th} (dashed line), 50^{th} (dotted line), 70^{th} (dotdash line) and 90^{th} (longdash line) percentiles of the Print Exposure Composite score.

an exploratory way, a wider range of individual differences scores.

2.3.3 Analysis 2b: interactions of transparency with individual differences measures

In this section we proceeded to test whether a greater range of verbal and other cognitive skills are engaged in semantic processing of compounds. We considered 13 different composite tests of verbal and non-verbal proficiency (including the previous analysis of the Print Exposure Composite score) which represent nine broader categories of verbal and non-verbal abilities. The format of the analysis proceeded as in Analysis 2a. Our diagnostic of the impact of the role of a given skill on the processing of compound transparency is whether it enters into a statistically significant interaction (after the family-wise correction measures) with either of the LSA scores. As there were 13 individual difference (ID) measures and six dependent measures (EM), we fitted the same model template to a total of 78 (13 x 6) different combinations of dependent measure and individual difference measure. As two measures of semantic transparency (LSA) are included in each model, there are 156 (2 x 78) different combinations of interactions between semantic transparency and individual differences tests across the eye-movement record. For number of trials per dependent variable (EM), see Part A of Table 2.1. Each model followed the same template as outlined in previous sections.

$\text{EM}^i \sim$
(LSA Modifier-Compound + LSA Head-Compound) * ID^j +
Compound Frequency +
Left constituent Frequency +
Right constituent Frequency +
Left constituent Family Size +
Right constituent Family Size +
Compound Length +
Plausibility +
Word position +
Trial number +

random effects +
error,

where $i \in \{1:6\}$ is the index of the eye-movement measure and $j \in \{1:13\}$ is the index of the skill test measures.

Out of all 156 interactions between semantic transparency and individual differences tests, 4 effects (3% of all potential effects) retained statistical significance after the application of either the family-wise probability or the false discovery rate false-detection measure. Table 2.4 summarizes the effects of those 4 models.

Table 2.4: Inferential statistics for the interactive effect (regression coefficient, standard error, t-value and p-value) and summary statistics (R^2, AIC, residual SD) for models where interactions between semantic transparency and individual differences measures reached statistical significance.

Skill test	Eye-movement	Transparency measure	$\hat{\beta}$	*SE*	t-value	p-value	R^2	model AIC	residual *SD*
Print Exposure Composite	FRR	Head-Compound LSA	-0.10	0.04	-2.78	0.032[a]	0.17	5378.30	1.00
Vocabulary Knowledge Composite	FRR	Head-Compound LSA	-0.10	0.04	-2.79	0.032[a]	0.17	5381.05	1.00
Years in education	FRR	Modifier-Compound LSA	-0.07	0.04	-1.98	0.048	0.17	5380.14	1.00
Years in education	TRT	Modifier-Compound LSA	-0.02	0.01	-2.87	0.025[a]	0.23	7877.17	0.51

[a]False discovery rate adjusted p-value

Both Modifier-Compound and Head-Compound LSA Similarity emerged as influential co-determiners of the eye-movement record. These transparency measures interacted with 3 unique component skills of reading out of a total of 13 different individual difference tests. As Figure 2.4 and Table 2.4 demonstrate, the tests that interacted with Modifier-Compound LSA Similarity and Head-Compound LSA Similarity were those that measure *experience with reading* directly (Print Exposure Composite, and Years in education) or have an obvious and well-established association with reading experience (Vocabulary Knowledge) (Cain & Oakhill, 2011; Cunningham and Stanovich, 1997).

Specifically, significant interaction effects were observed between Modifier-Compound LSA Similarity and years in education in first-pass regression rate [$\hat{\beta}$ = -0.074; *SE* = 0.037; p = 0.048] and also in total reading time [$\hat{\beta}$ = -0.023; *SE* = 0.008; p = 0.025]. Moroever, in addition to the interaction with Print Exposure Score reported in Analysis 2a, Head-Compound LSA Similarity also interacted with Vocabulary knowledge in first-pass regression rate [$\hat{\beta}$ = -0.099; SE = 0.035; p = 0.032]. All p-values are given after the family-wise correction for multiple comparisons. Tables C.2, C.3, and C.4 in Appendix C provide the full specifications of these models.

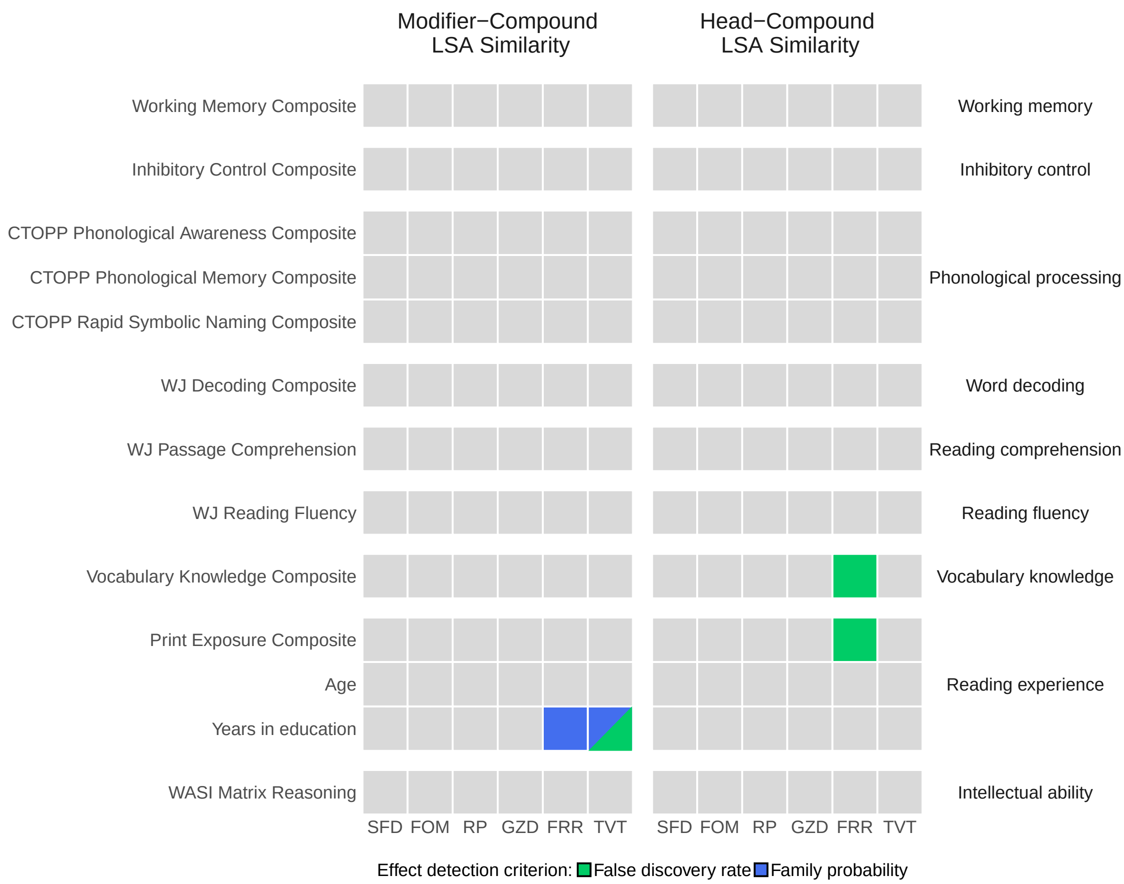

Figure 2.4: Heatmap visualizing presence of significant (after correction) interaction effects between individual differences measures (y-axis) and both Modifier-Compound LSA similarity (left panel) and Head-Compound LSA similarity (right panel) throughout the time-course of reading (eye-movements on x-axis).

The qualitative patterns of these interactions were consistent with the results reported in Analysis 2a. The difference in durational measures and regression rates between relatively experienced and inexperienced readers was maximal for highly transparent compounds. Greater semantic transparency served to increase the probability of regressive saccades amongst individuals with the least reading experience, fewer years in education, and least vocabulary knowledge. However, the same lexical property of high semantic transparency yields shorter total reading time fixation durations and lower regressive saccade rates in more experienced and proficient readers. Figure 2.5 illustrates these patterns by plotting the partial effects of semantic transparency by 10^{th}, 30^{th}, 50^{th}, 70^{th} and 90^{th} percentiles of educational experience and vocabulary knowledge scores.

Our results are compatible with the prediction that reading experience, as gauged by exposure to print, vocabulary knowledge and years of education, plays a dominant role during the processing of semantic transparency. On the whole, these patterns suggest a trade-off of the effect of conjunctive semantic activation and meaning discrimination during the processing of transparent compounds. However, when broken down by transparency relation type, the patterns of the interactive effects indicate a variation in the degree to which the facilitation of semantic boosting can be offset by the magnitude of the discrimination cost. This difference can be observed by comparing the plots depicted in Figure 2.3 and Panel (c) of Figure 2.5 with Panels (a) and (b) of Figure 2.5.

Firstly, the interactions between Head-Compound semantic transparency with reading experience and vocabulary knowledge on first-pass regression rates match the hypothesized pattern illustrated in Pattern D of Figure 2.1. In the scenario of Pattern D, the strength of the semantic transparency boost is offset by the cost of meaning discrimination for all but very proficient readers. This cost increases as individual reading experience becomes progressively lower, such that semantically transparent compounds inhibit word processing relative to opaque compounds, and especially so in relatively inexperienced readers.

Secondly, the interactive effect of Modifier-Compound semantic transparency with years in education on first-pass regression rate and total reading times most closely resembles Pattern C of Figure 2.1. In Pattern C, as in Pattern D, a discrimination cost is lower for more experienced readers and so the semantic boost remains present for these readers. However, in Pattern C, the strength of the semantic transparency boost is progressively attenuated only until it is ‘levelled out’ amongst the least experienced readers. This pattern characterizes the situation in which the force of

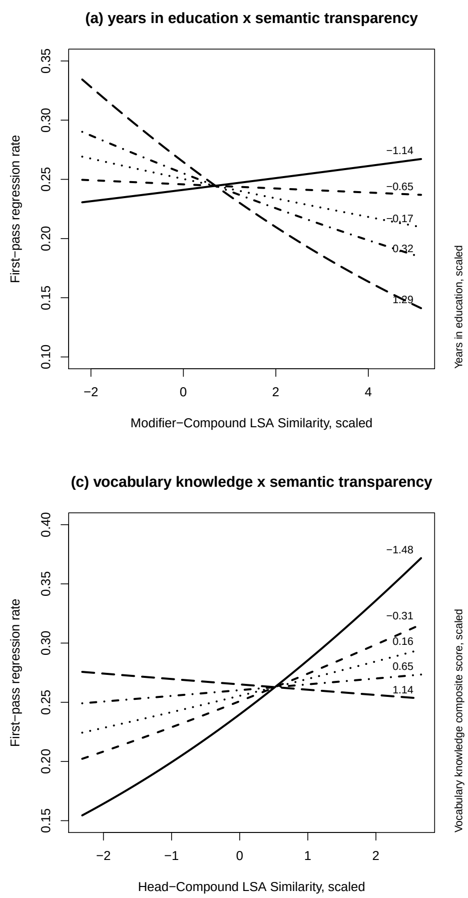

Figure 2.5: Interaction effects between semantic transparency and individual differences tests. *Upper panels*: partial effects of Modifier-Compound LSA similarity (scaled) modulated by years in education, on first-pass regression rate (a) and on total reading time (b). *Lower panel*: partial effects of Head-Compound LSA similarity (scaled) on first-pass regression rate, modulated by Vocabulary Knowledge Composite score (c). Scaled values of each test score are provided in the right margin and report, where available, the 10^{th} (solid line), 30^{th} (dashed line), 50^{th} (dotted line), 70^{th} (dotdash line) and 90^{th} (longdash line) percentiles of each test score.

the discrimination cost is roughly equal to that of a semantic boost. The implications of these two trade-off patterns and their apparent association with separate semantic transparency measures will be addressed in the General discussion.

2.4 General discussion

This paper 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 due to 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, i.e., the effort that is required to dissociate a target meaning of a compound from all other meanings (including those of constituents) that are activated by largely overlapping orthographic cues. The Naive Discriminative Reader (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 sub-strings that each denote distinct meanings of their own. For example, when reading a 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*. It so happens that, according to LSA, the words *rain* and *rainstorm* occur in very similar lexical environments (i.e., are more substitutable), as do the words *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, i.e., 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 co-activation of associated lexical representations, leading to an advantage in processing. However, under the NDR account, the same compound would trigger an increased cognitive load due to the requirement of discriminating between closely related meanings. In Figure 2.1, 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 the amount of individual exposure to print (as a proxy of one's reading experience) will be influential in that more experienced readers will be more proficient in resolving the discriminability challenge of transparent compound word processing (Analysis 2a). We showed that the effects of language experience are most likely to be demonstrated in the community population that we examined, where reading experience is heterogeneous, as compared to a convenience participant pool of university students (see Figure 2.2). In Analysis 2b, we explored a larger space of 13 component skills of reading (including print exposure) to establish whether other verbal and cognitive measures are preferentially implicated in semantic processing.

Analysis 1 showed no evidence for the main effect of semantic transparency measures (gauged by Modifier-Compound, and Head-Compound LSA scores) in the eye-movement record. Further analyses (Analysis 2a and 2b) revealed that the null effects of Analysis 1 represent the collapse of a rich interactive pattern into a flat line. This alone might account for the discrepant findings in the prior literature, where the presence or absence of the critical effect of semantic transparency would depend on the prevalent proficiency of their participants.

Based on the reported results, we outline three empirical contributions that this paper offers:

1. Individual reading experience and other component skills of reading modulate the effect of semantic transparency during the visual identification of compound words.

2. Experience with and exposure to language affects the ability to discriminate between meanings during compound word reading, such that the discrimination cost is lower in experienced readers. This finding is predicted by the Naive Discriminative Reader account of word recognition (Baayen et al., 2011). Furthermore, the semantic transparency of the modifier-compound (e.g., *rain-rainstorm*) and head-compound (e.g., *storm-rainstorm*) relations facilitated compound word reading.

3. The trade-off between the processing benefit and cost differed across each transparency relationship, Modifier-Compound vs. Head-Compound.

We will now address each of these findings in turn. Firstly, across all interactions, we found a gradient processing outcome. Readers with impoverished reading experience, limited vocabulary knowledge and less years in education exhibited either inhibited or equivalent processing when reading highly transparent compounds as compared to opaque compounds. For readers with greater reading experience, vocabulary knowledge and more years in education, greater transparency facilitated processing throughout. Thus, interactions revealed that highly transparent compounds caused the maximal divergence in the patterns of reading behaviour across the span of reading experience. The visual recognition of transparent compounds was facilitated among more experienced readers, whereas transparent compounds inhibited the processing of less experienced readers. The present results therefore demonstrate unequivocal evidence for the pivotal role of individual differences on the processing of semantic transparency during compound word reading.

As we argue in the Introduction, neither the "semantic boost" account nor the NDR model 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 skill), while the latter only predicts a bigger or smaller cost (and no benefit) for resolving the discrimination challenge. The exact trade-off patterns that we observe in Figures 2.3 and 2.5 match the theoretical prediction illustrated in either Panels C or D of Figure 2.1. Under these scenarios, the pattern observed for transparent compounds can only be true if, for readers with impoverished language skill, the facilitatory boost caused by conjunctive activation is weaker than or equal to the inflation in processing time associated with meaning discrimination. Thus, for less proficient readers, the resulting effect is that of relatively inflated processing times for the most transparent compounds. However, for skilled readers, the cost of discriminating transparent compounds is lower and does not offset the advantage afforded by the semantic boost. This reduces the cognitive load implicated in the processing transparent compounds as compared to opaque compounds, and is the reason for the observed facilitatory effect of relatively transparent compounds. As explained above, opaque compounds such as *bootleg* or *hogwash* offer no boost and impose little discrimination cost and so the difference between more or less skilled readers in processing opaque compounds is less related to their overall reading proficiency.

Our second point asks how exactly might the cost of conceptual discrimination 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 these modifications are driven by the accumulation of language experience over time (MacDonald, 2013; Ramscar, Hendrix, Shaoul, Milin & Baayen, 2014).

The above conjecture is supported by the phenomenon that we have observed empirically. Analysis 2b reveals that all of the interactive effects involved measures of exposure to print, vocabulary knowledge or years in education, that is, measure that straightforwardly tap into reading experience (see Figure 2.4 and Table 2.4). Unsurprisingly, it is not the case that less directly related skills such as phonological awareness, intellectual ability, grapheme-to-phoneme decoding ability or working memory are engaged during the processing of semantic transparency. We assume that this finding is a reasonable indicator of the dominant role of printed language experience during the processing of semantic transparency, the role that we hypothesized in line with the NDR account.

Ours is not the first study to report effects of inhibition associated with co-activations of closely related concepts. 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). For example, Mirman and Magnuson (2009) reported that refixation likelihood was increased among presentations of words defined as near semantic neighbours compared to words that are distant semantic neighbours (semantic neighbourhood density calculated over counts of semantic feature ratings of words). Additionally, in a lexical decision study, Mirman and Magnuson (2008) reported slowed processing of words with near neighbours and speeded processing to words with distant semantic neighbours. Although methodologically distinct from the current study, these studies report results that are analogous to the effects we observe for the least skilled readers when reading highly transparent compounds.

Thirdly, 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 that is at work. For both transparency measures, a higher degree of semantic similarity facilitated the visual recognition of skilled readers. However, for least skilled readers, Panels (a) and (b) of Figure 2.5, indicate in a weaker effect of transparency. This interactive effect (analogous to Pattern C of Figure 2.1) may indicate that among the least skilled readers, high Modifier-Compound transparency exerts a cost of discriminating between closely related meanings that is roughly equal to the strength of conjunctive semantic boosting.

Conversely, the pattern for the interactive effect of Head-Compound similarity and reading skill in Figure 2.3 and Panel (c) of Figure 2.5 shows that the least skilled readers are demonstrably inhibited when processing compounds with higher transparency relative to opaque compounds, i.e., Pattern D of Figure 2.1. Thus, for the least skilled readers, we may conclude that the inhibitory force of the discrimination challenge is more harmful specifically when processing compounds with a high semantic overlap between the head of the compound and the whole compound word.

The difference in the strengths of counteracting forces of discrimination and boosting may point to an important function of the compound head during semantic processing. From a structural perspective, the head of a compound is the conceptual kernel of the whole word, thus it is reasonable to assume that its relationship with the whole compound word and the concept that modifies it is likely to pose a more acute meaning discrimination problem during reading. While this may be true, 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) nor the issue of compositionality (e.g., Schmidtke, Kuperman, Gagné, & Spalding, 2016). To sum up, our results do not only enable us to posit the simultaneous operation of two cognitive mechanisms, but also identify their relative strengths across the range of individual variability and transparency measures.

Finally, it is clear that future investigation is needed to more precisely track the exact influence of reading experience on the reader's sensitivity to co-occurrence based metrics such as LSA. For example, our LSA scores were calculated over a corpus of text that is not a precise representation of the lexical knowledge of either poor or skilled readers. A future inquiry may be able to incorporate LSA scores that are computed over texts which are more characteristic of the lexicons of individuals of varying reading skill. In addition, we have used experience with printed materials and other component skills of reading proficiency as an index of greater discrimination skill. Even though NDR predicts that discrimination improves with reading experience, we did not independently nor directly test individual differences in semantic discrimination skill. Future work will need to examine this skill more finely and demonstrate its relationship with reading experience, and also its co-influence with compound semantics, as predictor of compound word recognition.

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 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*(1), 278–289.

Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2009). The cognitive and behavioral characteristics of children with low working memory. *Child Development*, *80*(2), 606–621.

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.

Andrews, S. (2012). Individual differences in skilled visual word recognition and reading. In J. Adelman (Ed.), *Visual word recognition volume 2: Meaning and context, individuals and development.* (pp. 151–172). Sussex UK: 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*(3), 279–296.

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *The Quarterly Journal of Experimental Psychology Section A*, *58*(6), 1065–1086.

Baayen, H., Kuperman, V., Shaoul, C., Milin, P., Kliegl, R., & Ramscar, M. (2013). *Decomposition makes things worse. A discrimination learning approach to the time course of understanding compounds in reading.* (Unpublished manuscript)

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

Baayen, R. H. (2014). Experimental and psycholinguistic approaches to studying derivation. In R. Lieber & P. Štekauer (Eds.), *Handbook of derivational morphology* (pp. 95–117). Oxford UK: 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*(4), 390–412.

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*(3), 438–481.

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

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*(3), 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*.

Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, *74*(4), 506–521.

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)*, 289–300.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*(2), 647–663.

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*(1), 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*(4), 977–990.

Cain, K., & Oakhill, J. (2011). Matthew effects in young readers reading comprehension and reading experience aid vocabulary development. *Journal of Learning Disabilities*, *44*(5), 431–443.

Cain, K., & Oakhill, J. (2012). Reading comprehension development from seven to fourteen years: Implications for assessment. In J. P. Sabatini, â. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability.* (pp. 59–76). Lanham, MD: Rowman & Littlefield Education.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, *96*(1), 31.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*(1), 42–100.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental psychology*, *33*(6), 934–945.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466.

Deater-Deckard, K., Mullineaux, P. Y., Petrill, S. A., & Thompson, L. A. (2009). Effortful control, surgency, and reading skills in middle childhood. *Reading and Writing*, *22*(1), 107–116.

De Beni, P., Palladino, F., Pazzaglia, C., & Cornoldi, R. (1998). Increases in intrusion errors and working memory deficit of poor comprehenders. *The Quarterly Journal of Experimental Psychology: Section A*, *51*(2), 305–320.

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

Demagistri, M. S., Richards, M. M., & Juric, L. C. (2014). Incidence of executive functions on reading comprehension performance in adolescents. *Electronic Journal of Research in Educational Psychology*, *12*(2), 343–370.

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). Cambridge University Press.

Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first-and second-language learners. *Reading Research Quarterly*, *38*(1), 78–103.

Dunn, L. M., & Dunn, L. M. (1981). *Manual for peabody picture vocabulary test-revised: PPVT-R.* Circle Pines: MI: American Guidance Service.

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, *9*(2), 167–188.

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*(1), 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*(6), 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*(4), 684–691.

Fiorentino, R., & Fund-Reznicek, E. (2009). Masked morphological priming of compound constituents. *The Mental Lexicon*, *4*(2), 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*(1), 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. (2004). Effect of relation availability on the interpretation and access of familiar noun–noun compounds. *Brain & Language*, *90*(1), 478–486.

Gagné, C. L., & Spalding, T. L. (2006). Conceptual combination: Implications for the mental lexicon. In *The representation and processing of compound words.* (pp. 145–168). Oxford UK: Oxford University Press.

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*(1), 20–35.

Gagné, C. L., Spalding, T. L., Figueredo, L., & Mullaly, A. C. (2009). Does snow man prime plastic snow? The effect of constituent position in using relational information during the interpretation of modifier-noun phrases. *The Mental Lexicon*, *4*(1), 41–76.

Garavan, H., Ross, T., & Stein, E. (1999). Right hemispheric dominance of inhibitory control: an event-related functional MRI study. *Proceedings of the National Academy of Sciences*, *96*(14), 8301–8306.

Gernsbacher, M. A. (1993). Less skilled readers have less efficient suppression mechanisms. *Psychological Science*, *4*(5), 294–298.

Goldstein, S., Naglieri, J. A., Princiotta, D., & Otero, T. M. (2014). Introduction: A history of executive functioning as a theoretical and clinical construct. In S. Goldsein & J. A. Naglieri (Eds.), *Handbook of Executive Functioning* (pp. 3–12). Springer.

Graesser, A. C., & Clark, L. F. (1985). *Structures and procedures of implicit knowledge: Advances in discourse processes.* New York City: New York: Ablex Publishing.

Huettig, F., & Altmann, G. T. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), 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*(4), 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*(1), 65–80.

Hulme, C., Roodenrys, S., Brown, G., & Mercer, R. (1995). The role of long-term memory mechanisms in memory span. *British Journal of Psychology*, *86*(4), 527–536.

Hyönä, J. (2012). The role of visual acuity and segmentation cues in compound word identification. *Frontiers in Psychology*, *3*, 188. Retrieved from `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*(4), 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*(4), 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: Elsevier.

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*(4), 1004–1019.

Kegel, C. A., & Bus, A. G. (2014). Evidence for causal relations between executive functions and alphabetic skills based on longitudinal data. *Infant and Child Development*, *23*(1), 22–35.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* Cambridge: Cambridge University Press.

Kuperman, V. (2013). Accentuate the positive: Semantic access in English compounds. *Frontiers in Psychology*, *4*(203). Retrieved from `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*(7), 939–966.

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

Kuperman, V., & Van Dyke, J. A. (2011b). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language*, *65*(1), 42–73.

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 802–823.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240.

Libben, G. (1998). Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment. *Brain & Language*, *61*(1), 30–44.

Libben, G. (2014). The nature of compounds: A psychocentric perspective. *Cognitive Neuropsychology*, *31*(1-2), 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*(1), 50–64.

Logan, G. (1994). On the ability to inhibit thought and action: A users' guide to the stop signal paradigm. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 189–239). San Diego: CA: Academic Press.

Long, D. L., & Golding, J. M. (1993). Superordinate goal inferences: Are they automatically generated during comprehension? *Discourse Processes*, *16*(1-2), 55–73.

Long, D. L., Oppy, B. J., & Seely, M. R. (1997). Individual differences in readers' sentence and text-level representations. *Journal of Memory and Language*, *36*(1), 129–145.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*, 226. Retrieved from `http://dx.doi.org/fpsyg.2013.00226`

Mandera, P., Keuleers, E., & Brysbaert, M. (in press). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*.

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*(6), 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*(4), 644–664.

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*(1), 65–79.

Mirman, D., & Magnuson, J. S. (2009). Dynamics of activation of semantically similar concepts during spoken word recognition. *Memory & Cognition*, *37*(7), 1026–1039.

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*(4), 496–512.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*(4), 357–383.

Perfetti, C., & Adlof, S. M. (2012). Reading comprehension: A conceptual framework from word meaning to text meaning. In J. P. Sabatini, â. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability.* (pp. 3–20). Lanham, MD: Rowman & Littlefield Education.

Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition.* (pp. 145–174). Hillsdale, NJ: Lawrence Erlbaum Associates.

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook.* (pp. 227–247). Oxford, UK: Blackwell.

Pham, H., & Baayen, R. H. (2013). Semantic relations and compound transparency: A regression study in CARIN theory. *Psihologija*, *46*(4), 455–478.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in s and s-plus.* New York: Springer.

Pollatsek, A., & Hyönä, J. (2005). The role of semantic transparency in the processing of Finnish compound words. *Language and Cognitive Processes*, *20*(1-2), 261–290.

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

Radach, R., & Kennedy, A. (2013). Eye movements in reading: Some theoretical context. *The Quarterly Journal of Experimental Psychology*, *66*(3), 429–452.

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

Ramscar, M., Dye, M., Popick, H. M., & O'Donnell-McCarthy, F. (2011). The enigma of number: Why children find the meanings of even small number words hard to learn and how we can help them do better. *PLoS ONE*, *6*(7), e22501. Retrieved from `http://dx.doi.org/10.1371/journal.pone.0022501`

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*(1), 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*(6), 909–957.

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*(4-5), 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*(6), 1090–1098.

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

Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist*, *43*(3), 151.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Crofts.

Sabatini, J., Albro, E., & O'Reilly, T. (2012). *Measuring up: Advances in how we assess reading ability.* Lanham, MD: Rowman & Littlefield Education.

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*(3), 529–567.

Schmidtke, D., Kuperman, V., Gagné, C. L., & Spalding, T. L. (2016). Competition between conceptual relations affects compound recognition: the role of entropy. *Psychonomic Bulletin & Review*, *23*(2), 556–570.

Share, D. L. (2008). Orthographic learning, phonological recoding, and self-teaching. In R. Kail (Ed.), *Advances in child development and behavior* (pp. 31–81). New York City: NY: Elsevier.

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, *24*(4), 402–433.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Burgess, S., & Hecht, S. (1997). Contributions of phonological awareness and rapid automatic naming ability to the growth of word-reading skills in second-to fifth-grade children. *Scientific Studies of Reading*, *1*(2), 161–185.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., & Garvan, C. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology*, *91*(4), 579–593.

Turney, P. D., Pantel, P., et al. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, *37*(1), 141–188.

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

van den Broek, P., & Espin, C. A. (2012). Connecting cognitive theory and assessment: Measuring individual differences in reading comprehension. *School Psychology Review*, *41*(3), 315–325.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, *45*(1), 2–40.

Verhoeven, L., van Leeuwe, J., & Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading*, *15*(1), 8–25.

von der Malsburg, T., & Angele, B. (2015). False positive rates in standard analyses of eye movements in reading. *arXiv preprint arXiv:1504.06896*.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processing: CTOPP.* Austin, TX: Pro-Ed.

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).

Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence (WASI).* San Antonia: TX: Psychological Corporation.

Whitford, V., & Titone, D. (2014). The effects of reading comprehension and launch site on frequency–predictability interactions during paragraph reading. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1151–1165.

Woodcock, R. W., McGrew, K., & Mather, N. (2001). *Woodcock-Johnson tests of achievement.* Itasca, IL: Riverside Publishing.

Zeno, S. M. (1995). *The educator's word frequency guide.* New York: Touchstone Applied Science Associates.

Zwitserlood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes*, *9*(3), 341–368.

Appendices

Appendix A: Descriptions of individual differences assessments

This appendix provides a detailed description of the tests of verbal and non-verbal skills that are summarized in Table 2.2. The skill test descriptions are organized by each domain. Composite scores were created by scaling and then summing the scores of each test that is included in the given composite.

1. *Working memory*: Working memory capacity characterizes the degree to which information is actively maintained and retrieved from long-term memory (cf. Cowan, Elliott, Saults, Morey, Mattox, Hismjatullina, & Conway, 2005; Alloway, Gathercole, Kirkwood, & Elliott, 2009). Working memory is hypothesized to play on important role in reading proficiency and is supported by findings indicating that individuals who exhibit suppressed abilities in reading comprehension also experience difficulties in working memory tasks that require the storage and processing of verbal materials (Cain, Oakhill, & Bryant, 2004; Daneman & Carpenter, 1980; Hulme, Roodenrys, Brown, & Mercer, 1995). In order to target individual differences in working memory we administered two tasks. The first is the Operation Span Task in which participants are asked to memorize letters while solving simple math problems. Participants perform a simple mathematical verification and then see a letter. A recall test follows a number of theses mathematical verification/letter pairs. For the recall component, participants are asked to choose the letters that were presented from a list of 12 possible letters, doing so in the order they were presented. The partial score is the number of correctly recalled letters which were recalled in the correct position, even if the entire trial was not correct. The second test of working memory is the Sentence Span Task is another working memory task. In this task participants listen to statement and decide whether the statement is true or false. After listening to 2-6 statements, participants are asked to recall the last word of each statement. All correctly recalled words are given one point, except if the first word recalled was the last word in the list of statements, in which case, they receive a score of zero for that word. The

Working Memory Composite is composed of the Operation Span Task and the Sentence Span Task.

2. *Inhibitory control*: Together with working memory, inhibition is a cognitive process that is part of a family of functions that together fall under the umbrella term of executive functioning (cf. Goldstein, Naglieri, Princiotta, & Otero, 2014). Inhibitory control is defined as an individual's capability to suppress irrelevant or interfering stimuli or impulses (Garavan, Ross, & Stein, 1999). Research has shown that inhibitory control predicts skills that are fundamental to reading, such as orthographic-phonological decoding and phonological awareness (Blair & Razza, 2007). Moreover, individual differences studies have revealed that attentional and inhibitory control predict performance not only in tasks targeting low-level reading processes, such as word-pseudoword discrimination tasks (Deater-Deckard, Mullineaux, Petrill, & Thompson, 2009) and orthographic knowledge (Kegel and Bus, 2014), but also tasks involving higher-order cognitive processing during reading such as comprehension ability (De Beni, Palladino, Pazzaglia, & Cornoldi, 1998; Demagistri, Richards, & Juric, 2014; Gernsbacher, 1993).

   The Inhibitory Control Composite is composed of the scores from three tasks to tap into the inhibitory control processes. Firstly, we administered two separate versions of the Stroop paradigm (Stroop, 1935), a group of tasks in which participants are asked to either name the colour of the font of a given stimulus or read the word that appears on the screen. In Stroop A, participants were required to name the colour of the font of the word when the stimuli is either a same or different colour word (e.g., the word *blue* presented in red font vs. the word *red* presented in red font). In Stroop C, participants are required to name the colour of the font when the stimuli presented is either not a word, where the colour matches the orthographic string, or when the stimuli presented is a different colour word (e.g., the orthographic string '@@@@@' presented in red font vs. the word *blue* presented in red font). In both tasks, participants are required to say the correct response out loud and then press the space bar. Responses are measured in milliseconds and any reaction times (RTs) 2 SD from the mean are discarded. After trimming, the mean for each condition is taken for each participant.

Secondly, we ran the Stop Signal Task (Logan, 1994). This task is designed to test the ability to override a habituated response. Participants build up a prepotent response to categorize words in a particular way. They are then asked to withhold their responses on a small proportion of trials during when they hear an auditory signal. In the task, participants are presented with either an X or an O and instructed to press either the left button or right button, respectively on a controller. They are asked to do so as quickly as they can, however if they hear a tone they should not press either button. The participants are instructed to not wait for the tone before making a response, but if they hear the tone before making a response they are asked to make a response. The dependent measure of interest is the mean Stop Signal reaction time (SSRT). The SSRT is computed as the mean response latency of every correct go response, i.e., the mean delay between onset of go signal and the onset of the stop signal.

3. *Phonological awareness*: Phonological processing plays an important role in reading ability, and has long been recognized as a skill that is absent or impoverished in non-skilled readers (Rayner et al., 2012). Increased phonological processing is characterized by the awareness of and capacity to form correspondences between graphemes and phonemes. Competency in phonological processing is demonstrated by success in utilizing these correspondences to efficiently recode written orthographic symbols into the sounds for which they stand during reading. Previous research has shown that deficits in the attainment and utilization of phonological decoding skills is one of the strongest predictive factors of reading difficulty, especially in developing readers (Torgesen, Wagner, Rashotte, Rose, Lindamood, Conway, & Garvan, 1999; Vellutino, Fletcher, Snowling, & Scanlon, 2004).

   To assess phonological processing we considered three composite scores from the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 1999) - Phonological Awareness Composite, Phonological Memory Composite, Rapid Symbolic Naming Composite - which target the separate abilities of phonological awareness, phonological memory and rapid naming respectively. The Phonological Awareness Composite comprises of the standard scores of three sub-tests, which assess skills in elision, word blending and phonetic isolation/segmentation. The Phonological Memory Composite combines the scores of a digit span task and a non-word repetition task. Lastly, the Rapid Symbolic Naming Composite consists of the combined standard scores of two sub-tests, which separately assess aptitude in rapid digit naming and rapid letter naming.

4. *Word decoding*: The ability to reliably and efficiently convert graphemes to phonemes is considered a fundamental component skill of reading proficiency (Ehri, 2005; Share, 2008). In order to target performance in the specific skill of fluid word decoding, we administered two standardized tests that measure accuracy of word naming during reading. These tests were drawn from the Woodcock-Johnson (WJ) Test of Cognitive Abilities III (Woodcock, McGrew, & Mather, 2001). The Letter-Word Identification sub-test requires participants to identify letters in bold type and then read a list of words of increasing difficulty. Participants are required to read aloud a list of words without any contextual aid. There are 76 potential trials. In order to establish a baseline score, participants begin halfway through the list and move backwards if an error is made. Participants advance through the list until they make six consecutive errors. The Word Attack sub-test measures a participant's ability to apply grapheme to phoneme conversion skills. This sub-test requires participants to produce sounds for a small set of single letters and then pronounce nonsense words of increasing complexity. The pseudowords are constructed such that they are phono-tactically legal in English, yet do not bear any meaning. There are a total of 32 trials in this subtest and participants begin halfway through the list moving backwards until six consecutive trials are read correctly. Participants then proceed from the starting point until 6 errors are made in a row. We combined the scores of the Letter-Word Identification sub-test and th Word Attack Subtest to create the WJ Decoding Composite Score.

5. *Reading comprehension and fluency*: According to Kintsch (1988), reading comprehension entails constructing a coherent and accurate memory representation of information encoded in a passage of text. Based on developmental insights (Cain & Oakhill, 2011; Sabatini, Albro & O'Reilly, 2012; van den Broek & Espin, 2012), cognitive models of reading comprehension posit that text comprehension is facilitated by the ability to extract the meanings of individual words and subsequently integrate these meanings with the local context in order to form a semantic whole (Perfetti & Adlof, 2012). This process is achieved as the reader continually updates the meaning representation of a text by connecting elements in the discourse and making inferences between these elements and wider world knowledge (Graesser & Clark, 1985; Long & Golding, 1993; Long, Oppy, & Seely, 1997).

   To target comprehension and reading fluency skill, we administered two subtests from the Woodcock-Johnson Test of Cognitive Abilities III (Woodcock et al. 2001); Passage Comprehension and Reading Fluency. Initial stages of the Passage Comprehension test require participants to match a picture symbol with an actual picture. The next list of items requires students to match a short phrase to the correct picture out of a choice of 3 pictures. The majority of items require a student to provide a missing word to sentences and paragraphs of increasing complexity. The final subtest from the Woodcock-Johnson Test of Cognitive Abilities III is Reading Fluency. In this task, participants have 3 minutes to read through a list of relatively short, easy sentences and decide if they are true. Reading fluency score is calculated by summing the number of sentences correctly marked as true or false.

6. *Reading experience*: As outlined earlier, the quality of word representations varies within and across individual lexicons. Perfetti (2007) argues that lexical quality develops when lexical representations are reinforced via repeated exposure to oral and printed forms of words. Thus, in order to gauge individual variability in the stability of participants' orthographic representations, we administered an updated version of Stanovich and West's (1989) Author Recognition Test (ART; Acheson et al., 2008). The ART elicits information about the reading habits of individuals via an author identification task. This

test estimates exposure to printed literature. Participants were presented with an intermixed list of 130 real and false author names from 20th Century literature printed on a single sheet of paper. 65 names are real authors (e.g., *George Orwell*) and the remaining 65 names are foils (e.g., *Beatrice Dobkin*). Participants are rewarded with one point for every author they guess correctly and penalized by one point for every author they guess incorrectly. The highest mark that could be achieved was 65. We also administered an adaptions of the ART; the Magazine Recognition (MRT), which substitutes author names with current magazine titles. In the present study, the ART and MRT test scores were combined to form the Print Exposure Composite Score.

7. *Vocabulary knowledge*: Another component factor of reading skill is vocabulary knowledge. Vocabulary knowledge is strongly associated associated to comprehension and phonological decoding skills. The relationship is straightforward: as vocabulary knowledge increases, so does phonological decoding and reading comprehension (Cain & Oakhill, 2011; Verhoeven, van Leeuwe, & Vermeer, 2011). The relationship is schematized in Perfetti's (1992) restrictive-interactive model in which extensive and precise lexical knowledge, consisting of high quality orthographic, phonological, and semantic representations, provide the foundation for skilled reading behaviour. The idea is that high lexical quality is advantageous when readers are confronted with new words which require new connections between orthographic forms and meanings (see also Perfetti, 2007). In skilled readers, these "crisp" lexical representations facilitate acquisition of new words by freeing-up processing resources that can be allocated to the task of learning new words. For example, it has been found that children with small vocabularies exhibit disrupted comprehension skill and recall of text (Beck, Perfetti, & McKeown, 1982; Perfetti, Landi, & Oakhill, 2005), while readers with larger vocabularies profit from larger lexicons during reading comprehension (Droop & Verhoeven, 2003; Torgesen, Wagner, Rashotte, Burgess, & Hecht, 1997).

   Vocabulary knowledge was assessed using the Peabody Picture Vocabulary Test 4th Edition (PPVT; Dunn & Dunn, 1981). The PPVT is administered such that for each trial the participant is presented with an array of four pictures.

The examiner orally produces a word which describes one of the pictures and the participant is asked to point out the picture that the word describes. We also measured vocabulary knowledge by administering the Vocabulary subtest of the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) - 2nd Edition. The WASI Vocabulary subtest is designed to to measure word knowledge and verbal concept formation. This subtest includes three picture items and 28 verbal items. For picture items, the participant is required to name the object that is presented visually. For verbal items, the participant is asked to define words that are presented both visually and orally. We combined the WASI Vocabulary subtest and the PPVT combined to create the Vocabulary Composite Score.

8. *Intellectual ability*: The Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) is a standardized, and norm-referenced individually administered test of intellectual ability. The WASI is composed of four subtests: performance in Vocabulary (discussed above) and Similarities form a Verbal Intelligence Quotient, and the combined scores in the Block Design and Matrix Reasoning tasks create a (nonverbal) Performance Intelligence Quotient (PIQ). Together, the four subtests form the Full Scale Intelligence Quotient (FSIQ), which was used in this study as an indicator of general intellectual functioning. As a measure of intellectual ability, we considered the Matrix Reasoning subtest.

Appendix B: Correlation matrices

Table B.1: Correlation matrix of lexical predictor variables.

	1.	2.	3.	4.	5.	6.	7.	8.
1. Modifier-Compound LSA Similarity								
2. Head-Compound LSA Similarity	0.16**							
3. Compound Frequency	-0.16**	-0.23***						
4. Compound Length	-0.06	-0.04	-0.01					
5. Left constituent Frequency	0.04	0.00	0.11*	0.08				
6. Right constituent Frequency	-0.14**	0.12*	0.05	-0.07	0.01			
7. Left constituent Family Size	0.02	0.00	-0.01	-0.21***	0.26***	-0.02		
8. Right constituent Family Size	-0.11*	0.06	0.01	-0.07	-0.05	0.58***	-0.01	
9. Compound Plausibility	-0.09	-0.02	0.11*	0.03	0.04	0.12*	-0.02	0.10

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

Table B.2: Correlation matrix of individual differences measures included in statistical models.

ID	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
1. Working Memory Composite												
2. Inhibitory Control Composite	-0.37***											
3. CTOPP Phonological Awareness Composite	0.53***	-0.14										
4. CTOPP Phonological Memory Composite	0.48***	-0.18	0.54***									
5. CTOPP Rapid Symbolic Naming Composite	0.34***	-0.25*	0.34***	0.24*								
6. WJ Decoding Composite	0.53***	-0.22*	0.67***	0.52***	0.33***							
7. WJ Passage Comprehension	0.61***	-0.19	0.59***	0.46***	0.32**	0.67***						
8. WJ Reading Fluency	0.64***	-0.38***	0.55***	0.48***	0.44***	0.65***	0.68***					
9. Vocabulary Knowledge Composite	0.67***	-0.36***	0.60***	0.46***	0.34***	0.70***	0.77***	0.76***				
10. Print Exposure Composite	0.49***	-0.25*	0.49***	0.34***	0.22*	0.50***	0.60***	0.62***	0.71***			
11. Age	0.07	0.01	0.11	0.11	0.07	0.07	0.18	0.18	0.16	0.27**		
12. Years in education	0.39***	-0.10	0.39***	0.24*	0.15	0.35***	0.47***	0.41***	0.53***	0.49***	0.66***	
13. WASI Matrix Reasoning	0.58***	-0.19	0.42***	0.42***	0.18	0.53***	0.61***	0.56***	0.69***	0.46***	0.09	0.33***

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

Appendix C: Model summaries

The following tables provide the fixed effects coefficients for models pertaining to the significant interactions between LSA measures and individual differences scores that are reported in Analysis 2a and 2b.

Table C.1: Fixed effects of the generalized linear mixed-effect model fitted to first-pass regression rate for the Print Exposure Composite by Head-Compound LSA Similarity interaction. The R^2 of the model is 0.17 and the standard deviation of the residual is 1. The standard deviation estimate for the random effect of Compound is 0.72. The standard deviation estimate for the random effect of Participant with random slopes for LSA Modifier-Compound is 0.09. The standard deviation estimate for the random effect of Participant with random slopes for LSA Head-Compound is 0.03. Number of trials = 4,766. Number of trials after trimming = 4,766.

	Estimate	Std. Error	z	p
Intercept	-0.924	0.101	-9.126	0.000
Print Exposure Composite	-0.011	0.035	-0.319	0.750
LSA Modifier-Compound	-0.059	0.055	-1.066	0.286
LSA Head-Compound	0.088	0.054	1.636	0.102
Compound Length	-0.064	0.053	-1.201	0.230
Compound Frequency	-0.256	0.055	-4.616	0.000
Left constituent Family Size	0.003	0.067	0.042	0.966
Right constituent Family Size	0.052	0.071	0.732	0.464
Left constituent Frequency	0.040	0.067	0.596	0.551
Right constituent Frequency	-0.109	0.072	-1.505	0.132
Word position	0.178	0.051	3.466	0.001
Trial number	-0.003	0.001	-2.020	0.043
Compound Plausibility	-0.021	0.052	-0.408	0.683
LSA Modifier-Compound * Print Exposure Composite	-0.070	0.037	-1.896	0.058
LSA Head-Compound * Print Exposure Composite	-0.099	0.036	-2.784	0.005

Table C.2: Fixed effects of the generalized linear mixed-effect model fitted to first-pass regression rate for the Vocabulary Knowledge Composite by Head-Compound LSA Similarity interaction. The R^2 of the model is 0.17 and the standard deviation of the residual is 1. The standard deviation estimate for the random effect of Compound is 0.72. The standard deviation estimate for the random effect of Participant with random slopes for LSA Modifier-Compound is 0.11. The standard deviation estimate for the random effect of Participant with random slopes for LSA Head-Compound is 0.04. Number of trials = 4,766. Number of trials after trimming = 4,766.

	Estimate	Std. Error	z	p
Intercept	-0.920	0.101	-9.075	0.000
Vocabulary Knowledge Composite	0.050	0.036	1.390	0.164
LSA Modifier-Compound	-0.057	0.055	-1.029	0.303
LSA Head-Compound	0.090	0.054	1.670	0.095
Compound Length	-0.063	0.053	-1.172	0.241
Compound Frequency	-0.254	0.055	-4.570	0.000
Left constituent Family Size	0.006	0.067	0.090	0.928
Right constituent Family Size	0.046	0.071	0.653	0.514
Left constituent Frequency	0.040	0.067	0.589	0.556
Right constituent Frequency	-0.107	0.073	-1.478	0.139
Word position	0.182	0.051	3.541	0.000
Trial number	-0.003	0.001	-2.051	0.040
Compound Plausibility	-0.026	0.052	-0.496	0.620
LSA Modifier-Compound * Vocabulary Knowledge Composite	-0.022	0.038	-0.589	0.556
LSA Head-Compound * Vocabulary Knowledge Composite	-0.099	0.035	-2.790	0.005

Table C.3: Fixed effects of the generalized linear mixed-effect model fitted to first-pass regression rate for Years in education by Modifier-Compound LSA Similarity interaction. The R^2 of the model is 0.17 and the standard deviation of the residual is 1. The standard deviation estimate for the random effect of Compound is 0.72. The standard deviation estimate for the random effect of Participant with random slopes for LSA Modifier-Compound is 0.09. The standard deviation estimate for the random effect of Participant with random slopes for LSA Head-Compound is 0.08. Number of trials = 4,766. Number of trials after trimming = 4,766.

	Estimate	Std. Error	z	p
Intercept	-0.922	0.101	-9.097	0.000
Years in education	0.048	0.035	1.369	0.171
LSA Modifier-Compound	-0.057	0.055	-1.043	0.297
LSA Head-Compound	0.092	0.054	1.693	0.090
Compound Length	-0.061	0.053	-1.140	0.254
Compound Frequency	-0.256	0.055	-4.615	0.000
Left constituent Family Size	0.004	0.067	0.066	0.947
Right constituent Family Size	0.046	0.071	0.647	0.518
Left constituent Frequency	0.044	0.067	0.653	0.514
Right constituent Frequency	-0.106	0.072	-1.464	0.143
Word position	0.182	0.051	3.553	0.000
Trial number	-0.003	0.001	-2.073	0.038
Compound Plausibility	-0.024	0.052	-0.464	0.643
LSA Modifier-Compound * Years in education	-0.074	0.037	-1.980	0.048
LSA Head-Compound * Years in education	-0.073	0.036	-2.028	0.043

Table C.4: Fixed effects of the linear mixed-effect model fitted to total reading time for Years in education by Modifier-Compound LSA Similarity interaction. The R^2 of the model is 0.23 and the standard deviation of the residual is 0.51. The standard deviation estimate for the random effect of Compound is 0.22. The standard deviation estimate for the random effect of Participant with random slopes for LSA Modifier-Compound is 0.001. The standard deviation estimate for the random effect of Participant with random slopes for LSA Head-Compound is 0.001. Number of trials = 4,755. Number of trials after trimming = 4,755.

	Estimate	Std. Error	t	p
Intercept	6.284	0.025	249.596	0.000
Years in education	-0.088	0.008	-11.135	0.000
LSA Modifier-Compound	-0.022	0.014	-1.512	0.131
LSA Head-Compound	0.004	0.014	0.262	0.793
Compound Length	0.032	0.014	2.279	0.023
Compound Frequency	-0.118	0.015	-7.986	0.000
Left constituent Family Size	0.008	0.018	0.457	0.648
Right constituent Family Size	-0.001	0.019	-0.029	0.977
Left constituent Frequency	-0.017	0.018	-0.944	0.346
Right constituent Frequency	-0.035	0.019	-1.807	0.072
Word position	0.064	0.014	4.609	0.000
Trial number	-0.001	0.000	-3.261	0.001
Compound Plausibility	-0.013	0.014	-0.916	0.360
LSA Modifier-Compound * Years in education	-0.023	0.008	-2.868	0.004
LSA Head-Compound * Years in education	-0.006	0.008	-0.709	0.478

3 Competition between conceptual relations affects compound recognition: the role of entropy

This chapter is reproduced with kind permission from Psychonomic Bulletin & Review as [Daniel Schmidtke, Victor Kuperman, Christina Gagné, and Thomas Spalding (2016). Competition between conceptual relations affects compound recognition: the role of entropy. *Psychonomic Bulletin & Review 23*(2), 556-570.]

Abstract

Previous research has suggested that the conceptual representation of a compound is based on a relational structure linking the compound's constituents. Existing accounts of the visual recognition of modifier-head or noun-noun compounds posit that the process involves the selection of a relational structure out of a set of competing relational structures associated with the same compound. In this article, we employ the information-theoretic metric of entropy to gauge relational competition and investigate its effect on the visual identification of established English compounds. The data from two lexical decision megastudies indicates that greater entropy (i.e., increased competition) in a set of conceptual relations associated with a compound is associated with longer lexical decision latencies. This finding indicates that there exists a competition between potential meanings associated with the same complex word form. We provide empirical support for conceptual composition during compound word processing in a model that incorporates the effect of the integration of co-activated and competing relational information.

3.1 Introduction

3.1.1 The relational structure of English compounds

The internal structure of endocentric compounds provides additional information above and beyond specifying the morphological role of each constituent. To illustrate, consider a compound such as *teacup*, which is composed of the modifying constituent *tea* and the head constituent *cup*. People seem to be able to know more than just that it is a cup that is in some way related to tea. Instead, they posit a particular connection between the constituents; the meaning of *teacup* can be paraphrased as "a cup for tea". Although relational structures are not often discussed in current theories of complex words (including compounds), such structures were prominent in earlier linguistic theories. Kay and Zimmer (1976), for example, noted that the semantic structure of nominal compounds is interesting in that "the relation between the two nouns is not explicitly present at any linguistic level, but rather is evoked by the construction itself" (pp. 29). As an example, consider that *olive oil* is oil derived from olives, but the same relation does not apply to *baby oil*. In this paper, we investigate whether relational links between constituents, which are not present in the orthography, contribute to compound word recognition.

There have been several attempts to characterize the specific relational link that exists between constituents of compounds (e.g., Downing, 1977; Finin, 1980; Lees, 1966; Levi, 1978; Li, 1971; Warren, 1978). Linguists and psycholinguists have proposed between 10 and 20 common relation categories that capture the majority of semantically transparent compounds (Downing, 1977; Gleitman & Gleitman, 1970; Kay & Zimmer, 1976; Lees, 1960; Levi, 1978; Warren, 1978). Examples of relations include MADE OF (*paper bag* = bag made of paper), FOR (*computer screen* = screen for a computer), and HAS (*chocolate muffin* = muffin that has chocolate). Although the nature of the categories vary, the assumption is that the underlying structure of compounds and modifier-noun phrases provides information about how the constituents are linked and that this structure plays an important role in determining the meaning of the whole compound/phrase. For example, Gleitman and Gleitman (1970) argue that modifier-noun phrases are derived from underlying relative clauses and that these underlying structures are recoverable. Levi (1978) makes a similar proposal, and claims that all complex nominals that are not derived by nominalization processes are derived from underlying semantic structures from which a predicate (e.g., CAUSE, HAVE, MAKE, FOR) has been deleted.

3.1.2 Conceptual relations and compound word recognition

In the psycholinguistic literature, there have been several findings that suggest the language system might attempt to compute meaning whenever morphemic representations become available. For example, Libben, Derwing, and de Almeida (1999) have examined the parsing of novel ambiguous compounds such as *clamprod*, which can be parsed as either *clam prod* or *clamp rod*. They examined whether participants would assign the first possible parse (that is, the parse which is first encountered using a left-to-right parsing strategy). They did this by presenting novel compounds to people and asking them to indicate where they think the compound should be divided. The results failed to show a preference for this left-to-right strategy; parsing preferences were equally divided between the two possible parses. However, the decisions about where to parse the compounds were not arbitrary; participants appeared to be selecting the parse based on the plausibility of the various parses. That is, parsing preference was correlated with the plausibility of the meaning of the various parses. This correlation indicates that the processing of novel compounds involves generating and evaluating multiple representations. This finding is especially interesting because it demonstrates that parsing is affected by the semantic fit between constituents rather than solely by properties of the compound.

More direct evidence for the involvement of relational structures in the processing of noun phrases and compounds has accumulated over the years (see Gagné & Spalding, 2014 for a review). For example, Coolen, van Jaarsveld, and Schreuder (1991) conducted a study in Dutch and found that it took longer to correctly respond to novel compounds that had been rated as being highly interpretable than it did to respond to novel compounds that had been rated as being less interpretable. This suggests that the easier it was to construct a meaning based on the constituents, the more difficult it was for people to reject the novel compound as being an existing word. To further examine this issue, Coolen et al. (1991) asked participants to provide paraphrases for the novel compounds. These paraphrases were then classified according to Levi's (1978) relations. Whether an item can be classified using Levi's relation was related to interpretability. High interpretability items were more likely to be paraphrased using one of Levi's relations than were low interpretability items. This observation led Coolen et al. (1991) to conclude that one aspect of the lexical decision process might rely on the meaning constructions based on a small set of semantic relations.

Subsequent research on novel compounds found that ease of interpretation was affected by the availability of the required relational structure (Gagné, 2000, 2001, 2002; Gagné & Shoben, 1997). Availability is affected by both general usage and recent usage. General usage refers to knowledge about how likely a particular relation is to be used with a constituent. For example, for the modifier *mountain*, the LOCATED relation is the most likely relation. Gagné and Shoben (1997) created a set of novel compounds and classified them in terms of relational categories. These classifications were used to calculate the frequency with which each modifier and head noun was used for each relation. Novel compounds that required a relation that was most likely for the modifier (e.g., *mountain cloud* uses the LOCATED relation) were processed more quickly (in a sense-nonsense task) than were novel compounds that required a relation that was not likely for the modifier (e.g., *mountain magazine* uses the ABOUT relation). Furthermore, recent usage also influences the availability of a relation. Several studies have demonstrated that it takes less time to make a sense-nonsense judgment to a novel compound when it has been preceded by a compound using the same modifier and the same relation than when preceded by a compound using the same modifier and a different relation (Gagné, 2000, 2001; Gagné & Shoben, 2002).

Relational structures also appear to be involved in the processing of established (i.e., familiar) compounds in that relational availability affects ease of processing. For example, Gagné and Spalding (2004; 2009) found evidence of relational priming: lexical decision latencies to a compound were faster when the compound had been preceded by a prime that used the same relation and the same modifier than when preceded by a prime that used a different relation. Other research indicates that both the relation selection and constituent assignment are involved because relational priming only occurs when the repeated constituent is used in the same position for both the prime and the target (Gagné, Spalding, Figueredo, & Mullaly, 2009). For example, responses were faster to *fur gloves* when preceded by *fur blanket* than by *fur trader*. However, there was no difference in response times following either *acrylic fur* or *brown fur*. This finding suggests that relational information is accessed/evaluated in the context of a constituent's morphosyntactic role.

Moreover, evidence also suggests that the nature of the priming effect is primarily competitive. The influence of relational competition demonstrated for novel compounds, for which semantic composition is obligatory (e.g., Gagné & Shoben, 1997; Gagné, 2001; Spalding, Gagné, Mullaly & Ji, 2010) is also found in the processing of lexicalized compounds. Spalding and Gagné (2011) found that having a prime with a different relation (e.g., *snowshovel* as a prime for the target *snowball*) slowed responses relative to a modifier-only prime (e.g., *snow*), whereas the related prime (e.g., *snowfort*) was equivalent to the modifier-only prime. This finding is suggestive of competition from the different relation prime rather than facilitation from the same relation prime.

In sum, research suggests that processing of both novel and familiar compounds is influenced by the availability of relational structures and this effect is specifically competitive in nature. These effects are predicted by the RICE theory of conceptual combination and its predecessor (the CARIN theory), which both propose that relational structures provide a gist-based representation that captures a simple paraphrase of the compound (Gagné & Shoben, 1997; Spalding et al., 2010). In particular, these theories claim that the interpretation of both novel and familiar compounds proceeds in three (partially overlapping) stages. First, the relations associated with the modifier compete to be selected as a potential interpretation, then relations associated with both modifier and the head are used (along with semantic information associated with both constituents) to select and verify a gist interpretation or paraphrase of the compound, and finally this gist interpretation can be elaborated (as needed) in order to derive fuller meanings of the compounds (see Spalding et al., 2010, for a detailed description and explanation). Thus, a key prediction of the CARIN and RICE theories is that during the interpretation of modifier-noun phrases and compounds, relational structures compete for selection during semantic composition. This specific prediction about the role of competition in compound interpretation has been verified several times, both in the sense that relations associated with a particular constituent compete with each other for selection (as shown by, e.g., Gagné & Shoben, 1997; Spalding & Gagné, 2008, 2011; Spalding et al., 2010), and in the sense that full relational interpretations compete with each other (as shown by, e.g., Gagné & Spalding, 2014; Spalding & Gagné, 2014). That is, it is not simply the case that frequent relations/interpretations are easy to derive, but that relations that are strong relative to other relations/interpretations are easy. In short, increased competition among relational interpretations produces increased processing difficulty.

3.1.3 Information theoretical approaches to relational structure

In the aforementioned experiments conducted by Gagné and colleagues, degree of competition was manipulated by using a prime that had either the same or a different relation, or by manipulating the constituent's availability as measured by the constituent's relational distribution. However, another way to evaluate competition – the way we adopt in the present paper – is in terms of the information-theoretic measure of entropy (Shannon, 1948). For the specific case of a paradigm of i semantic relations, each with its own probability of association with a given compound p_i, entropy H is defined as $H = -\Sigma p_i \log_2 p_i$. Thus, in the present case of gauging the competition between activated relational links during compound processing, entropy measures the expected amount of information in the probability distribution of semantic relations, and - for a specific compound - estimates the average amount of uncertainty in choosing any of i relations to be associated with the compound's relational meaning. Entropy increases when more semantic relations are associated with a compound and also when the probabilities of those relations are closer in value to each other. These mathematical properties make entropy a valuable tool for assessing competition between relations, which indeed is expected to be more effortful when more relations are available and none of them has a clear dominance over others.

Prior research has highlighted the utility of entropy and related measures for characterizing competition within morphological paradigms, which we illustrate using only two of many available examples (see Milin, Kuperman, Kostić & Baayen, 2009, and Milin, Đurđević, & Moscoso del Prado Martín, 2009, for a more complete survey of applications of information-theoretic tools to morphology). As a first example, Moscoso del Prado Martín, Kostić, and Baayen (2004) calculated entropy based on inflectional information using statistics of the base frequency of an inflectional paradigm (a series of inflected morphologically complex forms sharing the same base morpheme e.g., *vote*, *votes*, *voted*, *voting*) and the surface frequency of a word. They predicted that inflectional entropy as a metric of competition should be negatively correlated with lexical decision latencies. In addition, they calculated entropy of derivational paradigms (a series of derived morphologically complex forms sharing the same base morpheme, e.g., *perform*, *performance*, *performer*) based on word base frequency and cumulative root frequency. Entropy was higher for morphological paradigms with more members than for paradigms with fewer members. Moscoso del Prado Martín et al. (2004) found morphologically-based entropy effects, in that words with the morphological family that had very few dominant members were processed more quickly than words with morphological family members that had many competing dominant members. That is, the more uncertainty (i.e., the higher the entropy) present in a word's morphological paradigm, the more difficult it was to process the word. Similarly, Kuperman, Pluymaekers, Ernestus, and Baayen (2007) found that entropy in the morphological family of a compound's head affected speech production of Dutch compounds with inter-fixes (-s- in *oorlogsverklaring* "announcement of war" and -en- in *dierenarts* "veterinary"). Higher entropy, indicating a larger amount of uncertainty in the head's morphological family, led to prolonged acoustic durations in the pronunciation of inter-fixes.

As well as the successful application of information-theoretical tools to psycholinguistic data, the aforementioned studies also offer a crucial theoretical point of connection with our own efforts to model the visual identification of compound words. That is, they provide theoretical insights which are of special interest to the present investigation of conceptual integration during compound processing. Interestingly, Moscoso del Prado Martín et al. (2004) argue that the effect of entropy of family size is driven by semantic similarity existing between competing family members within a derivational morphological paradigm. The idea is that larger and more established families (i.e., sets of compounds sharing constituents) facilitate recognition of family members, arguably by boosting “semantic resonance” via simultaneous activation of multiple, typically semantically related words. This claim is also supported by findings of De Jong (2002) and Moscoso del Prado Martín, Deutsch, Frost, Schreuder, De Jong and Baayen (2005) in Dutch, English and Hebrew. Furthermore, studies investigating the purported effects of inflectional paradigm size on the processing of case-marking inflected forms, such as those in Serbian (Milin et al., 2009b), have also drawn the conclusion that the morphological family size effect is likely to play a role at the semantic level of lexical processing.

Importantly, these and most other applications of entropy to psycholinguistic data are based on the distributional characteristics of ‘visible’ aspects of words, such as the orthographic forms of compound constituents (see however Hahn & Sivley, 2011). As Moscoso del Prado Martín et al. (2004) note, the family size effect is not influenced by individual relations between pairs of words, but rather by the frequency-derived structural relations between morphological paradigms. Therefore, while this stream of research may well be validly capturing the so-called semantic “entanglement” of complex words (Baayen, Milin, Đurđević, Hendrix & Marelli, 2011), the reported effects are based on distributional measures of surface form characteristics, which are only indirectly related to the semantic properties of words. On the other hand, a measure of competition among relational structures of compounds is a variable that is unequivocally semantic in nature. As discussed earlier in the Introduction, relational structures are not explicitly stated in the surface form of the compound, but rather are implied structures. Therefore, detecting an effect of competition between semantic relations during compound word processing, as gauged by entropy, would provide a novel window into the semantic processing of complex words, and would do so without recourse to lexical measures derived from surface form characteristics.

Thus far, the implementation of information-theoretic measures in the study of relation-based competition has been promising. Pham and Baayen (2013) considered a measure which was based on Gagné and Shoben’s (1997) finding that the relative number of conceptual relations within the compound’s modifier family affected compound processing. Pham and Baayen (2013) calculated entropy over the probability distribution of the conceptual relations that exist within a modifier family. In other words, for a given compound (e.g., *snowball*) they calculated entropy over the distribution of all the conceptual relations associated with the modifier (*snow*) in the language, including the conceptual relation tied to that specific compound (*ball* MADE OF *snow*). This measure affected lexical decision times, such that greater entropy slowed down lexical decision response times. This finding indicates that when reading a compound word, competition exists between the relative strength of the relations associated with a modifier. Moreover, unlike Gagné and Shoben's (1997) measure, Pham and Baayen's measure of competition takes into account the probability distribution of all conceptual relations associated with a given modifier, and not just its three most frequent relations. Thus, Pham and Baayen (2013) demonstrated that lexical processing is affected by the divergence of the relation of modifier in a compound from the distribution of relations for the modifier defined over *all* compounds in that modifier's family. Crucially, Pham and Baayen (2013) derived their distribution of relations from a corpus in which each compound was coded with a specific semantic relation. However, entropy can also be calculated over distributions of conceptual relations using data generated from a *possible relations* task, upon which we will now elaborate.

In the possible relations task, participants are presented with a compound consisting of two words and are asked to pretend that they are learning English and know each of the two words, but have never seen the words used together. Their task is to choose the most likely literal meaning for a pair of nouns (e.g., *snow ball*). The choice is made out of a set of possible relational interpretations that participants are provided with (e.g, *ball* CAUSES *snow*, *ball* CAUSED BY *snow*, *ball* HAS *snow*, *ball* MAKES *snow*, *ball* FROM *snow*, *ball* MADE OF *snow*, *ball* IS *snow*, *ball* USED BY *snow*, *ball* USES *snow*, *ball* LOCATED *snow*, *snow* LOCATED *ball*, *ball* FOR *snow*, *ball* ABOUT *snow*, *ball* DURING *snow*, and *ball* BY *snow*). The set of relations was used in Gagné and Shoben (1997) and was an adaptation of Levi's (1978) original set of relations. The possible relations task generates a distribution of possible relational interpretations of a compound. Per compound, each relational interpretation is associated with a frequency with which that relational interpretation has been selected.

Until now, two information theoretic measures have been computed using data from a possible relations task (see Gagné & Spalding, 2014). The first is Relational

Diversity, defined as the number of distinct relational interpretations given to a compound. Another is Relational Relative Entropy, which measures the degree to which the probability distribution of relations identified for a particular compound differs from the probability distribution of relations estimated across a larger set of compounds. More specifically, the Relational Diversity Ratio is calculated by obtaining, for each item, the number of relations that were attested by ten or more participants (that is, by at least 10% of the participants who judged the item), and dividing that number by the total number of relations chosen by any participant for that item. Relational Relative Entropy for a given item is calculated as the probability of a given relation for that item multiplied by the binary logarithm of the probability of that relation for that item divided by the probability of that relation in the total relational distribution, summed across all 16 relations. Relational Diversity Ratio and Relational Relative Entropy were entered as predictors in a linear mixed effects model fitted to lexical decision latencies. Relational Diversity Ratio interacted with Relational Relative Entropy, and the relationship between the diversity measure and Relational Relative Entropy was different for semantically opaque and transparent compounds. In particular, for opaque compounds, the effect of Relational Diversity Ratio was attenuated when Relational Relative Entropy was low (i.e., when the item's relational distribution was close to the overall relational distribution). For transparent compounds, when Relational Relative Entropy was low, the effect of diversity was similar to the effect seen with opaque compounds. However, when Relational Relative Entropy was high, the effect of diversity was opposite to the effect seen with opaque compounds. High Relational Diversity and high Relational Relative Entropy were associated with slower response times for transparent compounds, whereas low Relational Diversity and low Relational Relative Entropy were associated with fast response times. These results demonstrate that lexical processing of transparent compounds is facilitated when a compound has a large number of potential relations but only a small number of them are strong candidates. On the other hand, the processing of opaque compounds is attenuated by the combination of high Relational Diversity and low Relational Relative Entropy, which channels lexical access towards a computed meaning which will not be the established meaning of the compound.

3.1.4 The current study

As we have summarized, several studies have shown evidence that both novel and established compounds are affected by availability of conceptual relations and that competition among relations influences ease of processing. Moreover, information-theoretic measures have been successfully used as indices of competition in terms of morphological forms, and, more recently, in terms of semantic relational structures. Thus, there is evidence that lexical processing is affected by both the diversity of the relational distribution of a given compound and its divergence from the relational distribution defined over all compounds in the set. However, this line of inquiry presently lacks one critical test of the competition within a relational distribution for a single compound. Namely, it remains to be tested whether *Entropy* (rather than Relational Diversity, Relational Relative Entropy or Pham and Baayen's (2013) measure of entropy) of the relational distribution of the compound itself influences the speed of processing.

It is important to test entropy because this measure differs in several ways from the measures that have already been tested. Relational Relative Entropy used by Spalding and Gagné (2014) gauges how much the probability distribution of relations for a single compound differs from the probability distribution calculated over relations of all compounds. In a similar fashion to Pham and Baayen's (2013) measure of entropy based on the relative frequency of relations in the compound's modifier family, Relational Relative Entropy can be broadly construed as a metric of how one's experience with possible interpretations for all compounds needs to be adjusted for interpretations available for a specific compound under recognition. Unlike entropy, this metric does not quantify how difficult it is to converge on one interpretation for a compound given the available set of relations, each with its own probability.

The distinction between Relational Diversity used by Spalding and Gagné (2014) and entropy can be illustrated by considering two different compounds, *floodlight* and *newsroom*. Both compounds have the same Relational Diversity value; they both have 9 relations that were chosen by more than one participant. Yet, for those 9 relations, the compounds' distributions of how often each relation was chosen tell a very different story. The compound *floodlight* has a distribution of responses that are apportioned equally among its 9 relations. For example, 3 relations within the distribution of 9 for *floodlight* (*light* FROM *flood*, *flood* IS *light* and *light* DURING *flood*) are all equiprobable, each with a selection frequency of 4 (i.e., each of these relations was chosen by 4 participants). On the other hand, *newsroom* has a very clear candidate for a relational interpretation (*room* FOR *news*), which has 66 responses (i.e., this relation was chosen by 66 participants). Thus, the FOR relation has the greatest share of responses and does not have an equally frequent competitor relation. Therefore, relative to *newsroom*, the average uncertainty in choosing any one relation to define the interpretation of *floodlight* is high and is expressed with a greater entropy value. The entropy of the distribution of conceptual relations can be operationalized in the present study as a precise measure of the competition among relation candidates that are engaged during compound word recognition.

The aim of the current project, then, is to further examine entropy as a measure of relational competition. In doing so, we are able to more directly test for evidence of the influence of relational information and, more specifically, of relational competition in the context of established compounds. Moving from Relational Diversity to entropy ensures that not only the number of distinct interpretations or the most frequent interpretation are accounted for, but also their balance of probabilities. Moreover, shifting focus from Relational Relative Entropy to entropy further gives prominence to the competition effect of item-specific relational structures, rather than an estimation of competition stemming from population-wide distributions of relational structures aggregated across multiple compounds.

As argued above, entropy is the most direct measure of relational competition within a morphological paradigm. Following the predictions of the CARIN and RICE theories of conceptual combination, we anticipate that higher entropy will reflect a stronger competition between available relations and will cause an increased processing effort in visual comprehension tasks. In this study, we test this prediction by using relational distributions for a number of compounds collected in two experiments (Gagné & Spalding, 2014; Spalding & Gagné, 2014), and their behavioural latencies attested in two lexical decision megastudies, English Lexicon Project (Balota, Yap, Hutchison, Cortese, Kessler, Loftis, Neely, Nelson, Simpson & Treiman, 2007) and British Lexicon Project (Keuleers, Lacey, Rastle & Brysbaert, 2012). We will thus test whether entropy can predict compound RTs in two separate lexical decision data sets.

3.2 Methods

3.2.1 Materials

The data set we considered was composed of the results of two separate experiments, which each had collected possible relations judgements for a number of compounds. One such source, from Spalding and Gagné (2014), included judgements for a total of 188 existing English compounds from 159 unique participants. The mean number of ratings per compound in this data source was 53 (range 52-54). The other source was a set of 56 existing English compounds, used in Gagné and Spalding (2014), which includes relation judgements to these compounds from a total of 111 unique participants (all participants contributed ratings for all compounds in this data source).

Once we combined both data sets, the resulting data source consisted of 232 unique compound words, each with a separate frequency distribution of possible relations judgements over 16 modifier-head relations. After combining both data sets, we found that 12 items overlapped across both sources. Because these compounds were present in two separate experiments, each of the 12 compounds was attested with a pair of differing judgement distributions. We decided to consider both judgement distributions in our statistical models. Thus, while there were 232 unique compounds, when taking into account the 12 duplicated items, we had a data source consisting of 244 different judgment distributions. Moreover, once the data from both experiments were combined, a total of 270 participants contributed responses and a median of 53 participants provided a judgement per compound. This stimulus data is provided as a supplementary data file. For further details on the procedure and stimuli selection, see Spalding and Gagné (2014), and Gagné and Spalding (2014).

3.2.2 Dependent variables

Trial-level lexical decision latencies were obtained from the English Lexicon Project [ELP] and the British Lexicon Project [BLP]. We only considered reaction times

(RTs) of trials for which there was a correct response. We also removed outlier responses by eliminating the top and bottom 1% of the RT distribution for both ELP and BLP samples. This led to a loss of 2.15% of the total data points in ELP and 2.04% of the total data points in BLP samples. In order to attenuate the influence of outliers, we used the inverse transform method to convert response times as indicated by the Box-Cox power transformation (Box & Cox, 1982). The percentage of incorrect responses to our stimuli of interest was too low (ELP: 9.91%, BLP: 16.69%) to warrant a separate investigation of the effect of critical variables on response accuracy. In both the ELP and BLP data sources, the compounds were all presented for lexical decision in a concatenated format (i.e. unspaced). Overall, of the 232 compounds for which we had possible relations judgements, 187 were present in the ELP data source and 143 were present in the BLP data source. A total of 130 items overlapped across the ELP and BLP data source, while 44 items from the original data pool did not occur in either ELP or BLP.

3.2.3 Independent variables

The critical variables of interest are ones related to the judgements of conceptual relations. One measure is Relational Diversity, estimated as the number of relations (out of the set of 16) that had been chosen by more than one participant for a given compound. The lower threshold of more than one participant was chosen to reduce the number of random or accidental (erroneous) responses. Another measure is Entropy calculated over the probability distribution of interpretations of conceptual relations for a given compound. The probability distribution was estimated *only* for relations that were selected more than once in the judgement task. Entropy is defined as $H = -\Sigma p_i \log_2 p_i$, where p_i is the probability of a relation within the respective distribution of possible relations for a given compound. Thus, for the compound *lawsuit*, the relations are FOR (selected 12 times), FROM (7), MADE OF (7), USES (6), CAUSES (4), USED BY (4), HAS (3), ABOUT (3), CAUSED BY (2) and BY (2). The resulting probability distribution for these relations is 0.24, 0.14, 0.14, 0.12, 0.08, 0.08, 0.06, 0.06, 0.04 and 0.04, which yields an Entropy value of 3.097. For both ELP and BLP data sources, the number of relations that were selected more than once per compound (Relational Diversity) ranged from 3 to 16 and the median number of relations that were selected per compound was 8.

In addition to the relational structure of a compound, a further morpho-semantic component of compound word recognition is semantic transparency, which is defined as the predictability of the meaning of the compound word given the meaning of its parts (see Amenta & Crepaldi, 2012 for a review of the effects of semantic transparency in lexical decision experiments). A highly transparent compound (e.g., *flashlight*) is composed of constituents with semantic denotations that are semantically similar to the meaning of the whole word (*flash* and *light*). An opaque compound (e.g., *brainstorm*), on the other hand, includes constituents (*brain* and *storm*) out of which at least one morpheme bears a meaning that is unrelated to the compound word. In order to control for the potentially confounding effects of semantic transparency we included measures of semantic transparency in our analysis. As a gauge of semantic transparency we employed the computational measure of Latent Semantic Analysis (LSA) (Landauer & Dumais, 1997), which is a statistical technique for analysing and estimating the semantic distance between words, based on the contexts in which the words have co-occurred in a corpus. Following previous research that has employed LSA as a metric of semantic transparency (Pham & Baayen, 2013; Marelli, Dinu, Zamparelli, & Baroni, 2015), we collected LSA scores for three types of semantic relationships: the left constituent (modifier) and the whole compound (Modifier-Compound; e.g., *flash* and *flashlight*), the right constituent (head) and the whole compound (Head-Compound; e.g., *light* and *flashlight*), and the left constituent and the right constituent (Modifier-Head; e.g., *flash* and *light*). The term-to-term LSA scores were collected from `http://lsa.colorado.edu` with the default setting of 300 factors: a higher score implies a greater semantic similarity between the pair of words under comparison. Modifier-Head LSA scores were available for all compounds, while Head-Compound and Modifier-Compound LSA scores were only available for 171 compounds.

Other control variables included measures that were demonstrated in prior research to affect compound processing: compound length (in characters), compound frequency, frequencies of the left and right compound's constituents, as well as the positional family size of the left and right constituents (defined as the number and summed frequency of compounds that share a constituent with the fixed position of either the left or right constituent of the target compound). Family-based estimates were calculated from the 18 million-token English component of the CELEX lexical database (Baayen, Piepenbrock & Van Rijn, 1995), with the help of its morphological parse. Word frequencies from the 51 million-token SUBTLEX-US corpus (Brysbaert & New, 2009), based on subtitles from US film and media, were obtained for the compounds and their respective constituents that were present in the ELP data source. Likewise, word frequencies from the 201 million-token SUBTLEX-UK corpus (van Heuven, Mandera, Keuleers & Brysbaert, 2014), based on UK television (BBC) subtitles, were obtained for the compounds present in the BLP data source. Frequency-based characteristics all pertain to compounds in their concatenated format. The possible relations experiment of origin for each item was included as a covariate. It did not show either a main effect or an interaction with any of the variables of interest, suggesting that the two data sources are equivalent for the purposes of the present study: we did not consider experiment of origin in further analyses. Distributional characteristics of all variables are reported in Table 3.1.

3.2.4 Statistical considerations

We fitted linear mixed-effects models to the reaction time latencies from ELP and BLP. We computed models using the `lmerTest` (Kuznetsova, Brockhoff & Christensen, 2013) package in the R statistical computing software program (R Core Team, 2014). Across all models we used restricted maximum likelihood (REML) estimations. All continuous independent variables were scaled to reduce collinearity. All models included by-item and by-participant random intercepts. We also included by-participant random slopes for trial and Entropy: according to the model comparison likelihood ratio tests, these random slopes did not significantly improve model fit and were therefore excluded from all models. Furthermore, across all analyses, we refitted models after removing outliers from both data sets by excluding standardized residuals exceeding 2.5 standard deviations. Final models are reported in Tables 3.3 and 3.4. Collinearity between compound and

Table 3.1: Descriptive statistics for the dependent and independent variables: Reported are the range, mean and standard deviations of the original and transformed variables after selection and trimming procedures.

Variable	**Original**			**Transformed**		
A. BLP *data source*	range	mean	sd	range	mean	sd
Reaction times (RT)	311:2019	648	185	-3.22:-0.5	-1.64	0.38
Entropy	1.08:3.65	2.52	0.56	-2.37:1.94	0	1
Relational Diversity	5:16	10.09	3.48	-1.63:3.03	0	1
Compound Length	6:11	8.5	1.87	-2.06:2.65	0	1
Compound Frequency (UK frequency)	2:8808	414	1065	-2.93:2.81	0	1
Left constituent Frequency (UK frequency)	102:358568	20329	38895	-2.73:2.48	0	1
Right constituent Frequency (UK frequency)	47:540985	34857	85612	-2.89:2.4	0	1
Left constituent Family Size	0:42	14	11	-0.78:4.01	0	1
Right constituent Family Size	0:56	17	14	-0.86:4.4	0	1
Modifier-Head LSA Similarity	-0.02:0.74	0.22	0.17	-1.36:4.19	0	1
Modifier-Compound LSA Similarity	-0.06:0.74	0.23	0.2	-1.42:3.41	0	1
Head-Compound LSA Similarity	-0.06:0.92	0.25	0.23	-1.37:4	0	1
B. ELP *data source*	**Original**			**Transformed**		
Reaction times (RT)	207:2956	761	279	-4.83:-0.34	-1.45	0.42
Entropy	1.18:3.65	2.54	0.54	-2:2	0	1
Relational Diversity	5:16	10.5	3.61	-1.66:3.04	0	1
Compound Length	6:11	8.5	1.87	-2.24:2.17	0	1
Compound Frequency (US frequency)	1:1142	93	146	-2.52:2.59	0	1
Left constituent Frequency (US frequency)	5:102467	5249	12507	-3.56:2.56	0	1
Right constituent Frequency (US frequency)	9:204428	9794	26164	-2.78:2.56	0	1
Left constituent Family Size	0:54	16	14	-0.73:6.12	0	1
Right constituent Family Size	0:155	22	28	-0.7:11.1	0	1
Modifier-Head LSA Similarity	-0.02:0.74	0.24	0.18	-1.42:4.14	0	1
Modifier-Compound LSA Similarity	-0.07:0.74	0.25	0.21	-1.49:3.16	0	1
Head-Compound LSA Similarity	-0.07:0.92	0.24	0.23	-1.36:4.37	0	1

constituent frequency-based measures was high (multicollinearity condition number > 30): importantly, it had no bearing on model estimates for our critical variable of Entropy. This is because Entropy and frequency-based measures correlated very weakly (all rs < 0.16). A correlation matrix of all independent variables is provided in Table 3.2.

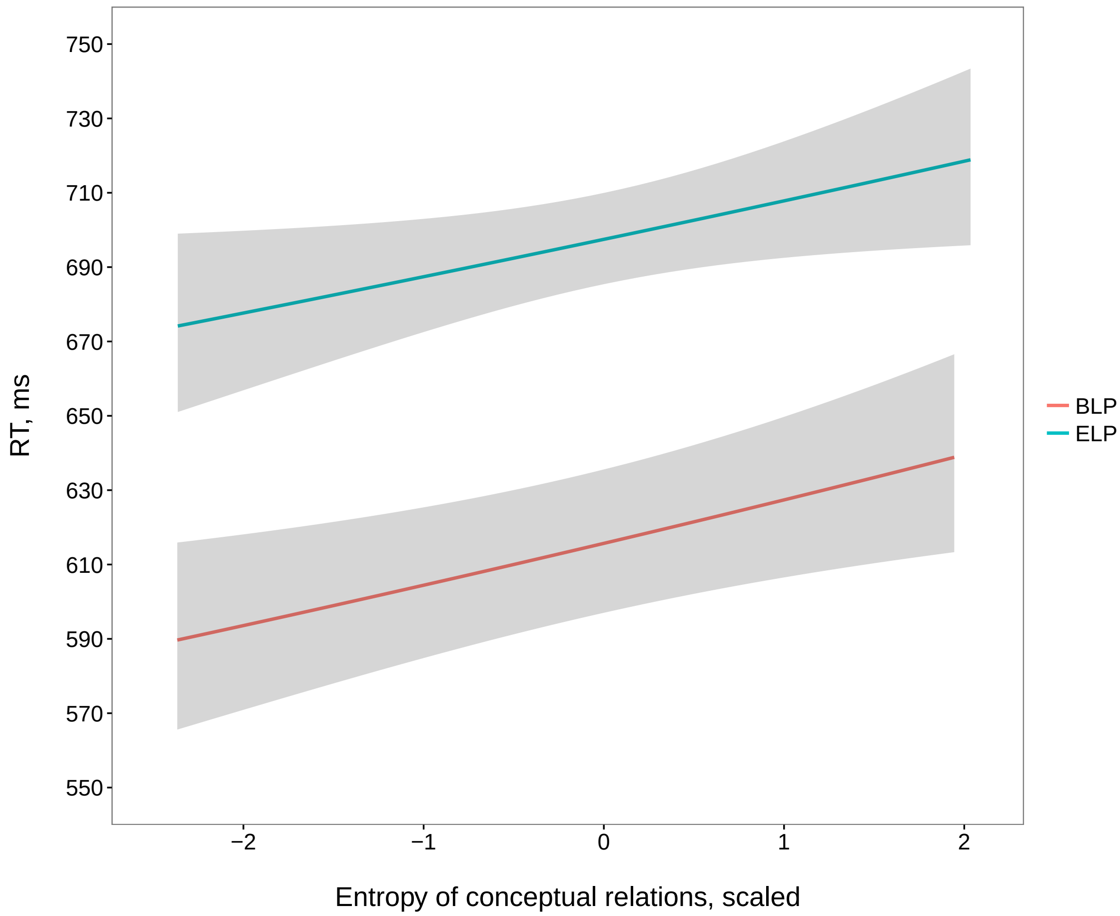

Figure 3.1: The partial effect of Entropy of relational competition (scaled) on RTs in ELP (English Lexicon Project) and BLP (British Lexicon Project) samples. Slopes represent predicted values of the linear-mixed effects models fitted separately to ELP and BLP samples. Grey bands represent lower and upper limit of 95% confidence interval for each model.

Table 3.2: Correlation matrix of all predictor (independent) variables.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1. Entropy													
2. Relational Diversity	0.86***												
3. Compound Length	-0.07	-0.17**											
4. Compound Frequency (UK)	0.03	-0.01	-0.06										
5. Left constituent Frequency (UK)	0.11	0.13*	-0.13*	0.07									
6. Right constituent Frequency (UK)	0.17**	0.19**	-0.16*	0.13	0.00								
7. Compound Frequency (US)	-0.11	-0.11	0.03	0.87***	0.07	-0.01							
8. Left constituent Frequency (US)	0.11	0.14*	-0.12	0.12	0.93***	0.02	0.17*						
9. Right constituent Frequency (US)	0.15*	0.16*	-0.12	0.09	0.00	0.95***	0.01	0.02					
10. Left constituent Family Size	-0.23***	-0.23***	-0.16*	0.20*	0.24***	0.02	0.00	0.21**	0.01				
11. Right constituent Family Size	0.07	0.07	-0.10	0.10	-0.02	0.26***	-0.06	-0.04	0.28***	0.04			
12. Modifier-Head LSA Similarity	-0.01	0.03	-0.03	0.09	0.07	0.10	0.05	0.10	0.07	0.03	-0.10		
13. Modifier-Compound LSA Similarity	-0.14	-0.12	0.10	-0.02	-0.01	0.01	-0.02	0.03	0.00	0.11	-0.11	0.21**	
14. Head-Compound LSA Similarity	-0.13	-0.08	-0.09	0.14	0.02	-0.03	0.07	-0.05	-0.04	0.17*	-0.13	0.30***	0.12

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

Table 3.3: Fixed effects of the linear mixed-effect model fitted to ELP lexical decision RTs. The R^2 of the model is 0.6 and the standard deviation of the residual is 0.25. The standard deviation estimate for the random effect of Compound is 0.26. The standard deviation estimate for the random effect of Participant is 0.11. Number of trials = 5,817. Number of trials after trimming = 5,727.

	Estimate	Std. Error	t	p
Intercept	-1.437	0.013	-111.478	0.000
Entropy	0.021	0.010	2.137	0.034
Compound Frequency	-0.078	0.010	-8.089	0.000
Compound Length	0.004	0.009	0.426	0.671
Left constituent Frequency	-0.012	0.010	-1.241	0.216
Right constituent Frequency	-0.018	0.010	-1.739	0.084
Left constituent Family Size	0.001	0.010	0.139	0.890
Right constituent Family Size	-0.004	0.009	-0.422	0.673

Table 3.4: Fixed effects of the linear mixed-effect model fitted to BLP lexical decision RTs. The R^2 of the model is 0.47 and the standard deviation of the residual is 0.27. The standard deviation estimate for the random effect of Compound is 0.105. The standard deviation estimate for the random effect of Participant is 0.211. Number of trials = 4,677. Number of trials after trimming = 4,624.

	Estimate	Std. Error	t	p
Intercept	-1.627	0.026	-62.759	0.000
Entropy	0.030	0.011	2.782	0.006
Compound Frequency	-0.087	0.010	-8.552	0.000
Compound Length	-0.009	0.010	-0.899	0.371
Left constituent Frequency	-0.015	0.011	-1.372	0.173
Right constituent Frequency	0.001	0.012	0.065	0.948
Left constituent Family Size	-0.008	0.012	-0.650	0.517
Right constituent Family Size	-0.031	0.012	-2.614	0.010

3.3 Results and discussion

The initial data pools consisted of 5,937 ELP trials and 4,747 BLP trials. We removed two compounds (*dustpan* and *tinfoil*) with an Entropy value greater than 2.5 standard deviations away from the respective mean of ELP and BLP Entropy distributions. We also removed two compounds with a log frequency of more than 3 standard deviations from the respective mean log frequencies of the ELP and BLP subset of compounds. These high frequency compounds (ELP: *boyfriend* and *breakfast*; BLP: *sunday*) were all over 1 standard deviation from the next highest frequency compound in each data set. The remaining ELP data pool contained 5,817 RTs to 184 unique compounds, and the remaining BLP data pool consisted of 4,677 RTs to 141 unique compounds. A total of 815 unique participants contributed RTs in ELP and 78 unique participants contributed to RTs in BLP.

Entropy of the probability distribution of conceptual relations chosen by more than one participant demonstrated an expected inhibitory effect on ELP and BLP response times. The effect indicated that a larger amount of uncertainty regarding the relational interpretation of a given compound led to a larger effort (i.e., longer response times) in responding to the compound in lexical decision (ELP: $\hat{\beta} = 0.02$, $SE = 0.01$, $t = 2.14$, $p = 0.03$; BLP: $\hat{\beta} = 0.03$, $SE = 0.01$, $t = 2.78$, $p = 0.006$). Tables 3.3 and 3.4 report the final mixed-effects models fitted to ELP and BLP RTs, respectively. The reported partial effects of Entropy are presented in Figure 3.1 (plots depict back-transformed values of response times, in ms, to aid interpretability). In addition, we also calculated Entropy over the complete probability distribution (i.e., not just the conceptual relations that were chosen more than once in the possible relations task). We included this Entropy measure in an identical model to those in which significant effects of the original Entropy measure were found. This particular measure did not exert any significant influence on response times in either ELP or BLP. This was likely due to the prevalence of random or accidental choices of irrelevant semantic relations by participants.

As well as revealing the novel effect of Entropy of conceptual relations, the models described in Tables 3.3 and 3.4 also report, for completeness, effects of other lexical characteristics. Mostly, they take the same direction as in prior literature and are small in magnitude, often failing to reach statistical significance. Compounds with larger morphological families were processed faster than words with smaller morphological families (cf. Juhasz & Berkowitz, 2011), especially in the BLP sample. Secondly, more frequent compounds and compounds with more frequent constituent morphemes were processed faster (cf. Andrews, Miller & Rayner, 2004 and Zwitserlood, 1994). These measures may not have reached significance because of high collinearity between the predictors. This collinearity could have been reduced, however the outcome of these variables was not the focus of the current study.

We also found that Entropy of conceptual relations was a more consistent and more robust predictor of lexical decision than Relational Diversity. Relational Diversity, i.e., the number of relations chosen by more than one participant, predicted RTs only in the BLP sample ($\hat{\beta} = 0.02$, $SE = 0.01$, $t = 2.57$, $p = 0.01$). This model indicated that increased diversity of conceptual relations was associated with slower response times. The model explained a negligibly smaller amount of variance (0.02%) than the model fitted to Entropy for the BLP data set. Moreover, the model fitted to Entropy produced a smaller Akaike information criterion (AIC) value when compared to the model including Relational Diversity as a predictor, indicating that the model containing Entropy of conceptual relations as a predictor was a slightly better fit.

Additionally, it was possible that the Entropy of conceptual relations effect was confounded with the semantic transparency of the compound. To investigate this, we tested the influence of the interaction of semantic transparency (we used LSA to gauge transparency, this is outlined in the Methods section) and Entropy (correlations between these variables are reported in Table 3.2). We first added each of the LSA variables to the fixed effect structure for models pertaining to both the ELP and BLP data samples. We found that none of the LSA variables influenced the regression coefficient or the statistical significance of Entropy (or Relational Diveristy in BLP). This was expected given weak correlations between Entropy and the semantic transparency measures. We then analyzed the interaction between all LSA measures and Entropy (with each LSA measure entered as a single multiplicative interaction with Entropy in three separate models, and with all three LSA measures simultaneously interacting with Entropy in one model) for the ELP and BLP data set. In all of these models, semantic transparency did not enter into a significant interaction with Entropy. We also repeated this analytical procedure with Relational Diversity, which also did not produce significant interaction effects. Thus, we conclude that two aspects of compound semantics (captured by the semantic transparency measures and entropy of conceptual relations) are unrelated and do not modulate each other's impact on compound recognition.

3.4 General discussion

Conceptual integration is demonstrably an important factor that codetermines the retrieval of compound word meaning (see Fiorentino & Poppel, 2007; Gagné & Spalding, 2014; Taft, 2003, for reviews). Indeed, prior studies have shown that the integration of visually-presented compounds involves semantic composition, such that, under experimental conditions, the processing of established compound words exhibits sensitivity to the availability of conceptual relations (Gagné & Spalding, 2004; 2009; Pham & Baayen, 2013). Gagné and Spalding (2009) presented evidence that a central component of the visual identification of compound words involves the construction of interpretive gists, whereby the cognitive system draws upon relational information linking compound constituents. Under this hypothesis, candidate relations are generated based on the characteristics of the modifier and head of the compound, and are then evaluated for plausibility. In addition, Gagné and Spalding (2006; 2014) and Gagné & Shoben (1997) revealed that relational structures compete for selection and that increased competition results in increased processing difficulty. Despite the insights presented in these studies, the precise locus of the effect of relational competition still remained unclear. It was still unknown whether the number of activated competitors or the relative probability of activated conceptual relations was driving the competition effect. The present study addressed this issue by introducing entropy calculated over the distribution of conceptual relations as a direct measure of relative competition.

In the current study we reanalyzed data obtained from a previous set of experiments in which participants were serially presented with a list of noun-noun compounds. For each compound, participants were asked to choose the most likely conceptual relation out of a possible 16 interpretations, which yields a frequency distribution of relational interpretations per compound. Previous research revealed that lexical processing is systematically affected by both the diversity of the relational distribution of a given compound and its divergence from the relational distribution of all compounds. Of particular interest here was to further investigate the influence of relational information on compound processing and also to test for evidence of compound-specific competition between conceptual relations. To test these hypotheses, we employed an information-theoretic measure as an index of competition among relational interpretations during lexical processing. The measure – the entropy of the distribution of responses per compound (Entropy) – thus served as a critical variable in a virtual experiment in which we examined its influence on visual lexical decision latencies obtained from two behavioural megastudies (the English and British Lexicon Projects, Balota et al., 2007; Keuleers et al., 2012).

We found robust evidence, replicated over two separate lexical decision data sources, that increased entropy of relational competition inhibited response times. These results demonstrate, through a more parsimonious measure of relational competition than ones employed earlier, that the relative difficulty of converging on any one interpretation of a compound translates into increased processing effort, i.e., longer response times in ELP and BLP. This difficulty is precisely what entropy calculated over a probability distribution of the compound's potential relations gauges. Interestingly, two very similar slopes were observed across ELP and BLP samples, indicating very similar effect sizes for the partial effect of entropy (range of predicted values in ELP = 45 ms; range of predicted values in BLP = 49 ms).

Another interesting aspect of this finding is that only a subset of available relations, and their entropy, affect word recognition, and not the entire set of theoretically possible relations. We saw effects of entropy on lexical decision times only when it was defined over relations that were selected for a compound by more than one participant; entropy calculated over a full set of relations (i.e., relations chosen just once or not at all), had no noticeable effect on RTs. In our data set, the number of selections per compound ranged between 4 and 16, out of a total of 16 of Levi's (1978) relations, which indicates that this is the range of relational interpretations among which semantic competition is possible. This finding might not seem so surprising given the analogous observation made by that information-theoretic measures based on morphological families only affected word recognition behaviour when based on relevant family members, i.e., ones that are semantically related to the shared meaning of the entire family (e.g., compare *bluebird* and *jailbird* as members of the family sharing *bird* as the second constituent). This implies that not only are higher-order interpretative processes able to navigate a wealth of semantic information, but also that they exploit only the semantic information that is *relevant* and has the potential to be selected as the meaning of the word that is being recognized. Thus, taken together with prior findings, our results suggest that there exists a competition between meanings associated with a single compound word. This competition is as real as the well-established neighbourhood competition effects between orthographic forms in word recognition.

In sum, we have shown that for established endocentric compounds, the ease of selecting a compound's relational interpretation (e.g., a *ball* made of *meat* for the compound *meatball*) out of an available relational set influences the ease with which the meaning of a compound is obtained. The more relational interpretations there are for a compound, and the more similar they are in their probability of being the most plausible interpretation for the compound, the longer it takes to identify the established meaning of that compound. We therefore conclude that semantic composition during the visual identification of existing compounds is a competitive process.

Situating our results within a theory of compound processing, our results are both concomitant with and extend the CARIN and RICE theoretical models of semantic integration developed by Gagné and Shoben (1997) and Spalding et al. (2010). In both theories (Gagné & Shoben, 1997; Spalding, et al., 2010), conceptual integration of compound word meaning involves the activation and competition of multiple relational gists. This process is summarized by Spalding and Gagné (2008), who posited that "ruling out any competitor relation likely requires some processing time, and ruling out more competitor relations should presumably require more time" (pp. 1576). In this paper, we have confirmed an extension to this hypothesis by showing that the time it takes to rule relations out is not only sensitive to the number of activated relations, but also to the degree of competition between them.

In addition to contributing to a wealth of evidence which suggests that co-activated conceptual relations affect compound word processing, we have also gauged conceptual competition using an information-theoretic scale that is, in principle, comparable to previous psycholinguistic studies that have provided empirical support for the central role of semantic access during complex word identification (Milin et al., 2009a; Milin et al., 2009b; Moscoso del Prado Martín et al., 2004; Pham & Baayen, 2013). The interpretation of our results - that competition at the conceptual level co-determines the visual identification of compound words - is thus tied to a perspective of complex word recognition that assumes "fast mapping" of form to meaning (Baayen et al. 2011). Under this view, the cost of processing arises when the language system attempts to single out one meaning among the many meanings that are activated by the word form that is under visual inspection. Accordingly, the successful resolution of form-meaning association is determined at the semantic level by higher-level cognitive processes. We believe that conceptual combination during compound word recognition, a process that is captured by the effect entropy of conceptual relations, is an example of one such high-level semantic process.

In summary, the success of deepening our understanding of complex word recognition depends on the ability to tap into the semantic components of complex words that have concrete psychological implications. We believe that entropy of conceptual relations is an important extension to previous work that has sought to understand semantic processing mechanisms underlying complex word recognition. The experimental psycholinguistic community has long relied upon the behavioural activity associated with frequency based measures, such as morpheme frequency or morpheme family size, as an index of meaning retrieval during compound word recognition. Another commonly used measure of compound semantics in psycholinguistics studies is semantic transparency, which is considered as a more direct measure of the conceptual composition of complex word. Similarly, this measure is also based on information that can be gleaned from the orthography of a complex word; transparency requires only the evaluation of the semantic similarity between the meaning denotations of two surface forms. Unlike these measures, entropy of conceptual relations appears to reliably drill down to an implicit source of morpho-semantic information. Thus, by supplementing the study of form with the study of meaning, we bring to bear a further lexical characteristic that meaningfully contributes to the information that a morphologically complex word carries, namely, 'relational competition'.

References

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.

Andrews, S., Miller, B., & Rayner, K. (2004). Eye movements and morphological segmentation of compound words: There is a mouse in mousetrap. *European Journal of Cognitive Psychology*, *16*(1-2), 285–311.

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*(3), 438–481.

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

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459.

Box, G., & Cox, D. (1982). An analysis of transformations revisited, rebutted. *Journal of the American Statistical Association*, *77*(377), 209–210.

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*(4), 977–990.

Coolen, R., Van Jaarsveld, H. J., & Schreuder, R. (1991). The interpretation of isolated novel nominal compounds. *Memory & Cognition*, *19*(4), 341–352.

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

Downing, P. (1977). On the creation and use of English compound nouns. *Language*, *53*(4), 810–842.

Finin, T. W. (1980). *The semantic interpretation of compound nominals.* (Unpublished doctoral dissertation).

Fiorentino, R., & Poeppel, D. (2007). Compound words and structure in the lexicon. *Language and Cognitive processes*, *22*(7), 953–1000.

Gagné, C. L. (2000). Relation-based combinations versus property-based combinations: A test of the CARIN theory and the dual-process theory of conceptual combination. *Journal of Memory and Language*, *42*(3), 365–389.

Gagné, C. L. (2001). Relation and lexical priming during the interpretation of noun–noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 236.

Gagné, C. L. (2002). Lexical and relational influences on the processing of novel compounds. *Brain & Language*, *81*(1), 723–735.

Gagné, C. L., & Shoben, E. J. (1997). Influence of thematic relations on the comprehension of modifier–noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 71.

Gagné, C. L., & Shoben, E. J. (2002). Priming relations in ambiguous noun-noun combinations. *Memory & Cognition*, *30*(4), 637–646.

Gagné, C. L., & Spalding, T. L. (2004). Effect of relation availability on the interpretation and access of familiar noun–noun compounds. *Brain & Language*, *90*(1), 478–486.

Gagné, C. L., & Spalding, T. L. (2006). Conceptual combination: Implications for the mental lexicon. In *The representation and processing of compound words.* (pp. 145–168). Oxford UK: Oxford University Press.

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*(1), 20–35.

Gagné, C. L., & Spalding, T. L. (2011). Inferential processing and meta-knowledge as the bases for property inclusion in combined concepts. *Journal of Memory and Language*, *65*(2), 176–192.

Gagné, C. L., & Spalding, T. L. (2014). Relation diversity and ease of processing for opaque and transparent English compounds. In *Morphology and meaning: Selected papers from the 15th International Morphology Meeting, Vienna, February 2012* (Vol. 327, pp. 153–162).

Gagné, C. L., Spalding, T. L., Figueredo, L., & Mullaly, A. C. (2009). Does snow man prime plastic snow? The effect of constituent position in using relational information during the interpretation of modifier-noun phrases. *The Mental Lexicon*, *4*(1), 41–76.

Gleitman, L. R., & Gleitman, H. (1970). *Phrase and paraphrase: Some innovative uses of language.* New York: Norton.

Hahn, L. W., & Sivley, R. M. (2011). Entropy, semantic relatedness and proximity. *Behavior Research Methods*, *43*(3), 746–760.

Juhasz, B. J., & Berkowitz, R. N. (2011). Effects of morphological families on English compound word recognition: A multitask investigation. *Language and Cognitive Processes*, *26*(4-6), 653–682.

Kay, P., & Zimmer, K. (1976). On the semantics of compounds and genitives in English. In *Sixth California Linguistics Association Proceedings* (pp. 29–35).

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*(1), 287–304.

Kuperman, V., Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2007). Morphological predictability and acoustic duration of interfixes in Dutch compounds. *The Journal of the Acoustical Society of America*, *121*(4), 2261–2271.

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2013). lmertest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). *R package version*, 2–0.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240.

Lees, R. B. (1966). *The grammar of English nominalizations.* (Unpublished doctoral dissertation).

Levi, J. N. (1978). *The syntax and semantics of complex nominals.* Academic Press, New York.

Li, C. N. (1971). *Semantics and the structure of compounds in Chinese.* (Unpublished doctoral dissertation). University of California, Berkeley.

Libben, G., Derwing, B. L., & de Almeida, R. G. (1999). Ambiguous novel compounds and models of morphological parsing. *Brain & Language*, *68*(1), 378–386.

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*(6), 1421–1439.

Milin, P., Kuperman, V., Kostić, A., & Baayen, R. H. (2009a). Paradigms bit by bit: An information theoretic approach to the processing of paradigmatic structure in inflection and derivation. In *Analogy in grammar: Form and acquisition* (pp. 214–252). Citeseer.

Milin, P., Đurđević, D. F., & Moscoso del Prado Martín, F. (2009b). The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian. *Journal of Memory and Language*, *60*(1), 50–64.

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*(4), 496–512.

Moscoso del Prado Martín, F., Kostić, A., & Baayen, R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition*, *94*(1), 1–18.

Pham, H., & Baayen, R. H. (2013). Semantic relations and compound transparency: A regression study in CARIN theory. *Psihologija*, *46*(4), 455–478.

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

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell Technical Journal*, *27*(4), 379–423.

Spalding, T. L., & Gagné, C. L. (2008). CARIN theory reanalysis reanalyzed: a comment on Maguire, Devereux, Costello, and Cater (2007). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1573–1578.

Spalding, T. L., & Gagné, C. L. (2011). Relation priming in established compounds: facilitation? *Memory & Cognition*, *39*(8), 1472–1486.

Spalding, T. L., & Gagné, C. L. (2014). Relational diversity affects ease of processing even for opaque English compounds. *The Mental Lexicon*, *9*(1), 48–66.

Spalding, T. L., Gagné, C. L., Mullaly, A. C., & Ji, H. (2010). Relation-based interpretation of noun-noun phrases: A new theoretical approach. *Linguistische Berichte Sonderheft*, *17*, 283–315.

Taft, M. (2003). Morphological representation as a correlation between form and meaning. In D. Assink & D. Sandra (Eds.), *Reading complex words* (pp. 113–137). Amsterdam: Kluwer.

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

Warren, B. (1978). Semantic patterns of noun-noun compounds. *Acta Universitatis Gothoburgensis. Gothenburg Studies in English Goteborg*, *41*, 1–266.

Zwitserlood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes*, *9*(3), 341–368.

4 Surviving blind decomposition: a distributional analysis of the time-course of complex word recognition

This chapter has been submitted to the Journal of Experimental Psychology as Human Perception & Performance as Daniel Schmidtke, Kazunaga Matsuki, and Victor Kuperman. Surviving blind decomposition: a distributional analysis of the time-course of complex word recognition.

Abstract

This study addresses a current debate between form-then-meaning and form-and-meaning theories of complex word recognition. We harnessed a non-parametric distributional technique of survival analysis (Reingold & Sheridan, 2014), to estimate how early orthographic, morphological and semantic variables exert influence on lexical decision reaction times. Results demonstrate that during recognition of English and Dutch derived words (e.g., *badness*; *bad* + *-ness*), and English pseudo-derived words (e.g., *wander*; *wand* + *-er*), semantic effects emerge early in the time-course of word recognition, mostly concurrently with morphological effects. Furthermore, contrary to form-then-meaning theoretical accounts of morphological decomposition, surface word frequency emerges as the earliest variable to exert a demonstrable influence on reaction times across all word conditions. Our results are convergent with a form-and-meaning account of complex word recognition.

4.1 Introduction

The question of how proficient readers recognize morphologically complex words in an apparently automatized and seamless manner has given rise to decades worth of psycholinguistic research (for a review, see Amenta & Crepaldi, 2012; Diependaele, Grainger, & Sandra, 2012). Despite advancements in the experimental and statistical resources available to psycholinguists, crucial components of the cognitive processes involved in complex word reading have been difficult to establish and sequence. This paper addresses a long-standing debate regarding the hypothesized stages during the time-course of word identification at which orthographic forms and meanings influence complex word recognition.

A dominant account in morphological processing predicts that the morpho-orthographic units of a complex word are analyzed by the language processing system before access to the word's meaning is able to proceed. Under this *form-then-meaning* account of complex word recognition (also known as *early morpho-orthographic decomposition* or *late-lexical-access*), complex words undergo orthographic segmentation into morphemic units in a serialized and semantics-blind manner (cf. Meunier & Longtin, 2007; Rastle & Davis, 2008; Solomyak & Marantz, 2010; Taft & Forster, 1976; Taft, 2004). According to this account, upon visual presentation of a complex word (e.g., *joker*), the word string is first decomposed into separate morphemes, i.e., *joke* + *er*. The decomposed orthographic form of the complex word then grants access to morphemic unit representations. Once the orthographic cues to the morpheme representations have been isolated, the morphemic representations are then extracted from the mental lexicon. This stage is referred to as either 'look-up' (i.e., the cognitive process of 'looking up' the component morphemes of complex word in the lexicon; Fruchter & Marantz, 2015) or 'licensing' (Meunier & Longtin, 2007). According to Meunier and Longtin (2007; pp. 469), the process of licensing "checks the appropriateness of morpheme combinations, for instance by assessing whether representations can be integrated on the basis of their subcategorization properties". At the final stage, the component morphemes of the complex word are recombined and the word meaning -'a person who likes telling jokes'- is obtained. One corollary of this theoretical view is that words with an apparent morphological structure (e.g., *corner*) undergo the same sequence of processing stages as complex words like *joker*, with difficulty arising at the recombination and meaning integration stages. In sum, for the processing of words with morphological structure, morpho-orthographic access is encapsulated from and *must* precede semantic access of the whole word.

4.1.1 Evidence from masked-priming research

There exists plentiful evidence in support of early morpho-orthographic decomposition among healthy adult populations. A main source of this evidence is based on the well-documented results of forward masked priming studies (for a review see Rastle and Davis, 2008). The masked priming paradigm is employed under the assumption that a briefly presented prime before the presentation of a target item, with a stimulus onset asynchrony (SOA) of < 60 ms, exploits unconscious early cognitive processing (Forster, Mohan, & Hector, 2003). Most priming studies find no reliable difference in the amount of priming between a condition in which a prime is morphologically and semantically related word to a simplex target word (i.e., semantically transparent; *trucker-TRUCK*) and a condition in which the prime only appears to be morphologically related but is not semantically related to the target (i.e., semantically opaque; *corner-CORN*[4]). This null effect of semantic transparency, relative to an orthographic control condition (i.e., no morphological relation between prime and target; *brothel-BROTH*) has been replicated across multiple experiments and languages, including English (e.g., Beyersmann, Ziegler, Castles, Coltheart, Kezilas, & Grainger, 2015; Marslen-Wilson, Bozic, & Randall, 2008; Rastle, Davis, & New, 2004), French (Longtin, Segui, & Hallé, 2003), Spanish (Lazaro, Illera, & Sainz, 2016), Korean (Kim, Wang, & Taft, 2015) and Russian (Kazanina, 2011). This pattern of results suggests that the semantics of either the entire prime (e.g., *trucker*) or that of its base morpheme (e.g., *truck*) is not accessed within the timeframe set by the duration

[4]This condition is labelled as semantically opaque relative to the transparent condition. However, as point out by Feldman, Milin, Cho, Moscoso del Prado Martín, and O'Connor (2015), many of the target stimuli used in this condition (see e.g., Rastle, Davis, & New, 2004) do not function as a true morphological stem. For example, *corn* does not serve as a true morpheme in *corner*, even though the latter includes a true suffix, -*er*. We therefore treat these prime-target pairs as a 'pseudo-suffixed' condition and the transparent prime-target pairs as a 'true derived' condition.

of the SOA. This result, statistically equivalent priming magnitudes across pseudo and true derived words, is taken as support for semantics-blind morpho-orthographic segmentation.

This body of evidence has been challenged by a growing number of studies implementing either masked priming or paradigms with very short SOAs. These studies demonstrate reliable semantic priming effects in Dutch, English, Finnish and Spanish (for extensive discussions, see Davis & Rastle, 2010; Feldman, O'Connor, & Moscoso del Prado Martín, 2009; Feldman, Milin, Cho, Moscoso del Prado Martín, & O'Connor, 2015; Van den Bussche, Van den Noorgate, & Reynvoet, 2009). Moreover, Duñabeitia, Kinoshita, Carreiras, and Norris (2011) employed a cross-case 'same-different' task. This task requires participants to assess the likeness of a lowercase and an uppercase letter string presented sequentially. They found no difference in priming magnitude across conditions containing true suffix strings (e.g., *trucker-TRUCK* and *corner-CORN*) and orthographic controls (e.g., *brothel-BROTH*), suggesting that semantics-blind morpho-orthographic segmentation is not obligatory. These results are also compatible with a host of theories of complex processing, that while varying in precise details, also assume that morphological activation is conditioned simultaneously by form and meaning characteristics of the complex word (Diependaele, Sandra, & Grainger, 2005; Diependaele, Duñabeitia, Morris, & Keuleers, 2011; Libben, 2005, 2014; Moscoso del Prado Martín, 2007; Plaut & Gonnerman, 2000), or that the activation of complex word meaning is not aided by a morphological level of representation at all (Baayen, Milin, Đurđević, Hendrix, & Marelli, 2011).

Despite contradictory results, the evidence accumulated using the masked priming technique supports the account that morpho-orthographic segmentation during complex word recognition is distinct from semantic access. However, as acknowledged by Rastle and Davis (2008, pp. 949), it may be "impossible" to infer the time-course of word recognition from a prime duration. That is, studies using masked priming with short SOAs are suboptimal for characterizing the absolute time-point at which these cognitive processes occur, such that an SOA of 40 ms does not indicate that decomposition occurs at 40 ms after presentation of the visual stimulus.

4.1.2 Evidence from neurophysiological studies

This lack of temporal precision is remedied by the use of paradigms that allow for an excellent temporal resolution of neural behaviour as it unfolds during visual word recognition, primarily magneto- and electroencephalography (MEG and EEG). Results of cross-linguistic MEG and EEG recording during masked priming or unprimed lexical decision of derived, inflected and compound words serve as a second source of empirical support for the form-then-meaning accounts of morphological processing. A number of experiments (Lavric, Clapp, & Rastle, 2007; Lavric, Elchlepp, & Rastle, 2012; Morris, Frank, Grainger, & Holcomb, 2007; Royle, Drury, Bourguignon, & Steinhauer, 2010; 2012; Morris & Stockall, 2012) that have combined event-related brain potentials (ERPs) with the masked morphological priming paradigm have reported results which converge with the observation that early components of neural activity are devoted to the processing of morpho-orthographic features. These studies showed that neural activity at 100-200 ms post-onset of the visual stimulus is only sensitive to visual features of the prime and the target, while later neural activity, between 300-500 ms post-onset of the visual stimulus, was affected by the semantic relatedness of the prime and target.

To take an example of one such experiment, in an unprimed lexical decision-ERP experiment, Lavric, Elchlepp, and Rastle (2012) reported an orthographic correlate of morphological decomposition taking place about 190 ms post-onset in the ERP signal, while differences between processing semantically transparent derived words (e.g., *darkness*) and pseudo-derived words (e.g., *corner*) were detected in neural activity approximately 70 ms later (see Zweig and Pylkkänen, 2009, for converging MEG data). In addition, Lavric, Rastle, and Clapp (2011) reported an ERP-priming study with long SOAs which contrasted true derived prime-target pairs (e.g., *magical-MAGIC*), pseudo-suffixed pairs (e.g., *compassion-COMPASS*) and orthographic control pairs (e.g., *brothel-BROTH*). The authors found effects of transparency (as gauged by Latent Semantic Analysis; Landauer & Dumais, 1997) at 400 ms post stimulus onset (a late N400 component). Based on these results, the authors concluded that, “The lateness of the effects of semantic transparency favours a single, orthography-based, mechanism of morphological decomposition ‘licensed’ at a late processing stage” (Lavric et al., 2011, pp. 684).

In line with the aforementioned ERP data, several MEG studies that record activity during lexical decision performance also report that effects of any and all semantic properties are only detectable in the late components of complex word recognition. MEG studies appear to converge on the M350 component (defined as a peak in neural activity approximately 350 ms after stimulus presentation) as the processing time window that is sensitive to lexical frequency, semantic transparency, morphological family frequency and size (the number and cumulative frequency of complex words sharing a morpheme), and lexical ambiguity (e.g., homonymy, heteronomy and polysemy), see Beretta, Fiorentino, and Poeppel (2005), Simon, Lewis, and Marantz (2012), Solomyak and Marantz (2009, 2010), and Lewis, Solomyak, and Marantz (2011); and also reviews in Fiorentino and Poeppel (2007) and Pylkkänen, Llinás, and Murphy (2006).

Of particular relevance to the current investigation, Solomyak and Marantz (2010) carried out a MEG study which indicated neural sensitivity to statistical properties of the stem and affix (M130 component), later parsing of the stem and affix (M170 component), which was then followed by an even later "lemma" activation stage (M350). The results of this study are well-suited to the time-course propounded in the semantics-blind obligatory morphological decomposition account, a prediction which is formalized by Solomyak and Marantz (2010) as follows:

1. In the first stage of processing, potential affixes are recognized by form, and parsing between stem and affix is attempted based on activation of the (visual) word form of the stem. Thus, we expect affix-specific variables such as affix frequency to correlate first with brain activity, followed by variables associated with parsing, such as the transition probability between stem and affix. Parsing of stem and affix leads to lexical access for the stem. At this stage, lexical variables including lemma frequency should be relevant, even for bound roots... Following lexical access to the stem, recombination of stem and affix occurs. Here, variables such as the transition probability between affix and stem and/or surface frequency should correlate with brain activity. (Solomyak & Marantz, pp. 2045)

With a similar set of predictions, Fruchter and Marantz (2015) recently investigated the time-course of English derived word recognition in a MEG experiment involving lexical decision to derived words. They reported left-temporal activity associated with derivational family entropy as early as 240 ms, followed by a facilitatory effect of surface frequency, in the same region, at a later time window of 430-500 ms. They also reported an effect of semantic coherence of the derived word at the left orbitofrontal region of interest, at 300-500 ms. The results of Fruchter and Marantz (2015), confirmed their prediction that "semantic effects, like the effects of surface frequency, should occur after the effects of derivational family entropy, which relate to stem lookup" (pp. 82). In addition, the authors also argued that these results are compatible with the timing of effects outlined in the obligatory decomposition model, in which base frequency effects "always" emerge at the early decomposition stage, while "surface frequency has its impact at the subsequent combination stage" (Taft, 2004, pp. 762).

In contrast to the above results, a line of ERP and MEG research demonstrates earlier (140-200 ms post-onset) effects of word frequency, lexicality, semantic coherence of a word's morphological family and other semantic lexical properties (Assadollahi & Pulvermüller, 2003; Hauk, Davis, Ford, Pulvermüller, & Marslen-Wilson, 2006; Sereno, Rayner, & Posner, 1998; Pulvermüller, 2002; Penolazzi, Hauk, & Pulvermüller, 2007). These findings appear to substantially bring forward the time window for the access of word meaning as compared to the neuroimaging studies surveyed above. Yet, to our knowledge, these effects are only reported for short morphologically simple words and may not hold for longer, more effortful processing of complex words. To sum up, the fine temporal resolution afforded by MEG and ERP techniques favours the outline of form-then-meaning theories of complex word recognition. That is, cortical activity associated with the processing of printed complex words suggests that orthographic segmentation into morphemic units is obligatory and proceeds in a strictly serial and semantics-blind manner.

4.1.3 Evidence from eye-movement research

The neuroimaging evidence provides an absolute estimate of the word processing time-course and also makes a claim regarding the relative order of effects, i.e., it suggests that semantic access to the whole word is licensed only after morpho-orthographic segmentation has occurred. Yet, eye-tracking – another paradigm offering a fine temporal resolution through the registration of eye-movements during silent reading – offers a number of results that are incompatible with both the absolute and the relative timelines of neuroimaging studies.

First, eye-movement studies report earlier effects (estimated by average eye fixation durations) than the time-points found during EEG or MEG recording. The registration of eye-movements during silent reading has revealed that a typical simplex word is fixated on only once with an average fixation time of 200-250 ms (see Rayner, 1998). This single fixation duration shows sensitivity to a host of lexical-semantic properties, such as dominant and subordinate meanings of the word, the emotional connotation of the word, imageability, concreteness, homonymy and polysemy (Rayner, 1998; Staub & Rayner, 2007; Schotter & Rayner, 2012). The finding that the behaviour of the visuo-oculomotor system is influenced by word meaning within 200-250 ms of the stimulus fixation is clearly incompatible with the 350-400 ms locus of semantic effects argued for in much of the MEG and EEG literature (for a discussion of this apparent paradox see Dambacher and Kliegl, 2007; Dimigen, Sommer, Hohlfeld, Jacobs, and Kliegl, 2011; Kliegl, Dambacher, Dimigen, Jacobs, and Sommer, 2012; Sereno and Rayner, 2003; Sereno, Rayner, and Posner, 1998).

Secondly, the sequentially late effect of surface frequency (the frequency of the whole word), reported by Fruchter and Marantz (2015) is seemingly incompatible with the results of a recent series of non-parametric distributional analyses (Staub, White, Drieghe, Hollway, and Rayner, 2010) which suggest that surface frequency shows a very early influence on word recognition speed. The authors re-analysed data obtained from two eye-movement experiments, in which participants silently read critical words that were embedded within sentences (Drieghe, Rayner, & Pollatsek, 2008; White, 2008). Distributional analyses revealed that word frequency exerts an influence on reading times across the full distribution of both first fixation durations (the duration of the reader's first eye fixation on the word) and gaze durations (the summed duration of all fixations before the eyes leave the word). To be exact, using the vincentile (Ratcliff, 1979; Vincent, 1912) plotting technique, Staub et al. (2010) showed that word frequency exerts an influence on eye-movement fixations (for first fixation duration and gaze duration) as early as 180-200 ms. Moreover, Pollatsek, Hyönä, and Bertram (2000; experiment 2) found no difference in the effect size of whole-word frequency on first fixation durations of (10-12 letter) Finnish compounds and matched monomorphemic words (average first fixation duration across (i) compounds = 198 ms, and (ii) monomorphemic words = 205 ms). This finding indicated that the influence of whole-word frequency was not delayed for compounds relative to the monomorphemes. These results and others (e.g., Rayner, Liversedge, White, & Vergilino-Perez, 2003; Rayner, Ashby, Pollatsek, & Reichle, 2004) demonstrate that the word frequency effect influences the very initial stages of lexical processing during reading and that the timing of the effect may be indifferent to the morphological complexity of the word.

Thirdly, eye-tracking studies of Dutch, English, Finnish and Italian compound words (Juhasz and Berkowitz, 2011; Kuperman, Bertram, and Baayen, 2008; Kuperman, Schreuder, Bertram, and Baayen; 2009; Marelli and Luzzatti, 2012) have revealed that lexical and semantic factors reliably affect the earliest eye-movements involved in reading. For example, Kuperman et al. (2009) found that the duration of first fixations (average = 270 ms) on long (8-12 letter) Dutch compounds was affected by lexical frequency of the compound as well as by the family size of the left constituent (*car* in *carwash*). Similar effects of compound frequency and left constituent family size are detected in the duration of the first fixation (average = 221 ms) on 10-18 letter-long Finnish compounds (Kuperman et al., 2008). Moreover, Juhasz and Berkowitz (2011) observed effects of left constituent family size on the likelihood of refixation on English compound words and on gaze durations (average gaze duration average across small and large family size conditions = 272 ms). This result indicates that morphological family size influences the decision to refixate at a very early time-point, i.e., the point at which the first of more than one fixations is made. Furthermore, early effects of the family size of the base word (e.g., *dream* in *dreamful*) and suffix productivity also influenced the duration of single fixations (average duration = 245 ms) on Dutch suffixed words (Kuperman, Bertram, & Baayen, 2010). This body of results aligns with De Jong's (2002) proposal that "the percept of the meaning of a word in the mind depends not only on the activation of its own meaning, but also on the co-activated meanings of its family members" (pp. 102). More importantly, the above studies suggest that such semantic processes are also engaged from the very earliest moments of complex word recognition.

Lastly, numerous recent studies have targeted semantic effects of complex words directly, and have found that they arise early in the eye-movement record. For example, Amenta, Marelli, and Crepaldi (2015), recently found evidence for early semantic access in an eye-tracking during sentence reading study which manipulated semantic transparency in Italian derived words. Namely, they found an effect of semantic transparency and stem frequency in first fixation durations in a dataset where the average first fixation duration was 243 ms. Amenta et al. (2015, pp. 1590) argue that their results indicate that early processing of morphologically complex words involves access to *form-and-meaning*. That is, morphological decomposition may involve concurrent access to meanings of the decomposed word form. Furthermore, in an eye-tracking during naturalistic reading study, Marelli and Luzzatti (2012) manipulated both the semantic transparency and the head position in Italian compounds. They reported a reliable influence of semantic transparency as early as 231 ms for first fixation durations averaged over quartiles of their semantic transparency measure. The data of Marelli and Luzzatti (2012) serve as evidence that access to the meaning of a complex word does not wait for complete access to the word's orthography during naturalistic word reading. Finally, Marelli, Amenta, Morone, and Crepaldi (2013; experiment 1) found further eye-movement effects that challenge the obligatory composition accounts. In a lexical decision experiment that measured eye-movements as a dependent variable (instead of a button press latency), Marelli et al. found support for an effect of semantic transparency. Using the same experimental conditions as used by Rastle et al. (2004), but for Italian derived words, they found a priming effect for transparent prime-target pairs.

The wealth of evidence of lexical and semantic influences on early eye-movement measures, observed across languages and word presentation formats, suggests two important characteristics of morphological processing. First, lexical properties such as complex word frequency, the family size of the left constituent and semantic transparency reliably affect complex word recognition within 200-250 ms post fixation of the visual stimulus, that is, within an average duration of the first-of-many or single fixation reported in the studies cited above. This is a substantial reduction as compared to 350-400 ms reported by neurophysiological studies as the temporal locus for complex word frequency and family size effects, cf. Pylkkänen et al. (2004). Second, effects of complex word frequency and family size on the first fixation precede full inspection of the complex word (typically word-final letters). This finding follows from the fact that the lengths of stimuli in work cited above ranged from 8 to 18 letters and the probabilities of refixation on the complex words were well above 60% (except in Kuperman et al., 2010). Based on these results, it would appear that the fine temporal resolution afforded by eye-tracking methodology places the semantic effects of words at an earlier absolute time-point. Moreover, the aforementioned results from Italian studies (e.g., Marelli and Luzzatti, 2012) suggest that semantic effects may emerge concurrently with orthographic effects. Crucially, though the absolute time-points of the effects of semantics are not consistent across the eye-tracking and the neuro-physiological literature, it would an estimation of the relative order of such effects that present the important test of the semantics-blind obligatory decomposition account.

4.1.4 The current study

Given the contradicting evidence in support of late semantic access (e.g., Crepaldi, Rastle, Coltheart, & Nickels, 2010; Fruchter & Marantz, 2015; Taft & Forster, 1976; 2004; Rastle et al., 2008) and early semantic access (e.g., Feldman et al., 2009; Feldman et al., 2015; Marelli & Luzzatti, 2012), the present research is motivated by the goal of providing further insight into the temporal sequencing of cognitive processes during printed complex word recognition. In this experiment, we employ a relatively novel non-parametric distributional approach to the analysis of reaction times (Reingold, Reichle, Glaholt, & Sheridan, 2012; Reingold & Sheridan, 2014). This approach, the non-parametric survival analysis technique, was developed with the purpose of establishing the earliest point in time at which a given variable exerts an influence over a chronometric response latency. The technique estimates a survival curve for two conditions, and through bootstrapped resampling of data points within each condition, establishes the point in time at which the two survival curves for each condition begin to diverge from one another. This divergence point is taken as an estimate of the earliest point in time at which the given variable has a discernible impact on response latencies.

In what follows, we first retrieve unprimed lexical decision latencies from the British Lexicon Project (BLP; Keuleers, Lacey, Rastle & Brysbaert, 2012) and Dutch Lexicon Project (DLP; Keuleers, Diependaele, & Brysbaert, 2010). With these lexical decision latencies serving as a dependent variable, we then employ the survival analysis technique to compute separate divergence point estimates for a range of continuous lexical variables, many of which have been used in prior studies on morphological processing (e.g., Fruchter and Marantz, 2015; Moscoso del Prado Martín, Kostić, and Baayen, 2004) and are considered to be diagnostic of the processes involved in morphological decomposition.

Survival analyses were conducted for separate categories of words in the lexical decision data pools. The categories of words represented three (A-C) different quasi-experimental conditions (A, derived words; B, pseudo-derived words; C, form control words). The three conditions were formulated in order to form a direct comparison with the conditions presented in previous masked priming lexical decision experiments (e.g., Rastle et al., 2004). Survival analysis for derived words (condition A) was estimated for lexical decision latencies in English (A1) and Dutch (A2). Thus, the survival analysis was computed for the following four sets of words: (A1) *English derived words*, where the word consists of an English stem and suffix combination (e.g., *badness*; *bad* and -*ness*); (A2) *Dutch derived words*, where the word consists of a Dutch stem and suffix combination (e.g., *duiker* "diver"; *duik* and -*er*); (B) *English pseudo-derived words*, cases where the word string contains a combination of a simplex word stem and an existing suffix, but where the stem does not function as a morpheme internal to the word structure (e.g., *wander*: *wand* and -*er*); and (C) *English form control words*, where the whole word is simplex, contains an embedded simplex word substring, and contains a final sequence of letters that do not represent a true suffix, (e.g., *ballad*; *ball* and -*ad*).

Operationally, the chronology of lexical effects in the current survival technique implementation is determined by the order of divergence point estimates for each lexical property across each word condition (derived, pseudo-derived, form control). Under the predictions of the semantics-blind obligatory decomposition hypothesis, one would expect that for derived word processing, variables associated with the form and meaning of the whole word string (e.g., surface frequency, semantic transparency and emotional valence; see below for definitions) to have divergence point estimates that occur later than those associated with morphological processing. To be precise, if one refers back to the timeline of cognitive effects outlined by Solomyak and Marantz (2010) in quote (1), the form-then-meaning account would predict the following sequence of divergence point estimates. Affix-specific variables, such as affix frequency, are expected to emerge first, followed by variables related to parsing of the stem and affix, such as transition probability between the stem and affix. The initial stages of recognition are also expected to be dominated by variables indicating access to orthographic form (such as orthographic neighbourhood density; Fruchter & Marantz, 2015). The influence of these properties would correspond to the initial stages of recognition, which is devoted to morpho-orthographic properties of the input. Once the morphological properties of the affix has been accessed, lexical access to the stem is expected to commence. The initiation of this process is assumed to be associated with the onset of the effect of lemma frequency, stem frequency or derivational family entropy (Fruchter & Marantz, 2015, pp. 83). Following this, recombination of the stem and affix is predicted to occur. This process is indexed by the emergence of the effect of transition probability between stem and affix (again) and/or surface frequency. The onset of LSA distance is also expected to emerge at this point also, as an indicator of the system computing the semantic fit between the stem and the derived word.

The semantics-blind obligatory decomposition accounts also predict that the sequencing of lexical effects for pseudo-derived (opaque) word processing to be equivalent to that of derived words. Naturally, form controls are not expected to demonstrate any influence of lexical properties associated with morphological processing (for any morphological characteristics that are able to computed for this class of words). Thus, the same sequence of divergence points are expected to emerge as outlined for derived and pseudo-derived words, except with an absence of morphological variables.

In contrast, form-and-meaning accounts (e.g., Feldman et al., 2009; Feldman et al., 2015; Marelli & Luzzatti, 2012) predict that divergence points of lexical properties associated with the meaning of the whole derived word to precede, or be contemporaneous with, the divergence points of lexical variables associated with morphological decomposition. Following from the evidence cited earlier (e.g., Assadollahi & Pulvermüller, 2003; Hauk et al., 2006; Sereno et al., 1998; Pollatsek et al., 2000; Pulvermüller, 2002; Rayner et al., 2004; Staub et al., 2010), the effect of surface frequency is expected to arrive relative early. These findings indicate that the timing of the onset of surface frequency would be indifferent to whether the word is complex or not. Therefore, for all conditions (true derived words, pseudo-derived words and form controls), form-and-meaning accounts expect to find early divergence point estimates for surface frequency and/or orthographic form variables (such as orthographic neighbourhood density) followed by divergence point estimates for semantic access, such as LSA distance, and other semantic variables pertaining to the whole word form (such as valence). Critically, for derived words and pseudo-derived words, this family of accounts predicts that access to meaning precedes without intervention of morphology-specific characteristics. Thus, variables such as lemma transition probability and derivational family entropy are expected to arrive after, or concurrently with, the onset of semantic access. Again, the timeline of cognitive processing for simplex form controls is expected to follow the same sequence as derived and pseudo-derived words, with the exclusion of morphological variables. This follows from the prediction that variables associated with morphological processing are not expected to play a role in the processing of simplex words.

4.2 Methods

4.2.1 Participants, Materials, and Procedure

Lexical decision results were retrieved from The British Lexicon Project (BLP; Keuleers et al., 2012) and the Dutch Lexicon Project (DLP; Keuleers et al., 2010), which are collections of lexical decision latencies to over 14,000 words and an equal number of non-words in English and Dutch. In both datasets, word stimuli were selected from lists of mono- and disyllabic words, representing a broad range of frequency of occurrence and phonological and morphological complexity: non-word stimuli were phono-tactically valid and closely matched the syllabic and phonological structure of word stimuli. In the BLP study, 78 students and employees of Royal Holloway University of London responded to all stimuli, using their dominant hand for word responses and the non-dominant hand for non-word responses: for further details on stimuli, procedure and apparatus see Keuleers et al. (2012). In the DLP study, 39 participants responded to all stimuli using a similar procedure: for full details, see Keuleers et al. (2010).

4.2.2 Response variables

The dependent variable in all analyses was the lexical decision latency. We only considered correct responses to word stimuli, which constrained the overall data pool of 2,240,940 responses to 8,48,108 data points for the BLP study and from the overall pool of 1,098,942 responses to 462,244 data points for the DLP study.

4.2.3 Predictor variables

The key question of this paper is the relative order of effects elicited by variables related to a derived word's morphological structure (including frequency-based measures), orthography and semantics. We discuss these groups of variables in turn.

Frequency characteristics

We used the 200 million-token corpus of British films and media subtitles, SUBTLEX-UK (Van Heuven, Mandera, Keuleers, & Brysbaert, 2014) to estimate frequencies of occurrence for both whole (derived, pseudo-derived and form control) English words as well as their (pseudo)stems or embedded strings (e.g., *rainy* and *rain*; *trumpet* and *trump*; *ballot* and *ball*). The 44 million-token corpus of Dutch film and media subtitles, SUBTLEX-NL (Keuleers, Brysbaert, & New, 2010) was used for estimates of the equivalent Dutch language frequency-based measures. We refer to the frequency of the whole word as surface frequency, and the frequency of the derived/pseudo/form control stem, as stem frequency. We opted to used stem frequency instead of lemma frequency. This is because surface frequency and lemma frequency are highly correlated (r = 0.934), whereas surface frequency and stem frequency correlate weakly

($r = 0.135$). Thus stem frequency is less likely to impose a confounding impact on the results of surface frequency, and vice versa. We discuss the impact of collinearity further in Study 1.

Morphological variables

We considered three morphological variables. The first is the lemma transition probability (TPL), which is defined as the ratio of each word's surface frequency to its lemma frequency (Solomyak & Marantz, 2010). This variable measures the conditional probability of encountering the whole word given the stem, and is thus taken as an ideal variable with which to detect processing effects of morphological parsing. In addition, we computed derivational family entropy for the true derived and pseudo-derived word conditions. This was achieved by estimating lemma frequencies of words that shared the (true or pseudo-)stem with the target (e.g., *zippy* and *zipping* for *zipper*). Derivational entropy was calculated as Shannon's entropy over the probability distribution p of the word's morphological family (obtained by dividing frequencies of family members by the cumulative family frequency): $H = -\Sigma log_2(p) * p$. Our definition of a morphological family included both derived and compound words: constraining families to just derivations did not appreciably alter results reported below. We also considered suffix productivity (SP), which was estimated as the number of word types that shared this suffix with the target word.

The metrics of morphological structure defined above (i.e., derivational entropy and suffix productivity) were quantified with the aid of the morphological parsing made available for 79,672 words in the English Lexicon Project (Balota, Yap, Hutchison, Cortese, Kessler, Loftis, Neely, Nelson, Simpson & Treiman, 2007) and associated word frequencies from the Hyperspace Analogue to Language corpus of English (HAL; Lund and Burgess, 1996), and the morphological parsing available from the CELEX lexical database for Dutch (Baayen, Piepenbrock, & Van Rijn, 1995).

Orthographic variables

To measure the orthographic neighbourhood of a word, we calculated the average Levenshtein distance (OLD20), which was defined as the mean orthographic distance from the 20 nearest orthographic neighbours. This measure was estimated for each target word in our stimulus list, using the library `vwr` (Keuleers, 2013). The SUBTLEX-UK corpus word list was used as the lexicon with which to estimate orthographic neighbourhood density for a given word in Engish, and SUBTLEX-NL in Dutch. We also calculated a measure of the transition between the root and the affix in all types of words (true derived, pseudo-derived and form control words). This variable is labelled here as the bigram transition probability (TPB; see Solomyak & Marantz, 2010). TPB is defined as the frequency of the first letter of the suffix given that the preceding letter (i.e., the last letter of the stem) appears in its position relative to the end of the word. We also considered word length in characters. Finally, we considered orthographic frequency of the string that represents the target word's suffix, regardless of its morphological status within a word (e.g., *-ly* in *rapidly* and *ply*). This measure of the frequency of the form of the suffix is referred to as form suffix productivity and FSP.

Semantic variables

Following Kuperman (2013), we considered both "relational" semantic properties of derived words, i.e., ones that are defined as a relationship between meanings of the whole complex word and its embedded word, and their "atomic" properties, i.e., ones that only require semantic access to either the stem or the whole word, and not the relationship between them both. As a relational property, we considered a computational measure of the semantic similarity (or transparency) between the stem and the derived word. We defined semantic similarity as the cosine between distributional vectors representing two words in a multi-dimensional lexical space. To this end, we collected pairwise estimates of semantic similarity using the Latent Semantic Analysis (LSA) from the UKWAC and SUBTLEX-UK corpus of English (available at http://zipf.ugent.be/snaut-english/, Mandera, Keuleers, & Brysbaert, in press). This application uses a 300-dimensional semantic space with CBOW embeddings, and a 6-word window for calculating co-occurrence statistics. The LSA solution for Dutch was trained on SONAR-500 and subtitle corpora, and used a 200-dimensional semantic space with CBOW embeddings and a window of 10 (available at http://zipf.ugent.be/snaut-dutch/, Mandera et al., in press). A greater LSA score indicates a greater dissimilarity between meanings of a pair of words.

Two "atomic" semantic properties were additionally examined: the psychological valence (positivity) of the whole word and, that of its stem. Valence estimates were obtained from a set of norms to 14,000 English lemmas (Warriner, Kuperman, & Brysbaert, 2013) and a set of norms for 4,300 Dutch lemmas (Moors, De Houwer, Hermans, Wanmaker, van Schie, Van Harmelen, De Schryver, De Winne, & Brysbaert, 2013). Words were rated on a scale of 1-9 (sad to happy) by about 20 raters each. In these norming studies, words were presented in isolation, without any information about word sense, word's part of speech, or supporting context: the average of these ratings was taken as the value of the word's semantic norm. Emotional valence is a semantic variable that has demonstrated effects in prior research. Experimental evidence has revealed that readers maintain automatic vigilance towards the emotional affect of words, such that non-neutral stimuli capture attention faster and automatically (Adelman & Estes, 2013; Kuperman, Estes, Brysbaert, & Warriner, 2014). Our choice of this connotative property of word meaning was also motivated by Kuperman's (2013) observation that valence was the only semantic property of morphological constituents that exerted an effect on lexical decision latencies to compounds. He found that compound words with more emotionally positive constituents (either left or right constituent) facilitated RTs to the whole compound word, over and above the effect of the valence of the whole compound (i.e., the compound word *dragonfly*, was affected independently by the emotional valence of its constituents *dragon* and *fly*).

4.2.4 Distributional analysis

The analytical method at the forefront of this study is non-parametric distributional survival analysis, which was introduced to psycholinguistic research by Reingold, Reichle, Glaholt and Sheridan (2012) and was further refined in Reingold and Sheridan (2014). Our presentation of the method below closely follows the overview in Reingold and Sheridan (2014): We refer readers to this paper for a detailed exposition of the technique. The distributional survival analysis method was developed in order to estimate the earliest point in time at which an experimental variable shows a discernible effect on a distribution of response latencies. At focus of the survival technique is the creation of separate survival curves for conditions which are formed by levels of the experimental variable (e.g., high vs. low frequency, or availability vs. unavailability of parafoveal preview, Reingold et al., 2012). A survival curve depicts, for every time-point t, the percentage of responses (fixation durations, lexical decision latencies or other) that have a duration longer than t: see Figure 4.1 for survival curves of low vs. high surface frequency in Study 1 (below). The depiction of the survival curve illustrates that immediately after stimulus presentation, all responses still remain out of the full distribution of responses as neither a button press (in lexical decision) nor the termination of a fixation (in eye-tracking) has been executed; thus for earliest time-point, survival proportion is 100%. However, as time moves on within the distribution of responses, lexical decision button presses are executed, or eye-fixations are terminated. Thus, greater values of t are associated with a progressively higher number of trials that have been terminated by a response decision. Eventually, the survival percentage (proportion of responses still un-executed) decreases, until it eventually reaches 0% at t equal to the longest observed response time. If experimental conditions lead to different distributions of response times, the difference between conditions would appear as two different survival curve functions, with a faster condition showing lower survival percents than a slower condition, at least at some time-points. The distributional analysis proposed in Reingold et al. (2012) and Reingold and Sheridan (2014) identifies the earliest divergence point between two survival curves. Critically, this divergence point is indicative of the earliest temporal locus of the experimental effect elicited by the contrast between two experimental conditions.

In this paper, we adopted the Confidence Interval Divergence Point Analysis (DPA) outlined by Reingold and Sheridan (2014), which produces both the divergence point estimates and confidence intervals for a comparison between experimental conditions. We computed the DPA analysis using the `RTSurvival` (Sheridan, Reingold, & Matsuki, 2016) package in the R statistical computing software program (R Core Team, 2014). The DPA procedure uses a bootstrap

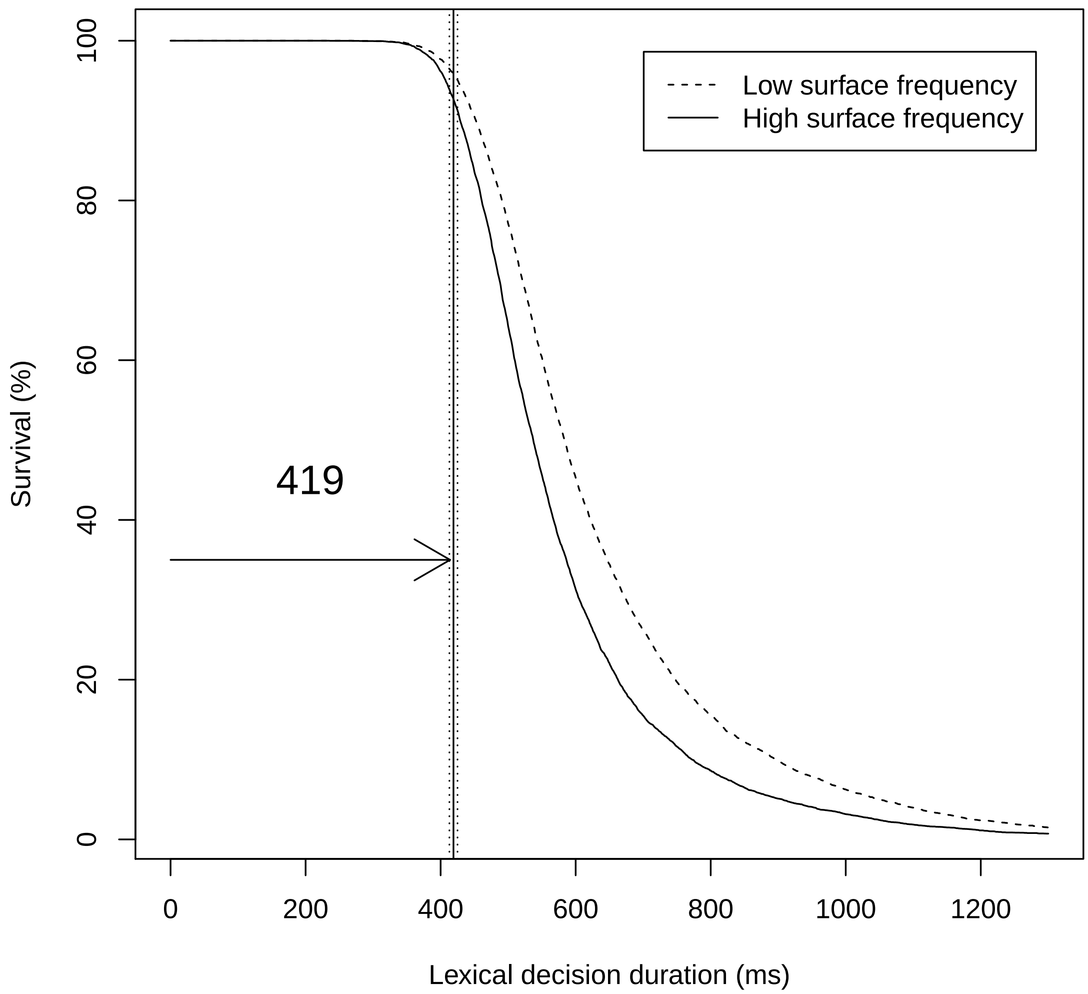

Figure 4.1: The divergence point estimate and its confidence interval for survival curves formed by the surface frequency effect in Study 1.

resampling technique (Efron & Tibshirani, 1994), which draws multiple random samples with replacement from an available pool of observations, calculating the statistic of interest at each iteration. As implemented by Reingold and Sheridan (2014), the DPA procedure runs 1,000 iterations of random resampling (with replacement) of response times for each participant and condition. For each iteration of the bootstrap procedure, survival curves are generated for each individual participant. Next, for each 1-ms bin ranging from 1 to 1200 ms, survival percent values are averaged across participants to produce the group survival curves and the divergence point is estimated. Specifically, for each iteration, the divergence point estimate is defined as the first 1-ms bin in a run of five consecutive bins in which the survival percentage in the slower (e.g., Low surface frequency) condition is greater than the survival percentage in the faster condition (e.g., High surface frequency) by a pre-defined percentage threshold. Next, divergence point estimates from the 1,000 iterations are sorted from the smallest to the largest value. The 25^{th} and 975^{th} values of this distribution of divergence points constitute the 95% confidence interval. In addition, the median of the 1,000 divergence point values is used as the divergence point estimate for the sample. Based on sets of simulations, Reingold and Sheridan recommend 1.5% as the minimum difference threshold between survival percentages. This value avoids a false detection of a divergence point due to noise or low statistical power (as low as 12 observations per participant per condition) and they remark that it is safe to lower this threshold in very large samples. Even though our datasets do not suffer from low statistical power (with about 50 to 500 observations per participant per condition), we stipulated a more conservative difference threshold of 3% between conditions. This ensures that we uncover relatively strong effects, and further reduces the possibility of false positive detection.

To illustrate an example from the current investigation, the analysis in Figure 4.1 indicates 419 ms as the divergence point for survival curves of low vs. high frequency derived words in Study 1, with a narrow 95% confidence interval of 415 to 426 ms. This distribution of divergence points marks the last point in a sequence of five adjacent 1-ms bins showing a divergence of 3% or more: it is observed after only 5.84% of responses were terminated.

Our overarching interest lies in using the technique just described to detect the onset of the effect for multiple lexical and sublexical variables. For each variable in turn, we generated experimental contrasts by splitting words at the median of that predictor (e.g., whole-word psychological valence), and estimating the divergence point between survival curves associated with words in these groups. Importantly, the non-parametric survival analysis is designed for true factorial manipulations, where all but one or two critical contrasts between stimuli (e.g., a frequency manipulation) or presentation conditions (e.g., the validity of parafoveal preview) are minimized through matching or repetition. No matching was applied to our data, which represent naturalistic distributions of variables in the language. Given the collinearity between many lexical predictors, distributions of variables other than predictor under consideration might also vary across the two groups formed by the median split of that predictor. Furthermore, distributions of variables in surviving responses change over time. We discuss this issue in Study 1 and show that, in our datasets, this potentially confounding influence is not a concern.

4.3 Results and discussion

4.3.1 Study 1: Derived English words

We selected 544 derived words ending in one of the following suffixes: -*ion*, -*y*, -*ful*, -*er*, -*ness*, -*ly*, -*en*, -*less*, -*ment*, -*ize*. The derived words were relatively transparent in that their stems were free-standing lexemes and were semantically related to the meaning of the whole word. The full list of stimuli, along with the orthographic, morphological and semantic properties of whole-words and stems, is reported fully in the online supplementary materials S1. After restricting the BLP to only correct responses to the 544 derived words, a total of 19,285 response times were available. Table 4.1 reports descriptive statistics of all predictors for selected words as well as response times to the words. Table A.1 of Appendix A reports pairwise Pearson correlations between predictor variables.

We split each predictor variable by the value of the median to create two quasi-experimental conditions. For example, splitting frequency at the median value provides two conditions, one with words that have high frequency values and one with

Table 4.1: Descriptive statistics for lexical variable across each word group.

Condition		RT	Surface freq.	Stem freq.	Valence of word	Valence of stem	LSA distance	Derivational Entropy	Suffix prod. (SP)	Lemma transition probability (TPL)	Word length	Orthographic neighbourhood density (OLD20)	Bigram transition probability (TPB)	Form suffix prod. FSP)
Derived	Min	236	5	31	1.85	1.53	0.296	0	5	0.117	4	1.000	0.0000	88
	1st Qu.	488	160	1878	3.90	4.50	0.572	0.8324	1214	0.677	6	1.400	0.0013	2056
	Median	558	352	9019	5.00	5.59	0.666	1.309	2276	1.000	6	1.650	0.0026	6646
	Mean	606	1362	32630	4.95	5.37	0.663	1.377	2276	0.834	6.59	1.732	0.0045	6448
	3rd Qu.	666	956	24990	5.86	6.47	0.757	1.93	3499	1.000	7	1.925	0.0062	13050
	Max	2213	89540	681500	8.21	8.21	0.995	3.857	3499	1.000	11	3.700	0.0495	13050
Derived (Dutch)	Min	347	6	8	1.67	1.38	0.368	0	56	0.400	5	1.100	0.0000	3
	1st Qu.	490	76	420	2.67	2.88	0.512	1.059	501.8	0.685	6	1.850	0.0016	41
	Median	548	291	1918	4.92	4.80	0.623	1.713	2991	0.800	7	2.100	0.0025	314
	Mean	578	983	6102	4.34	4.30	0.642	1.985	3187	0.793	7.39	2.195	0.0056	575
	3rd Qu.	632	882	4772	5.75	5.50	0.756	2.919	3097	0.952	8	2.600	0.0049	335
	Max	1315	15500	139900	6.42	6.42	1.022	5.289	10330	1.000	11	3.700	0.0683	5508
Pseudo-derived	Min	281	26	116	1.90	2.65	0.549	0	24	0.147	4	1.000	0.0002	225
	1st Qu.	477	336	606	4.49	4.61	0.809	0	24	0.691	5	1.112	0.0023	833
	Median	547	1291	2246	5.38	5.36	0.863	0.2476	2276	0.863	6	1.475	0.0046	6646
	Mean	591	6460	13900	5.22	5.19	0.856	0.6655	1829	0.805	5.82	1.452	0.0074	5737
	3rd Qu.	648	5585	10020	5.99	5.82	0.915	1.248	3499	0.997	6	1.700	0.0094	6646
	Max	1877	73400	362900	7.71	7.60	1.008	3.658	6254	1.000	9	2.600	0.0495	13050
Form	Min	272	9	25	1.95	2.12	0.361			0.185	5	1.000	0.0001	7
	1st Qu.	484	189	532	4.25	4.25	0.807			0.617	6	1.500	0.0032	416
	Median	552	434	1851	5.30	5.21	0.873			0.841	6	1.725	0.0054	983
	Mean	600	2599	18220	5.22	5.15	0.858			0.775	6.19	1.687	0.0112	2054
	3rd Qu.	660	2154	13210	6.12	5.99	0.932			0.989	6	1.850	0.0097	3098
	Max	2328	28430	717100	8.05	8.34	1.079			1.000	9	3.050	0.2975	18170

words that have low values. The Divergence Point Analysis (DPA) was applied to survival curves for each condition in order to detect the onset of the effect that the predictor elicited: see Statistical considerations for the description of the procedure, and Figure 4.1 for the estimated temporal locus of the derived word frequency effect (419 ms, CI 95% [415, 426]).

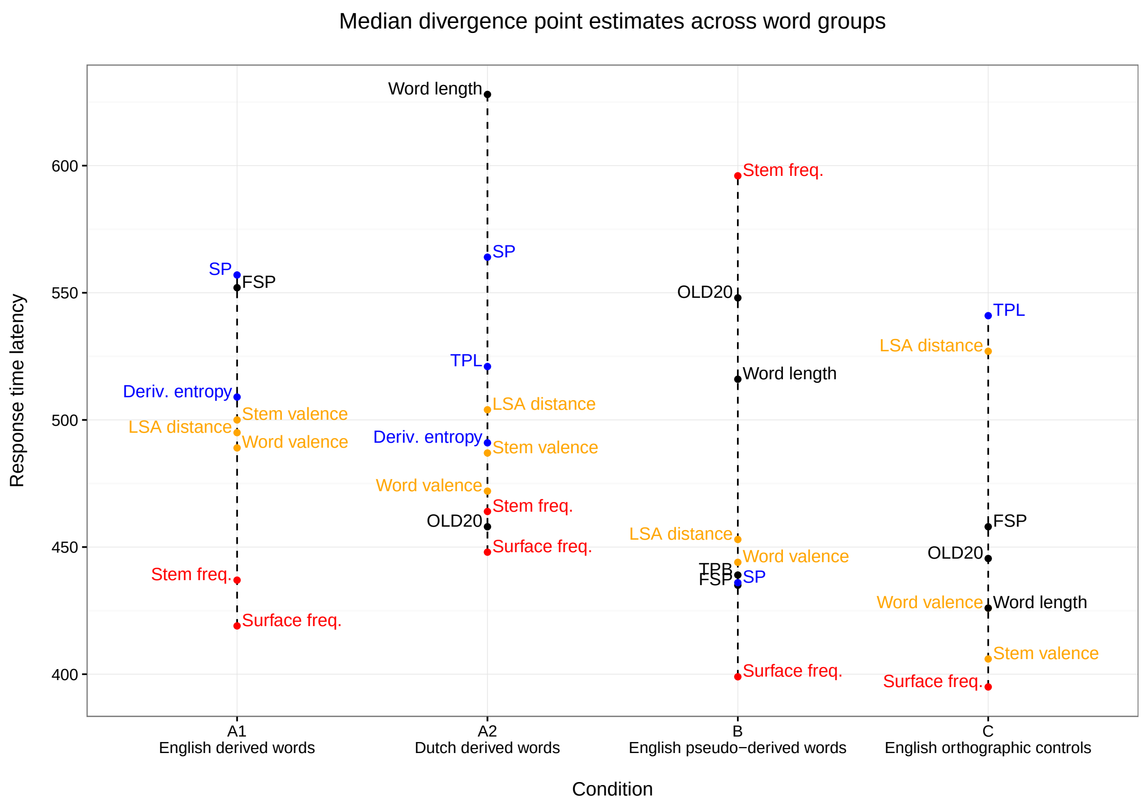

Figure 4.2: Plot of median divergence points for English derived words (A1), Dutch derived words (A2), English pseudo-derived words (B), and form control words (C). Colours represent groups of variables: frequencies of the whole word and stem (red); as well as orthographic (black), morphological (blue) and semantic (orange) predictors. Only variables that elicited a detectable divergence point between survival curves in at least 700 out of 1,000 iterations with a 3% minimal contrast are plotted (other effects were deemed too unstable for an estimate). Suffix productivity is abbreviated to SP, form suffix productivity is abbreviated to FSP, lemma transition probability is abbreviated to TPL and bigram transition probability is abbreviated to TPB.

The plot in Figure 4.2 (timeline A1) enables the visual inspection of median divergence point distributions across all lexical predictor variables. For English derived words, the plot reveals that effects of predictors under consideration occupy a broad temporal range, with median divergence points from 419 ms (a point at which 5.66% of responses are terminated) to 552 ms (48% of responses terminated). Surface frequency showed the earliest effect with a median divergence point of 419 ms, followed by the effect of stem frequency (437 ms). This suggests that access to the full-form precedes morphological stem look-up. Semantic effects show their onset next, in a contiguous succession of median divergence point estimates (between 475 and 500 ms). Semantic access to the connotation of the whole derived word precedes that of the connotation of the stem (as indicated by effects of whole-word valence and stem valence respectively), and overlaps in time with the effect of semantic transparency. The divergence point estimates that occured next in time were those that pertain to morphological and orthographic effects (derivational entropy, suffix productivity, and orthographic suffix probability). These effects were the only morpho-orthographic effects that passed our selection criteria and emerged late, with median divergence points between 509 ms and 557 ms. These findings suggest that orthographic cues do not play a major role in decomposing the stem and the suffix in relatively transparent "true" derived words. Moreover, access to the suffix, which should be granted as soon as the hypothesized process of early obligatory decomposition has been completed, is in fact very late, subsequent to access of the semantics of the whole-word.

What is the impact of collinearity?

As outlined in Statistical considerations, the quasi-experimental groups that are formed by the median split of a given predictor (henceforth predictor A) are not matched in our data on means or distributions of other predictors. Such a mismatch might not be problematic if the distributions of potential confounding variables remained stable over groups of predictor A across the entire time-course of word recognition. However, as word recognition unfolds, distributions of all predictors change, and may potentially do so at different rates. Thus, what appears to be an onset of the effect of predictor A, may in fact be a reflection of a change in the distributions of other predictors (B, C, D, E ... etc.) at that time-point. We examined whether this was indeed the case for pairs of predictors that were correlated with each other at $|r| > 0.1$. As an example, consider whole-word valence, which we will take as predictor A, and surface frequency, which we will take as the potentially confounding predictor B variable. In this data set, these variables correlated at $r = 0.159$. Since negative and low-frequency words take longer to respond to (for valence see Kuperman et al., 2014), distributions of valence and frequency are expected to gravitate towards a larger proportion of negative and infrequent words as time unfolds (lexical decision latencies protract) and survival percentage decreases. The critical test is whether the distribution of frequency is significantly different for the dataset where $t_1 = 1$ ms onwards, i.e., when no response is terminated, compared to the dataset at t_2 and onwards (here, 489 ms), when the median divergence point for whole-word valence is detected.

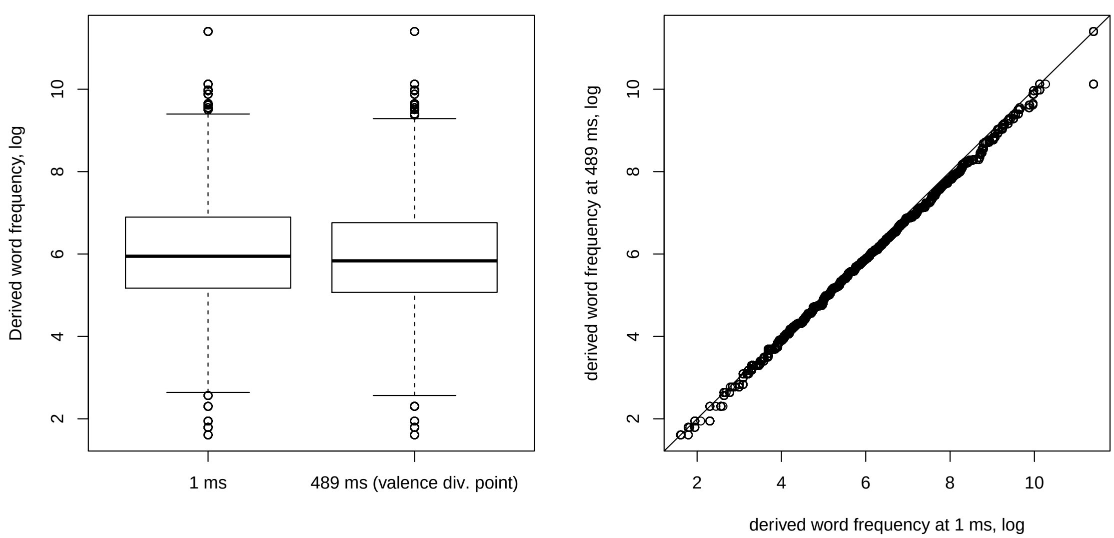

Figure 4.3: Boxplots (left) and the quantile-quantile plot (right) of the distributions of word frequency for positive derived words in English. The solid line indicates the expected relationship for identical distributions.

Distributions of word frequency (i) between 1 ms and the median divergence point of word valence (489 ms) and (ii) between 489 ms and onwards are presented in box plots (Figure 4.3, left panel) and in a quantile-quantile plot (right panel).

For illustrative purposes, we log-transformed frequencies: formal distribution tests below are applied to raw frequencies. Figure 4.3 demonstrates that differences in frequency distributions are in the expected direction (lower frequency words have a higher survival rate at 489 ms and onwards) and are minor. A two-sample Kolmogorov-Smirnov non-parametric distribution test was applied to percentiles of the two distributions (at 1 and 489 ms) and revealed no reliable difference ($D = 0.038$, $p = 0.47$).

Furthermore, for all remaining permutations of pairs of lexical variables (serving as predictors A and B) in this dataset, we did not detect a reliable difference between the distributions of predictor B when the dataset was split at the median divergence point of predictor A. After applying the same distributional check across the remaining datasets in this study, we found that only 1 out of 442 permutations of variable pairs showed a potentially confounding effect (using a Bonferroni correction for multiple comparisons). For orthographic control words, the Kolmogorov-Smirnov non-parametric distribution test revealed a reliable difference in the distribution of whole-word frequency when the data set was split at the divergence point estimate (541 ms) of lemma transition probability ($D = 0.11$, $p = 0.004$). The correlation of these two variables ($r = 0.153$) suggests that greater probability of encountering the form control affix (e.g., -*el*) given its bogus stem (e.g., *broth*) is associated with a higher frequency of the whole word (e.g., *brothel*). The difference in distributions of whole-word frequency indeed indicates that at time-point t_2 (541 ms) onwards, word frequency of form control words may indeed be driving the effect of lemma transition probability. We acknowledge that our conclusions regarding this pair of predictors in English form control words may be biased due to collinearity.

Aside from the single potential confounding variable reported above, we conclude that a confounding influence of a "third variable" is unlikely to explain the relative order of estimated divergence points. Possibly, this is because most correlations were in the weak range ($|r| < 0.4$), the range that is also considered harmless for regression estimates (Belsley, Kuh, & Welsch, 2005). We note, however, that the non-parametric survival analysis that we employ here is not designed to offer a principled statistical control over confounding factors, and this improvement – likely requiring parametric solutions – is a desideratum for future research.

4.3.2 Study 2: Dutch derived words

This study followed the same procedure as in Study 1, yet this time examined derived words in Dutch. We selected 74 derived words that were inflected with one of the following suffixes: -*aal*, -*aar*, -*aard*, -*ant*, -*baar*, -*dom*, -*en*, -*er*, -*heid*, -*ig*, -*ing*, -*lijk*, -*loos*, -*nis*, -*s*, -*schap*, -*sel*, -*st*, -*te*, -*vol*, -*vrij*, -*zaam*. As in the English dataset, the Dutch derived words were relatively transparent in that their stems were free-standing lexemes and were semantically related to the meaning of the whole word (e.g., *blijheid*, meaning 'happiness', is composed of the stem *blij* 'happy' and the suffix *heid*). The full set of Dutch derived words, along with orthographic, morphological and semantic properties of whole-words and stems, is reported in the online supplementary materials S2. A total of 2,794 response times were available in the DLP for the set of 74 stimuli. Table 4.1 reports descriptive statistics of all predictors for selected words as well as response times to the words. Table A.2 of Appendix A reports pairwise Pearson correlations between predictor variables.

As can be seen in timeline A2 of Figure 4.2, the results of the survival analysis revealed that surface frequency had the earliest median divergence point estimate (448 ms), which is then followed at 458 ms and 464 ms by divergence point estimates for orthographic neighbourhood density and stem frequency respectively. Convergent with the results of English derived words, this sequence of divergence point estimates suggests that access to the full-form precedes morphological stem look-up. Closely following access to the stem is the onset of a succession of three semantic effects, which begins with semantic transparency (472 ms), is followed by whole-word valence (487 ms) and ends with the valence of the stem (504 ms). The earliest morphological effect, derivational entropy (491 ms), overlaps with the time window occupied by these three semantic variables. The onset of derivational entropy occurs after both whole-word and stem valence have been accessed. Following derivational entropy is the remainder of morphological (suffix productivity and lemma transition probability), and orthographic (word length) effects that passed the selection criterion. These morpho-orthographic effects occurred within a broad temporal range of between 521 ms and 628 ms.

Much in common with the results of English derived words in Study 1, timeline A2 in Figure 4.2 indicates surface and stem frequency influenced response times at earlier time-points than most other lexical variables. Unlike the English dataset, orthographic neighbourhood density shows an influence, and does so early, after access to the full form of the word, as indexed by surface frequency. Nevertheless, in common with English derived words, access to the suffix (suffix productivity and lemma transition probability) or the family size of the morphological family (as indexed by derivational entropy) takes place after access to the meaning of the whole words and the stem respectively. Interestingly, the sequence of the block of semantic effects and the block of morphological effects were not temporally discrete, such that derivational entropy shows an onset that occurs at a time-point that is approximate to the onset of the effect of stem valence and LSA distance. This finding may suggest that although morphological decomposition does not precede semantics temporally, the processes of decomposition and semantic access may be initiated concurrently.

In sum, these findings demonstrate, in conformity with the results from the English language dataset, that morphological parsing takes place only once access to the whole word form has been activated. Indeed, it is logical to assume that the orthographic identification of a word string (either the surface or the stem) is a necessary initial step toward accessing a word's meaning. Crucially, the results so far suggest (for Dutch and English derived words) that access to a complex word's meaning does not wait until after obligatory decomposition of the word string has taken place. Instead, morphological decomposition of the complex word appears to occur, at its very earliest, simultaneously with access to the whole word meaning.

4.3.3 Study 3: Pseudo-derived words

The study to which we now turn focuses on semantically opaque (pseudo-)derived words, e.g., words containing an existing suffix and a semantically unrelated stem (*broth* and *-er* in *brother*). Identifying divergence point estimates for this variable was motivated by the presence of the same condition in masked priming studies (see e.g., Rastle, Davis, & New, 2004). The original reasoning behind the inclusion of this condition in these experiments was to test the hypothesis that morphological decomposition operates independently of semantic information. Thus, according to Rastle et al. (2004), "Any stimulus bearing a morphological surface structure ... would be decomposed, irrespective of its semantic transparency or etymological characterization" (pp. 1091).

We proceeded to test this hypothesis by identifying 102 pseudo-derived English words that had available behavioural latencies in the BLP. We selected words which had a range of sublexical and lexical predictors for the stem, suffix and the entire word. These lexical characteristics were the same as were collected for the English and Dutch derived words. The full list of English stimuli, along with the properties of whole-words and stems, is reported in the online supplementary materials S3. We did not perform a comparable analysis in Dutch because the pool of pseudo-derived words in Dutch data was too small for consideration. After trimming described above, these words yielded a pool of 4,976 response times from the BLP. Table 4.1 reports descriptive statistics of all predictors for selected words as well as response times to the words. Table A.3 reports pairwise Pearson correlations between predictor variables.

Firstly, as is revealed in Figure 4.2 (timeline B), the range of median divergence points is larger than in all other conditions. In addition, the onset of the earliest variable to influence response times takes place before that of both of the true derived word conditions. Once again, surface frequency emerges as the earliest variable to exert an influence, at 399 ms. The very last variable to demonstrate a divergence in lexical decision times is the frequency of the stem, some ~200 ms later, at 596 ms. The late influence of stem frequency may be indicative of the reduced role of the meaning of the stem for the process of decomposition. This finding will be discussed further in the General discussion.

Following the onset of surface frequency, the next group of variables to demonstrate an impact on lexical decision pertain to either the orthographic or morphological characteristics of the word: form suffix productivity (435 ms), the frequency of the suffix's status as a true morphological string (suffix productivity, 436 ms) and the probability of encountering the first letter of the affix given the last letter of the root (bigram transition probability, 439 ms). Crucially, the temporally compact onsets of these variables coincide with the emergence of the effect of whole-word valence (444 ms) and semantic similarity (453 ms). The simultaneous accompaniment of morphological and semantic effects is further visualized in the vincentile plot

(Ratcliff, 1979; Vincent, 1912) of Figure B.3 in Appendix B (for full details on plotting technique, please consult explanation in Appendix B). Unlike Figure 4.2, which plots the median divergence point estimate for each variable, Figure B.3 in Appendix B depicts, for each variable, the full range of divergence point estimates within the 95% confidence interval. The plot shows that, even at the earliest possible divergence, morpho-orthographic and semantic effects show considerable temporal overlap. Moreover, this overlap remains stable across the full range of divergence points. To sum up, the virtually contemporaneous onsets of lexical characteristics linked to morphological parsing, and those related to the semantics of the whole word, suggests that morphological decomposition of pseudo-derived words is not indifferent to the role of semantics.

4.3.4 Study 4: Form control words

To recap, previous masked priming studies have detected no priming differences between true derived and pseudo-derived (opaque) words, while the magnitude of priming in these conditions was greater than for form control words (Rastle & Davis, 2008). Although this study does not follow the same design and methodology as a masked priming study, we opted to include a baseline condition with which to compare divergence point estimates with those of true and pseudo-derived words. To this end, this study considered a set of 158 words that are not morphologically complex (i.e., do not contain an affix), but embed another simplex word. For example, the simplex word *chapel* contains the substring *chap*. Two constraints were applied during the selection of these words. The first was that the whole word string was, at maximum three letters longer than the embedded word. The second was that those final letters (e.g., -*el* in *chapel*) were not true English suffixes. There were also two variables from the previous three studies that could not be estimated for the current set of words, by virtue of these words not being morphologically complex. These variables were derivational entropy and suffix productivity.

A total of 5,350 response times were available in the BLP for the selected stimuli. Table 4.1 reports descriptive statistics of all predictors for selected words as well as response times to the words. Table A.4 of Appendix A reports pairwise Pearson correlations between predictor variables. The full list of stimuli, along with the properties of whole-words and stems, is reported in the online supplementary materials S4. We did not perform an analysis in Dutch because the pool of simplex form control words in Dutch data was too small for consideration.

Comparable with the pseudo-derived word condition, Figure 4.2 (timeline C) reveals that the first divergence point is earlier than for both true derived word conditions. Moreover, as with all conditions, surface frequency is the first variable to exhibit an influence over lexical decision latencies, with a median divergence point estimate of 395 ms. Surface frequency is then followed by the valence of the whole word at 409 ms, and then LSA similarity at 426 ms. As the vincentile plot in Figure B.4 of Appendix B shows, the valence of the whole word and word length share often overlapping onsets throughout the full distribution of divergence point estimates. That is, word length has a median divergence point estimate that is identical (426 ms) to the divergence point estimate of whole word valence. However, Figure B.4 of Appendix B also shows that the earliest divergence point estimate for word length (399 ms) precedes that of word valence (411 ms). Following this pairing of variables, are the median divergence point estimates of two orthographic variables: orthographic density (445 ms) and form suffix productivity (458 ms). Finally, the onsets of the effect of stem valence and lemma transition probability (defined as the ratio of the word's surface frequency to the frequency of its embedded string) arrive last, with median divergence points of 527 ms and 541 ms respectively.

Overall, it is apparent in the summary plot in Figure 4.2 (timeline C) that only one variable associated with morphological parsing passed the criteria that we set for the survival analysis. This result is expected as all but one of the morphological variables was not available for this word group from the outset. When the effect of lemma transition probability (the only 'morphological variable') does arrive, it does so late. This is also not surprising; because the substring is not morphologically related to the whole word, the computation of the frequency of the substring relative to the frequency of the whole word is not a high priority for the language system. It appears that in this condition too, semantic access appears relatively early in the time-course, and again suggests that once frequency of the whole form has been activated, language processing immediately becomes sensitive to what the letter string may signify. This is shown by the early onset of the effect of the valence of the word, which is then followed by LSA distance. This suggests that even for simplex words, the meaning of an embedded substring is evaluated during the recognition of the whole word. This finding converges with the semantic effects of word strings embedded within larger simplex words reported by Bowers, Davis, and Hanley (2005). Nevertheless, it is intriguing that, in convergence with the pseudo-derived condition, post-onset of the surface frequency effect, access to the semantics of the word is not mediated by the frequency of the stem. Whereas Study 3 eventually suggested a late effect of stem frequency, the influence of this variable did not pass the criteria that we set for the survival analysis here. This absence may be indicative of the difference in decomposition processes across conditions.

To sum up, the key results of this study is that surface frequency demonstrates the earliest detectable effect in the survival analysis, which is a finding that is constant across all studies in the current paper. Moreover, most semantic variables show early divergence points that are either simultaneous with or precede those of orthographic variables. Finally, lemma transition probability shows the latest influence on lexical decision latencies out of all lexical variables.

4.4 General discussion

Several studies of derived word recognition using MEG or EEG report evidence support the early decomposition plus late recombination scenario (e.g., Lavric et al., 2011; 2012; Fruchter & Marantz, 2015; Solomyak & Marantz, 2010). In these studies, late effects of semantic variables and surface frequency, relative to effects driven by morphological variables, are taken as a demonstration of the form-then-meaning account of complex word recognition (e.g., Taft, 2004). The neurophysiological evidence purportedly elaborates on the findings from the masked-priming literature (e.g., Rastle, Davis, & New, 2004; Marslen-Wilson, Bozic, & Randall, 2008; Beyersmann, Ziegler, Castles, Colheart, Kezilas, & Grainger, 2015), which is also taken to suggest that early processing of morphological information does not require semantics as a supplementary information source. The aim of this study was to harness a novel distributional approach to the analysis of reaction time data in order to provide further scrutiny of the account that these results support.

A semantics-blind account of obligatory decomposition predicts that the recognition of any word ending in a decomposable suffix follows the same staged sequence of cognitive processes, and does so whether that word is truly derived (e.g., *dreamer*), or is opaque (e.g., *corner*). Empirical support for this sequence requires that specific lexical variables show influence at certain stages of the timeline of complex word recognition. The diagnostic variables, and the stages at which their influence is predicted to arrive during the time-course of word recognition are outlined by Solomyak and Marantz (2010; see quote 1. in the Introduction) and Fruchter and Marantz (2015). Under the form-then-meaning account, affix-specific variables are predicted to emerge first in the time-course, such as affix frequency. This stage is then succeeded by variables related to parsing of the stem and affix, such as transition probability between the stem and affix. Thus, the initial stages of word recognition are expected to be dominated by variables indicating access to orthographic form and morphological structure. The semantics-blind morpho-orthographic segmentation account predicts that the next stage is devoted lexical access of the stem. The initiation of this process is assumed to be associated with the onset of the effect of lemma frequency, stem frequency or derivational family entropy (Fruchter & Marantz, 2015). Following morphological parsing and stem look-up, recombination of the stem and affix begins to take place. According to Solomyak and Marantz (2010), this stage may be indexed by the emergence of an additional effect of transition probability between stem and affix, LSA distance (as an indicator of the system computing the semantic fit between the stem and the derived word) and ultimately surface frequency. Crucially, this final recombination stage is where the semantics of the word is accessed and word recognition has been achieved.

The above pieces of evidence that are required to substantiate the semantics-blind obligatory decomposition hypothesis found no support in our data. As revealed by the current survival analysis, the temporal emergence of semantic effects precede those of morphological parsing during the processing of English and Dutch derived words. In Study 1 and Study 2, early divergence point estimates were obtained for relational semantic properties (i.e., LSA distance; indicating access to meanings of the stem and the whole word) and atomic semantic properties (i.e., whole-word and stem valence; indicating access to meanings of individual lexemes) for lexical decision latencies for English and Dutch derived words. Contrary to the theoretical assumptions underlying the semantics-blind morphological decomposition account, the onsets of these divergences appeared early relative to those estimated from a complex word's quantified morphological characteristics (e.g., derivational family entropy and suffix productivity). Indeed, in Study 3, processing of English pseudo-derived words also revealed onsets of semantic effects that were contemporaneous with onsets of lexical characteristics linked to morphological parsing and the statistical properties of letter strings. Additionally and critically, across all studies, surface frequency emerged as the earliest source of divergence in reaction time survival rates. We take these findings to challenge the highly influential claim that during visual recognition, a complex word is, without exception, first decomposed into its constituents without influence from semantics, followed later in time by the process of recombination (see, e.g., Fruchter & Marantz; Taft & Forster, 1976; Taft, 2004; Rastle & Davis, 2008).

To dissect the present results in further detail, the survival analyses reported here contribute two main pieces of evidence that are problematic for the form-then-meaning account. The first piece of evidence, and perhaps the most controversial, is the late appearance of morphological variables, relative to semantic variables (see the divergence point time-course A1, A2 and B in Figure 4.2). Critically, for derived words, we found no intervening divergence point estimates pertaining to morphological statistical properties between divergence estimates of frequency measures and those of variables pertaining to the semantics of the whole word (i.e., whole word valence and LSA distance). For derived word processing, derivational family entropy began to diverge at time-points subsequent to, but near-contemporaneous with, the onset of semantic access to the whole word (and to the emotional connotation of the stem in English). This finding suggests that at the stage subsequent to accessing form, the lexical properties relevant to processing may also be semantic and not just morphological in nature. That is, the independent emotional connotations of the word and the stem, the semantic relationship between the stem and the whole word, and information about the family of words sharing the stem of the derived word, are all accessed concurrently and not at different stages.

The second piece of evidence that poses a challenge to the form-then-meaning account is the early divergence point estimate for surface frequency, which emerges as the earliest divergence point across derived, pseudo-derived and form control words. This is incompatible with the hypothesized role of surface frequency as a marker of the final recombination stage within the decomposition account; the point at which the suffix and the root of a complex word have been successfully combined in the mental lexicon and access to meaning has been granted. According to a semantics-blind account of obligatory decomposition, the initial lexical search stage involves access to the stem of the complex word (e.g., *dream*), and a later stage involves access to the whole-word unit (e.g., *dreamer*). Thus, within such accounts, surface frequency and constituent frequency are two lexical distributional measures that play pivotal diagnostic roles (for more discussion, see Baayen, 2014). Constituent frequency effects are attributed to early stages of morphological decomposition, whereas surface frequency effects are attributed to subsequent recombination and semantic access (Solomyak & Marantz, 2010; Taft, 2004). In the current study, divergence point estimates of surface frequency arrive at the earliest time out of all variables, and naturally, precede divergence point estimates of stem frequency in Dutch derived words, and in English derived and pseudo-derived words (stem frequency does not influence reaction times in the form control condition). We argue that the early response to surface frequency, across all word types, reflects the rapid and automatic response to the reinforcement of the associations between letter combinations within the whole word unit and the meanings that those letter combinations may denote (e.g., Rayner et al., 2003; Rayner, et al., 2004). Furthermore, for derived and pseudo-derived words, responding to the surface form of the word first indicates that the whole word form, and not its stem, is the initial unit of analysis that provides cues to the meaning of the complex word.

Another crucial finding is that the overall pattern of results for pseudo-derived words are largely convergent with the results of the true derived word conditions. For pseudo-derived words, the results indicate a rapid response to the form of the whole word (surface frequency), which is then followed by concurrent access to semantics (whole-word valence and LSA distance) and to morpho-orthographic variables (form suffix productivity and suffix productivity). Interestingly, unlike true derived words, the results suggest that frequency of the stem is not registered until very late and access to the family of the stem (derivational family entropy) is not attempted (see

Figure 4.2). It is possible that these effects are evidence of a weak influence of the stem meaning during the processing of the whole opaque word (i.e., accessing the meaning of *trump* in *trumpet*). This speculation is in line with the findings that show that semantic opacity modulates complex word processing (see e.g., Juhasz, 2007; Feldman et al., 2009; Marelli & Luzzatti, 2012).

There is a growing consensus within the field that semantic transparency is a gradient phenomenon (e.g., Marelli & Luzzatti, 2012; Pastizzo & Feldman, 2003). One criticism of previous masked priming studies is how meanings of complex words and pseudo-complex words are discretized in order to form experimental group comparisons. For example, Baayen et al. (2011, pp. 465-466) questioned the stimuli used in the pseudo-derived word condition of Rastle et al.'s (2004) masked-priming study. They argued that the pseudo-complex words that were used in this condition varied in the extent to which the suffix transparently contributes to the meaning of the complex word. For example the etymological origin of "archer" (someone who wields a bow) is the Latin form *arcus* ('bow'). Although the meaning of the pseudo-stem is not "synchronically visible", the whole complex word is still structurally similar to "trucker" (an individual who drives trucks). Beyersmann et al. recently responded to these criticisms and conducted a masked priming study in which the pseudo-complex words did not contain pseudo suffixes that actually carry their true meaning/function in the pseudo-suffixed words. They reported similar priming magnitudes across derived and pseudo-derived words, and thus provided further support for the semantics blind decomposition account. Following Beyersmann et al.'s (2015) criticism, we also imposed a more stringent set of criteria for our selection of pseudo-complex stimuli. Before performing the survival analysis, we verified that the words in this condition were indeed fully opaque, in that we excluded all "archer" examples. Furthermore, to confirm that we had indeed selected pseudo-complex words, we compared semantic transparency across each condition and found that there is no concern that pseudo-complex words carry a transparent meaning (please see Appendix C, where this analysis is reported). Nevertheless, even though we observe differences between conditions in their semantic transparency estimates, we observe that semantic transparency, as gauged by LSA, exerts an influence on lexical decision survival rates across all word conditions. This indicates that the internal distribution of semantic transparency of each condition is enough to influence word processing (as determined by the distributional analysis).

An additional contribution of this paper is the addition of valence to the palette of semantic variables with which one can explore the time-course of semantic access during complex word recognition. Our findings corroborate with Kuperman's (2013) finding that the emotional connotation of a compound word and those of its constituents play a role in recognizing the whole word. As well as demonstrating that semantic access to the atomic properties of compound word extends to derived words, these findings also supplement current theoretical treatments of the effects of emotion on word recognition. Irrespective of morphological complexity, a number of experiments have shown that negative words, such as *vomit*, elicit slower responses as compared to neutral stimuli, such as *nunnery* (e.g., Citron, 2012; Wentura, Rothermund, & Bak, 2000; Kuperman, Estes, Brysbaert, & Warriner, 2014). These results are typically attributed to the human organism's *automatic vigilance* to negative stimuli (Erdelyi, 1974). This hypothesis proposes that negative words capture attention for longer and thus generate slower lexical decision responses. The relatively early time at which a divergence between positive and negative stimuli exerts an influence on lexical decision times demonstrates the rapidity of the automatic response to the emotional content of linguistically encoded stimuli. That these effects are early for both simplex and complex words, and across Dutch and English, demonstrates the privileged role of valence during word recognition.

The success of detecting the early semantic effects of valence, and LSA similarity, is also due to the novel application of survival analysis to this line of research. The main advantage of non-parametric survival technique is that it provides insight into the sequencing of cognitive processes involved in complex word recognition. The distributional survival technique adopted here provides information about the time-course of cognitive processes during complex word recognition which is independent of the strength of the relationship between the lexical variable and the dependent variable. In order to test whether the strength of an effect is associated with its divergence point estimate, we correlated the strength of the relationship between each lexical variable and lexical decision RTs. If the relationship between the effect strength of a given lexical variable (estimated by a Spearman correlation coefficient) and the variable's divergence point time is strong, then the temporal order may be a replica of the strength of a particular effect on RTs. We found no relationship between effect strength and divergence point estimates across all data sets. This finding suggests that the ordering of divergence points estimated by the survival technique are not just estimations of the strength of the variable as a predictor of RTs.

Moreover, another advantage of this statistical technique is that it is able to elucidate stages of processing from a lexical decision response, rather than generate a finding based on the final outcome (i.e., the latency of the lexical decision response). In all our analyses, all divergence points were detected at time-points when less than 50% of responses were terminated, and in a vast majority of cases were shorter than the mean response time in the dataset. Thus, the non-parametric survival technique adopted here is not constrained to using a measure of central tendency, such as the mean, as a basis for discovering behavioural patterns in the data. For example, an examination of mean reaction time across English derived, Dutch derived, English pseudo-derived and English form derived words (606 ms, 578 ms, 589 ms, and 600 ms respectively) is not informative about the timing of cognitive processes, whereas the distributional approach adopted here is able to offer an alternative perspective of a chronometric data source, and is one that is well-suited to the study of the sequencing of cognitive processes during word recognition.

However, there are limitations to this distributional approach. Firstly, although one is able to infer the onset of an effect on word recognition, this analysis is currently unable to evaluate the relative importance of a given variable on the survival rate (whether a lexical decision has been made or not) throughout the entire time-course. For example, whereas surface frequency might influence the very earliest responses within a distribution of RTs, we do not know whether surface frequency retains influence for the slowest responses. Secondly, we are also unable to control for idiosyncratic variation that is due to individual differences in participants, or for the items themselves. Lastly, in the current survival implementation, each lexical variable is considered independently of one another. Although we addressed the issue of collinearity (see Study 1), there is also a need for future parametric solutions to take into account the combined effect of lexical variables (and also variables pertaining to individual differences in reading skill) on the word recognition time-course. Ultimately, more complex solutions to modelling survival rates of lexical decision responses will be able remedy many of these issues. We also argue that more such distributional analyses may provide a promising tool for comparing behavioural data with the results from data derived from the continuous electrical or magnetic signal associated with the behavioural decision (i.e., MEG and ERP).

4.4.1 Conclusion

Our findings provide evidence that morphological decomposition during complex word recognition does not precede access to the semantics of the complex word. Our finding that access to meaning does not follow an obligatory stage of morphological decomposition is in line with other behavioural studies that also challenge the early decomposition (see, e.g., Juhasz, 2007; Feldman et al., 2009; Juhasz & Berkowitz, 2011; Marelli & Luzzatti, 2012), as well as several EEG studies (Assadollahi & Pulvermüller, 2003; Hauk et al., 2006; Sereno et al., 1998; Pulvermüller, 2002). Indeed, for derived and pseudo-derived words, we find evidence that decomposition is attempted, but this occurs at least as early as access to the semantics of the whole complex word. We suggest that the early detection of surface frequency followed by stem frequency (and OLD20 for Dutch derived words) reflects the detection of orthographic cues and their relative activation strength. The subsequent effects of whole-word valence and LSA similarity reflects access to the semantics of the derived word unmediated by attempts at morphological parsing. It is only at the point at which semantics is accessed that there is evidence of access to the morphological family of the stem, which is followed later by access to the statistical properties of the suffix. In sum, our results provide cross-linguistic support for a form-and-meaning account of complex word recognition.

References

Adelman, J. S., & Estes, Z. (2013). Emotion and memory: A recognition advantage for positive and negative words independent of arousal. *Cognition*, *129*(3), 530–535.

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*(5), 1587–1596.

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

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

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*(3), 438–481.

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

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459.

Belsley, D. A., Kuh, E., & Welsch, R. E. (2005). *Regression diagnostics: Identifying influential data and sources of collinearity.* (Vol. 571). John Wiley & Sons.

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

Beyersmann, E., Ziegler, J. C., Castles, A., Coltheart, M., Kezilas, Y., & Grainger, J. (n.d.). Morpho-orthographic segmentation without semantics. *Psychonomic Bulletin & Review*, *23*(2), 533–539.

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*(1), 131–143.

Citron, F. M. (2012). Neural correlates of written emotion word processing: a review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain & Language*, *122*(3), 211–226.

Crepaldi, D., Rastle, K., Coltheart, M., & Nickels, L. (2010). 'fell' primes 'fall', but does 'bell' prime 'ball'? Masked priming with irregularly-inflected primes. *Journal of Memory and Language*, *63*(1), 83–99.

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

Davis, M. H., & Rastle, K. (2010). Form and meaning in early morphological processing: Comment on Feldman, O'Connor, and Moscoso del Prado Martín (2009). *Psychonomic Bulletin & Review*, *17*(5), 749–755.

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

Diependaele, K., Duñabeitia, J. A., Morris, J., & Keuleers, E. (2011). Fast morphological effects in first and second language word recognition. *Journal of Memory and Language*, *64*(4), 344–358.

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). Cambridge University Press.

Diependaele, K., Sandra, D., & Grainger, J. (2005). Masked cross-modal morphological priming: Unravelling morpho-orthographic and morpho-semantic influences in early word recognition. *Language and Cognitive Processes*, *20*(1-2), 75–114.

Drieghe, D., Rayner, K., & Pollatsek, A. (2008). Mislocated fixations can account for parafoveal-on-foveal effects in eye movements during reading. *The Quarterly Journal of Experimental Psychology*, *61*(8), 1239–1249.

Duñabeitia, J. A., Kinoshita, S., Carreiras, M., & Norris, D. (2011). Is morpho-orthographic decomposition purely orthographic? Evidence from masked priming in the same–different task. *Language and Cognitive Processes*, *26*(4-6), 509–529.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

Erdelyi, M. H. (1974). A new look at the new look: perceptual defense and vigilance. *Psychological review*, *81*(1), 1–25.

Feldman, L. B., Milin, P., Cho, K. W., del Prado Martín, F. M., & O'Connor, P. A. (2015). Must analysis of meaning follow analysis of form? A time course analysis. *Frontiers in Human Neuroscience*, *9*. Retrieved from `http://dx.doi.org/10.3389/fnhum.2015.00111`

Feldman, L. B., O'Connor, P. A., & Moscoso del Prado Martín, F. (2009). Early morphological processing is morphosemantic and not simply morpho-orthographic: A violation of form-then-meaning accounts of word recognition. *Psychonomic Bulletin & Review*, *16*(4), 684–691.

Feldman, L. B., & Pastizzo, M. J. (2003). Morphological facilitation: The role of semantic transparency and family size. *Trends in Linguistics Studies and Monographs*, *151*, 233–258.

Fiorentino, R., & Poeppel, D. (2007). Compound words and structure in the lexicon. *Language and Cognitive processes*, *22*(7), 953–1000.

Forster, K. I., Mohan, K., & Hector, J. (2003). The mechanics of masked priming. In S. Kinoshita & S. Lupker (Eds.), *Masked priming: State of the art* (pp. 3–37). New York: Psychology Press.

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.

Hauk, O., Davis, M. H., Ford, M., Pulvermüller, F., & Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage*, *30*(4), 1383–1400.

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: Elsevier.

Juhasz, B. J., & Berkowitz, R. N. (2011). Effects of morphological families on English compound word recognition: A multitask investigation. *Language and Cognitive Processes*, *26*(4-6), 653–682.

Kazanina, N. (2011). Decomposition of prefixed words in Russian. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1371–1390.

Keuleers, E. (2013). vwr: Useful functions for visual word recognition research. *R package version 0.3.0.*. Retrieved from `http://CRAN.Rproject.org/package=vwr`

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*(3), 643–650.

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*(1), 287–304.

Kim, S. Y., Wang, M., & Taft, M. (2015). Morphological decomposition in the recognition of prefixed and suffixed words: Evidence from Korean. *Scientific Studies of Reading*, *19*(3), 183–203.

Kliegl, R., Dambacher, M., Dimigen, O., Jacobs, A. M., & Sommer, W. (2012). Eye movements and brain electric potentials during reading. *Psychological Research*, *76*(2), 145–158.

Kuperman, V. (2013). Accentuate the positive: Semantic access in English compounds. *Frontiers in Psychology*, *4*(203). Retrieved from `http://dx.doi.org/10.3389/fpsyg.2013.00203`

Kuperman, V., Bertram, R., & Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes*, *23*(7-8), 1089–1132.

Kuperman, V., Bertram, R., & Baayen, R. H. (2010). Processing trade-offs in the reading of Dutch derived words. *Journal of Memory and Language*, *62*(2), 83–97.

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: valence and arousal affect word recognition. *Journal of Experimental Psychology: General*, *143*(3), 1065–1081.

Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading polymorphemic Dutch compounds: toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 876–895.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240.

Lavric, A., Clapp, A., & Rastle, K. (2007). ERP evidence of morphological analysis from orthography: A masked priming study. *Journal of Cognitive Neuroscience*, *19*(5), 866–877.

Lavric, A., Elchlepp, H., & Rastle, K. (2012). Tracking hierarchical processing in morphological decomposition with brain potentials. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 811–816.

Lavric, A., Rastle, K., & Clapp, A. (2011). What do fully visible primes and brain potentials reveal about morphological decomposition? *Psychophysiology*, *48*(5), 676–686.

Lázaro, M., Illera, V., & Sainz, J. (2016). The suffix priming effect: Further evidence for an early morpho-orthographic segmentation process independent of its semantic content. *The Quarterly Journal of Experimental Psychology*, *69*(1), 197–208.

Libben, G. (2005). Everything is psycholinguistics: Material and methodological considerations in the study of compound processing. *The Canadian Journal of Linguistics/La revue Canadienne de linguistique*, *50*(1), 267–283.

Libben, G. (2014). The nature of compounds: A psychocentric perspective. *Cognitive Neuropsychology*, *31*(1-2), 8–25.

Longtin, C.-M., Segui, J., & Halle, P. A. (2003). Morphological priming without morphological relationship. *Language and Cognitive Processes*, *18*(3), 313–334.

Lund, K., & Burgess, C. (1996). Hyperspace analogue to language (HAL): A general model semantic representation. *Brain and Cognition*, *30*(3), 5–5.

Mandera, P., Keuleers, E., & Brysbaert, M. (in press). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*.

Marelli, M., Amenta, S., Morone, E. A., & Crepaldi, D. (2013). Meaning is in the beholder's eye: Morpho-semantic effects in masked priming. *Psychonomic Bulletin & Review*, *20*(3), 534–541.

Marelli, M., & Luzzatti, C. (2012). Frequency effects in the processing of Italian nominal compounds: Modulation of headedness and semantic transparency. *Journal of Memory and Language*, *66*(4), 644–664.

Marslen-Wilson, W. D., Bozic, M., & Randall, B. (2008). Early decomposition in visual word recognition: Dissociating morphology, form, and meaning. *Language and Cognitive Processes*, *23*(3), 394–421.

Meunier, F., & Longtin, C.-M. (2007). Morphological decomposition and semantic integration in word processing. *Journal of Memory and Language*, *56*(4), 457–471.

Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., van Schie, K., Van Harmelen, A.-L., . . . Brysbaert, M. (2013). Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behavior Research Methods*, *45*(1), 169–177.

Morris, J., Frank, T., Grainger, J., & Holcomb, P. J. (2007). Semantic transparency and masked morphological priming: An ERP investigation. *Psychophysiology*, *44*(4), 506–521.

Morris, J., & Stockall, L. (2012). Early, equivalent ERP masked priming effects for regular and irregular morphology. *Brain & Language*, *123*(2), 81–93.

Moscoso del Prado Martín, F. (2007). Co-occurrence and the effect of inflectional paradigms. *Lingue e linguaggio*, *6*(2), 247–263.

Moscoso del Prado Martín, F., Kostić, A., & Baayen, R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition*, *94*(1), 1–18.

Penolazzi, B., Hauk, O., & Pulvermüller, F. (2007). Early semantic context integration and lexical access as revealed by event-related brain potentials. *Biological psychology*, *74*(3), 374–388.

Plaut, D. C., & Gonnerman, L. M. (2000). Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes*, *15*(4-5), 445–485.

Pollatsek, A., Hyönä, J., & Bertram, R. (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 820–833.

Pulvermüller, F. (2002). *The neuroscience of language: on brain circuits of words and serial order.* Cambridge University Press.

Pylkkänen, L., Llinás, R., & Murphy, G. L. (2006). The representation of polysemy: MEG evidence. *Journal of Cognitive Neuroscience*, *18*(1), 97–109.

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

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

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*(6), 1090–1098.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, *86*(3), 446–461.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372-422.

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: implications for the ez reader model. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(4), 720–732.

Rayner, K., Liversedge, S. P., White, S. J., & Vergilino-Perez, D. (2003). Reading disappearing text cognitive control of eye movements. *Psychological science*, *14*(4), 385–388.

Reingold, E. M., Reichle, E. D., Glaholt, M. G., & Sheridan, H. (2012). Direct lexical control of eye movements in reading: Evidence from a survival analysis of fixation durations. *Cognitive Psychology*, *65*(2), 177–206.

Reingold, E. M., & Sheridan, H. (2014). Estimating the divergence point: a novel distributional analysis procedure for determining the onset of the influence of experimental variables. *Frontiers in Psychology*, *5*. Retrieved from http://dx.doi.org/10.3389/fpsyg.2014.01432

Royle, P., Drury, J. E., Bourguignon, N., & Steinhauer, K. (2012). The temporal dynamics of inflected word recognition: A masked ERP priming study of French verbs. *Neuropsychologia*, *50*(14), 3542–3553.

Schotter, E. R., & Rayner, K. (2012). Eye movements and word recognition during reading. In J. S. Adelman (Ed.), *Visual word recognition: Volume 2. meaning and context, individuals and development* (pp. 73–75).

Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends in cognitive sciences*, *7*(11), 489–493.

Sereno, S. C., Rayner, K., & Posner, M. I. (1998). Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. *Neuroreport*, *9*(10), 2195–2200.

Sheridan, H., Reingold, E., & Matsuki, K. (2016). Rtsurvival: Survival analysis and divergence point estimation methods for reaction time data. *R package version 0.1*.

Simon, D. A., Lewis, G., & Marantz, A. (2012). Disambiguating form and lexical frequency effects in MEG responses using homonyms. *Language and Cognitive Processes*, *27*(2), 275–287.

Solomyak, O., & Marantz, A. (2009). Lexical access in early stages of visual word processing: A single-trial correlational MEG study of heteronym recognition. *Brain & Language*, *108*(3), 191–196.

Solomyak, O., & Marantz, A. (2010). Evidence for early morphological decomposition in visual word recognition. *Journal of Cognitive Neuroscience*, *22*(9), 2042–2057.

Staub, A., & Rayner, K. (2007). Eye movements and on-line comprehension processes. In M. G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 327–342). Oxford, UK: Oxford University Press.

Staub, A., White, S. J., Drieghe, D., Hollway, E. C., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1280–1293.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology Section A*, *57*(4), 745–765.

Taft, M., & Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior*, *15*(6), 607–620.

Van den Bussche, E., Van den Noortgate, W., & Reynvoet, B. (2009). Mechanisms of masked priming: a meta-analysis. *Psychological Bulletin*, *135*(3), 452.

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

Vincent, S. (1912). The function of the vibrissae in the behavior of the white rat. *Animal Behavior Monographs*.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191–1207.

Wentura, D., Rothermund, K., & Bak, P. (2000). Automatic vigilance: the attention-grabbing power of approach-and- avoidance-related social information. *Journal of Personality and Social Psychology*, *78*(6), 1024–1037.

White, S. J. (2008). Eye movement control during reading: Effects of word frequency and orthographic familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(1), 205–223.

Zweig, E., & Pylkkänen, L. (2009). A visual M170 effect of morphological complexity. *Language and Cognitive Processes*, *24*(3), 412–439.

Appendices

Appendix A: Correlation matrices for lexical variables

Table A.1: Correlation matrix of lexical variables in the English derived words condition.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
1. Surface freq.		0.135	0.16	0.142	-0.129	0.073	0.01	0.043	-0.055	-0.041	0.051	-0.03
2. Stem freq.	0.002		0.165	0.249	0.153	0.22	-0.073	-0.003	0.052	0.022	-0.051	-0.171
3. Valence word	0	0		0.589	-0.092	0.137	0.05	-0.058	-0.054	-0.033	-0.027	-0.012
4. Valence stem	0.001	0	0		0.012	0.205	-0.033	0.015	0.045	0.05	0.005	-0.057
5. LSA distance	0.003	0	0.033	0.778		0.069	-0.024	0.058	0.055	-0.011	-0.049	-0.019
6. Deriv. entropy	0.089	0	0.001	0	0.11		-0.203	-0.083	0.208	0.145	-0.042	-0.332
7. SP	0.809	0.09	0.24	0.437	0.581	0		-0.413	-0.446	-0.656	0.122	0.525
8. TPL	0.315	0.942	0.177	0.731	0.177	0.052	0		0.058	0.284	-0.036	0.072
9. Word length	0.197	0.227	0.209	0.291	0.201	0	0	0.177		0.781	-0.03	-0.682
10. OLD20	0.345	0.613	0.441	0.243	0.806	0.001	0	0	0		-0.155	-0.583
11. TPB	0.238	0.233	0.53	0.912	0.25	0.33	0.004	0.401	0.482	0		0.122
12. FSP	0.483	0	0.774	0.182	0.651	0	0	0.093	0	0	0.005	

Table A.2: Correlation matrix of lexical variables in the Dutch derived words condition.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
1. Surface freq.		0.333	0.218	0.197	-0.195	-0.094	-0.015	0.2	-0.058	-0.163	0.12	-0.095
2. Stem freq.	0.004		0.091	0.049	-0.092	0.294	-0.038	0.202	0.059	-0.046	-0.094	0.04
3. Valence word	0.062	0.44		0.793	0.11	-0.054	-0.03	0.076	-0.039	-0.001	-0.111	-0.268
4. Valence stem	0.093	0.678	0		-0.044	0.012	0.027	0.059	-0.001	-0.02	-0.051	-0.263
5. LSA distance	0.095	0.434	0.351	0.713		0.153	-0.197	-0.265	-0.121	-0.023	0.064	0.1
6. Deriv. entropy	0.425	0.011	0.646	0.921	0.194		0.145	0.065	-0.158	-0.172	-0.062	0.207
7. SP	0.898	0.75	0.801	0.819	0.092	0.216		0.003	-0.322	-0.583	0.045	0.403
8. TPL	0.087	0.085	0.519	0.616	0.022	0.585	0.977		0.233	0.052	-0.239	-0.298
9. Word length	0.623	0.619	0.739	0.994	0.305	0.179	0.005	0.046		0.719	-0.22	-0.479
10. OLD20	0.165	0.696	0.995	0.866	0.846	0.144	0	0.659	0		-0.128	-0.483
11. TPB	0.31	0.428	0.345	0.668	0.586	0.597	0.705	0.04	0.059	0.275		0.371
12. FSP	0.419	0.734	0.021	0.023	0.395	0.076	0	0.01	0	0	0.001	

Table A.3: Correlation matrix of lexical variables in the English pseudo-derived words condition.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
1. Surface freq.		0.165	0.314	0.004	-0.068	0.207	0.086	0.131	-0.14	-0.147	-0.092	0.148
2. Stem freq.	0.098		0.2	0.215	-0.328	0.122	0.097	-0.021	0.211	0.137	-0.054	0.045
3. Valence word	0.001	0.044		0.274	0.043	0.164	-0.032	0.083	0.141	0.127	-0.166	-0.088
4. Valence stem	0.969	0.03	0.005		0.062	0.094	-0.054	-0.029	-0.071	-0.004	-0.108	-0.005
5. LSA distance	0.496	0.001	0.666	0.533		-0.261	0.004	-0.037	-0.099	-0.031	-0.068	-0.049
6. Deriv. entropy	0.037	0.22	0.1	0.349	0.008		-0.069	0.011	0.219	0.171	-0.105	-0.058
7. SP	0.391	0.331	0.747	0.592	0.97	0.49		-0.103	-0.172	-0.335	0.033	0.648
8. TPL	0.189	0.838	0.405	0.773	0.709	0.911	0.304		0.01	0.035	0.068	0.138
9. Word length	0.16	0.033	0.157	0.481	0.322	0.027	0.084	0.923		0.804	0.008	-0.502
10. OLD20	0.14	0.17	0.203	0.969	0.761	0.086	0.001	0.73	0		-0.182	-0.472
11. TPB	0.356	0.591	0.095	0.279	0.494	0.292	0.744	0.495	0.933	0.067		0.203
12. FSP	0.138	0.653	0.381	0.962	0.622	0.565	0	0.168	0	0	0.041	

Table A.4: Correlation matrix of lexical variables in the English form control condition.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Surface freq.		-0.039	0.416	0.108	-0.183	0.14	-0.104	-0.105	-0.026	0.033
2. Stem freq.	0.623		0.026	0.173	-0.008	0	-0.034	-0.091	-0.029	-0.028
3. Valence word	0	0.749		0.103	-0.204	0.166	-0.169	-0.12	-0.091	0.134
4. Valence stem	0.175	0.03	0.199		-0.146	0.017	-0.051	0.04	0.162	-0.096
5. LSA distance	0.022	0.918	0.01	0.067		0.049	0.069	0.12	0.007	-0.065
6. TPL	0.079	0.995	0.037	0.837	0.538		0.026	0.1	0.003	-0.107
7. Word length	0.196	0.672	0.034	0.526	0.392	0.745		0.614	-0.028	-0.252
8. OLD20	0.189	0.258	0.134	0.62	0.134	0.211	0		0.001	-0.356
9. TPB	0.744	0.715	0.253	0.042	0.932	0.971	0.726	0.987		-0.094
10. FSP	0.683	0.728	0.093	0.231	0.415	0.181	0.001	0	0.241	

Appendix B: Vincentile plots

The DPA procedure applied to all predictors retrieved distributions of divergence points based on 1,000 iterations. To visualize the full distribution of DPA points within the confidence interval range, we used a vincentile plotting technique (Ratcliff, 1979; Vincent, 1912). All point-wise estimates of the divergence point for a given predictor were ordered from smallest to greatest and broken down into quintiles, i.e., five bins at 20% intervals. For example, the vincentile plot in Figure B.1 depicts the median estimate of the divergence point for each quintile and for each predictor. The dashed line shows divergence point estimates for the third quintile (ranging over 40^{th} - 60^{th} percentile of the distribution and is plotted at the 50^{th} percentile). These estimates are virtually identical to the the divergence point estimates obtained from the entire distribution in the DPA procedure, e.g., the time-point marked by the arrow in the individual DPA plot for derived word frequency (see Figure 4.1).

Colours represent groups of variables: frequencies of whole word and stem (red); as well as orthographic (black), morphological (blue) and semantic (orange) predictors. The vincentile plots here and below only represent variables that elicited a detectable divergence point between survival curves in at least 700 out of 1,000 iterations with a 3% minimal contrast (other effects were deemed too unstable to estimate).

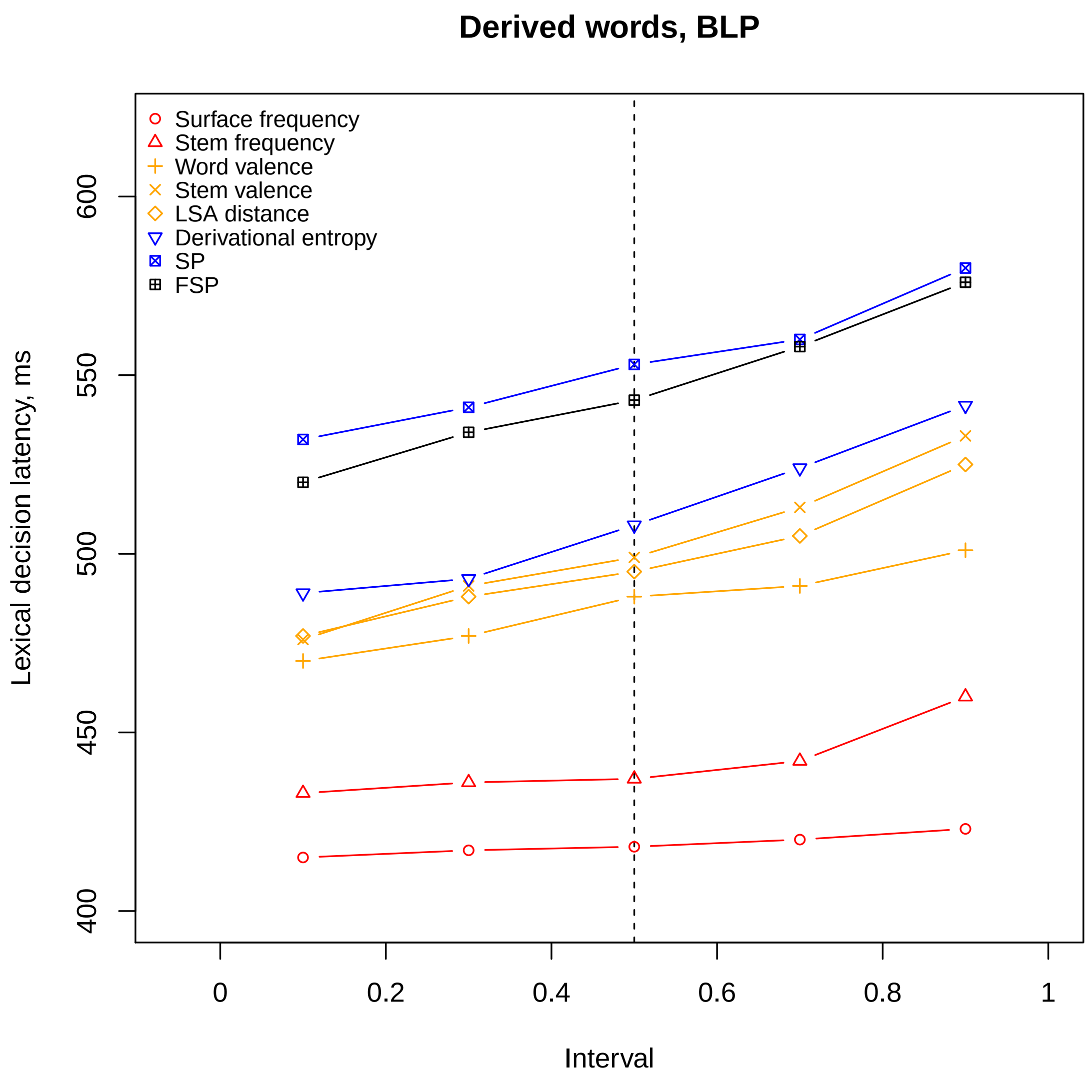

Figure B.1: Vincentile plot of divergence point intervals for English derived words. Colours represent groups of variables: frequencies of whole word and stem (red); as well as orthographic (black), morphological (blue) and semantic (orange) predictors. Only variables that elicited a detectable divergence point between survival curves in at least 700 out of 1,000 iterations with a 3% minimal contrast are plotted (other effects were deemed too unstable to estimate).

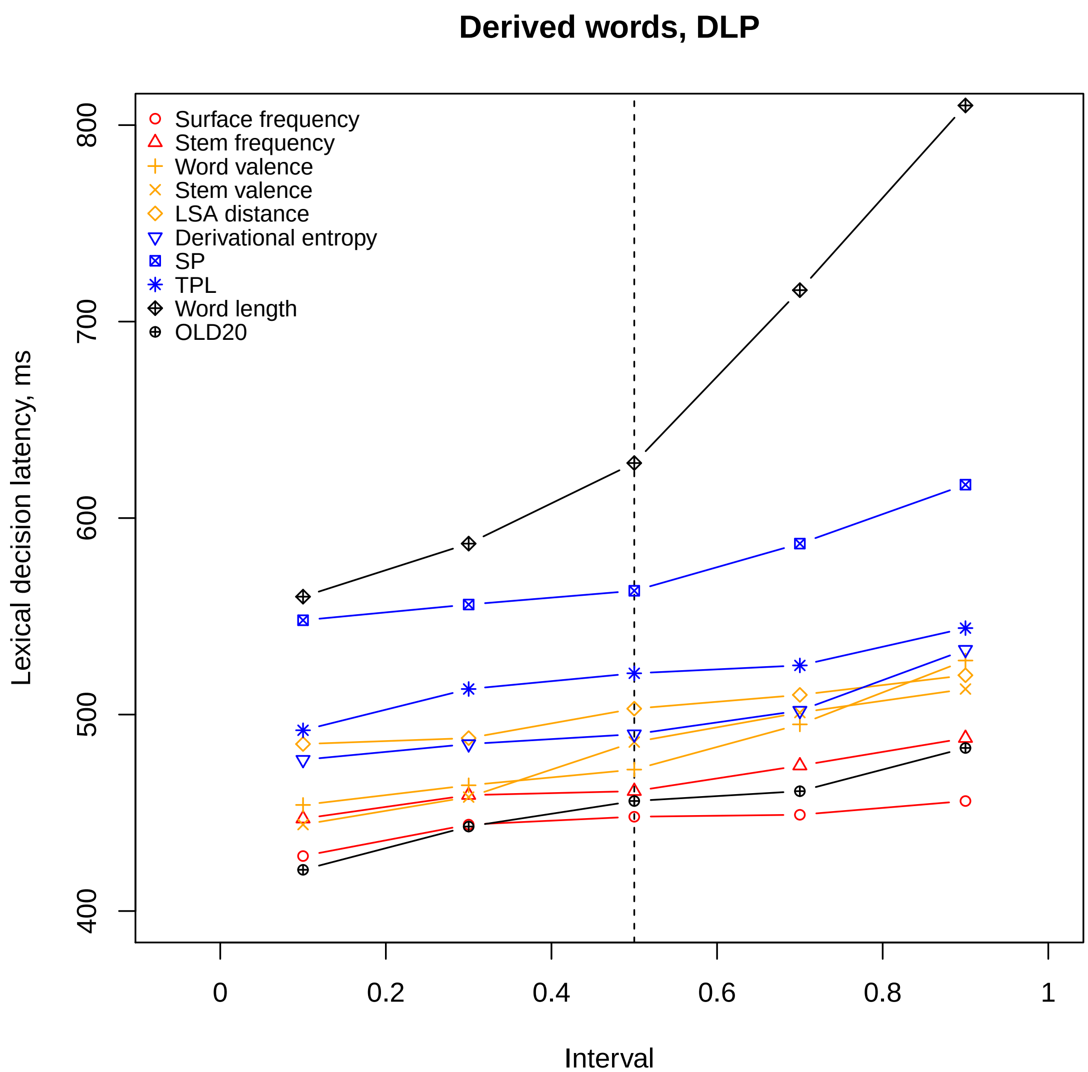

Figure B.2: Vincentile plot of divergence point intervals for Dutch derived words. Colours represent groups of variables: frequencies of whole word and stem (red); as well as orthographic (black), morphological (blue) and semantic (orange) predictors. Only variables that elicited a detectable divergence point between survival curves in at least 700 out of 1,000 bootstrap iterations with a 3% minimal contrast are plotted (other effects were deemed too unstable to estimate).

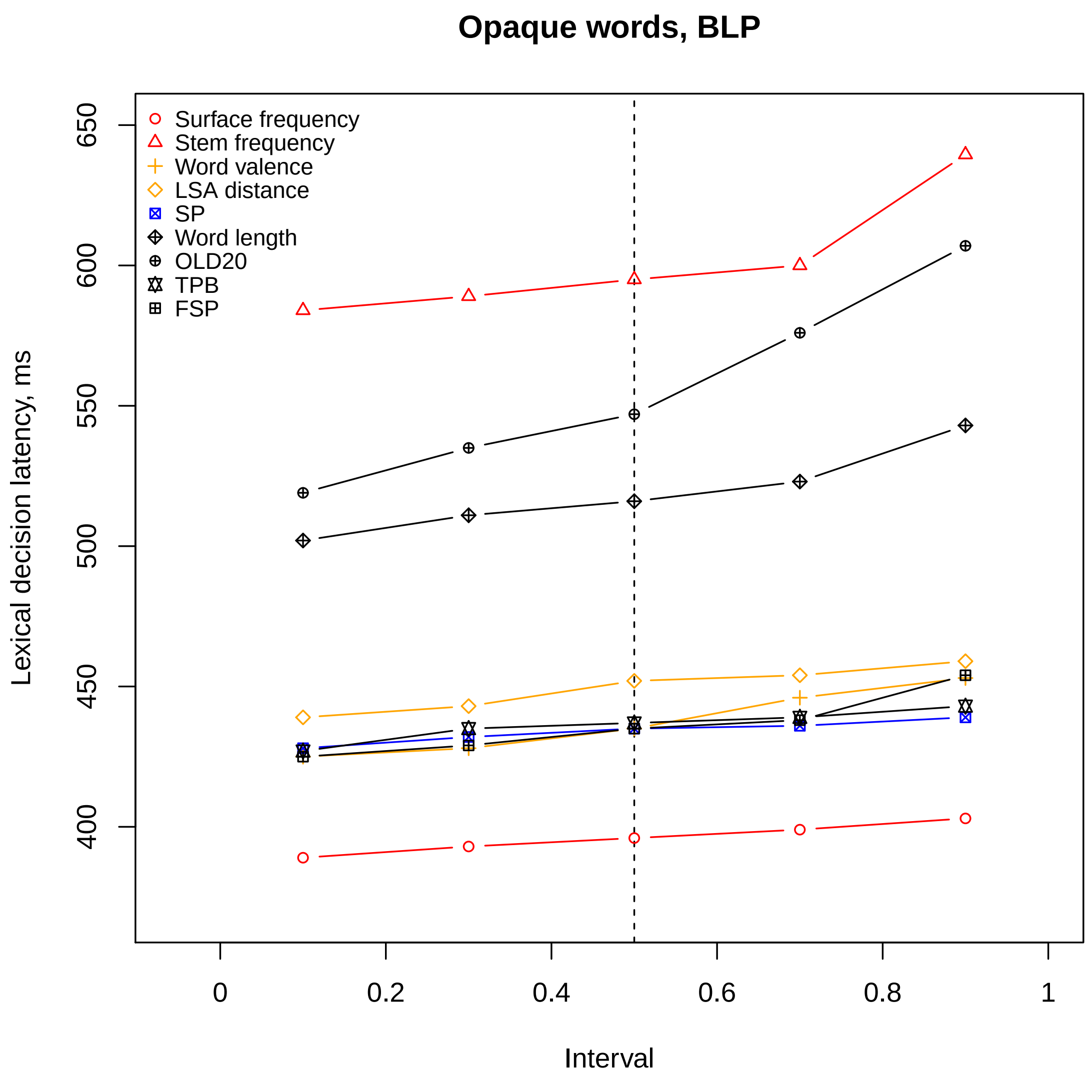

Figure B.3: Vincentile plot of divergence point intervals for pseudo-derived English words. Colours represent groups of variables: frequencies of whole word and stem (red); as well as orthographic (black), morphological (blue) and semantic (orange) predictors. Only variables that elicited a detectable divergence point between survival curves in at least 700 out of 1,000 iterations with a 3% minimal contrast are plotted (other effects were deemed too unstable to estimate).

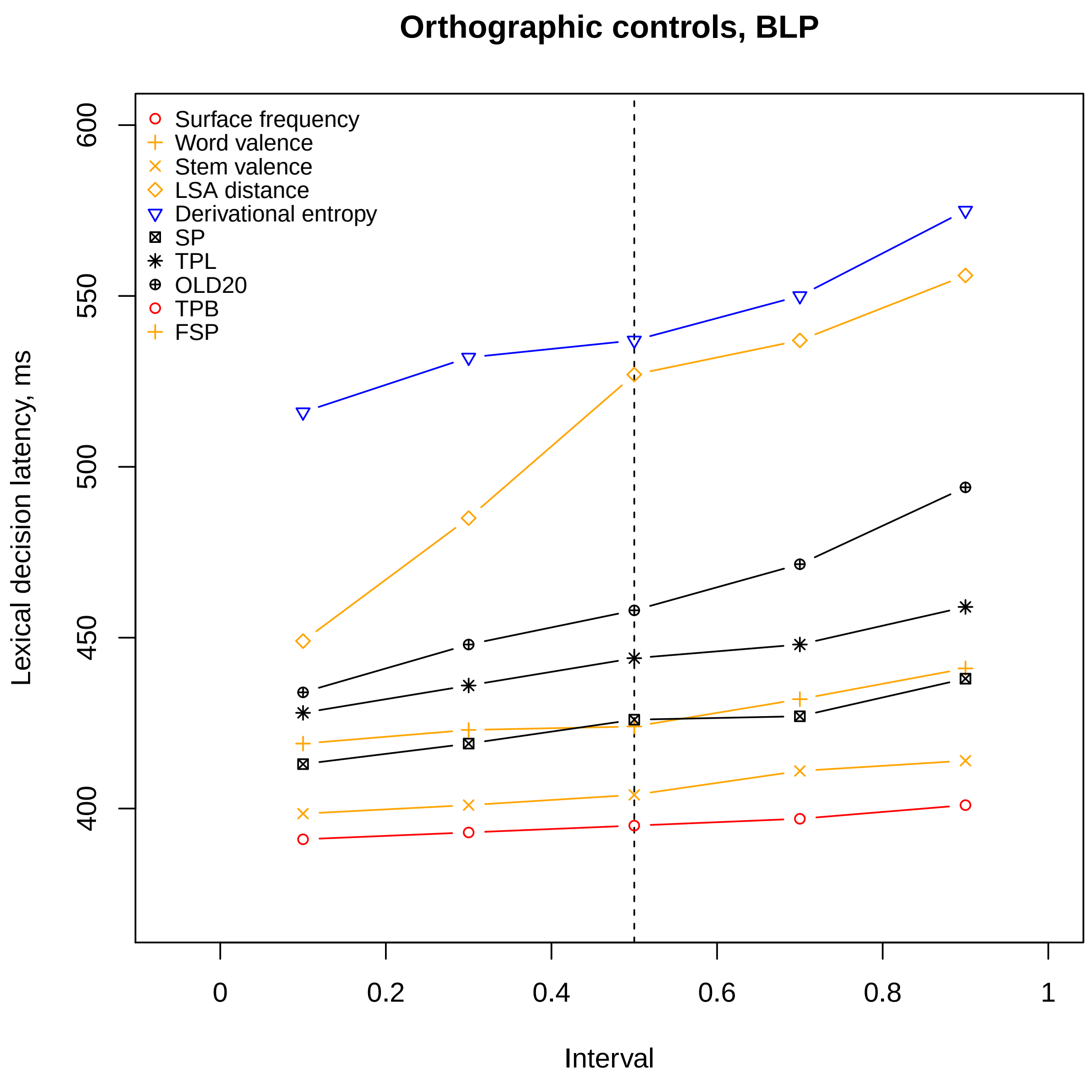

Figure B.4: Vincentile plot of divergence point intervals for English form control words. Colours represent groups of variables: frequencies of whole word and stem (red); as well as orthographic (black), morphological (blue) and semantic (orange) predictors. Only variables that elicited a detectable divergence point between survival curves in at least 700 out of 1,000 iterations with a 3% minimal contrast are plotted (other effects were deemed too unstable to estimate).

Appendix C: Evaluating semantic transparency

We tested the distributions of semantic transparency across conditions by assessing the LSA similarity of the pseudo-derived word with its pseudo-stem. LSA scores provide a measure of semantic similarity of the stem and the whole word. We compared the LSA similarity of these scores with those in the Dutch derived word condition and the English derived word condition. The distributions of LSA scores across conditions is depicted in Figure C.1. The independent Mann-Whitney U-test confirmed that the difference between the LSA scores of the English derived words ($M = 0.662$) and the English pseudo-derived words ($M = 0.864$) was reliable ($W = 4756.5$, $p < 0.0001$). The same was true of the comparison between the LSA scores of Dutch derived words ($M = 0.643$) and English pseudo-derived words ($W = 815$, $p < 0.0001$). Meanwhile, the difference in LSA scores between both derived word conditions across languages did not yield a statistically reliable difference ($W = 22050$, $p = 0.18$). In addition, Figure C.1 suggests that form control words are as opaque as pseudo-derived words, such that LSA similiarity between the whole word and the embedded word is similar across conditions. A further independent Mann-Whitney U-test confirmed that the difference between the distribution of LSA scores for the pseudo-derived word condition ($M = 0.864$) and the form control condition ($M = 0.858$) was not statistically significant ($W = 7938$, $p = 0.84$). In summary, we are confident that words we selected possess transparency characteristics that are expected.

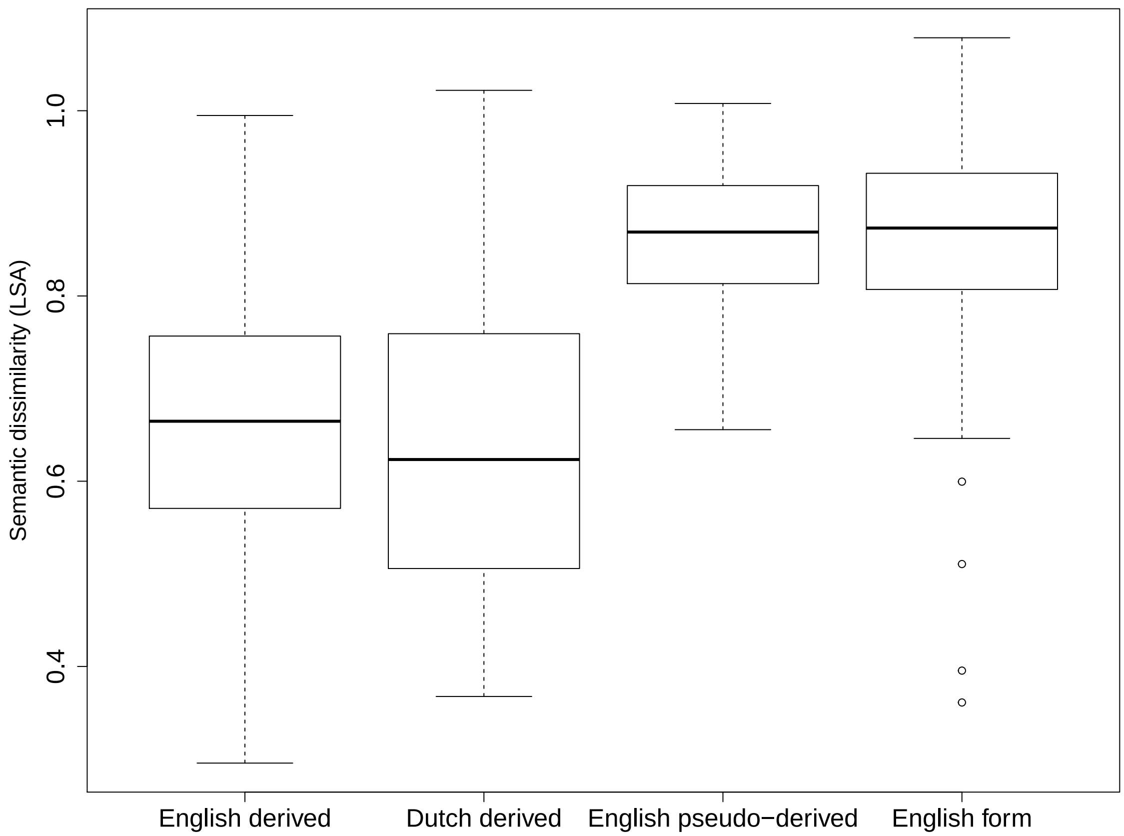

Figure C.1: The distributions of semantic similarity across each word condition. Semantic similarity was gauged by the LSA distance between derived word/pseudo-derived word/form control and derived stem/pseudo-stem/form control stem. A greater LSA score indicates a greater dissimilarity between meanings of a pair of words.

5 Summary and Conclusions

The goal of this thesis was to further understand how the human mind uses the semantic information carried by morphemes and the whole-words in which they are embedded during complex word reading. We targeted this research aim from three different angles: (i) the interactions between various reading skills and the semantic transparency of compound words as co-determinants of compound word recognition, (ii) the role of competition between conceptual relations during compound word recognition, and (iii) the time-course of the activation of stem and whole-word meanings during the visual comprehension of derived words.

In a series of experiments on the visual recognition of compound and derived words, we found evidence that lexical processing of complex words involves the activation of many sources of semantic information pertaining to morphemes and the whole words in which they are embedded. Emerging from the evidence that we present is that these semantic cues play a central role in the comprehension of complex words. The aim of this chapter is to spell out this claim by presenting an overview of the findings of this thesis.

In what follows, we first provide a summary of the contents and implications of each of the preceding chapters. We then continue with three sections, the first of which ties together the theoretical implications and the broader significance of the individual chapters of this thesis, the second one draws a general conclusion of this thesis, and the third outlines possible avenues of future research.

Summary of findings

This thesis consists of three separate experimental studies. Here we outline the key findings of each study.

Retrieving compound word meanings is dependent on reading skill

In Chapter 2, we broadened the scope of research on the semantic transparency effect on compound word processing. We document an eye-tracking study which examined the influence of semantic transparency and individual differences on the silent reading of compound words embedded within natural sentence contexts. We investigated the effect of semantic transparency on a sample of 100 young adult non-college-bound readers from the community, who were tested on a large battery of different verbal and general cognitive skills. The individual differences test battery measured a range of component skills of skilled reading, such as phonological awareness, simple and complex memory span, rapid automatized naming, word and nonword reading (decoding), reading comprehension, and reading experience.

At the outset of the study, we demonstrated that the variability in the amount of reading exposure in a non-college bound community was greater than in a convenience sample of university students. This finding motivated our choice to examine compound processing in the non-college bound community, such that greater variability in reading skill would increase the likelihood of detecting larger potential discrepancies in processing behaviour. The summary of the key findings documented in Chapter 2 is as follows:

- We report interactions between individual reading experience and the effect of semantic transparency on eye-movements. We found a gradient processing outcome that was consistent and robust across all interactions. Greater semantic transparency served to (i) increase the probability of regressive saccades amongst individuals with lower exposure to printed materials and lower vocabulary knowledge, and (ii) slow down total reading times of individuals with fewer years in education. However, the same lexical property of high semantic transparency yielded a reverse effect in more skilled readers, such that greater semantic transparency of compounds was associated with (i) lower regressive saccade rates in participants with greater exposure to print and more vocabulary knowledge, and (ii) shorter total reading time fixation durations for participants with more years in education.

- We found no evidence for a main effect of semantic transparency in the eye-movement record. The presence of interactions with reading skill (see above) reveals that the null main effect of semantic transparency represents the collapse of a rich interactive pattern into a flat line. We argue that this alone might account for the discrepant findings in the prior literature (e.g., Frisson, Niswander-Klement, & Pollatsek, 2008; Pollatsek & Hyönä, 2005), where the presence or absence of the critical effect of semantic transparency would depend on the prevalent proficiency of their participants.
- The extent of the processing cost differed across each transparency relationship. For the least skilled readers, the inhibitory effect of transparency was more harmful specifically when processing compounds with a greater semantic overlap between the meanings of the head and the whole compound word.

In summary, the results of Chapter 2 uncover the previously unknown interaction between an individual's amount of language experience and compound transparency. As we discuss further in this chapter, the pattern of results reported in Chapter 2 can be explained by the simultaneous operation of two independent processing mechanisms: conjunctive activation of multiple related meanings and discrimination learning (Baayen, Milin, Đurđević, Hendrix, & Marelli, 2011).

Compound recognition involves competition between conceptual relations

Chapter 3 sought to understand the exact nature of the hypothesis advocated by Gagné & Shoben (1997), that relational structures of compound words (e.g., *snowman* is a man MADE OF snow) compete for selection during compound word recognition. A key development of this theory was the finding that lexical processing is systematically affected by the diversity of the distribution of potential relations for a given compound. That is, Gagné & Spalding (2014) found that a greater absolute number of potential relational interpretations for an existing compound word slows down its word recognition time. However, it was still unknown whether just the number of activated conceptual relations, or the *relative strength* of activated conceptual relations was driving the hypothesized competition effect.

In Chapter 3, we reanalyzed the lexical data obtained from the experiment reported by Gagné & Spalding (2014), in which participants were serially presented with a list of noun-noun compounds. For each compound, participants chose the most likely conceptual relation out of a possible 16 interpretations. With a resulting frequency distribution of relational interpretations per compound, in Chapter 3 we calculated competition between conceptual relations using the information theoretic measure of Entropy (Shannon, 1948). Entropy measures the amount of uncertainty within a probability distribution. Applied to data from the possible relations task, greater entropy indicates a larger amount of uncertainty regarding the relational interpretation of a given compound.

Across two independent lexical decision data sources, we found the expected effect that increased entropy of conceptual relations for English compound words inhibited response times. These results demonstrate, through a more parsimonious measure of relational competition than ones employed earlier, that the relative difficulty of converging on any one interpretation of a compound translates into increased processing effort, i.e., longer response times. Thus, we confirmed that for established endocentric compound words, the ease of selecting a compound's relational interpretation relative to other potential relations in the distribution (e.g., a ball MADE OF meat vs. a ball FOR meat in the compound *meatball*) influences the ease with which the meaning of a compound is retrieved from memory.

In summary, the results of Chapter 3 provide further support for the RICE model (Spalding, Gagné, Mullaly & Ji, 2010) of compound word recognition. Specifically, we provide evidence for Stage 2 of the RICE theory, that possible conceptual relations linking the modifier *and* the head are concurrently activated. Once these conceptual relations are activated, they compete for selection as the interpretation of the compound word.

Semantic access of complex words is early

The aim of Chapter 4 was to address the hotly contested debate between form-then-meaning and form-and-meaning theories of decomposition during complex word recognition. We obtained lexical decision latencies from English and Dutch mega-studies for (i) derived words, where the word consists of a decomposable stem and suffix combination (e.g., *badness*; *bad* + *-ness*); (ii) pseudo-derived words (e.g., *wander*; *wand*, + *-er*); and (iii) form control words (e.g., *ballad*; *ball*, + *-ad*). We also collected several lexical characteristics of these words, including those pertaining to the orthographic and morphological structure (e.g., lexical neighbourhood density derivational family entropy, suffix productivity), frequency (e.g., stem and surface frequency) and meaning (e.g., semantic similarity between the stem and whole word, valence of the stem, and valence of the whole word).

We applied a distributional survival analysis (Reingold & Sheridan, 2014) to lexical decision latencies in order to obtain an estimate of the onset of the influence of an experimental variable on reaction times. The survival technique was employed to provide insight into the time-course of complex word recognition. The key findings were as follows:

- Across all conditions, surface frequency was the earliest variable to exert and influence on reaction times.
- The onset of the effects of whole-word and stem valence precedes that of any lexical variable that captures the morphological structure of Dutch and English derived words.
- The onset of the effect of semantic transparency of the whole-word (gauged by the LSA distance between the whole-word and stem) precedes access to the onset of any variable that is assumed to be a hallmark of decomposition of the derived word (e.g., suffix productivity or the transition probability between the root and the affix of the complex word).
- The temporal onset of derivational family entropy is the earliest ‘morphological’ effect to emerge for derived words in English and Dutch.

In sum, the findings of Chapter 4 provide a novel methodological and theoretical contribution to the literature on the time-course of lexical-semantic access and the nature of morphological decomposition. As we will discuss further, our results are in line with accounts of complex word processing that claim concurrent access of form-and-meaning (e.g., Feldman, Milin, Cho, Moscoso del Prado Martín, & O'Connor, 2015).

Theoretical significance and broader implications

Discrimination strain: the price of impoverished reading experience

The significant contribution of Chapter 2 is the novel approach to examining the consequences of variability in reading skill on the processing of semantic transparency, an exploration which has been overlooked in previous research. Based on the results presented in Chapter 2, it now seems likely that individual reading experience and the semantic characteristics of the compound word co-influence the visual identification of compound words during reading. The pattern of results for poorer readers in Chapter 2 can be interpreted well within the principles of the Naive Discriminative Reader (NDR) model of word learning and reading (Baayen, Milin, Đurđević, Hendrix, & Marelli, 2011).

Based on mathematical formulations of error-driven human learning propounded by Rescorla and Wagner (1972), the naive discriminative learning mechanism proposes that learning, in a discriminative sense, serves to minimize the uncertainty between a set of cues and a distribution of predicted outcomes (Baayen, Hendrix & Ramscar, 2013). In the case of reading, cues in the NDR model can be conceived of as orthographic input, and the outcome of the model can be conceived of as the level of activation that a word's symbolic meaning representation or "lexome" receives in the mind of the language user (Baayen, Shaoul, Willits, & Ramscar, 2015). Within this framework, the results presented in Chapter 2 support the notion that limited reading experience serves to maintain, and not minimize, the uncertainty between the compound cue and the meaning that is associated with it. For transparent compounds, we show that this uncertainty is especially costly, and results in imprecision when dissociating the target meaning of a compound from all other meanings (including those of constituents) that are activated by largely overlapping orthographic cues.

How exactly does the learning system work for compound words? In the NDR model, it is assumed that the meanings that are cued by the head, the modifier and the compound are in a state of constant competition for the target outcome (i.e., the meaning of the compound). The idea is that during learning, the intersection of words appearing in the context of the modifier or head, and those appearing in the context of a compound with that modifier or head, creates a competition among the meanings that the orthographic cues of the compound denote. Whenever the words in the shared contexts co-occur with a constituent (which occur much more frequently in the language), weights to that constituent are strengthened, while at the same time the weights to the compound are weakened. Conversely, when the words in the shared learning environments of the compound and its constituents are more likely to occur with the compound, the weights on their connections to the compound are strengthened, whereas those to the constituent are weakened (see section 2.1.3 The Naive Discriminative Reader theory in Chapter 2 for discussion on how the Latent Semantic Analysis measure we adopt to gauge semantic transparency is related to the predictions of the NDR model).

Why is limited reading experience especially costly for the processing of transparent compound words? Recall from above that the NDR account assumes that compound learning is driven by the overlap in the environments in which the head, the modifier and the compound itself are found. For semantically transparent words, the amount of overlap in the learning environments is substantial relative to opaque compounds. In other words, transparent compounds embed morphemes that are learned from more or less the same contexts as the compound themselves. Consider reading a transparent compound like *shellfish*. There is considerable overlap in the contexts of the compound itself, *shellfish* and its head, *shell*. For example, these contexts may include words like, *sea*, *rock*, *sand* and *shark*. In addition, *shellfish* and *fish* also shares the same contexts as words sharing overlapping significant orthography, such as *catfish*, *goldfish* and *starfish*: one discriminating between fish that have large protruding whiskers and other fish, one discriminating between fish that are typically orange coloured and other fish, and one discriminating between a marine echinoderm with five or more radiating arms and other marine life.

As transparent words like *shellfish* are learned from very similar environments as their constituents, there is a potential for a particularly high level of uncertainty in the mapping of forms to meanings at the very beginning of learning. Specifically, the high number of cues in the shared contexts of the compound and its constituents (e.g., *sea*, *rock*, *sand*, *shark*, *catfish*, *goldfish*, *starfish* etc.) weaken the weights to the compound meaning, while at the same time strengthening the weight to the constituent meaning. This process causes greater ambiguity in associating the compound with its intended meaning, which in turn translates into a greater challenge of discriminating between between all activated meanings. Ultimately, this cost of reading experiences imposes increased cognitive load for the reading of transparent compounds.

However, at the same early stage of reading, this challenge will not be as acute for opaque compounds. An opaque compound, such as *hogwash*, is predicted to impose relatively little meaning discrimination cost (and no facilitation) at an early stage of learning. Following the same line of reasoning, the constituents *hog* or *wash* do not occur in similar lexical contexts as the whole word, and so the uncertainty in discriminating between the meanings cued by *hogwash* and *hog* or *hogwash* and *wash* is not as great with any amount of reading exposure. This reduced uncertainty does not impose the same level of discrimination strain as a transparency compounds during reading, and thus imposes relatively little cognitive load[5].

Importantly, the key proposal of the NDR model is that the strengths of the activations between form and meaning that learners acquire from experience are in constant flux. In other words, both the context of words *and* the prior experience of the reader influence the strength of the activation between the lexome outcomes and the linguistic cues. Chapter 2 presents a major step in provides empirical support for this theory.

Experiential learning and lexical quality

Remarkably, the individual differences tests that interacted with semantic transparency in Chapter 2 each provide an approximation of an individual's accumulation of experience with printed language over the lifespan, i.e., the number of books an individual has read, the number of words an individual knows, and the number of years an individual has been in school. The finding that skilled word reading is associated

[5]We thank Harald Baayen for his help in clarifying how the NDR model handles compound word learning.

with amount of reading experience also meshes with the Lexical Quality Hypothesis of reading development (Perfetti, 1985; 2007; Perfetti & Adloff, 2012).

The Lexical Quality Hypothesis argues that weak lexical representations are strengthened with increased exposure to words across many different contexts. A high-quality lexical representation is assumed to comprise of a strong memory trace of the word's orthography, phonology and semantic representations. For lexical development over the lifespan of the individual, the Lexical Quality Hypothesis entails that with limited experience of a word, its lexical representation is impoverished and thus slows down word recognition. However, with increasing exposure to language, word representations become more entrenched in memory via strengthened connections between the orthographic, phonological and semantic components of lexical representation.

Considered jointly, the statistical learning frameworks outlined in the NDR model and Lexical Quality Hypothesis both point to a key ingredient for effective discrimination of word meanings from other word meanings: *reading experience*. Moreover, the pseudo-developmental findings presented in Chapter 2 dovetail with a number of studies have shown that individual differences in reading comprehension and experience modulate sensitivity to the effects of distributional information present in the orthography of printed language. For example, a series of experiments have demonstrated that less proficient readers are most affected by the influence of low frequency words, the frequency of embedded morphemes in derived words, the probabilistic bias for the presented format of compound words (e.g., spaced vs. concatenated), and predictability (cf. Ashby, Rayner, & Clifton, 2005; Falkauskas & Kuperman, 2015; Kuperman & Van Dyke, 2011a, 2011b, 2013; Whitford & Titone, 2014). The results of Chapter 2 supplements this body of work by demonstrating how language experience modifies cognitive entrenchment at the conceptual level of lexical representation.

The semantic neighbourhood effect

The principles of the NDR perspective (Baayen et al., 2011) are well-suited to capture the semantic transparency effects we observe for experience-impoverished readers in Chapter 2. Moreover, our results for inexperienced readers are also in line with several results obtained from a different strain of research on semantic competition effects. Recently, work by Mirman and colleagues (Chen & Mirman; 2012; Mirman & Magnuson, 2008; Mirman, 2011) and others (Huettig & Altmann, 2005; Huettig & McQueen, 2007; Huettig, Quinlan, McDonald & Altmann, 2006) has shown inhibitory effects of dense semantic neighbourhoods on word recognition. For example, in a word recognition task, Mirman and Magnuson (2008) found that words with distant semantic neighbours, i.e., words that denote concepts with fewer semantic features, facilitated word recognition. On the other hand, in the same study they found that words denoting concepts with many semantic neighbours inhibited visual word recognition. Furthermore, Mirman (2011) replicated this effect on language production, in a picture naming study on aphasic and non-aphasic controls.

In series of computational experiments, Chen and Mirman (2012) reported that opposite effects of near and distant semantic neighbours can be explained in a single unified model of parallel distributed processing. Chen and Mirman (2012) proposed an Interactive Activation Competition model of lexical processing. As in other applications of Interactive Activation Models (e.g., Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; McClelland & Elman, 1986; McClelland & Rumelhart, 1981), the network models the graded, parallel activation of multiple similar representations across three representational layers (i.e., phonemes/letters, words and concepts). Across a range of parameter settings in multiple simulations of the network, they found a consistent core 'computational principle' that determined whether or not neighbourhood effects at the level of semantics were facilitative or inhibitory. Chen and Mirman reported that strongly active neighbours had a net (i.e., total activation received in the simulation) inhibitory effect, and weakly active neighbours had a net facilitative effect. These findings are akin to those of inexperienced readers we observed in Chapter 2. Therefore, it could be speculated that the net activation of closely related concepts (i.e., those activated when reading a highly transparent compound; *boot* and *lace* in *bootlace*) relative to distantly related concepts (i.e., concepts denoted by opaque compounds; *boot* and *leg* in *bootleg*) is one of the reasons behind the qualitative pattern of eye-movement results for poor readers.

Despite the theoretical insights that are gained by the IAC and NDR models, neither framework alone is able to fully account for the results of Chapter 2. In other words, the opposite effects of inhibition and facilitation when processing competing semantic representations can only be partly explained by the IAC or NDR model. The IAC approach does not explicitly address how neighbourhood effects change as learning increases, and thus does not make any assertions about differences in processing across individuals with differing amounts of language experience. Moreover, the NDR model makes explicit predictions about language learning over the lifespan and addresses word learning specifically, yet it is not yet known whether it is mathematically plausible for the NDR model to predict faster processing of highly transparent compounds as language experience accumulates.

We argue in Chapter 2 that the reversal from inhibition to facilitation can be explained by an underlying additive effect of (i) either the principles of the IAC or NDR models, and (ii) the psychological mechanism of conjunctive activation. The conjunctive activation or "semantic boosting" 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), but inhibit processing when the semantic representations are weakly associated (i.e., for the case of opaque compounds). The finding that facilitation of word processing is associated with dense and highly interconnected semantic representations is robust (Buchanan, Westbury, & Burgess, 2001; Duñabeitia, Avilés, & Carreiras, 2008; Locker, Simpson, & Yates, 2003; Siakaluk, Buchanan, & Westbury, 2003; for a discussion of this effect in morphological families, see De Jong, 2002). Since this theory does not make any predictions about how reading proficiency affects the conjunctive activation of similar conceptual representations (as far as we know), we assume that conjunctive activation for transparent compounds relative to opaque compounds is present across the entire continuum of reading experience. While this effect remains stable throughout all levels of reading experience, we posit that the effect of conjunctive activation is offset by the additive effect of the independent psychological mechanism of discrimination learning, or by semantic neighbourhood principles outlined by Chen and Mirman (2012).

Relational properties of morphemes during complex word recognition

This thesis has touched upon a number of sources of semantic information that are encapsulated within complex words. As we have discussed so far, one source of semantic information stems from the semantic similarity between constituent morphemes and whole-word meanings of compound words. However, the study reported in Chapter 3 addressed the role of an additional relational characteristic encoded in the percept of a compound word: *conceptual relations*. The novel contribution of Chapter 3 is the implication that there can be many active conceptual relations associated with a compound in conceptual memory, and some interpretations are stronger semantic competitors relative to other ones. These findings demonstrate that the process of meaning combination is bound by the interpretive 'jists' that are computed from a specific set of relations (e.g., Spalding and Gagné, 2014). While this may be the case, how can the findings of semantic transparency in Chapter 2 and those of conceptual relations in Chapter 3 be understood jointly?

So far, the processing inhibition associated with competition between concurrently activated meanings (i.e., semantic transparency) has been interpreted within a naive discrimination learning perspective (see Chapter 2). However, as Baayen et al. (2015, pp. 23) point out, an NDR framework "explicitly rejects the conceptualization of language as a formal calculus". Unlike the unmediated mapping of form to meaning propounded by the NDR account, the RICE theory of compound recognition (Spalding, Gagné, Mullaly & Ji, 2010) implies that higher order compositional processes guide compound processing. Thus, the implications of the lexical decision experiment in Chapter 3 does not fit neatly with the calculus-absent discrimination learning perspective.

Perhaps a way to unify these two strands of research is to foreground the cognitive process of word meaning retrieval that is common to the results of Chapters 2 and 3: *semantic competition*. Though the NDR account does not explicitly accommodate a meaning computation stage of compound processing, it does predict that lexical processing involves selecting one out of the many potential meanings that are activated by the orthographic code of the compound word. It could be that the relational information that is central to the RICE account of compound word processing constitutes an additional source of competitive conceptual information.

This above interpretation is supported by the operationalization of conceptual combination in Chapter 3. The advantage of entropy calculated over the distribution of conceptual relations associated with a compound is that it shifts the focus from the relational structure denoted by the linguistic signal, to the underlying probability distribution of the potential meanings that are conveyed by the signal. The entropy measure defined in Chapter 3 thus characterizes conceptual relations as a probabilistic source of information that is similar to the uncertainty that exists between a distribution of meanings and orthographic cues in an NDR network. Thus, the shift toward an information theoretical approach of understanding conceptual combination perhaps provides a major step in reconciling the discrimination view invoked in Chapter 2 with RICE theory foregrounded in Chapter 3.

In sum, we may conclude that there are two sources of semantic uncertainty stemming from the meaning relations of constituents during compound word recognition. The first involves successfully discriminating between the entanglement of meanings that are cued by the orthographic code of the compound (e.g., identifying the meaning of *rainstorm* and not *rain* or *storm*). The second involves selecting the most likely relational interpretation for the compound (e.g., *storm* HAS *rain*, *storm* CAUSES *rain*, or *rain* DURING *storm*)[6]. Thus, compound word recognition may involve the automatic activation of many conceptual representations, including the general purpose mechanism of discriminating between the meaning of the whole word and meaning of the morphemes, with subsequent or parallel activation of the conceptual 'whole' of the compound word, which may include selecting between many different potential relational interpretations.

The time-course of complex word processing: relational and atomic semantic information is accessed early during derived word recognition

Whereas Chapter 2 and Chapter 3 focussed on two relational lexical semantic properties of compound words, Chapter 4 turned to derived words and examined the time-course of access to both relational *and* atomic lexical semantic properties during derived word recognition. For the case of derived words, relational semantic

[6] Indeed, the view that compound recognition involves both the effects of semantic transparency and the automatic computing conceptual relations is hinted at in other theoretical accounts (e.g., Ji, Gagné, and Spalding, 2011).

properties were estimated based on the evaluation of the meaning of the derived word's free morpheme in relation to the entire derived word (e.g., the semantic similarity between a derived word and its stem). On the other hand, atomic semantic properties were estimated for the individual free morpheme and whole complex word separately (e.g., the valence of both the derived word and of the stem)[7].

In Chapter 4 we collected valence ratings for derived words and their free standing stems, the semantic similarity between derived words and their stems, frequency-based measures (e.g., stem frequency and whole-word frequency), measures of orthographic properties (e.g., character length and orthographic neighbourhood density), and properties pertaining to morphological structure (e.g., derivational family entropy and suffix frequency). With the aid of a non-parametric survival analysis of reaction time data (Reingold & Sheridan, 2014), we then constructed a timeline of onsets for each of the above lexical variables during visual word recognition.

The survival analysis of lexical decision latencies revealed a robust, similar, temporal pattern of effects the pose a challenge for models that require that complex words are decomposed in an obligatory and semantics blind manner:

- The effect of whole-word frequency is typically considered as an indication of full-form lexical access at the post-decomposition phase of complex word recognition (Solomyak & Marantz, 2010; Taft, 2004). The results of Chapter 4 contradict this theoretical stance by showing that whole-word frequency had the *earliest* onset of all lexical effects across derived words in English and Dutch, and also for pseudo-derived and monomorphemic words.

  However, the exact nature of the whole-word frequency effect is still not clear. As Kuperman (2013) notes, some accounts consider whole-word frequency to be indicative of semantic access (e.g., Baayen, Feldman, & Schreuder, 2006; Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004). We argue that the early response to surface frequency, consistent across mono-morphemic and complex words, reflects the rapid and automatic detection of the orthographic code of the whole-word unit (Rayner, Liversedge, White, & Vergilino-Perez, 2003; Rayner,

[7] See Kuperman (2013) for further discussion of relational and atomic properties of compound words.

Ashby, Pollatsek, & Reichle, 2004; Staub, White, Drieghe, Hollway, & Rayner, 2010).

- For English and Dutch derived words, the onsets of variables indicating access to semantic information preceded those of lexical variables implicated in morphological parsing. The finding that the processing of relational semantic properties (i.e., LSA distance; indicating access to meanings of the stem and the whole word) and atomic semantic properties (i.e., whole-word and stem valence; indicating access to meanings of individual lexemes) is not delayed until after morphological processing run counter to any account that advocates that access to morphological structure must precede access to semantics (e.g., Rastle, Davis, & New, 2004; Fruchter & Marantz, 2015).

- Our findings corroborate Kuperman's (2013) findings of compound word processing that the emotional connotation of a complex word and those of its constituents independently contribute to the efficiency in recognizing the whole multi-morphemic string. We extend Kuperman's findings to derived words (across two languages), and also demonstrate that 'atomic' morphemic and whole-word properties are rapidly activated during the recognition of complex (and also simplex) words. Considering the present findings jointly with Kuperman's (2013) results, we argue that valence should be considered as an important variable which can be used to diagnose access to morpheme and whole-word meanings during the time-course of complex word recognition. This is an especially intriguing result because as, until now, surface frequency and base frequency have typically been used as diagnostics of whole-word and morpheme access.

- Chapter 4 suggests that the earliest and most consistent morphological variable across English and Dutch derived words was related to the morphological family of the complex word. This finding corroborates the findings across many languages (including English and Dutch) that complex words may be organized and retrieved from the mental lexicon on the basis of morphological families (cf. De Jong, Feldman, Schreuder, Pastizzo, & Baayen, 2002; Juhasz & Berkowitz, 2011, Kuperman, Bertram, & Baayen, 2008; Moscoso del Prado Martín, Kostić,

& Baayen, 2004).

Taken together, our analysis in Chapter 4 captures an early effect of the meaning(s) that are encapsulated within a complex word. The temporal aspect of this finding is noteworthy because it runs counter to the characterization of visual word recognition as a strictly staged process involving independent access to (morpho-orthographic then morpho-semantic) lexical cues (e.g., Meunier & Longtin, 2007; Rastle & Davis, 2008; Solomyak & Marantz, 2010; Taft & Forster, 1976). The larger picture that emerges from the data is that following processing of orthographic form, processing of semantic and morphological cues occurs in parallel. Moreover, our findings suggest that atomic properties are accessed in concert with the relational properties of derived words.

It is important to note that the data presented in Chapter 4 do not suggest that there is no place for a process of morphological decomposition in a model of complex word recognition. As we argue in the discussion and conclusion of Chapter 4, after access to the form of a derived word, conceptual and morphological units of representation may be processed concurrently. Thus, while the results of Chapter 4 contradict the relative ordering of the timeline proposed by accounts of semantics-blind orthographic decomposition (e.g., Fruchter & Marantz, 2015), we do not deny that morphological parsing occurs during the visual recognition of complex words. On the face of it, the results of Chapter 4 are in favour of a multi-route model of complex word reading (a development of the dual-route model by Schreuder and Baayen, 1995), which assumes early and concurrent access to multiple sources of lexical information (Kuperman, Bertram, & Baayen, 2008; Kuperman, Schreuder, Bertram, & Baayen, 2009).

In summary, we contend that, although not without methodological drawbacks, the application of the distributional survival technique provides an important and fruitful method of exploring time-course issues during word recognition. Based on the results reported in Chapter 4, we argue that a processing stage that is devoted solely to the morpho-orthographic properties of the input, while remaining blind to semantic information, unnecessarily deprives the word recognition process with a potentially important source of information.

Conclusions: Finding meaning in the Hungry Lexicon

There is no clear consensus about how morphemic structure is represented in the mental lexicon. As a result, the field of psycholinguistics currently accommodates a surfeit of processing models that make conflicting predictions about the precise nature of morphological processing. Historically, much of the debate in the morphological processing literature has concerned whether complex words are mentally represented as lexical units (as assumed in so-called full-listing models, e.g., Bradley, 1980; Butterworth, 1983; Lukatela, Gligorijević, Kostić, & Turvey, 1980) or by means of their single constituents (as assumed in decomposition models, e.g., Aronoff, 1994; Fruchter & Marantz; 2015; Taft & Forster, 1976). Full listing maximizes computational efficiency: the process of composition would only be required when a compound or derived word is produced or comprehended for the first time. In contrast, decomposition maximizes storage efficiency, since only monomorphemic items would be represented in the lexicon. In this case, the computational cost of decomposition is greater since composition would be required at any time a compound or derived word is produced or understood.

More recent psycholinguistic theories have claimed that neither one of these approaches is suitable to fit experimental data and that morphological processing is balance of lexical storage and computation. A family of recent models of morphological processing (Baayen, Wurm, & Aycock, 2007; Ji et al., 2011; Kuperman et al., 2008; Kuperman et al., 2009; Laudana & Burani, 1995; Libben, Gibson, Yoon & Sandra, 2003, Marslen-Wilson, Tyler, Waksler, & Older, 1994; Pollatsek, Hyönä, & Bertram, 2000; Schreuder & Baayen 1995), though varying in precise details, claim that the language system activates as many representations as possible during the comprehension of compound or derived words. These theories suggest that in order to maximize the opportunity of obtaining the meaning of a complex word, readers will attempt to utilize all available processing cues. Thus, when considered jointly, the findings of this thesis align with Libben's (2006) 'maximization of opportunity' principle of morphological processing.

The semantic transparency of compounds is often used as a key diagnostic of access to the semantic representations of the constituents of compound words (Libben, 2005). Not all compounds can be derived from the meanings denoted by their constituents and therefore the impact of transparency during online processing sheds light on whether compound meanings are computed from their constituents when the constituents do relate in meaning to the whole word. The eye-movement study in Chapter 2 shows that semantic transparency affects word processing, and indicates that, in conjunction with the language experience of the reader, it should be included as an essential variable in future compound processing studies. The study reported in Chapter 2 has attempted to broaden the empirical and theoretical scope of the semantic transparency effect by moving from examining its effect at the aggregate level (i.e., the average effect among a population) to its effect as conditioned by individual variability in reading skill. Along with other studies (e.g., Falkauskas & Kuperman, 2015), we argue that it is incumbent on future researchers to consider the impact of individual language experience when studying the nature of compound word processing.

Together, the results of Chapters 2 and 3 further highlight the importance of the conceptual relationships between morpheme meanings in compound words (Fiorentino & Poeppel 2007), and also the consequences of the activation of semantically similar morpheme and whole-word meanings (Libben, 2005; 2006). Emerging from the findings of Chapters 2 and 3 is the conclusion that a complete model of compound word learning and reading must include the independent contributions of competition between compositional processes (e.g., RICE theory; Gagné & Spalding, 2014) and must also account for the consequences of experiential learning on the ability to handle concurrent semantic activation of constituents and whole-words (e.g., NDR; Baayen et al., 2011). Despite the fact that both findings were obtained from separate methodologies, we show that the conceptual integration of morpheme representations serves as component of compound recognition which may reduce the negative consequences that arise from the ambiguity caused by the concurrent activation of closely related meanings.

The results of the survival analysis in Chapter 4 suggest that semantic relations between morpheme meanings and whole words, the atomic meaning properties of stems and whole words, distributional properties pertaining to morphological structure, frequency of the stem and the whole word, and orthographic properties themselves are *all* assessed during complex word processing. Importantly, the sequence of cognitive processes presented in Chapter 4 for derived word processing places the retrieval of the semantic representations of derived words and their morphemes at the heart of the recognition process (e.g., Feldman, Milin, Cho, & Moscoso del Prado Martín, 2015), and does not demote their significance to an end product of decompositional processes (e.g., Rastle & Davis, 2008; Solomyak & Marantz, 2010; Taft & Forster, 1976). In sum, the results of Chapter 4 presents an important piece of evidence that morphological processing does not unfold in a way that conforms with a strictly serial form-then-meaning view (see discussion above).

To sum up, the overall results of this thesis support the general notion of the 'Hungry Lexicon' propounded by Libben (2005). The hungry lexicon is a figurative impression of the theoretical position advocated at the beginning of this section, which is that the language processing system makes use of many information sources during complex word processing. In support of this outlook, Libben states that,

> the morphological processing system is not designed to get the right answer. It is not organized to generate exactly the correct morphological parse, nor is it organized to generate exactly the correct lexical excitation. Rather, it is designed to make the greatest number of potentially useful representations and analyses available to other components of the cognitive system: in this sense then, the lexicon may be said to be hungry. (2005, pp. 271)

The findings presented in this thesis have offered several new perspectives on morphological processing which are characteristic of the hungry lexicon. Through the studies reported here, I have found evidence of four separate components of semantic access during the visual comprehension of complex word meanings. These components are that access to complex word meanings (i) depends on the language experience of the reader, (ii) relies on selecting a conceptual representation out of a set of competing conceptual relations, (iii) is sensitive to the emotional connotations of the stem and whole word, and (iv) arrives early in the time-course of word recognition. I hope that these distinctive findings will be absorbed into future theoretical and empirical approaches to the study of semantic access during the visual recognition of complex word forms.

Topics for Further Research

There are several lines of research that follow naturally from the results of this thesis and the limitations therein:

Individual differences in the time-course of complex word processing: A limitation of Chapter 4 is that the time-course of complex word processing does not take into account individual differences. A straightforward first step to achieve a better understanding of differences in processing across different levels of reading skill would be to implement the Individual Participant divergence point analysis procedure (Reingold & Sheridan, 2014). This technique could be extended to derive a sequence of cognitive processes from eye-movements, as well as lexical decision latencies, to compound and derived words. The aims of this research project will be to determine whether the same sequencing of cognitive processes hold across different reading skill levels, and also to find out whether the time intervals between the onset of each process are different across reading skill levels. This project would supplement research on the time-course issue of complex word processing, and would also provide further insight into how reading skill affects the strength of the connections between orthographic and conceptual levels of word representation (e.g., Perfetti, 2007).

The interplay between learning and discrimination: Though Chapter 2 provides empirical evidence for the behavioural effects of discrimination learning, there remains a number of crucial questions that need to be answered. Perhaps the most pressing question revolves around the need to achieve theoretical parsimony in the implications that are drawn from the evidence presented in Chapter 2. The conclusions of Chapter 2 posit a processing trade-off that is contingent on the relative contribution of two cognitive mechanisms: semantic boosting and discrimination cost. The first mechanism, semantic boosting, is accounted for by a theory which assumes that processing facilitation is propelled by conjunctive activation between related semantic entities. The second, discrimination inhibition, is accounted for by the simple learning principles that are mathematically formalized in the Rescorla-Wagner equations. A scientific explanation of data must strive for maximal theoretical parsimony. Thus, one must ask the question of whether the processing trade-off observed in Chapter 2 can be explained within one unifying cognitive theory, instead of the combination of two.

A starting point would be to investigate whether the Rescorla-Wagner equations outlined in the NDR model can predict the experience-contingent gradual reversal of inhibition to facilitation for the task of discriminating between closely related concepts. An approach that examines whether the facilitation boost can be explained by discrimination learning, instead of the other way round, is perhaps the most promising option because the NDR model (i) provides a mathematically falsifiable claim and (ii) makes explicit and computationally tractable predictions about word learning. Thus, a future research challenge is to see whether the pattern of results observed in Chapter 2 can be modelled on the basis of the already developed theoretical construct of naive discriminative learning. Within this approach, there could be two promising research projects:

In principle, an initial step would be to perform a computer simulation of the experiment reported in Chapter 2. The goal of the simulation study would be to create an artificial series of *learning events*, i.e., the contexts in which compound words are embedded, and apply the Rescorla-Wagner learning algorithm to these simulated learning events. Importantly, the simulation would need to incorporate the assumption of the NDR model that processing differences between transparent and opaque compounds are driven by the degree to which the contexts of their constituents and their whole-word forms overlap. Next, with a toy corpus in place, one can simulate the gradual learning of compound words with both transparent and opaque properties.

A second method could involve training the Rescorla-Wagner model on a *real* language corpus and use the activations produced by this training to predict actual reading times. This method could be easily implemented on the compounds and their associated latencies found in the dataset presented in Chapter 2. This endeavour may involve slightly redefining an operationalization of transparency, such that one would substitute corpus derived LSA scores with activations generated by a Rescorla-Wagner network. The activation (i.e., the amount of corpus training) that a compound receives for a specific individual would be proportional to the amount of relative experience of the individual language user in our dataset (or another individual differences measure present in the data set, such as vocabulary knowledge).

The advantage of the aforementioned technique is that a measure of transparency is calculated based on the principles of discrimination learning. No NDL-derived measure is implemented in Chapter 2. Another advantage of this approach is that the computational model is trained on a natural language corpus and not an artificial one. However, this opens up further questions about which corpus should be used as a representation of the texts that are read by the sample of the population used in our study. Note also that we have only concerned ourself with NDL modelling, one could also carry out a similar experiment on LSA scores (or any other distributional semantics model). Ultimately, an abundance of semantic-association measures will need to be calculated on multiple corpora and pitted against one another. The corpus-derived measure which best fits the experimental data will represent the optimal model of semantic memory during compound learning and reading.

Individual differences in conceptual combination and discrimination: A key prediction of the Naive Discriminative Learning perspective (Baayen et al., 2011; Baayen et al., 2015) is that experience drives learning. In Chapter 2, we revealed that attributes of an individual's language experience (as indexed by exposure to print, vocabulary knowledge and years in education) modulate the transition from inhibition to facilitation during the processing of semantic transparency. Further support for the NDR model within the current strand of compound reading research could be provided by verification that processing compound transparency involves the cognitive skill(s) implicated in discriminating between closely related concepts. Similarly, the RICE theory of compound processing predicts that compound recognition involves integrating morpheme meanings via the process of conceptual combination. A promising avenue of future research may therefore involve examining individual performance on a number of skill tests that target the ability of discriminating between closely related concepts and conceptual combination (see e.g., the Relatedness Judgement Task and the Semantic Category Probe Task which was designed and implemented by Freedman and Martin, 2001). This project could then examine whether population-wide variation in these skills interact with the processing of semantic transparency and conceptual relations during compound word recognition. Answering this question will help shed light on the nature of the cognitive skill-set(s) involved in conceptual discrimination and integration and provide insight into its relationship with language processing and learning.

Eye-tracking and entropy of conceptual relations: The model of compound word recognition that includes a measure of entropy of conceptual relations (Chapter 3) needs to be expanded to eye-movement behaviour. This could be achieved by performing the same analytical procedure as is developed in Chapter 2, i.e., fitting models that exclude the effects of individual differences and fitting models that include interactions with individual differences and entropy of conceptual relations. This project may also include a battery of individual differences tests that specifically target verbal and non-verbal skills in conceptual combination (see above).

In conclusion, across three experiments, this thesis has demonstrated that morphological processing appears to involve rapid and concurrent access to many sources of conceptual information. These findings suggest that readers attempt to utilize all semantic cues at their disposal during the visual recognition of complex word forms. I hope that future explorations into the visual recognition of complex word forms will build upon the discoveries that have been presented in this thesis.

References

Aronoff, M. (1994). *Morphology by itself: Stems and inflectional classes* (No. 22). MIT press.

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *The Quarterly Journal of Experimental Psychology Section A*, *58*(6), 1065–1086.

Baayen, R., Shaoul, C., Willits, J., & Ramscar, M. (2015). A discrimination-driven model for the acquisition of lexical knowledge in auditory comprehension. under revision.

Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, *55*(2), 290–313.

Baayen, R. H., Hendrix, P., & Ramscar, M. (2013). Sidestepping the combinatorial explosion: An explanation of n-gram frequency effects based on naive discriminative learning. *Language and Speech*, *56*(3), 329–347.

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*(3), 438–481.

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*(3), 419–463.

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

Bradley, D. (1980). Lexical representation of derivational relation. *Juncture*, 37–55.

Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review*, *8*(3), 531–544.

Butterworth, B. (1983). *Lexical representation.* na.

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological review*, *119*(2), 417–430.

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

De Jong, N. H., Feldman, L. B., Schreuder, R., Pastizzo, M., & Baayen, R. H. (2002). The processing and representation of Dutch and English compounds: Peripheral morphological and central orthographic effects. *Brain and Language*, *81*(1), 555–567.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological review*, *104*(4), 801–838.

Duñabeitia, J. A., Avilés, A., & Carreiras, M. (2008). NoA's ark: Influence of the number of associates in visual word recognition. *Psychonomic Bulletin & Review*, *15*(6), 1072–1077.

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*(6), 1607–1627.

Feldman, L. B., Milin, P., Cho, K. W., del Prado Martín, F. M., & O'Connor, P. A. (2015). Must analysis of meaning follow analysis of form? A time course analysis. *Frontiers in Human Neuroscience*, *9*. Retrieved from `http://dx.doi.org/10.3389/fnhum.2015.00111`

Fiorentino, R., & Poeppel, D. (2007). Compound words and structure in the lexicon. *Language and Cognitive processes*, *22*(7), 953–1000.

Freedman, M. L., & Martin, R. C. (2001). Dissociable components of short-term memory and their relation to long-term learning. *Cognitive Neuropsychology*, *18*(3), 193–226.

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*(1), 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., & Shoben, E. J. (1997). Influence of thematic relations on the comprehension of modifier–noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 71.

Gagné, C. L., & Spalding, T. L. (2014). Relation diversity and ease of processing for opaque and transparent English compounds. In *Morphology and meaning: Selected papers from the 15th International Morphology Meeting, Vienna, February 2012* (Vol. 327, pp. 153–162).

Huettig, F., & Altmann, G. T. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), 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*(4), 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*(1), 65–80.

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*(4), 406–430.

Juhasz, B. J., & Berkowitz, R. N. (2011). Effects of morphological families on English compound word recognition: A multitask investigation. *Language and Cognitive Processes*, *26*(4-6), 653–682.

Kuperman, V. (2013). Accentuate the positive: Semantic access in English compounds. *Frontiers in Psychology*, *4*(203). Retrieved from `http://dx.doi.org/10.3389/fpsyg.2013.00203`

Kuperman, V., Bertram, R., & Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes*, *23*(7-8), 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*(3), 876–895.

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

Kuperman, V., & Van Dyke, J. A. (2011b). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language*, *65*(1), 42–73.

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 802–823.

Laudana, A., & Burani, C. (1995). Distributed properties of derivational affixes: Implications for processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing.* Hillsade, NJ: Lawrence Erlbaum.

Libben, G. (2005). Everything is psycholinguistics: Material and methodological considerations in the study of compound processing. *The Canadian Journal of Linguistics/La revue Canadienne de linguistique*, *50*(1), 267–283.

Libben, G., Gibson, M., Yoon, Y. B., & Sandra, D. (2003). Compound fracture: The role of semantic transparency and morphological headedness. *Brain & Language*, *84*(1), 50–64.

Locker, L., Simpson, G. B., & Yates, M. (2003). Semantic neighborhood effects on the recognition of ambiguous words. *Memory & cognition*, *31*(4), 505–515.

Lukatela, G., Gligorijević, B., Kostić, A., & Turvey, M. T. (1980). Representation of inflected nouns in the internal lexicon. *Memory & Cognition*, *8*(5), 415–423.

Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the english mental lexicon. *Psychological Review*, *101*(1), 3–33.

McClelland, J. L., & Elman, J. L. (1986). The trace model of speech perception. *Cognitive psychology*, *18*(1), 1–86.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. an account of basic findings. *Psychological review*, *88*(5), 375.

Meunier, F., & Longtin, C.-M. (2007). Morphological decomposition and semantic integration in word processing. *Journal of Memory and Language*, *56*(4), 457–471.

Mirman, D. (2011). Effects of near and distant semantic neighbors on word production. *Cognitive, Affective, & Behavioral Neuroscience*, *11*(1), 32–43.

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*(1), 65–79.

Moscoso del Prado Martín, F., Kostić, A., & Baayen, R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition*, *94*(1), 1–18.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*(4), 357–383.

Perfetti, C., & Adlof, S. M. (2012). Reading comprehension: A conceptual framework from word meaning to text meaning. In J. P. Sabatini, â. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability.* (pp. 3–20). Lanham, MD: Rowman & Littlefield Education.

Perfetti, C. A. (1985). *Reading ability.* Oxford University Press.

Pollatsek, A., & Hyönä, J. (2005). The role of semantic transparency in the processing of Finnish compound words. *Language and Cognitive Processes*, *20*(1-2), 261–290.

Pollatsek, A., Hyönä, J., & Bertram, R. (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 820–833.

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., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, *11*(6), 1090–1098.

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: implications for the ez reader model. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(4), 720–732.

Rayner, K., Liversedge, S. P., White, S. J., & Vergilino-Perez, D. (2003). Reading disappearing text cognitive control of eye movements. *Psychological science*, *14*(4), 385–388.

Reingold, E. M., & Sheridan, H. (2014). Estimating the divergence point: a novel distributional analysis procedure for determining the onset of the influence of experimental variables. *Frontiers in Psychology*, *5*. Retrieved from http://dx.doi.org/10.3389/fpsyg.2014.01432

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black &

W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Crofts.

Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing.* (pp. 345–364). Hillsade, NJ: Erlbaum.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell Technical Journal*, *27*(4), 379–423.

Siakaluk, P. D., Buchanan, L., & Westbury, C. (2003). The effect of semantic distance in yes/no and go/no-go semantic categorization tasks. *Memory & Cognition*, *31*(1), 100–113.

Solomyak, O., & Marantz, A. (2010). Evidence for early morphological decomposition in visual word recognition. *Journal of Cognitive Neuroscience*, *22*(9), 2042–2057.

Spalding, T. L., & Gagné, C. L. (2014). Relational diversity affects ease of processing even for opaque English compounds. *The Mental Lexicon*, *9*(1), 48–66.

Spalding, T. L., Gagné, C. L., Mullaly, A. C., & Ji, H. (2010). Relation-based interpretation of noun-noun phrases: A new theoretical approach. *Linguistische Berichte Sonderheft*, *17*, 283–315.

Staub, A., White, S. J., Drieghe, D., Hollway, E. C., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1280–1293.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology Section A*, *57*(4), 745–765.

Taft, M., & Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior*, *15*(6), 607–620.

Whitford, V., & Titone, D. (2014). The effects of reading comprehension and launch site on frequency–predictability interactions during paragraph reading. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1151–1165.