CARNEGIE MELLON UNIVERSITY

DEPARTMENT OF PSYCHOLOGY

Dissertation

MORPHOLOGICAL PROCESSING DURING VISUAL WORD RECOGNITION: MECHANISMS OF DYNAMICS AND ACQUISITION

by

PATIENCE STEVENS

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2022

ACKNOWLEDGMENTS

A resounding thank-you to my advisor, Dr David Plaut, without whom none of this research would have been possible. Thank you for your guidance and your trust throughout this process, and for setting an example as a thoughtful scientist and dedicated teacher.

I would also like to thank the rest of my thesis committee, Drs Anna Fisher, Nazbanou Nozari, and Charles Perfetti, each of whom has made a significant and unique contribution to my growth as a researcher over the last five years. Anna, thank you for the steady supply of insights, advice, and humor, and for your attentive advising in my first year; Bonnie, for your investment as a committee member and for reminding me to take myself seriously; Chuck, for sharing your knowledge of and enthusiasm for the science of reading, and helping me connect with other reading scientists in Pittsburgh and beyond.

I am very grateful to the PIER learning community, and particularly to Professors David Klahr and Sharon Carver, for their dedication to providing young researchers with the tools and knowledge they need to apply their work to education. I am honored to be part of a network of such creative and altruistic people. Extra thanks to Sharon, for doing each of the impossible number of things she does with such care and thoroughness, and still always making time to check in on how everyone is doing.

Heartfelt thanks to Jessie Brown, who significantly contributed to the design and implementation of the developmental study detailed in Chapter 8.

I'm indebted to the researchers who shared data and materials with me, particularly Professors Debra Jared, Sally Andrews, and William Nagy. Thank you also to the JSPsych support staff, who always replied promptly with extensive troubleshooting support when I ran into issues designing these remote studies.

None of this research would have been possible without the participants; thank you to all individuals who participated in these studies, and to the participant databases that made recruitment possible, especially the Children Helping Science site and the CMU Psychology Undergraduate Research Pool.

Thank you to the VisCog lab meeting group, for a steady stream of stimulating talks and discussions, and for listening to too many presentations about what's going on during semantically opaque morphological priming.

Thank you to all the department staff, but especially Erin Donahoe, Audrey Russo, Becky Finkel, Ginger Placone, and Nick Pegg, for their timely help and friendliness.

I was lucky to pursue this degree alongside three of the most brilliant, compassionate, and delightfully strange people I have ever had the pleasure of getting to know. Jeanean, Alexandria, and Phoebe: I think this is really just the start of our friendship, and I can't wait to see where we all go next.

Finally, I would like to express love and gratitude for my parents, Douglas and Teresa, for their support of my lifelong educational pursuits, and for my husband, Ian Chesser, whose last name is a pseudo-morphological word.

The research reported here was supported, in whole or in part, by The Institute of Education Sciences, U.S. Department of Education, through grant R305B150008 to Carnegie Mellon University. The opinions expressed are those of the author and do not represent the views of the Institute or the U.S. Department of Education. The Center for Behavioral and Decision Research Graduate Student Small Grants Program also provided funding to support projects reported in this dissertation.

MORPHOLOGICAL PROCESSING DURING VISUAL WORD RECOGNITION: MECHANISMS OF DYNAMICS AND ACQUISITION

PATIENCE STEVENS

Carnegie Mellon University, Department of Psychology, 2022

Advisor: David C. Plaut, PhD

ABSTRACT

The recognition of complex words while reading is facilitated by form-to-meaning regularities present in written words, such as the shared letters in TEACHING and TEACHER or MINIATURE and MINISKIRT. This dissertation explores the mechanisms driving sensitivity to such morphological structure in skilled and developing readers, using computational and empirical approaches to strengthen the distributed theoretical account of these phenomena. Simulation work demonstrates that sensitivity to morphological structure can be acquired from natural language input, without providing explicit information about morphemes, and that the dynamics of morphological processing can be captured by a straightforward neural network architecture when time pressure, recurrence, and biologically plausible constraints on connectivity are incorporated. Empirical work focuses on whether sensitivity to embedded strings during word processing can be predicted by the string's orthography-semantics consistency (OSC), for which a novel variation on a prior metric is proposed. This metric approximates how semantically informative a given letter string is, across the words in which it appears. Although OSC effects prove to be difficult to detect in skilled readers, they are evident in younger readers with less experience. Additionally, skilled readers trained to learn novel stems with manipulated semantic consistency showed greater sensitivity to more consistent than to less consistent stems a week later. Finally, novel predictions are generated regarding the developmental trajectory of sensitivity to words that only appear complex, such as DEPARTMENT, to be tested in future research. The current work strengthens the distributed account of morphological processing by simulating known phenomena, particularly those argued to conflict with the distributed view, and investigating untested but core predictions of the model. Additionally, new insights are provided regarding how morphological sensitivity in word recognition emerges over reading acquisition, strengthening a much needed connection between the adult and developmental literatures on how complex words are recognized.

CONTENTS

CHAPTER 1

Introduction

1.1 MOTIVATION

A multitude of cognitive processes are simultaneously at play to enable successful reading comprehension. Consider this short excerpt from Brontë's *Jane Eyre* (p. 127), a book often taught in high school English literature classes:

> *'I mentally shake hands with you for your answer, despite its inaccuracy; and as much for the manner in which it was said, as for the substance of the speech; the manner was frank and sincere; one does not often see such a manner.'*

For these lines to take on meaning, the reader must decode the phrase "mentally shaking hands" by abstracting the social implications of a handshake away from the physical action. Simultaneously, the reader must understand that "your answer" refers back to the last utterance of the other speaker, understand that "its" and "it" also refer to that utterance, adjust to the antiquated abundance of semicolons, and additionally contextualize these comments within the characters' social context to appreciate the speaker's ironic tone.

Underlying all of these operations essential for full comprehension, though, is a much more basic capacity: The reader must be able to recognize all—or at least most[1]—of the words in the sentence. If identification of each word is laborious, the reader's working memory will be taxed in trying to accomplish the aforementioned tasks simultaneously, or at all. Furthermore, a reader's ability to integrate across word meanings to comprehend passages depends on the quality of the word representations that are retrieved (Perfetti, 2007). Thus, successful and efficient word recognition is the foundation upon which all more advanced reading comprehension skills are built.

[1] See Hu & Nation (2000) and Schmitt et al. (2011) for discussions of how many words must be recognized for a passage to be comprehended.

The task of word recognition essentially requires mapping from a word's visual representation to its meaning. Traditionally, the relationship between the surface form of a word and its meaning is described as arbitrary (e.g., Hockett & Hockett, 1960): for example, the words BEACH and PEACH look and sound alike but mean very different things, and thus perceptual similarity must be disregarded in deriving their meanings. However, this is not the case for morphologically complex words, such as BEDROOM and BEDDING, or MINIATURE and MINIMIZE. In these words, perceptual similarity does convey similarity in meaning. Complex words are the result of a productive morphology: a system, dependent on the language in question, by which recurring sequences of letters or phonemes can be combined or transformed to convey more varied and nuanced meanings. In some languages, the written form of complex words can provide more information about the words' meaning than the spoken form. In English, for example, the stems of the words MAGICIAN and HEALTH are more salient when read than when heard (see Rastle, 2019a). The morphological structure of complex words makes the form-to-meaning mappings that are learned during reading acquisition more systematic, and this regularity presents opportunities for more efficient access of word meanings.

Ample evidence indicates that, during the time-pressured task of reading, the visual system does make use of morphological structure to process words. For example, after learning a list of words that includes the word CAR, English readers recognize a morphologically related word like CARS more quickly than a word that is merely visually similar, like CARD (Murrell & Morton, 1974). Uncommon complex words in Italian are recognized more quickly if they have high-frequency as opposed to low-frequency stems, suggesting more common morphemes better facilitate the recognition of the word (Burani et al., 1984; Taft, 1979). Researchers have found evidence of sensitivity to morphological structure in reading in a wide variety of languages and writing systems, including French (Grainger et al., 1991), Spanish (Duñabeitia et al., 2008), Dutch (Drews & Zwitserlood, 1995), Hebrew (Frost et al., 1997), Chinese (Tsang & Chen, 2013), Japanese (Clahsen

& Ikemoto, 2012), and Serbian (Feldman, 1994). Despite the variability in morphological systems across these languages, morphological structure is consistently learned and used during lexical processing.

Given widespread evidence for morphological sensitivity during visual word recognition, much attention has been paid to understanding how a word's morphological structure affects the way it is represented and processed. Two prominent and somewhat opposing views regarding the cognitive mechanisms of morphological processing are the *decomposition* and *distributed* views.

1.2 THEORETICAL APPROACHES TO COMPLEX WORD PROCESSING

Theories of morphological processing are diverse, and have evolved significantly over the past few decades. An enduring divide among these approaches, however, is whether decomposition or distributed mechanisms are relied upon to explain sensitivity to word-form structure.

1.2.1 Decomposition-based theories

Decomposition theories propose that complex words are decomposed into their constituent morphemes, prior to recognition and access of the word's overall meaning (e.g., Murrell & Morton, 1974; Stanners et al., 1979; Taft & Forster, 1975). Advocates of decomposition theories vary in their descriptions of how decomposition is executed. In Rastle & Davis (2008), decomposition is described as "morpho-orthographic", applied to any word that has the appearance of morphological complexity (e.g., applied indiscriminately to both the pseudo-complex word CORNER and the truly complex word TEACHER). Crepaldi et al. (2010a) proposed that, in addition to an initial level of morpho-orthographic representations, a lemma level exists to allow for more abstracted identification of constituent morphemes. Grainger & Ziegler (2011) emphasized the inhibition of affixes to allow for stem recognition as the primary means of decomposition, whereas

Taft (1994) argued for the simultaneous activation of both stem and affix representations. Despite such variation in descriptions of decomposition, the common element among them is that, within some phase or route of word processing, the word is represented as a combination of independent morphemic units.

An entirely combinatorial approach to processing complex words, however, would work only for a language with a perfectly systematic morphology. This certainly does not hold for English: several words appear complex but aren't (such as CORNER and WITNESS), whereas others are morphologically structured but the morphemes' contributions to word meaning must be interpreted more loosely (e.g., RECALL doesn't mean "to call again"). Similar challenges arise in most other languages (e.g., in French, -ETTE means "little" in FILLETTE, "little girl", but not in BAGUETTE). Irregular words, such as TAUGHT from TEACH, additionally complicate the form-to-meaning mapping by introducing variability in complex words' appearances. Thus, a system for processing complex words must be able to take advantage of systematicity where relevant, but also handle words where morphological structure is less informative or even misleading.

Decomposition theories satisfying these constraints are mostly combinations of, or compromises between, listing and parsing theories. Taft (1979) refined his prefix-stripping theory to assert that words are stored *both* by their root morpheme, produced by splicing off affixes, and by their whole form. Clahsen (1999) distinguished routes for irregular versus regular processing of inflected words: irregular words are processed via structured lexical entries, whereas regular words are subject to affix-stripping. The Parallel Dual-Route Model proposed by Schreuder & Baayen (1995) posits two parallel access routes: one that processes the whole word directly and a second that splits morphological constituents and constructs meaning from them, with reaction times determined by the fastest route for a particular word (also see Baayen et al., 1997; Grainger & Ziegler, 2011). Despite differing in important ways, these models all assume that morphemes are explicitly represented during complex word processing, and give a central role to the decomposition of at least some complex words into constituent morphemes Rueckl (2010). Decomposition theories have shaped the framing of research on morphological processing for decades (see reviews by Amenta & Crepaldi, 2012; Diependaele et al., 2012; Marelli et al., 2020) and continue to feature prominently in current discussions of this subfield (e.g., Fleischhauer et al., 2021; Ciaccio et al., 2020; De Grauwe et al., 2019).

1.2.2 Distributed theories

Distributed theories of cognition provide an alternative theoretical perspective on morphological processing. Such theories are most commonly implemented in the form of artificial neural networks (also known as *connectionist* models). Neural networks consist of a system of connected units, typically arranged in layers. The activation of one unit is determined by the activations of the units connected to it, which are weighted by the strengths of their respective connections and summed. This sum is then passed through a nonlinear function. The values of connections between units (referred to as the *weights*), are adjusted with training to optimize performance on a given task. Neural networks are often mischaracterized as being purely associative, only able to learn surface-level statistical regularities. In fact, such networks can learn to make use of abstract structural relationships and perform complex operations, if doing so is useful for the task at hand (see Elman et al., 1996; Rumelhart et al., 1986, for overviews on using neural networks to understand cognition).

Distributed theories that tackle the processing of complex words trace back to a controversial chapter by Rumelhart & McClelland (1986). A neural network free of rules or explicit morpheme representations was presented as a potential account of how English speakers generate the past tense of verbs from base forms. Rumelhart & McClelland's model, unlike other linguistic accounts of conjugation at that time, did not require discrimination between regular ("wanted", "sounded") and irregular ("went", "threw") forms, nor any explicit representations of suffixes or phonological transformations. Instead, the only morphological information the model received was implicit in the structure of the task it was trained to perform: phonological representations of the base forms of a small set of verbs were presented to a neural network, which was trained via adjustment of unit-connecting weights to output the phonological representation of the past tense form. Within this simple learning environment, the model learned the base-to-past mappings well and was able to generalize to novel verbs (e.g., "wug" to "wugged"). In some phases of training it overgeneralized the "-ed" rule to irregular forms ("goed"), an observed tendency in children learning language. The chapter spurred an "intellectual firestorm" among linguists and cognitive scientists regarding the nature of linguistic knowledge (see Seidenberg & Plaut, 2014).

To explore morphological sensitivity in the context of word comprehension (as opposed to in production), Rueckl & Raveh (1999) trained a neural network with one hidden layer to map from orthographic representations to semantic representations for two artificial languages. In one language, morphological regularity was captured by constructing words from invented stems and suffixes, concatenating stem and suffix orthographic representations and combining their semantic representations. The other language was constructed by shuffling the first language's mappings from form to meaning, such that words' morphological structures did not inform their meanings. Rueckl & Raveh found that the language with morphological regularity required less training to perform well and could accommodate larger vocabularies, demonstrating how morphology enables efficient word learning. Additionally, they found evidence of compositionality in network processing after training: words containing a particular stem had similar patterns of hidden layer activation. Critically, this intermediate sensitivity emerged by applying general learning mechanisms to the whole-word form-to-meaning mappings the authors constructed, and no explicit morphological knowledge was encoded in the model itself.

Plaut & Gonnerman (2000) also trained neural networks on form-to-meaning mappings, and demonstrated that the stronger overt priming by semantically opaque words (i.e., words that appear to be but are not complex, such as DEPARTMENT and CORNER) in Hebrew compared to English can be attributed to the languages' differences in morphological systematicity. Networks were trained to map orthographic representations to semantic representations for the same set of artificial words, together with a remaining vocabulary that was either entirely transparent in how morphological structure informed words' meanings (morphologically *rich*) or entirely opaque (morphologically *impoverished*). When trained in the context of a morphologically rich language, likened to Hebrew, the network was faster to settle on a representation of a word after being presented with a morphologically but not semantically related word (i.e., opaque priming occurred). In the context of a morphologically impoverished language, likened to English, the network did not show such facilitation, as is seen in overt priming contexts in English. Plaut & Gonnerman also simulated morphological priming effects with diminishing magnitude for complex words whose meanings were semantically close versus distant from the target. This provides a concrete account for graded contributions of morphological structure sensitivity, as seen in Jared et al. (2017) and Gonnerman et al. (2007). Both Rueckl & Raveh (1999) and Plaut & Gonnerman (2000) demonstrate how theories of morphological processing, as well as the emergence and variation of morphological processing across languages, can be explored and understood using a distributed approach.

1.2.3 Distinguishing distributed and decomposition theories

Perhaps the key theoretical contrast between decomposition and distributed accounts is the following: Whereas, on a decomposition account, a given word either does or does not contain a morpheme, on a distributed account, the notion of "containing" a morpheme—as a recurring string of letters or phonemes—or of "sharing" a morpheme with another word, is entirely a matter of degree. Thus, on the latter, the representation of DRESSER doesn't contain that of DRESS; rather, the contribution of the letter string DRESS to the internal representation of DRESSER is highly similar to—but not identical to—the varying contributions DRESS makes to the representations of other words (e.g., DRESSING, REDRESS). In light of this contrast, decomposition and distributed theories are often pitted against each other in discussions of morphological processing mechanisms (e.g., Jared et al., 2017; Marelli et al., 2020; Fleischhauer et al., 2021). It is worth noting, however, they are not always in conflict, and can be difficult to distinguish empirically. In the context of a language with a perfectly systematic morphology, they could even be thought of as describing the same phenomenon at differing levels of detail. As demonstrated by Plaut & Gonnerman (2000) in the morphologically rich condition, a fully-trained distributed system can give rise to decomposition-like phenomena for all complex words. In a perfectly systematic language, hidden layers' activations in response to complex words might be mostly divisible into independent contributions of constituent morphemes, and a morpheme's contribution to these representations could be nearly identical, and hence functionally equivalent, across its appearances in different words. In this case, discriminating between distributed and decomposition theories might feel a bit like splitting hairs—at least with regard to the performance of skilled adult readers—and it would be natural to characterize morphological processing as involving decomposition. The "rule" of decomposition would have emerged from the sensitivity of local learning mechanisms to highly reliable regularities. Differences between the two accounts become more critical, however, when considering the prevalence of quasi-regularity (i.e., rules that are broken to varying degrees, all the way to semantically opaque words like COURTEOUS and DEPARTMENT) and nonlinearity (i.e., context-dependence of morpheme contributions) in morphological systems across languages. Even here, though, it is worth keeping in mind that graded effects can often be captured by differential and dynamic weighting among an ensemble of explicit, discrete representations, as in so-called "localist" connectionist models (e.g., Dell, 1986; McClelland & Rumelhart, 1981; Taft, 1994), although formulating an effective learning procedure for how such representations are identified, weighted, and updated remains a challenge.

In chapter 2, we consider a range of established empirical findings on morphological processing in visual word recognition, and discuss how well they align with the decomposition and distributed accounts. We conclude that while the distributed approach better captures the flexible and graded nature of morphological effects, there are also several phenomena that suggest a decomposition-like mechanism which have not been addressed by distributed simulations. Additionally, some direct predictions of the distributed account have yet to be observed empirically. As such, a good deal of research remains to be done in order to resolve these distinct accounts.

1.3 OVERVIEW OF DISSERTATION AIMS

The work presented in this dissertation deepens the field's understanding of morphological processing, both its dynamics over the time-course of visual word recognition and its emergence with reading experience, while simultaneously strengthening the distributed account of these phenomena. This is accomplished via neural network simulations, trained to map from orthographic (spelling) to semantic (meaning) representations of words, as well as empirical studies of word recognition and word learning in English. Chapter 2 provides a review of existing empirical findings on morphological processing that motivate and provide context for the current work. The specific aims of the research reported in subsequent chapters are listed below.

1.3.1 Develop a dynamic and authentic model of morphological processing in English.

To investigate how sensitivity to morphological structure can be acquired from natural language without explicit morphological information being provided, a feedforward network is trained to map from form to meaning for a large, developmentally-authentic English vocabulary (Chapter 3). Simulation of the time-course of morphological processing is explored via a recurrent excitatory-inhibitory neural network, trained on artificial morphologically-structured vocabularies (Chapter 4). Testing is conducted at the end of training, as well as at intermediate points in training for comparison with priming results from developing readers.

1.3.2 Contextualize opaque morphological priming within a distributed account.

Semantically opaque masked priming effects (e.g., seeing the word CORNER priming the recognition of CORN, although these words are not truly morphologically related) are often cited as key evidence for decomposition theories. Thus, a distributed explanation for their occurrence, particularly in morphologically impoverished languages like English, is needed. A measure of *orthography-semantics consistency* (OSC) is generalized to apply to any letter string (Chapter 5). Using this measure, a masked primed lexical decision study with skilled readers investigates the occurrence of opaque effects in word pairs whose shared letters vary in their OSC (Chapter 6). The interaction of OSC and semantic transparency is examined to determine whether opaque priming only occurs in the case of word pairs with high OSC. Additionally, a time-course study of semantic transparency effects in English (Rastle et al., 2000) is replicated to determine the robustness of these findings (Chapter 7).

1.3.3 Investigate role of orthography-semantics consistency in the emergence of morphological processing.

If morphological sensitivity is driven by regularity in the form-to-meaning mapping, as predicted by distributed accounts, it follows that sensitivity would emerge first for letter strings with more consistent meaning across their occurrences (i.e., words with higher OSC). Two studies test this prediction. First, a cross-sectional study with developing readers, spanning 4th to 9th grade, investigates which words show earliest morphological masked priming effects (Chapter 8). Second, an artificial language learning study with skilled readers examines the effect of OSC on learning new morphemes (Chapter

9). Novel words were constructed to contain either semantically consistent or inconsistent stems, and participants learned to recognize these words and associate them with particular meanings.

The dissertation concludes with a general discussion of the findings, implications, limitations, and conclusions of the reported research (Chapter 10).

1.4 SIGNIFICANCE

The presented work strengthens and extends the distributed account of complex word processing, via simulations focused on how morphological sensitivity affects complex word recognition processes, and how it is acquired. Additionally, predictions of the distributed perspective are tested, and phenomena seemingly at odds with this perspective are examined more closely. Thus, this dissertation aims to resolve theoretical controversies surrounding the mechanisms of morphological processing, to enhance our understanding of how the brain makes use of morphological information to facilitate accurate and efficient recognition of complex words.

Complex words are encountered in abundance in everyday reading, and making use of morphological information provides readers a notable advantage for recognizing and comprehending unfamiliar words. Given the centrality of morphological processing for successful word recognition and, consequently, comprehension, the lack of agreement among psycholinguists regarding how they unfold warrants close examination.

CHAPTER 2

Review of Empirical Findings in Morphological Processing

There are many reviews of the literature on morphological processing (e.g., Amenta & Crepaldi, 2012; Diependaele et al., 2012; Hay & Baayen, 2005; Marelli et al., 2020; Milin et al., 2018), in part because the relevant findings are extensive and sometimes inconsistent. The goal of the current review is to provide a comprehensive summary and theoretical discussion of empirical evidence regarding the nature of morphological processing in visual word recognition. The selection of publications to be included was delineated accordingly. First, this review focuses entirely on the visual aspects of word processing. This means that studies aiming to investigate auditory language processing or language processing that is not modality-specific (e.g., results obtained using cross-modal priming) were not included unless they provide an informative contrast with visual effects. Additionally, to focus on processing during the mapping from visual word forms to their meanings, only evidence from tasks that promote semantic access were included. Such tasks include lexical decision (determining a stimulus status as a word or nonword), determining lexical category (e.g., is this a noun or an adjective?), and sentence reading.[1] Finally, the most useful theoretical model would be one that explained not only how processing occurs but also how those mechanisms emerge with language experience. Thus, the current review will emphasize papers that highlight differences in morphological processing across languages or over the course of reading acquisition.

A rich body of empirical findings exists regarding the recognition of complex words. Below, we discuss results in light of how they contribute to understanding mechanisms of morphological processing. We start with a consideration of factors that characterize individual morphemes.

[1]Word naming was not included, due to previous findings suggesting that semantic characteristics of complex words show strong effects for lexical decision but not for word naming, while phonological characteristics play a stronger role in word naming (e.g., Baayen et al., 2007; Burani et al., 1999).

2.1 MORPHEME CHARACTERISTICS

2.1.1 Frequency

In a lexical decision task, high-frequency words are responded to more quickly than low-frequency words Forster & Chambers (1973); Scarborough et al. (1977); Shapiro (1969). Frequency effects are often interpreted as evidence that the representations of words that are seen or heard more frequently are easier to access (e.g., Morton, 1979). Taft (1979) argued that, correspondingly, if morphemes are also represented during complex word processing, a morpheme's cumulative frequency across different words should predict how quickly a reader recognizes a word containing it. To demonstrate this effect in English, Taft carefully selected prefixed words with similarly low surface frequencies[2] (e.g., RECLINE, DEMOTE) whose stems occur in other contexts with either higher frequency (e.g., in INCLINE and DECLINE) or lower frequency (PROMOTE, EMOTE). For such stimuli, readers make lexical decisions to words with high-frequency stems more quickly than words with low-frequency stems. Stem frequency effects—also referred to as base frequency effects—have been found for both prefixed and suffixed words Taft (1979), as well as for words that are inflected, derived Bradley (1979) and compounded (e.g., HEADSTAND vs. LOINCLOTH; Taft & Forster, 1976). In addition to English, they are found in Italian (Burani et al., 1984), Dutch (Baayen et al., 1997), French (Colé et al., 1989), Finnish (Kuperman et al., 2008), and Chinese (Myers et al., 2006).

Stem frequency effects also interact with words' surface frequency: high-frequency stems facilitate recognition for less-frequent words, but slightly inhibit recognition of more-frequent words Baayen et al. (2007). Along similar lines, surface frequency is less predictive of lexical decision latencies for low-frequency complex words than for low-frequency simple words Alegre & Gordon (1999). Suffix frequency has also been found to moderate stem frequency effects Burani & Thornton (2011). These results suggest that

[2] *Surface frequency* refers to the frequency of the whole word, in contrast to frequencies of the word's morphological constituents.

information at the word level and at the morpheme level supplement each other during complex word processing: Very familiar words can be recognized rapidly regardless of the identification of sublexical structures, whereas less-familiar words benefit from the facilitation that those structures provide. Such an interpretation aligns with previously described theories for the parallel roles of decomposition and listing access mechanisms (e.g., Schreuder & Baayen, 1995).

However, decomposition is not the only explanation for stem frequency effects. Burani et al. (1984) argued that such effects can be explained within a listing model, in which words are accessed directly by their surface form and not by their stem. In this model, when a word's representation is activated, the morphologically related words are subsequently activated to a lesser degree. High-frequency words will be easiest to access as a result of their regular occurrence (perhaps due to a lowered threshold of activation; Morton, 1979), but words that are morphologically related to high-frequency words will also be affected due to this lateral activation. Thus, both surface and stem frequency would impact the ease of a word's activation, with the relative import of either predictor depending on its frequency and that of its morphological cousins.

Frequency effects can be found not only for morphemes and simple and complex words, but also for short phrases (e.g., "don't have to worry"; Arnon & Snider, 2010). Given the broad range of linguistic grain sizes over which frequency matters, requiring unique representations for all informative linguistic units seems unwieldy and inefficient. The distributed approach resolves this issue by attributing frequency effects to differences in cumulative weight changes, as opposed to differences in activation thresholds of representations. Linguistic entities that are encountered more frequently have a greater impact on the learned weights of a network, and so the weights are more customized to their accurate and speedy retrieval Seidenberg & McClelland (1989). When presented with a high-frequency complex word, then, it is likely that the most influential weights are fine-tuned to that specific word. The recognition of a low-frequency word consisting of common morphemes, on the other hand, will make more use of weights tuned to sublexical structures to benefit from the word's frequent constituents.

As noted above, Baayen et al. (2007) found that higher-frequency stems actually slightly inhibit recognition of words with high surface frequency. *Inverse* stem frequency effects can also be prompted by certain lexical processing tasks: Taft (2004) found that complex words with higher frequency stems were more slowly classified than those with medium-frequency stems when nonword foils contained real stems (e.g., MIRTHS, REDLY), but not when nonwords contained nonsense stems (KOSSLED, JUXING). Amenta et al. (2015) also found inverse stem frequency effects on first fixation durations when words were read in opaque sentence contexts ("His efforts were FRUITLESS") relative to transparent sentence contexts ("The tree was FRUITLESS"). In both cases, the contexts that led to inverse frequency effects were ones in which information from the stem needed to be overridden or disregarded in order to accomplish the task at hand. Taft (2004) interpreted these effects as indication of an obligatory morphological decomposition mechanism, rather than a dual-pathway model in which words are processed either via their decomposed or surface form. Amenta et al. (2015) made a similar proposal, and furthermore posited that morpheme semantics are accessed as part of initial combinatorial processing. Within a distributed framework, mechanisms that are optimized for typical situations (e.g., stem meaning access facilitating sentence comprehension or lexical decision) can be suboptimal, and thus inhibitory, in atypical situations. These effects are stronger for higher frequency stems because the impact of such stems on processing is stronger, as described above. Distributed models that could simulate task-prompted inverse frequency effects would need to move beyond static form-to-meaning mappings to incorporate higher level semantic processes.

Much of the controversy regarding how to interpret stem frequency effects centers around when these effects occur: prior to lexical access (as suggested by decomposition models) or after lexical access (as suggested by listing models). From a distributed perspective, the answer to this question might be best summarized as “a bit of both”. Word and morpheme characteristics should determine the degree of morphologically-influenced processing occurring early on, but lateral and feedback activation that is sensitive to morphological factors occurs throughout processing as well.

The timing of frequency effects in complex word processing has been explored in experiments using masked primed lexical decision. In this paradigm, a word or nonword prime is presented extremely briefly (around 50 ms) following a mask (e.g., hash marks; #######) so as to be weakly processed but not consciously perceived. Following the masked prime and a brief delay, a second word is presented as a target for lexical decision. If participants decide on a target more quickly following a related prime than following a control prime, this strengthens the case for facilitation due to visual word processing, rather than due to conscious reasoning following recognition of the word.

Masked priming experiments in French do not detect stem frequency effects on the magnitude of morphological priming (Giraudo et al., 2016; Giraudo & Grainger, 2000). However, primes with high surface frequency do yield stronger morphological priming than primes with low surface frequency (Giraudo & Grainger, 2000). Although a comparable study in English found no such difference (McCormick et al., 2009), Amenta et al. (2020a) recently provided evidence that the effect in English may depend on orthography-semantics consistency (see section below). In particular, Amenta et al. found that high-consistency targets showed a positive effect of prime frequency on masked priming magnitude, whereas low-consistency targets showed a negative effect. This suggests that, among morphologically related words that share more semantically useful morphemes, the effects reported by Giraudo & Grainger (2000) can be found in English (which is less morphologically rich than French). Overall, these results favor listing models over decomposition models. Additional neural network simulations would be beneficial to determine whether a positive effect of prime frequency on masked morphological priming is also amenable to the distributed perspective.

2.1.2 Family size and entropy

In addition to a morpheme's frequency, the number of different words in which it appears also impacts complex word processing. Variations in *morphological family size*, or the number of compound and derived words sharing a particular stem, predict lexical decision latencies: larger families lead to faster responses (Bertram et al., 2000; Schreuder & Baayen, 1997). Family size effects can be found across a range of languages, including Dutch (Schreuder & Baayen, 1997), German (Lüdeling & De Jong, 2002), Finnish (Moscoso Del Prado Martín et al., 2004), and Hebrew (Moscoso del Prado Martın, 2003).

Schreuder & Baayen (1997) suggested that family size might be the underlying factor driving stem frequency effects, since more frequent stems are also likely to have larger morphological families. In the context of a factorial design with Dutch words, De Jong et al. (2000) did not find an effect of family frequency (frequencies of morphological relatives minus the frequency of the standalone stem) on lexical decision response latencies when family size is controlled, but did find an effect of family size when family frequency is controlled. However, Ford et al. (2010) demonstrated with a correlational design that stem frequency and morphological family size independently facilitate lexical decision responses in English, with family size effects being somewhat weaker. De Jong et al. (2002) found similar results for lexical decision with both Dutch and English compounds: family size of the left constituent (e.g., BANK in BANKROLL) predicts response latencies separately from and more weakly than family frequency. De Jong et al. (2000) found that family size estimates are stronger predictors when irregular forms, such as TAUGHT for the stem TEACH, are included; likewise, these estimates improve when semantically opaque forms, like WITNESS from WIT, are excluded (Schreuder & Baayen, 1997; Bertram et al., 2000). Thus, whereas stem frequency could be a more orthographically-driven effect, stem family size may relate more to semantic processing.

Complex word recognition also appears to be impacted by its *secondary family size*: that is, the number of words that share a morpheme with any word in the primary morphological family. For example, TROLLEY is compounded in TROLLEY CAR, TROLLEY-BUS, and TEA-TROLLEY; CAR, BUS and TEA occur in 16, 3 and 25 compound words, respectively, so the secondary family size of TROLLEY is 44 (Baayen, 2010). Secondary family size has a slightly inhibitory effect on lexical decision and word naming when the semantically dominant constituent has a small family size and the compound is not generally strongly connected to other compounds (Baayen, 2010; Mulder et al., 2014). This is presumably because the activation of information corresponding to secondary family members via spreading activation leads to semantic processes unrelated to that of the target word, which can be inhibitory if not overwhelmed by support from direct relatives.

Family size effects have also been observed cross-linguistically: Moscoso Del Prado Martín et al. (2004) and Moscoso del Prado Martín et al. (2005) found that lexical decision latencies for a word in a language known to participants could be predicted by the morphological family size of the word's translation in a second language that is *unknown to them*. This "isomorphism" effect was found between Finnish and Dutch, and between Dutch and Hebrew, both pairings of very different morphological systems (and, in the latter case, of unrelated etymology). This suggests that morphological family size may be, at least partly, a proxy for richness of the word's semantic domain: larger morphological families are more likely in rich semantic domains, shared across languages and cultures. Thus, a word that happens to have a small morphological family in English yet falls in a semantically rich domain, would be likely to have larger morphological family sizes in other languages and also be responded to more quickly. Modern distributional semantics methods could be useful in testing this explanation (see Amenta et al., 2020b).

In addition to the size of a morphological family, the relative frequencies of words within that family also impact recognition latencies. *Entropy*, a concept from information theory, describes the distribution of usage over the forms of a word and is lower if certain words within the family occur much more frequently than others. For example, the inflectional entropy of ANT is lower than that of WASP, because ANTS appears more frequently than ANT, while WASPS and WASP are used with relatively equal frequency (Baayen et al., 1995). Higher entropy is correlated with faster lexical decision reaction times in Dutch (Moscoso Del Prado Martín et al., 2004) and in English (Baayen et al., 2006). Additionally, Milin et al. (2009) found in Serbian that response latencies are slower for base words with atypical distributions of usage frequency across inflectional forms. In other words, the typicality of a morpheme's inflection usage profile relative to other nouns or other verbs facilitated recognition of a word containing it. This relative entropy effect in Serbian was replicated in a sentence reading task, suggesting that it is robust and relevant in more natural reading contexts (Baayen et al., 2011).

Family size and entropy effects favor a view of morphological processing in which a morpheme is not represented independently, but rather the other words containing it are, to a lesser degree, involved in the given word's processing. This would explain why the number of words in which a morpheme appears, and the frequencies of these words relative to each other, play a role in predicting morphological sensitivity after accounting for morpheme frequency. Such an interpretation aligns most closely with listing and distributed models. However, the nature of secondary family size effects, cross-linguistic family effects, and the contribution of irregular words to family size suggests that some such effects may come into play primarily at the level of semantic processing. If so, understanding the range of known family size and entropy effects may require a more advanced treatment of semantics than those provided in current form-to-meaning models.

That said, the only explicit accounts of family size and entropy phenomena are grounded in distributed processing mechanisms. Moscoso del Prado Martín et al. (2004) trained a neural network to produce past-tense forms from present tense for almost 3,000 Dutch verbs. The number of similar words with a particular ending (akin to family size, for suffixes) predicted how likely their model was to choose that ending for a novel word, regardless of those words' frequencies. Baayen et al. (2011) trained the Naive Discriminative Model to map from orthographic to semantic representations for 3,003 simple and derived English words, and found that stem frequency, morphological family size, affix family size and surface frequency each explained independent variance in the model's response latencies. Frequency, family size, and inflectional entropy explained independent variance after training on 2,314 simple and inflected English words, as did constituent family size and frequency, surface frequency, and compound entropy after training on 921 compound words. Baayen et al. (2011) also found weak evidence of secondary family size effects in their compound word demonstration. Together, these simulations demonstrate that morpheme frequency, family size, and entropy effects can all be explained by a single distributed processing mechanism, "without any explicit parsing process being involved" (Baayen et al., 2011, p. 49).

2.1.3 Semantic consistency

Family size effects have been shown to be particularly driven by the family members for which the stem makes a meaningful contribution Schreuder & Baayen (1997); Bertram et al. (2000). For example, when counting the family members of WIT, excluding semantically unrelated words such as WITNESS yields a stronger predictor of lexical decision latencies. This suggests that it is not simply the number of words containing the stem, but the number of words using the stem in a semantically consistent manner, that best predicts morphological facilitation.

Along similar lines, Marelli et al. (2015) devised a metric of stems' *orthography-semantics consistency* (OSC), calculated as the frequency-weighted mean cosine similarity between the stem's meaning as a standalone word and the meanings of all words containing it. For example, WHISK has a lower OSC than CHEER, because words containing WHISK (e.g., WHISKING, WHISKEY, WHISKER) are overall less semantically related to WHISK than words containing CHEER (CHEERING, CHEERFUL, CHEERY) are to CHEER. OSC captures the degree to which a simple word is semantically similar to any words containing it, making it a better metric of the simple word's consistency than of the morpheme's consistency. However, OSC explains additional variance in morphological priming magnitudes from masked primed lexical decision experiments after controlling for target family size, orthographic neighborhood, length, and frequency (Amenta et al., 2020a). Additionally, the relationship between prime frequency and priming magnitude is positive for high-consistency targets, but negative for low-consistency targets. This suggests that OSC influences the nature of priming that is occurring: morphological facilitation in the context of high-frequency primes with high-OSC stems, and orthographic indifference or slight inhibition in the context of high-frequency primes with low-OSC stems. Variance in unprimed lexical decision reaction times for standalone stems can also be partially explained by their OSC (Marelli & Amenta, 2018; Marelli et al., 2015).

Although we are not aware of any neural-network simulations of OSC effects on morphological processing, the existence of such effects follows most naturally from a distributed account. If sensitivity to morphological structure emerges during the process of learning to map from written words to their meanings, semantically consistent morphemes would be expected to influence learned weights more strongly than their less-consistent counterparts. Put another way, the presence of a consistent morpheme is more likely to be useful for the task of activating meaning than the presence of an inconsistent morpheme, and thus sensitivity to such morphemes should manifest earlier and more strongly.

The role of orthography-semantics consistency is less clear in models that rely on decomposition as a primary mechanism. As discussed in Dawson et al. (2021), more-consistent morphemes may have better consolidated and thus presumably more easily activated representations. However, the details of how less- versus more-consolidated representations differ, and the manner in which such differences impact the magnitude of morphological effects, are not well-specified.

2.2 MORPHOLOGICAL PROCESSING IN PSEUDO-COMPLEX NONWORDS

Another source of evidence on the mechanisms of complex word recognition involves the processing of nonwords (i.e., pronounceable strings of letters that are not real words). Analyses of lexical decision tasks typically focus on how quickly participants can confirm a word's lexical status. Studies that are focused on nonwords instead measure how long it takes participants to reject the target. Taft & Forster's seminal study using this approach in 1975 compared correct rejection latencies for bound stems from prefixed words (e.g., JUVENATE from REJUVENATE) with those from non-prefixed words (PERTOIRE from REPERTOIRE). Rejection latencies were slower for stems from prefixed words, suggesting greater sensitivity to the presence of stems that appear in morphologically informative contexts. Following up on this work, Taft & Forster (1976) demonstrated that participants reject polysyllabic nonwords more slowly when the first syllable is a real word (in other words, FOOTMILGE is rejected more slowly than MOWDFLISK). However, the same effect is not found for the second syllable (TROWBREAK and MOWDFLISK are rejected at comparable speeds). These findings imply the existence of position-weighted sensitivity to embedded strings in a novel word.

Subsequent nonword lexical decision studies have typically investigated morphological processing using nonwords constructed from novel combinations of real morphemes. Italian nonwords composed of a stem and a suffix (e.g., CANTEVI, analogous to a nonword like BUYED in English) show the slowest rejection times relative to a non-stem with a suffix, a stem with a non-suffix, or a non-stem with a non-suffix (Caramazza et al., 1988). Similar delayed latencies for pseudo-suffixed nonwords have been found in Finnish (Leinonen et al., 2009) and English (Crepaldi et al., 2010b). The presence of morphological effects despite the lack of established whole-word meanings suggests that morphological sensitivity is present prior to or without lexical access, supporting a decomposition view.

However, Günther et al. (2020) demonstrated that at least when the nonword is presented overtly, the presence of independent morphemes is not the sole source of interference. Using compositional distributional semantic models (Marelli et al., 2017), they quantified the semantically interpretability of morphological combinations, and found that this metric had a significant, inhibitory effect on how quickly novel German compounds composed of two real words were rejected. This finding illustrates that although morphological sensitivity does not require access to lexical representations, it does not stop at accessing constituents' independent meanings. Instead, the degree to which that morpheme *combination* is meaningful impacts how difficult it is to reject, suggesting that even in pseudo-complex nonwords, morphological information is processed with respect to its context as would be predicted by a distributed account. Semantic representations in published distributed models of morphological processing are too simplistic to capture such nuanced effects.

One challenge for evidence from unprimed lexical decision studies is that nonwords may be processed in a distinct manner from real words (as emphasized by Burani et al., 1984). If so, investigating effects for pseudo-complex nonwords in a task that encourages identification of lexical status may not provide the most relevant insights into morphological processing mechanisms.

Longtin & Meunier (2005) used the masked primed lexical decision paradigm to look more closely at the time-courses for word and nonword processing in French. Lexical decisions were made on real, simple target words (RAPIDE) following briefly presented suffixed real words (RAPIDEMENT, similar to RAPIDLY in English), pseudo-suffixed nonwords (RAPIDIFIER, like RAPIDIFY), and non-suffixed nonwords (RAPIDUIT, like RAPIDEL), relative to unrelated controls. Note that participants never made a lexical decision on, or even consciously perceived, the nonwords of interest, mitigating concerns regarding distinct nonword processes. Longtin & Meunier (2005) found that pseudo-suffixed nonwords primed their stems as strongly as suffixed real words, but found no significant priming by non-suffixed nonwords. Beyersmann et al. (2016a) also found equivalent morphological masked priming with word and nonword primes, and extended the finding to include prefixed and pseudo-prefixed primes. Masked priming experiments in English also show significant priming by pseudo-complex nonwords, similar in magnitude to that by complex words, for both pseudo-suffixed (McCormick et al., 2009) and pseudo-compound (Fiorentino et al., 2015) primes.

Contrary to the initial findings by Longtin & Meunier (2005), Beyersmann and colleagues (Beyersmann et al., 2016a; Beyersmann & Grainger, 2018; Grainger & Beyersmann, 2020) found significant non-suffixed nonword priming (RAPIDUIT) under certain conditions (also see Morris et al., 2011; De Rosa & Crepaldi, 2021). Beyersmann et al. (2016a) showed that non-suffixed nonword priming is more strongly related to the participant's word reading proficiency than other priming effects, possibly because it is a weaker phenomenon that only emerges with a great deal of word recognition practice. In Beyersmann & Grainger (2018), such effects were modulated by the family size of the stem. Grainger & Beyersmann (2020) demonstrated that non-suffixed nonword priming is stronger for stems that are more frequently followed by a derivational suffix. Thus, although the pairing of a stem and a suffix best primes the stem's recognition in nonword priming contexts, a stem followed by non-suffix letters can also be somewhat facilitatory, depending on characteristics of the stem and the reader.

The pattern of nonword priming effects varies between masked and overt priming paradigms. For example, suffixed nonword masked priming appears equally strong regardless of whether the stem-suffix was compatible (e.g., SPORTLESS, where -LESS can follow nouns) or incompatible (SPORTATION, where -ATION usually modifies verbs; Longtin & Meunier, 2005). This is not the case with similar stimuli in a cross-modal priming context: an auditorily presented suffixed nonword prime significantly primes recognition of its written stem only if the stem-suffix combination is compatible (Meunier & Longtin, 2007). In the case of compound words, novel compounds (DRUGRACK) and novel pseudo-embedded words (SLEGRACK) are similarly strong primes in a masked priming paradigm, but novel compounds facilitate significantly more strongly when priming is overt (Fiorentino et al., 2015). Contrasting results in overt and masked priming contexts demonstrate a key trend in morphological processing studies: word-level semantic properties, such as stem-suffix compatibility and lexical status, appear to matter more in contexts where the word is presented for a longer duration and is consciously perceived. This phenomenon is also found in investigations of semantic transparency effects, discussed below.

Most results from studies using nonwords align well with decomposition models. In particular, masked priming results with pseudo-suffixed nonwords demonstrate that, for at least a brief period of time, their componential structure facilitates subsequent word processing to a degree that is comparable with real suffixed words, despite having no previous surface form exposure or established meaning.

The presence of weaker non-suffixed nonword priming effects, moderated by stem derivability, stem family size and reader proficiency, tips the evidence scale somewhat away from certain models within the decomposition view. If decomposition is triggered in instances of morphological structure but not in the case of embedded words, as suggested by Rastle et al. (2004) and Longtin et al. (2003), priming by non-suffixed nonwords must be explained by a separate mechanism. Beyersmann & Grainger (2018) suggested that this type of priming proves that embedded stems can be extracted without the need for morpho-orthographic decomposition or affix splicing, by means of recognition and activation of edge-aligned words and their corresponding morphological representations. When edge-aligned words appear in real words such as CASHEW, they clarified, activation of the full word inhibits the embedded word and prevents this type of priming from occurring.

Evidence from morphological processing studies using nonwords is also compatible with the distributed account. Neural networks learn to make use of componential information as much as possible, while maintaining good performance in cases where such information is not useful (Plaut & Gonnerman, 2000). Thus, the processing of novel morpheme combinations in a manner that hinders their rejection as words and facilitates recognition of related words is plausible within this framework. Facilitation by nonsuffixed nonword primes can be explained similarly. As an added benefit, meaning conveyed by sub-lexical orthographic features that might not even be regarded as morphological (e.g., see Hendrix & Sun, 2021) also has the opportunity to impact nonword processing within a distributed account, as information at multiple grain-sizes may play simultaneous roles. However, neural network simulations of known effects with pseudo-complex nonwords have not yet been conducted.

2.3 SEMANTIC TRANSPARENCY

Semantic transparency refers to how directly a complex word's meaning can be derived from its morphological structure. Transparent words include DARKNESS and UNHAPPY because the meanings of these words can easily be inferred from their constituent morphemes. In contrast, as the meanings of RELEASE and DEPARTMENT are unrelated to those of LEASE and DEPART, respectively, these items are semantically opaque. [3] As pointed out by Marslen-Wilson et al. (1994), semantic transparency is an important lexical-level characteristic for morphological processing in English: in a cross-modal primed lexical decision task, auditorily-presented complex primes facilitate lexical decision for their stems only if the primes are semantically transparent. For example, hearing "punishment" primes recognition of PUNISH but hearing "casualty" does not prime recognition of CASUAL. Transparent priming is also stronger than opaque priming when primes are overt (not masked) and presented visually (Feldman & Soltano, 1999; Rastle et al., 2000; Feldman et al., 2004; Rueckl & Aicher, 2008).

[3] The term *semantically opaque* typically refers to complex words that were historically morphologically related to their stems, although modern usage does not make this relationship clear (e.g., COURTEOUS). In contrast, *pseudomorphological* words only appear to be morphologically structured and have no historical morphological relatedness (e.g., CORNER was never morphologically related to CORN). In this review we treat opaque and pseudomorphological words as one category, referred to as opaque, for simplicity.

When visual prime duration is brief and the word is not consciously perceived (i.e., 35-60 ms masked priming; see Forster & Davis, 1984), semantically opaque prime-target pairs exhibit significant priming. Results originally found by Longtin et al. (2003) and Rastle et al. (2004) suggest that opaque primes facilitate recognition of their pseudo-stems (CORNER-CORN) as much as transparent primes facilitate recognition of their actual stems (TEACHER-TEACH), and significantly more than orthographic primes (BROTHEL-BROTH). A subsequent meta-analysis conducted by Rastle & Davis (2008) found that this pattern of results held across 19 masked morphological priming studies in Indo-European languages. Researchers concluded from these findings that word forms with apparent morphological structure are treated similarly to actual complex words during initial visual processing, supporting morphological decomposition as an early word processing phase prior to recognition. In other words, interpreting the opaque masked priming effect as evidence that CORN is represented briefly when the word CORNER is shown implies the existence of a semantics-blind morphological decomposition mechanism (Rastle & Davis, 2008). Similar patterns of effects have been shown in English with derived-derived priming (Feldman & Soltano, 1999), stem-derived priming (Marslen-Wilson et al., 2008), and compound word priming (Fiorentino & Fund-Reznicek, 2009).

As Rastle et al. (2004) found in English, readers in French and Serbian show equivalent morphological priming with transparent and opaque primes when the prime is masked, but transparent priming is stronger when the prime is overt (Longtin et al., 2003; Feldman et al., 2002). Complex Chinese words with a common morpheme that makes either a similar semantic contribution (e.g., the words for "public park", and "the public", for which the same character conveys "public" in both cases) or a distinct semantic contribution ("public park" and "rooster", sharing the same character) showed equivalently strong facilitation in masked, but not overt, priming contexts (Tsang & Chen, 2013). Stem homographs in romance languages, such as CERRO ("hill") and CERRAR ("to close") in Spanish, appear to be forms of the same stem, but they are semantically and morphologically unrelated. Such words inhibit each others' recognition relative to merely orthographically related words (CERDO, "pig") when the prime is overt (Laudanna et al., 1989), but are facilitatory relative to orthographic when the prime is masked (Badecker & Allen, 2002). Masked priming studies in Dutch and Bangla found opaque priming to be greater than orthographic, but significantly weaker than transparent priming, although these results may be attributable to slightly longer masked prime durations (Diependaele et al., 2009; Dasgupta et al., 2015). In the cases of Hebrew (Frost et al., 2000), Arabic (Boudelaa & Marslen-Wilson, 2001), and German (Smolka et al., 2009), facilitatory opaque morphological priming continues to be strong even when the prime is overt[4]. Cross-linguistic differences in semantic transparency effects will be discussed in more depth later in this review. Generally speaking, studies in languages other than English have also found significant evidence of facilitatory opaque priming beyond what would be expected from orthographic similarity, particularly in a masked priming context. The remainder of this section focuses on English experiments, as a great deal of the discussion regarding opaque priming is grounded in these results.

Initial results in English (Rastle et al., 2004) suggested that facilitation due to opaque morphological priming is not only greater than orthographic priming, but equivalent in magnitude to that of transparent priming. Some subsequent studies have found similar results: for example, Marslen-Wilson et al. (2008) found no significant difference in facilitation by transparent, quasi-transparent, and opaque priming, and this held for prime durations of 36, 48 and 72 ms. Lavric et al. (2007) and Lavric et al. (2012) found ERP evidence for similar initial treatment of opaque and transparent stimuli. However, some studies have found effects from opaque primes to be significantly weaker than those from transparent primes. Morris et al. (2008) observed stronger effects, both behaviorally and with ERP measures, for transparent primes relative to opaque (next strongest) and ortho-

[4]Semantically opaque morphological priming by unmasked primes has been found in English (Libben et al., 2003) and Dutch (Zwitserlood, 1994), in the context of constituent priming during a lexical decision task with compound words. Thus, the type of complex stimuli and task in question may also impact whether opaque morphological priming effects are observed within a language.

graphic (weakest). Similarly, Diependaele et al. (2011), Andrews & Lo (2013) and Feldman et al. (2015) all found intermediate effects of opaque priming relative to transparent and orthographic priming. Such differences may be found inconsistently in part because they are small: in the meta-analysis conducted byRastle & Davis (2008), 8 of the 12 studies in English that included a transparent condition showed numerically stronger priming for transparent than for opaque conditions, suggesting that perhaps this effect is present but not easily detected statistically.

Similar effect sizes for opaque and transparent priming may also be due to confounding factors across the prime-target pairs in these conditions. Feldman et al. (2009), finding significantly stronger priming for transparent than opaque pairs when affixes are matched across conditions, posited that lack of accounting for affixes during stimulus selection may contribute to confusing results. These authors suggested that, in multiple studies, a disproportionate use of affixes with low productivity (-ILE) or an inconsistent semantic role (-ER; as in GREATER versus TEACHER) in the transparent condition may have decreased those words' priming, making them more comparable to opaque priming magnitudes[5]. Another potential confound in studies of transparency effects in lexical decision was pointed out by Marelli et al. (2015), relating to the orthography-semantics consistency (OSC) of target words. They demonstrated that the targets used in several previous masked priming studies of semantic transparency differ significantly in OSC across conditions, and argued that this drives commonly observed faster response times to transparent than to opaque targets when preceded by unrelated primes. In other words, lower-OSC words like WHISK, often used in opaque conditions following WHISKER, are responded to more slowly than higher-OSC transparent targets like CHEER (primed by CHEERFUL), regardless of what prime precedes them. It is possible that, because the targets in transparent conditions are already being recognized quickly, the degree to which

[5]However, as pointed out by Davis & Rastle (2010), the results of Feldman et al. (2009) are also in question due to a higher incidence of non-morphological orthographic transformations from prime to target (e.g., BLISTERY-BLISS and COYNESS-COIN) in the opaque condition than in the transparent condition.

their recognition could be further facilitated is reduced, constraining transparent priming to a level more comparable with opaque priming.

Jared et al. (2017) provided perhaps the most informative recent empirical contribution to the debate over early transparency effects in English. Using the same masked primed lexical decision paradigm implemented in many prior studies, they consistently found graded effects of semantic transparency across six experiments, using both behavioral and ERP measures. In their design, each condition consisted of substantially more prime-target pairs than were used in prior studies investigating semantic transparency (104, compared to the typical 40 to 50), and a quasi-transparent condition (e.g., DRESSER-DRESS, BOOKISH-BOOK) was included to examine intermediate transparency effects. Two notable discrepancies across previous studies' methods—relatedness of primes for nonword distractors and masked prime duration—were addressed and tested. Neither of these factors changed the pattern of results: for both ERP and behavioral results, transparent masked primes showed significant facilitation across all experiments, and opaque effects were weak and never significantly stronger than the orthographic control. Quasi-transparent effects fell between those of transparent and opaque conditions in magnitude. Jared et al. (2017) also replicated ERP evidence of graded transparency effects with a masked primed *semantic* decision task ("Is this a sea creature or not?"), and demonstrated that color boundaries between morphemes yielded a boost in ERP signals only for the transparent and quasi-transparent conditions[6]. Jared et al. (2017) not only repeatedly demonstrates the effect of semantic transparency on masked morphological priming magnitudes, but also suggests that opaque priming in English may be notably weaker than previously suspected, or possibly nonexistent. Similar results were found recently by Chee & Yap (2022); for both semantic categorization and lexical decision tasks, opaque

[6]In the *illusory conjunction* paradigm, a word is presented partly in one color and partly in another: the boundary between the colors either aligns with a boundary between components of proposed psychological import (e.g., morphemes) or not. If an advantage in recognition is observed when boundaries align, the components are thought to be cognitively represented (see Rapp, 1992, for another example of this paradigm being used to investigate complex word processing).

masked priming was found to be no stronger than orthographic priming, and significantly weaker than transparent priming.

Despite some counterexamples and potential confounds, opaque priming has been found in a wide range of studies in English. Thus for the subsequent discussion of these effects with respect to decomposition and distributed theories, we will presume that opaque morphological facilitation in English masked priming studies is significantly greater than orthographic priming, though weaker than transparent morphological priming.

In decomposition theories of morphological processing, semantically opaque priming is interpreted as evidence that early in visual word recognition, morphological decomposition is applied to all words that have the appearance of being complex Rastle & Davis (2008). Furthermore, these effects have been cited as evidence against the distributed view, characterizing distributed models as only being able to explain morphological effects for words that are semantically transparent (e.g., Rastle & Davis, 2008; Duñabeitia et al., 2007; Beyersmann et al., 2012b). However, simulations relying on distributed mechanisms do provide explanations of opaque effects. Using the Naive Discriminative Model, Baayen et al. (2011) simulated transparent and opaque priming effects that were not significantly different, using the same stimuli employed in Rastle et al. (2004). However, this result relied on how these items were represented when input to the model: Baayen et al. (2011) argued that for several of the opaque primes used in Rastle et al. (2004), the suffix plays a functional role in the words' meaning (e.g., -ER in ARCHER denotes "one who", with ARCH etymologically referring to a bow). The orthographic inputs of prime words for which Baayen et al. (2011) judged this to be a concern were thus associated with the meaning of the whole word (ARCHER) and that of the suffix (-ER), and those words drove the strong facilitation observed in priming simulations for the opaque condition. Although the transparency of certain prime affixes may strengthen the opaque priming magnitudes found in some studies, Beyersmann et al. (2016b) showed that this argument was insufficient to entirely account for opaque priming. Using a much more tightly controlled set of opaque prime-target pairs for which no affix or stem could be argued to be contributing semantically to the surface meaning, they still found opaque priming to be significantly greater than orthographic priming, and not significantly different than transparent priming.

Plaut & Gonnerman (2000) provided an alternative distributed explanation of why opaque priming occurs. After a neural network was trained on a morphologically rich artificial language (likened to Hebrew), it reached stable semantic representations for target inputs faster when preceded by an opaque morphological prime than an orthographic prime— in other words, opaque priming was observed to be significantly stronger than orthographic priming. The network's learned weights that handle opaque words' inputs were more strongly influenced by the input's components due to the strong role of morphological structure present in the rest of the language. Sensitivity was not just acquired for particular stems in this case, but for morphological structure in word forms more generally. For the morphologically impoverished language, in which no morphological structure existed save in the testing words shared with the rich language, Plaut & Gonnerman (2000) did not find significant opaque priming. In this case, learned processing of opaque words was less affected by their constituent morphemes, as morphology had less impact on the overall learning of the language and thus on the manner in which word forms, regardless of their transparency, were processed.

In retrospect, considering an artificial language almost entirely lacking in morphological structure to be analogous to English was an exaggeration, and thus the fact that the impoverished artificial language did not show opaque priming does not preclude an account of subsequent findings of opaque priming in English (Rastle et al., 2004). Presumably, opaque priming can be expected to occur with a degree of strength or robustness corresponding to the language's morphological richness, and English may be sufficiently morphologically rich to account for opaque priming that has been observed in masked priming contexts. If this is the case, Plaut & Gonnerman (2000) does provide an explanation of opaque priming, even in English, via distributed processing, and additionally explains the graded transparency effects found by Jared et al. (2017) and others.

The fact that opaque priming is often only observed in masked priming contexts, not when the prime is overt, is sometimes interpreted as evidence that the effect occurs during faster, earlier phases of word processing; however, this interpretation implies that processing is halted the moment the stimulus changes, which does not align with a general understanding of visual processing. As argued by Tzur & Frost (2007, p. 323), masked priming duration might be better interpreted as "the amount of energy that is provided to the cognitive system for the perception and identification of the stimulus". If word presentation initiates a wave of propagated activation, the duration of the prime could determine the strength of that activation. Consistent with this view, Tzur & Frost (2007) demonstrated that both the luminance of the prime and the amount of contrast between the prime and the background yield variability in effects comparable to that produced by manipulating prime duration. For example, a 20 ms prime with brighter luminance showed similar effects on lexical decision latencies as a 40 ms prime presented more dimly. From this perspective, graded semantic transparency effects as shown by Jared et al. (2017) and others could be interpreted, not as semantic, word-level processing happening "early", but as weak lexical input being sufficient to activate semantic information to some extent. Likewise, opaque effects during masked but not opaque priming in English can be explained as the weak influence of the language's overall morphological structure on how the word form is processed, an influence which is overwhelmed when the signal is strong enough to more thoroughly activate semantic and surface-form level information.

To be clear, under this interpretation of masked priming effects, opaque and graded transparency effects in masked morphological priming are not inconsistent with morpho-orthographic decomposition. It could be that words with apparent morphological structure (pseudo or actual) are in fact split along morphemic boundaries regardless of those boundaries' semantic significance, but that this purely morpho-orthographic phase is too brief to be detectable in the absence of semantic feedback effects. This claim would be difficult to falsify. The distributed account, however, provides an alternative explanation: that simultaneous influence of both word-level and morpheme-level characteristics occurs even during early morphological processing, with the relative weighting of each dependent on the word, morpheme, and language in question. Thus, instead of providing evidence against the distributed view, opaque morphological priming can be explained by either perspective, while evidence of graded transparency effects in even masked priming contexts is more simply accounted for by the distributed view.

As a final note regarding semantic transparency, in favoring a view of morphological processing that does not regard morphemes as functional units, the distributed view provides an alternative means of quantifying and conceptualizing semantic transparency. Instead of viewing the contributions of constituent morphemes, particularly affixes, in terms of a linear addition (e.g., the meaning of TEACHER = the meaning of TEACH + the meaning of -ER), affixes' semantic role can be conceptualized as functions which transform the meaning of the stem in typically predictable ways (Marelli & Baroni, 2015). Once these functions are learned for each affix, compositional meanings for derived words and novel words can be constructed to depict what the word's morphology conveys about its meaning (e.g., the compositional meaning of SUMMER is "one who sums"). Furthermore, the similarity between the word's actual meaning, represented via traditional distributional semantics techniques (e.g., Landauer & Dumais, 1997), and its compositional meaning can be construed as a new measure of semantic transparency (Marelli & Baroni, 2015). Using compositional semantics techniques on compound words, Günther & Marelli (2018) demonstrated that a higher-dimensional treatment of semantic transparency yields better predictors of lexical decision latencies (also see Günther et al., 2020). The construction and comparison of compositional meanings has promise as an improved approach to understanding the role of semantic processing in learning and recognizing complex words (Amenta et al., 2020b; Baayen et al., 2019). Such tools can also be applied to quantifying the overall transparency of an entire language's morphology (Günther et al., 2019).

2.4 TASK EFFECTS

The predominant task used across morphological processing studies is the lexical decision task. The use of a standard task allows for ease of comparison across many studies' results, an advantage exploited in this paper and other reviews. However, variation in task is essential for determining what aspects of morphological processing are unique to, or at least more common for, a specific paradigm. Marelli et al. (2013) and Duñabeitia et al. (2011) both demonstrated, with tasks emphasizing more semantic and orthographic aspects of word processing, respectively, that morphological effects are quite sensitive to the goal of the reader.

Marelli et al. (2013) used a variation on the masked priming paradigm in which, following a morphologically transparent, morphologically opaque, or orthographic prime, a word is presented where the reader is fixating. Simultaneously, a number is shown elsewhere on the screen. The participant is motivated to recognize the presented word as quickly as possible so as to have time to look at the number. After some trials, a prompt appears asking about the most recent word or number (e.g., "Is it a tool?" or "Is it even?"), requiring that the participant access and retain semantic information about the stimuli. In this context, transparent priming significantly shortens word fixation duration, but opaque and orthographic priming do not. However, opaque priming is detected in a lexical decision paradigm with the same stimuli (Marelli et al., 2013).

Duñabeitia et al. (2011) investigated both transposed-letter and semantic transparency effects in the context of "same-different" judgments: participants are presented with two words (a reference and a target) and are asked to determine if they are the same or different. In the context of a lexical decision task, Duñabeitia et al. (2007) had found that masked primes with two transposed letters (e.g., AMNOG from AMONG) prime the original word more strongly than do masked primes with two replaced letters (AMELG from AMONG), but that this effect goes away when the transposed letters cross a morphological boundary (WALEKR). In the context of a same-different task, however, the transposed-letter effect does not weaken when crossing a morphological boundary (Duñabeitia et al., 2011). In a second study, same-different judgments were paired with transparent, opaque and orthographic masked primes, comparable to Rastle et al. (2004) and Jared et al. (2017). Same-different judgments yielded similarly strong priming across all three conditions, in contrast to the pattern of decreasing magnitudes found using lexical decision.

Marelli et al. (2013) and Duñabeitia et al. (2011) both demonstrate that sensitivity to morphological structure during word processing varies with the nature of the task. Critically, opaque and orthographic effects are undetectable when necessary information is semantic in nature ("Is it a tool?") whereas they are close in magnitude to transparent effects when orthographic information is needed ("Are these words the same?"). This suggests that the processes giving rise to decomposition phenomena are not applied indiscriminately, as stated in Rastle & Davis (2008), but are sensitive to the needs of the context (although some contexts can yield similar results; see Chee & Yap, 2022). In a distributed account, these influences would arise via top-down feedback from higher-level task representations (e.g., Cohen et al., 1990; Gilbert & Shallice, 2002), although such effects have not been simulated in a morphological processing context. Task-varying effects could also be accommodated in a decomposition model—although perhaps somewhat less naturally—by allowing for alternative processing routes whose use varies with context. Kuperman et al. (2008) made a proposal along these lines to account for eye tracking data recorded while reading compounds (also see Kuperman et al., 2009): in their model, multiple routes simultaneously processing different types of information (compound frequency, right constituent family size, etc.) interact with each other and impact behavior to the degree necessary for efficient recognition in the given context.

2.5 ORTHOGRAPHIC FLEXIBILITY AND SPECIFICITY

The orthographic constructions of complex words vary widely, both within and across languages. In some cases, the manner in which an orthographic form might be decomposed into its constituents is easy to imagine: In English, SINGER can be split into SING and ER. In Chinese, a two-character word can be split into its constituent characters. Even in languages with nonconcatenative morphologies, such as Arabic and Hebrew, word forms can typically be separated into interleaving letter subsets corresponding to each morphological component (e.g., TIZMORET into root ZMR and pattern TI_O_ET; Frost et al., 1997). However, words for which morphemes are not easily split apart illuminate the role of orthographic form in morphological processing. Complex words that cannot be perfectly split into their morphological constituents are not merely exceptional cases, to be viewed in contrast to an easily-split norm. In the English language, for example, around 39% of complex words fall into this category (Baayen et al., 1995). Studying such cases is essential for getting a complete picture of morphological processing.

It has been demonstrated in several experiments that morphological sensitivity is robust to lost, added, and replaced letters. For example, some derivations in Setswana are formed using letter replacement as well as concatenation (Kgolo & Eisenbeiss, 2015). MOREKI, "buyer", and THEKO, "a manner of buying" are derived from REKA, "to buy", and MORERI, "preacher", and THERO, "sermon", are derived from RERA, "to preach". Although it's not clear how these derivations would be divided into stem and affix if decomposed, these words still show significant morphological masked priming in lexical decision experiments (Kgolo & Eisenbeiss, 2015). In English, ADORABLE primes ADORE despite a missing -E-, LOVER primes LOVE despite a shared -E-, and DROPPER primes DROP despite the added -P- (McCormick et al., 2008). Importantly, semantically opaque words with similar transformations such as BADGER-BADGE and FETISH-FETE also show morphological priming despite not being perfectly segmentable into stems and affixes.

To accommodate orthographic flexibility in a decomposition view of morphological processing, McCormick et al. (2008) proposed that stem representations activated during the morpho-orthographic decomposition phase are underspecified (also see Taft, 1979). In other words, mutable letters like E at the end of LOVE are missing or optional in stem representations so as to be compatible with the decomposition of LOVING, whereas the M at the end of DRUM can be optionally duplicated to map onto DRUMMER and DRUMMING. Although this theory fits some English results nicely, it is challenged by more complicated morphological transformations such as the Setswana examples above. The orthographic transformation needed to generate THEKO from REKA and THERO from RERA would result in REKA and RERA's underspecified representations consisting of only -EK- and -ER-, respectively, which would likely risk confusing these stems with other linguistic entities and eliminate orthographic information (R-) that is useful for other derivations (e.g., MOREKI and MORERI). A theory reliant on underspecified stem representations would benefit from context-sensitivity when applied to Setswana, such that the flexible interchanging of a beginning R- with TH- is contingent upon the interchanging of an ending -A with -O.

The instances in which morphological operations (e.g., dropping -E and adding -ABLE) do *not* appear to generalize can also inform our understanding of morphological processing mechanisms. For example, idiosyncratic inflections, such as FELL from FALL, show masked priming effects, but word pairs which follow that orthographic transformation but are not morphologically related do not (e.g., BELL-BALL; Crepaldi et al., 2010a). This finding has two implications: first, morphological masked priming for FELL and FALL can't be purely driven by the appearance of morphological relatedness (or the same effect would occur for BELL and BALL). Second, sensitivities to morphological transformations that occur only in one or very few contexts aren't learned to the extent that they impact opaque word processing. Perhaps if the past tenses of ENTHRALL, STALL, and APALL were ENTHRELL, STELL and APELL, a priming effect would be found between BELL and BALL.

Weak and defective roots in Hebrew provide another informative instance where morphological priming does not occur (Velan et al., 2005). Most derived forms of weak roots in Hebrew contain only two of the three formal root consonants, whereas derived forms of defective roots contain the third consonant sometimes, but not always. In a masked primed lexical decision paradigm, derivations of weak roots are facilitated by the two letters that appear across all forms, but not by the three letters of the formal root. Conversely, recognition of defective-root derivations are facilitated by both two- and three-letter primes, even if the target contains only two of the three root letters. These findings imply that morphological processing is sensitive to the variability of morphological family members' orthographic forms.

Customized orthographic flexibility is also seen in the context of position-specificity. Although morphologically structured nonwords have been found to be rejected more slowly than other nonwords, inverted suffixed words like NESSKIND are rejected as quickly as suffix-less forms like NELSKIND (Crepaldi et al., 2010b), suggesting that suffixes are processed as morphologically significant only if they appear at the end of a word. Similarly, although BEGNESS is rejected more slowly than non-suffixed BEGNUSS, the inverted NESSBEG is not rejected more slowly than NUSSBEG (Crepaldi et al., 2010b). Position-specific effects are also found for both prefixes and suffixes in Spanish nonwords (Carden et al., 2019). On the other hand, Taft (1985) showed that transposed English compound words (e.g., DROPRAIN) were rejected more slowly than other pseudo-compound nonwords in a lexical decision task, suggesting that morpheme representations may be less position-sensitive when the morphemes can occur in varying positions or as standalone words (e.g., DROP in RAINDROP, GUMDROP, DROPLET, DROPPER). Similarly, Duñabeitia et al. (2009) found masked priming between Basque compound words whether the shared constituent was in the same position (LANORDU and LANPOSTU) or in a different position (SUMENDI and MENDIKATE). Thus, morphological processes are position-specific where that is sufficient (e.g., suffixation and prefixation) but can be more flexible for morphemes that appear in a variety of positions.

Overall, morphological sensitivity persists despite changes in a morpheme's appearance. Even opaque priming is robust to orthographic alterations in some cases, but only if such changes are common across derivations or for that particular morpheme. To account for orthographic flexibility of morphological processing effects, Crepaldi et al. (2010a) proposed a variation of the decomposition view in which lemma representations are activated following morpho-orthographic representations. Inflected variations activate the same lemma, whereas derived forms have distinct lemmas due to their more varied semantic and syntactic roles. Within their model, irregular inflected priming, such as that seen for FELL and FALL but not BELL and BALL, can be attributed to FELL and FALL activating of the same lemma despite having different morpho-orthographic representations. The lemma modification explains the described results in English, but does not explain the uniquely flexible priming of defective roots in Hebrew or differences in position-specificity across morphemes. Carden et al. (2019) suggested that differences in position-specificity are due to inherent differences in affix and stem representations (see also Grainger & Beyersmann, 2017). However, the nature of those differences are underspecified. Additionally, it is unclear where prefixes that are more stem-like in nature (e.g., PSYCHO- and ANTHRO-) would fall within this framework.

In a distributed account, additional features or phases are not needed to account for the orthographic flexibility of morphological processing effects. More complex and abstract operations are learned by a neural network in contexts in which they are frequent and useful, such as dropping Es and doubling consonants in English. For simpler instances in which straightforward orthography-semantics regularity is present (e.g., the single case of FELL and FALL, or among Hebrew words with a root that always contains all three consonants), this is all that is learned. The complex-when-necessary learning style of neural networks aligns well with the empirical results surrounding orthographic flexibility of morphological processing, while also providing a mechanistic account of how these effects might emerge with reading experience. Both Rumelhart & McClelland (1986) and Baayen et al. (2011) used neural network simulations to demonstrate irregularity effects in morphological processing, showing that these models can account for such phenomena without the need for separate routes. However, neither paper captures the mapping of orthographic representations to distributed meaning representations: Rumelhart & McClelland's 1986 network trained on production of the past-tense instead of semantic access, and Baayen et al.'s 2011 network used localized lemma representations as output. Additionally, both simulations only examined inflected words in isolation. More simulations using neural networks are necessary to complete the distributed account of orthographic flexibility in morphological processing. Simulating masked priming effects such as those presented by Crepaldi et al. (2010a), Velan et al. (2005) and McCormick et al. (2008) would be particularly enlightening.

Difficulties with dividing complex words into morphological constituents have prompted several prominent linguists to reject the morpheme as a useful unit of meaning. Anderson (1992) emphasized the lack of isolable morphological constituents in contexts such as reduplication in Tagalog (e.g., MAGLALAKBAY from MAGLAKBAY and PAGBUBUKSAN from BUKSAN) and apophonic relations in English (SANG from SING and DOVE from DIVE). He argued for an approach to morphological structure focused on relational processes rather than components. For example, instead of viewing the A in SANG as an indicator of past, past is associated with the *process* of changing the I in SING to an A. Similarly, Blevins (2016) advocated for the "Word and Paradigm" view, stating that morphological structure provides information via patterns of change across words, not via separable contributing constituents. The difficulty of the morpheme unit with respect to explaining the nature of complex word forms has driven linguists as well as psychologists to pursue more flexible accounts.

2.6 EMERGENCE AND DIVERGENCE OF MORPHOLOGICAL PROCESSING MECHANISMS

As demonstrated in the sections above, morphological processing effects are found to be very sensitive to the morpheme's context and its general usage in the language in question. The pervasive importance of morpheme usage characteristics, from orthography-semantics consistency to inflectional entropy, highlights the need for a better understanding of how the statistics of language drive the emerge of these mechanisms through experience. Broadening our scope to consider how morphological processing effects differ across languages further underscores this need: How does the visual system of a reader adapt to the distinct morphological systems of a certain language and develop appropriate processing mechanisms? Below, we review salient differences in morphological processing known to exist across languages, as well as existing evidence concerning morphological processing development in both young readers and second-language learners.

2.6.1 Differences across languages

Although cross-linguistic in its scope, to this point our review has focused on characteristics of morphological processing that are similar across languages; however, noting where processes differ can provide essential insights into how language characteristics shape these mechanisms.

As mentioned previously, in some languages semantically opaque morphological priming persists even when the prime is consciously perceived. Bentin & Feldman (1990) and Frost et al. (2000) showed in Hebrew that when morphologically related primes were presented overtly, lexical decisions for visual targets were significantly facilitated even if the prime wasn't semantically related to the target. In fact, opaque priming wasn't even significantly weaker than transparent priming in this context. Similar results have been found in Arabic (Boudelaa & Marslen-Wilson, 2001). If the effect were found only in these two languages, one might attribute its emergence to a unique characteristic of the Semitic language family, such as embedded morphology; however, German participants also show strong opaque priming by an overt visual prime (Smolka et al., 2015).

The language characteristics driving such differences in semantic transparency effects have been discussed and assessed in some depth. An idea previously put forward by Plaut & Gonnerman (2000) attributed overt opaque priming in certain languages to the greater morphological systematicity of those languages. In other words, languages in which a word's meaning can usually be inferred directly from its morphological structure might show greater opaque morphological priming, because broadly applied morpho-orthographic sensitivity would, in this context, have more advantages than drawbacks. Günther et al. (2019) set out to quantify morphological systematicity in both English and German, to determine whether it differed sufficiently to account for the contrasting transparency effects. Using a distributional semantics approach (Landauer & Dumais, 1997), they calculated "observed" meanings for complex words based on what other words they co-occurred with in a large text corpus for each language. Linear functions to capture how word meanings are transformed when a derivational affix (e.g., -NESS or -ITY) was added to a free-standing stem (e.g., DARK or PROSPER) were also approximated, and used to calculate what the "compositional" meaning of each derived word would be (i.e., the compositional meaning of DRESSER might be roughly "one who dresses"). Comparisons of compositional and observed semantic vectors showed evidence of a more systematic morphology in German relative to English, supporting the hypothesis of Plaut & Gonnerman (2000). This is only one example of how variations in morphological systems across languages can shape how readers process complex words.

Differences in the productivity of a language's morphology also appear to impact how complex words are processed. A cross-linguistic study comparing lexical decision for derived words found greater sensitivity to stem frequencies in English than in Finnish

(Vannest et al., 2002). Given the rich and productive morphological system in Finnish, this pattern of stem frequency effects, often interpreted as a measure of morphological sensitivity, is initially surprising. However, the large number of words containing more than one affix in Finnish, resulting in many multimorphemic words and large morphological family sizes, may explain this result. For example, the suffix -TÖN in Finnish is followed by an additional suffix in 87.5% of its occurrences, while the suffix -ABLE in English is followed by an additional suffix in only 11.8% of its occurrences (Vannest et al., 2002). Moscoso Del Prado Martín et al. (2004) demonstrated that for languages in which morphological family sizes are large, statistics for subsets of the family that are more semantically similar predict processing effects better than those of the whole family. For example, the word TYÖTÖNMYYS ("unemployment") might be more informatively considered relative to other words beginning with TYÖTÖN ("workless" or "unemployed"), not all words beginning with TYÖ ("work"). Thus, TYÖTÖN is functionally a more relevant "stem" than TYÖ in this context. Finnish words with a single derivational suffix, such as those used for lexical decision by Vannest et al. (2002), may be perceived in a manner more similar to monomorphemic stems in English, explaining their weak stem frequency effects.

One final cross-linguistic difference that is not yet well understood can be found in the morphological processing of dyslexic readers. Dyslexic readers in orthographically transparent languages (in which letter-to-sound correspondences are consistent), such as Italian and Spanish, appear to take advantage of morphological structure to support their word recognition. Conversely, in more orthographically opaque languages, such as English and Chinese, dyslexic readers show a deficit in morphological processing (see review by Deacon et al., 2019). This finding is somewhat surprising, as cases where spelling-to-sound mappings are less clear in written English words often correspond to a boost in spelling-to-meaning regularity (Rastle, 2019a). For example, JEALOUS, NERVOUS, BONUS, and NECKLACE all end in the same sound, but the -OUS spelling uniquely conveys adjective status (Berg & Aronoff, 2017). As argued by Rastle (2019b), sensitivity to these meaningful variations in spelling may be particularly essential for reading acquisition in English, yet dyslexics do not generally appear to benefit from it. One explanation for this relies on grain-size theory (Ziegler & Goswami, 2005; Law & Ghesquière, 2021; Marelli et al., 2020). For languages in which small visual units (one or two letters) reliably map to speech sounds, developing readers' progress towards direct mapping of form to meaning requires a large jump in grain-size (letter to word). Morphemes may provide an efficient intermediate stepping stone for readers in these languages, but less so in orthographically opaque languages in which the grain-size for phonological decoding is already large. This wouldn't imply that morphological information does not play an important role in English reading acquisition, but rather that it is not salient to dyslexic readers in the same way that phonological information is not. Understanding why the role of morphology varies for dyslexic readers across languages would enhance our understanding of the emergence of morphological processing and how it interacts with other well-studied aspects of word processing.

Across all three of these examples, it is clear that the manner in which complex words are processed is customized to the language in question. As a given individual might have been born into any linguistic community, a complete theory of complex word recognition processes must account for these cross-linguistic differences in terms of how language characteristics shape morphological representations and processes during language and reading acquisition. As illustrated by Plaut & Gonnerman (2000), neural network simulations are uniquely well-suited to addressing such questions.

2.6.2 Morphological processing during reading acquisition

As children learn to read, they come to know an increasingly large set of words by sight (Ehri, 2005). With reading experience, even their earliest learned orthographic representations continue to evolve and become more specific. For example, Castles et al. (2007)

found in a longitudinal study that although 3rd graders showed orthographic priming of real words (e.g., PLAY) from both replaced-letter (RLAY) and transposed-letter masked primes (LPAY), they showed only transposed-letter priming two years later. This suggests increasing specificity of orthographic representations, moving towards no orthographic priming in either condition, as is seen in adults (Castles et al., 2007; Castles, 1999).

Early readers appear to be sensitive to the presence and meaning of embedded words. Seven-year-olds reading in English are slower to reject words as belonging to a certain category (e.g., "body parts") if the words have embedded strings that pertain to that category, such as SHIP (containing HIP) and CLIP (containing LIP; Nation & Cocksey, 2009). Sensitivity to embedded strings, potentially a byproduct of less precisely specified orthographic representations, could play an important role in the detection and exploitation of orthography-semantics regularities as the system is refined. Hasenäcker et al. (2021) extended this finding to Italian children (ages 8 to 11 years) and adults, and additionally investigated the effect of embedded strings on category rejection for both words with suffix and non-suffix endings (BURRONE versus RAPACE, comparable to CORNER versus PEACE in English). They found that participants at all ages show similar effects of the embedded string in slowing rejection latencies, regardless of the word's ending, but interestingly younger children are more likely to make errors for words with affix endings.

Indeed, children develop sensitivity to the morphological structure of words quite early. In lexical decision tasks, early readers (ages 7 to 10) are less accurate in rejecting morphologically structured nonwords relative to non-suffixed or non-morphological nonwords in Italian (Burani et al., 2002), English (Dawson et al., 2018), and French (Casalis et al., 2015). McCutchen et al. (2009) found in an overt priming task that 5th and 8th graders recognized English words (e.g., PLAN) more quickly following a morphologically related word (PLANNER) than following one that was only orthographically (PLANET) or semantically (STRATEGY) related. French 4th graders, similarly, showed stronger morphological priming relative to orthographic priming in an overtly primed lexical decision task (Casalis et al., 2009). When primes were presented more briefly (but still visible; 75 ms), morphologically and orthographically related primes facilitated children's lexical decision latencies equally, suggesting that recognition of morphological relatedness at this phase of reading acquisition takes longer and is perhaps more reliant on semantic similarity (Casalis et al., 2009).

Developmental masked priming studies that manipulate semantic transparency (i.e., including opaque primes such as CORNER) reinforce the importance of semantic factors in the development of morphological priming. In English, Beyersmann et al. (2012a) found that 3rd and 5th grade readers demonstrated transparent but not opaque morphological priming in a masked primed lexical decision task, whereas adult participants showed significant priming in both conditions. Dawson et al. (2021) replicated this finding with participants across a broader range of ages to show the gradual emergence of opaque effects with increased word reading proficiency. Likewise, in Hebrew, transparent morphological primes facilitate lexical decision for both 4th and 7th graders, whereas only 7th graders showed a marginal opaque morphological priming effect (Schiff et al., 2012). In German, Fleischhauer et al. (2021) found no masked morphological priming in 1st and 2nd graders, only transparent morphological priming in German 3rd graders, and both transparent and opaque priming in 4th graders and adults. Thus, in English, Hebrew, and German, opaque morphological priming appears to emerge later than transparent priming, and only after a great deal of reading experience.

In contrast, both Quémart & Casalis (2015) and Quémart et al. (2011) found evidence for opaque morphological priming effects in relatively young French children. Quémart & Casalis (2015) found equally strong transparent and opaque masked priming effects for typically developing French readers in 4th and 8th grade, although dyslexic 8th graders showed only transparent priming effects. In Quémart et al. (2011), French children in 3rd, 5th, and 7th grade showed opaque priming (e.g., BAGUETTE, "breadstick", priming BAGUE, "ring") for both short (60 ms) and long (250 ms) prime durations, whereas adults showed such priming only for the short duration. Their results suggest that French children may acquire sensitivity to morphological form separate from meaning earlier than children reading in English and Hebrew. The persistence of opaque priming for long prime durations in developing French readers could potentially be due to less-developed semantic processing or slower orthographic inhibition in children. All of the developmental morphological priming studies run in languages other than French (described above) only included one prime duration (between 50 and 65 ms). It may be that overt prime durations would also reveal longer-lasting opaque priming for children in these languages, a possibility also supported by the recent results of Hasenäcker et al. (2021) in Italian. Additional developmental studies comparing morphological processing effects for different prime durations would clarify these findings' significance.

Direct cross-linguistic comparisons of morphological processing development suggest that children reading in French show earlier sensitivity to morphological structure than children reading in German (Beyersmann et al., 2021) or English (Casalis et al., 2015). The described difference across languages is somewhat puzzling: French has a richer morphological system than English but a less-productive one than German; additionally, its orthography is opaque, different from German but similar to English. The early onset of morphological processing effects in French children could be due to the joint pressures of a morphologically rich and orthographically opaque language, but further investigation and simulation work is merited to fully understand this difference.

The emergence of semantic transparency effects has also been investigated among adult second language learners. Silva & Clahsen (2008) found that morphological masked priming effects in non-fluent English language learners (whose native languages were German, Chinese, or Japanese) are much weaker or nonexistent compared with those of native English speakers, reinforcing that these effects rely on extensive written word experience. Low-proficiency Chinese-English bilingual participants show transparent, but not opaque, morphological masked priming (Li et al., 2017). However, proficient Spanish-

English, Dutch-English, and Italian-English bilinguals show both transparent and opaque morphological masked priming, although opaque effects are weaker than transparent effects (Diependaele et al., 2011; Viviani & Crepaldi, 2019). Interestingly, both of these studies also found that opaque and transparent effects were of more similar magnitudes in mostly proficient bilinguals than in highly proficient bilinguals, suggesting that opaque priming does not necessarily become steadily stronger with experience, but instead may peak and then lessen relative to transparent priming. This could be due to the differing impacts of masked presentations on readers, or additive effects of morphological and orthographic priming (in that non-native native readers showed orthographic priming whereas native readers did not).

Studies with adults have found morphological effects to be surprisingly robust to orthographic variability (as discussed previously; e.g., McCormick et al., 2008). Thus, another question of interest in the developmental literature is how morphological effects for irregular and less orthographically consistent paradigms emerge with development. It appears that the automatic recognition of morphological relatedness for irregular forms comes later: Quémart & Casalis (2015) showed that neither 4th nor 8th grade French children's lexical decisions were facilitated by masked primes with an orthographic shift (e.g., SOIGNEUX, "careful", and SOIN, "care"). Interestingly, those same children *did* show significant opaque morphological priming effects in the same paradigm (Quémart & Casalis, 2015, , as discussed above), so morphological processing effects not supported by typical form-base similarity may emerge more slowly than those not supported by semantic transparency. In Hebrew, masked priming experiments with 3rd and 7th grade children using stimuli derived from defective roots showed similar sensitivity to orthographic variation (Schiff et al., 2008). Significant facilitation was found when both prime and target contained all three letters of the root (e.g., [NiGuV] and [leNaGeV]), and when both prime and target contained only two letters of the root ([maGeVet] and [maGaV]), but not when one contained two letters of the root and the other contained three. Overall, abstraction away from precise orthographic representations of morphological paradigms during visual word processing appears to emerge relatively late in reading acquisition.

In summary, among early readers of a given language, masked priming effects for semantically opaque and orthographically inconsistent primes are generally weak or nonexistent. Thus, these effects might be best thought of as late properties of a well-trained system, not as foundational features of morphological processing across all readers. Within a distributed view, the emergence of morphological processing could be explained in the following manner: regularity between words' forms and their meanings provides initial informative associations that are relatively simple. With more practice and exposure to a wider variety of words, orthographic processing becomes increasingly sensitive to frequent and informative higher-order structures, to the point that even simple words containing those structures (e.g., CORNER) are affected. The described evidence of changes over acquisition mostly aligns with this sequence. However, relatively few distributed simulations of morphological processing have focused on developmental phenomena, and so emergence has not been adequately demonstrated. Doing so could illuminate the need for more authentic approaches to model learning, such as increasing exposure to complex words or allowing orthographic representations to change over the course of training.

As mentioned previously, accounts of the emergence of morphological sensitivity within decomposition frameworks also look to the statistics of the language. Rastle & Davis (2008) discussed three methods by which a developing reader might learn which letter sequences to represent on morphemes, derived from strategies discussed in the speech segmentation literature: 1) identifying boundaries from low probability sequences, 2) noting letter sequences that occur with high probability, and 3) detecting more meaningful letter groupings via form-meaning regularity. A separate proposal put forward by Grainger & Beyersmann (2017) focuses on the representation of affixes, reasoning that as embedded words are already salient due to their appearance as standalone words (flanked by spaces), the differentiation of BROTHER and BROTHEL relies on affix recognition. By activating embedded words alongside full words, complex words receive a semantic boost if the embedded word is related, and the accompanying letter string's semantic and orthographic representations are strengthened. In some ways, these developmental accounts are not so different from that of distributed accounts: regularities across contexts lead to learning of higher-order structures which impact processing. However, the proposal by Grainger & Beyersmann (2017) is limited in that it applies only to languages for which the vast majority of complex words contain only two morphemes (e.g., English and French) and for which words are delineated by spaces (which is not the case for Chinese). Additionally, the mechanisms described by both Rastle & Davis (2008) and Grainger & Beyersmann (2017) rely on a concatenative morphology (not present in Arabic and Hebrew) and isolable morphological constituents (violated in English by apophonic relations such as SANG-SING and DOVE-DIVE; Anderson, 1992). Although both accounts describe statistical features that likely contribute to the emergence of morphological sensitivity in certain languages, the distributed view provides a more fully specified and general explanation for how sensitivity to morphological structure is learned.

More research is needed in order to fully understand how morphological processing emerges over development. As noted by Dawson et al. (2021), as readers become more proficient they also recognize words more quickly, which means that a 50 ms prime presented to less experienced readers may yield a weaker or less informative signal than it would if presented to more experienced readers. Additionally, orthographic representations are developing and differentiating at the same time as morphological processing, and the timing with which these emerge relative to each other, and how they interact at different phases, is not well-specified. These challenges could be addressed by running masked priming studies that use multiple prime durations for several age groups (as done in Quémart et al., 2011), and always including an orthographic priming condition. Additionally, measurements of participants' word recognition proficiency may be a more precise metric of their phase of reading development than their age or grade level, particularly for the purposes of visual word processing research (Andrews & Lo, 2013; Dawson et al., 2021; Kahraman & Kırkıcı, 2021). Taking such measurements as part of the experiment can help explain variability among participants of purportedly the same level, clarifying when certain phenomena occur. More generally, detailed studies of morphological effects in reading acquisition may provide key insights into how experience can shape mature, language-specific representations and processes.

2.7 DISCUSSION

The picture that emerges from our review of studies of morphological processing in visual word recognition is of a flexible processing system that is very fine-tuned to the statistics of the language (e.g., relative entropy effects in Serbian), demonstrating graded degrees of morphological sensitivity depending on context (e.g., word frequency, word transparency, task goal) and morpheme characteristics (e.g., stem frequency, position-specificity). Morphological priming when the prime is not consciously perceived demonstrates greater sensitivity to the appearance of morphological structure, while also being robust to common orthographic variations. Morphological sensitivity in semantically opaque and orthographically inconsistent contexts emerges only after a great deal of reading experience.

In our view, the distributed account is better equipped to explain the full range of morphological processing effects. Distributed processing via propagated activation through optimally-adjusted weights can accommodate the graded, language- and context-specific nature of these results. Additionally, an explanation of how morphological sensitivity that is customized to statistics of the language emerges with experience is inherent in this framework, providing a fruitful lens for understanding both acquisitional and cross-linguistic comparisons.

Decomposition-based accounts, on the other hand, often need to be extended with added paths or features to address new challenges (e.g., Baayen et al., 1997; Crepaldi et al., 2010a; Dawson et al., 2021; McCormick et al., 2008; Taft, 1979). Because these accounts for the most part have not been implemented computationally, they are in many respects also underspecified (Rueckl, 2010). For example, at intermediate points in the acquisition of morpheme representations described by Grainger & Beyersmann (2017), what constitutes a partially-formed affix representation, and how does that lead to weaker or stronger decomposition? An answer to this question would be necessary to understand the gradual increase in opaque priming magnitudes over acquisition seen in Dawson et al. (2021).

The primary shortcoming of the distributed account is that it is not yet complete. Few neural-network simulations have been published that directly explore known effects in morphological processing in visual word recognition, and thus many questions regarding whether certain empirical phenomena can be demonstrated within this framework have not been given a satisfactory response. For example, there have been no simulations of effects surrounding nonword priming, orthography-semantics consistency, or orthographic variability in the context of morphological processing. Neural network simulations that account for word processing dynamics, capturing the distinctions between effects in masked versus overt priming contexts, are essential for determining whether the phenomena that most strongly suggest a role for decomposition can also be generated via distributed processes. Recurrent neural networks have been successful in capturing word processing dynamics in other contexts (e.g., Armstrong & Plaut, 2016; Cheyette & Plaut, 2017), so applying such methods to morphological questions is feasible.

More adequate treatment of semantic processing and semantic representations in distributed models would also advance our understanding of morphological processing. In particular, authentic depictions of the sparse and dynamic nature of lexical semantics and the nature of morphological contributions to word meaning could greatly enhance our understanding of morphological family size effects, task- and context-varying phenomena, and pseudo-complex nonword reading. Such improvements would also enhance the relevance of morphological processing models to higher-level processes (e.g., sentence comprehension).

In addition to its suitability for explaining the existing range of empirical morphological processing effects, some more general considerations favor transitioning to a distributed view. These are discussed briefly below.

2.7.1 Aligning with a graded view of morphology

The distinction between morphemes and non-morphemes is less discrete than it may appear at first glance. For example, the bound stem -MIT-, though etymologically considered a Latinate morpheme meaning "send", makes a very unclear semantic contribution across its appearances in ADMIT, COMMIT, PERMIT, and other words (Aronoff, 1976). Meanwhile, phonaesthemes such as GL- (GLIMMER, GLEAM, GLINT) and SN- (SNORT, SNIFF, SNEEZE) are not counted as morphemes, despite their prevalence in words with certain meanings and their morpheme-like priming effects (Bergen, 2004)[7]. There are a variety of ways in which spellings convey meaning, and the categorization schemes proposed by structuralist linguistic theories do not determine how morphological processing emerges; instead, these mechanisms are shaped by regularities learned via reading exposure.

Marelli et al. (2020) directly challenges the "morpheme-as-unit" assumption that is prevalent in morphological processing studies, arguing that "psycholinguistic studies have found evidence for morphological units because they were designed to look for that evidence" (p. 566; also see Gonnerman et al., 2007; Hay & Baayen, 2005; Blevins, 2016; Anderson, 1992). However, some experimentalists in morphological processing have already begun to shift towards a more graded approach to morphological information in their study designs. Ulicheva et al. (2020) investigated regularities between spelling and lexical category (e.g., -OON occurs predominantly in nouns such as SPOON, NOON,

[7]See Baayen et al. (2011) for simulations of these effects using the Naive Discriminative Reader.

and MOON). They quantified two such regularities—diagnosticity and specificity—and demonstrated that skilled readers are sensitive to the magnitudes of these graded metrics. Amenta et al. (2020a) also used a graded metric to demonstrate the spectrum between meaningful and meaningless letter strings: a high-frequency masked prime facilitates recognition of its stem (characteristic of morphological priming) if the stem has high orthography-semantics consistency, but inhibits recognition of its stem (characteristic of orthographic priming) if it has low orthography-semantics consistency. Such graded characterizations of regularity in the form-to-meaning mapping have become more possible in recent years thanks to advances in computational approaches such as distributional semantics (see Amenta et al., 2020b), and they are essential for more informative characterization of the impact of morphological information on word processing. However, it is worth noting that both of these efforts still quantify degree of morphological information with respect to units of contiguous letters, an approach that cannot be generalized to all cases (e.g., Hebrew or Setswana, as discussed above). Overcoming this barrier will likely require more involved metrics of shared orthographic structure.

Decomposition accounts of morphological processing are less naturally reconciled with a graded view of morphological information. Splitting a word in some cases and not in others would require a threshold judgment Rastle & Davis (2008) leading to unnecessarily lost information from less informative letter strings such as -OON and GL-. For an interpretation of decomposition in which the word is not necessarily split but constituent morphemes are represented, determining which letter strings demonstrate sufficient regularity to be represented, and how to represent potentially innumerable letter strings, remains a challenge. In a distributed account, all letters within a word can contribute to its processing via stronger and weaker weights, customized to useful regularities previously detected in the language. Separate representations of all slightly meaningful letter combinations are not necessary.

2.7.2 Integrating into broader theories of cognition

Written word recognition is primarily a task accomplished by the visual system, and several established orthographic effects, such as bigram frequency effects, have been shown to occur for other types of complex visual stimuli as well (Vidal et al., 2021). This suggests that many mechanisms involved in word recognition are also applicable to vision more generally. Neural network simulations have proved an essential tool for understanding the visual system (Kriegeskorte, 2015; Yamins & DiCarlo, 2016), and for appreciating complex word recognition within this broader context (Hannagan et al., 2021). Shared use of a distributed approach may allow insights from morphological processing research to inform our understanding of other visual phenomena, and vice versa.

Many researchers studying morphological processing are not doing so out of an interest in componential processing or mid-level visual processing per se, but because of its relevance to reading and language more generally. Integrating morphological processing into broader theories of reading and language requires a theoretical approach that can be easily combined with our understanding of orthographic, phonological, semantic, sentence and discourse processing. Neural network models rely on general learning mechanisms that can be applied to any domain or modality, and have been shown to be applicable in many of these contexts (e.g., Cheyette & Plaut, 2017; Plaut et al., 1996; Rabovsky & McClelland, 2020).

Distributed approaches represent a general theory of cognition, approximating how principles of neural computation and learning can give rise to complex behavior. Although this may not be the most effective way of capturing every higher-level cognitive phenomenon at present (e.g., conscious problem solving, or social interactions), the evidence we have reviewed suggests that it is the most productive approach to understanding morphological processing. Contextualizing stimuli- and task-specific phenomena within a broader common theory of learning and cognition promotes integration across separately studied phenomena. For example, the relationship between morphological processing in visual and auditory modalities might be better understood with a distributed approach.

2.7.3 Compatibility with natural language processing research

One final notable advantage of embracing a distributed account of language processing is that current state-of-the-art systems for natural language processing (NLP) rely on similar mechanisms and assumptions. For example, GPT-3 is a language model developed to produce high-quality human-like text (see Brown et al., 2020; Radford et al., 2019). The model was generated by training a larger Transformer architecture (similar to recurrent neural networks, but parallelized for faster training; see Vaswani et al., 2017) on a massive text corpus, to predict natural language sequences. Without being trained on any of the tasks explicitly, GPT-3 can translate, answer questions, and perform simple arithmetic.Although substantial gaps remain between the performance of such models and that of humans (Floridi & Chiriatti, 2020), it is worth appreciating that the most successful methods to date for processing and generating language in a human-like manner rely on graded and distributed learning mechanisms. The insights we can gain from NLP relevant to human word recognition are currently limited, as most efforts in this area take tokenized words as input (for some interesting exceptions, see Cao & Rei, 2016; Liu et al., 2017; Zhang et al., 2015). Additionally, NLP research is generally not focused on understanding the dynamics of language processing, so data comparable to reaction times and priming magnitudes are not examined. However, greater alignment in scientists' and engineers' understanding of core mechanisms of language cognition could lead to a more productive exchange of information between these perspectives.

2.8 CONCLUSION

Our review of the morphological processing literature suggests that, although both decomposition and distributed models may be consistent with known empirical effects, the distributed approach provides a more straightforward account, while also allowing for contexts in which decomposition-like phenomena occur. A distributed approach also better explains cross-linguistic similarities and differences in morphological processing, and how the underlying mechanisms can be acquired via language experience. Contextualizing decomposition effects within a graded and distributed view of morphological processing, not only in our approach to experimental design but also in our theoretical models of underlying mechanisms, has benefited and will continue to benefit this subfield. Additional simulation work with neural network models, particularly with dynamic, recurrent models, will be essential to reaching a deeper understanding of how readers make use of morphological structure in written language. New research efforts in this vein are described in the next two chapters.

CHAPTER 3

A Feedforward Neural Network Model of Morphological Processing Trained on an Authentic English Vocabulary

Distributed computational models of morphological processing are few in number; many questions regarding how known phenomena of morphological processing occur, and how they emerge with reading experience, remain unaddressed within this framework. A strong foundation exists on which new simulation work can build: for example, Plaut & Gonnerman (2000) demonstrated that sensitivity to morphological structure, and more specifically the emergence of opaque priming, are driven by morphological regularity of the language as a whole. However, this model consists of a simple feedforward network trained on artificial languages, and thus is quite removed from authentic contexts of word recognition. The Naive Discriminative Learning (NDL) model (e.g., Baayen et al., 2011) demonstrated that established morphological effects, such as the impact of family size, entropy, and semantic transparency on complex word recognition, can arise by training on real language. However, the model's structure is even simpler than that of Plaut & Gonnerman (2000), and the semantic representations used in the NDL simulations are not distributed (as will be discussed further shortly). These models need to be expanded and assessed with respect to a broader range of results in order to demonstrate the sufficiency of a distributed perspective for explaining observed morphological effects. Additionally, exploring how justifiable alterations to a model can close an initial gap between simulated and empirical effect patterns will enrich our understanding of how certain aspects of processing give rise to known phenomena. This chapter and the following one aim to strengthen the distributed account of morphological processing by taking steps toward a more ecologically and biologically plausible model of complex visual word recognition.

A central question for the view of morphological sensitivity as a result of a language's statistical structure is whether statistics are *sufficient* (Seidenberg & Gonnerman, 2000, p. 359). In other words, how much of human behavioral effects related to morphological processing can be explained by the statistical structure of the language, without needing to introduce additional mechanisms to the process? Note that "additional mechanisms" could refer to the explicit morpheme identification and word-splitting processes which define decomposition-based theories (e.g., Taft & Forster, 1975; Rastle et al., 2004; Rastle & Davis, 2008; Beyersmann et al., 2016a), or to added complexity in the network structure or representations used in a distributed model. The work presented here starts to test the sufficiency of language statistics by determining whether, and to what degree, a simple feedforward network with two hidden layers can produce morphological effects known to occur in skilled adult readers when trained on an actual English vocabulary. Additionally, we look at the change in pattern of effects over model training to determine if the trajectory of their emergence aligns with what we would predict given existing developmental results.

The NDL model, as already mentioned, maps from orthographic to semantic representations of words from an authentic vocabulary, and successfully simulates a variety of known morphological effects (Baayen et al., 2011; Milin et al., 2017; Baayen & Smolka, 2020). This computational work is a notable demonstration the power of sensitivity to a language's statistics as an explanation of many complex word processing phenomena. However, there are three aspects of this model that make it an inadequate answer to the question of sufficiency. First, some noteworthy morphological effects were not assessed in this model. For example, the effects of orthography-semantics consistency (Marelli et al., 2015) and spelling diagnosticity and specificity (Ulicheva et al., 2020) on word recognition latency were not included because these effects more most directly demonstrated after Baayen et al. (2011) was published. Robustness to orthographic variation (e.g., priming HAPPY with HAPPINESS; McCormick et al., 2008) was also not demonstrated. Second, the representations used for semantics in Baayen et al. (2011) were localist rather than distributed. The orthographic representation of HANDS, for example, might map to semantic units corresponding to the concept “hand” and the concept “plural”. This categorical approach to semantics makes the model's sensitivity to morphological structure less surprising, because many morphologically related words are mapping to the same semantic unit (e.g., HAND, HANDS, and RIGHT-HANDED all map to the semantic unit for "hand"). This approach to semantic representation additionally resulted in their account of opaque morphological priming in English being called into question. For several items used to make up the opaque priming condition for simulating semantic transparency effects, the prime word's semantic representation included activation of the stem meaning (e.g., ARCHER activated "arch") because the authors deemed those words to be sufficiently related. Beyersmann et al. (2016b) pointed out that even after those prime-target pairs are removed, opaque masked priming is still observed empirically, while those pairs undoubtedly drove observations of opaque priming in the context of the NDL model. This example demonstrates why more objectively-generated semantic representations, as well as more distributed ones, would be preferable. The third challenge for the NDL model is that it is unable to learn non-linear relationships. The model structure is comparable to that of a neural-network with no hidden layers: weights connect directly from orthographic to semantic units. As such, it is a stricter test of whether language statistics are sufficient for driving morphological effects; a neural network with hidden layers would allow for non-linear learning (e.g., STICK- conveys adhesion when followed by -Y or -ER but not when followed -MAN) and thus allow more complex relationships to be learned from language structure without providing any additional information. The current modeling work aims to address all three of these limitations.

Aligning with the goals of this thesis, the acquisition of flexible morphological representations will be explored by investigating the effects of semantic transparency (e.g., Rastle et al., 2004) on morphological priming magnitudes.

As discussed more thoroughly in Chapter 2, numerous studies spanning almost two decades paint a complex picture of the effects of semantic transparency in English. However, evidence generally demonstrates that, in adult skilled readers, semantically opaque morphological primes (e.g., DEPARTMENT priming DEPART) facilitate word processing in a masked primed context, though more weakly than do semantically transparent primes (TEACHER priming TEACH) (see Rastle & Davis, 2008; Feldman et al., 2009; Jared et al., 2017). To investigate the developmental emergence of semantic transparency effects in English, Beyersmann et al. (2012a) conducted a cross-sectional masked priming lexical decision study with 7- to 9-year-olds, 9- to 11-year-olds, and adults, also comparing the reaction times for transparent, opaque, and form-related pairs. Stimuli were based on those used in Rastle et al. (2004), the seminal demonstration of opaque morphological priming in English-speaking adults, but included more high-frequency words to maximize the number of words younger participants recognized. Beyersmann and colleagues found transparent but not opaque morphological priming in the two younger age groups, while opaque priming was present but weaker than transparent priming in adults. A more recent version of this study using identical methodology and stimuli but recruiting participants from a much broader age range (Dawson et al., 2019) confirmed that opaque priming emerges later and more weakly than morphological priming. Dawson and colleagues additionally observed that age and word reading ability measures are both strong predictors of opaque priming effect size. Given that semantically opaque priming has been a focal point for proponents of decomposition-based theories of morphological processing (e.g., Rastle & Davis, 2008), and that it was not satisfactorily explained by the NDL model (Baayen et al., 2011; Beyersmann et al., 2016b), it is particularly important that semantic transparency effects be included in simulation testing.

In the simulation work described below, we take an initial step towards explaining the developmental emergence of morphological effects, driven by language statistics, by investigating to what degree a neural network trained on a developmentally authentic English vocabulary demonstrates similar patterns of performance to those observed in developing and adult readers. Specifically, we trained a network to map from orthographic representations to semantic representations of words for a vocabulary based on the Touchstone Applied Sciences Association (TASA) corpus of educational texts Zeno et al. (1995). We then tested it for sensitivity to semantic transparency over the course of training. Neither the orthographic nor the semantic representations presented to this model contained any explicit morphological information, and they were derived directly from the words' spelling and usage. The network was tested on procedures comparable to primed and unprimed lexical decision, using prime-target word pairs from actual studies to enable direct comparison with empirical results.

3.1 METHOD

3.1.1 Network Architecture

The LENS neural network simulator (Rohde, 2003) was used to build, train and test the network, which consisted of 780 input units, two 2000-unit hidden layers, and a 200-unit output layer. Hidden units used a rectified linear unit function (ReLU) in an effort to speed up learning, given the multiple hidden layers and large example set. The output units used a sigmoid unit function to avoid over-activation. At the start of training, all units from one layer were connected unidirectionally to the units of the downstream layer. Additionally, all units in the output and hidden layers received input from a bias unit (with fixed activation of 1.0).

3.1.2 Training

3.1.2.1 Vocabulary

The vocabulary presented to the network for training was drawn from the Touchstone Applied Sciences Association (TASA) corpus of school texts used from first grade through the end of high school (Zeno et al., 1995), so as to avoid over-representing adult-oriented texts which do not accurately reflect word frequency distributions during reading acquisition. The TASA corpus was compiled for the purpose of approximating the frequency with which school-age students encounter particular words (e.g., Landauer & Dumais, 1997; Bhide et al., 2014). Words with more than 8 letters were removed to account for the fact that longer words may require multiple fixations to recognize, potentially impacting the orthographic processing involved. Any words not also contained within the American Heritage Dictionary (Watkins, 2000) were also excluded to account for aberrant words (such as character names in literature textbooks) and typos. Finally, the vocabulary was limited to types appearing 30 or more times within the TASA corpus, resulting in training set of 9,970 words.

3.1.2.2 Orthographic representations

To approximate the orthographic input received when looking at a word, a variation of the Overlap Model, proposed by Gomez et al. (2008), was used. Letters within a word were presented as overlapping normal distributions of activation, as opposed to more common slot-filling approaches to orthographic representations of words in connectionist models of reading (e.g., Plaut & McClelland, 1993; Plaut et al., 1996; Zorzi et al., 1998). This more continuous approach to orthographic representations prevents morphologically related words from being recognized by identical beginnings. For example, the first two columns of Figure 1 show that although KIND and KINDLY have the same first 4 letters, there is no simple operation by which to identify morphological relatedness from their orthographic representations. Additionally, the probabilistic presentation of letters accounts for lower-level word reading phenomena such as the confusability of adjacent letters (Gomez et al., 2008) which might impact the ease with which reoccurring subsets of letters are detected.

Unlike in the original overlap model, in which letters closer to the start of a word had sharper distributions, letters distributions in our representations had the same standard deviation of 2 regardless of their position in the word. This was meant to reflect the eye movement behavior of a novice reader, as fixating closer to word onset (facili-

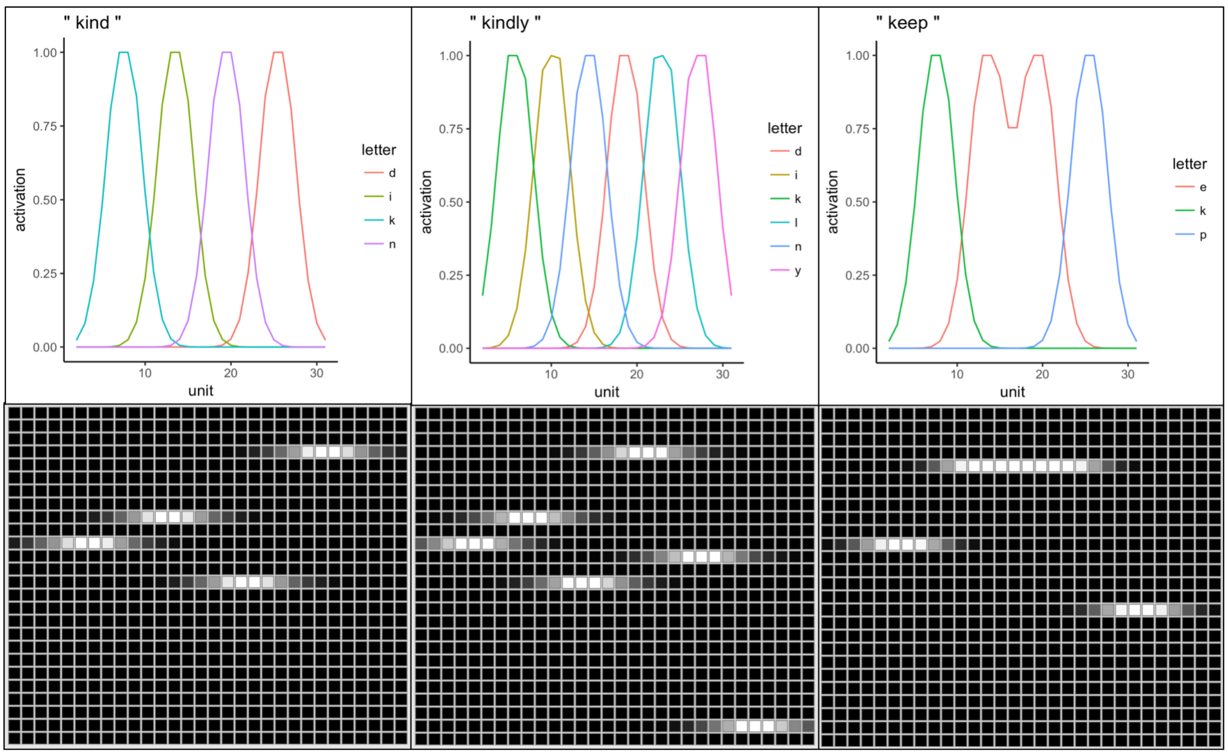

Figure 3.1: Orthographic representations visualized graphically (above) and unit-by-unit (below) for "kind", "kindly" and "keep".

tating greater acuity for earlier letters) is presumably learned through large amounts of visual word recognition experience (as suggested by differing fixation patterns across languages, e.g., Alhama et al., 2019). The center of each letter distribution was determined by word length and letter position: spacing between letters was calculated such that shorter words had slightly more spread between their letters than longer words[1], and all words were centered within the 30-unit space. If the same letter appeared multiple times in a word, the activations from normal distributions with different centers were summed to yield the total activation across the units for that letter, with activation capped at 1.0 (the rightmost column of Figure 3.1 illustrates this summation for the repeated E in KEEP).

3.1.2.3 *Semantic representations*

Meaning representations for words in the training vocabulary were based on GloVe, an algorithm for learning vector space representations of words using co-occurrence information from text corpora that has been shown to capture human performance on semantic judgments well relative to similar methods (Pennington et al., 2014). Real-valued 300-dimensional semantic vectors generated from the Common Crawl internet text corpus were converted to 200-dimensional binary vectors using a binary multidimensional scaling algorithm (Rohde, 2002). Binary rather than real-valued vectors were used because it is easier for a network to drive sigmoid unit activations to extreme values as compared to intermediate values. The binarization process greatly reduced the dimensionality of the semantic representations while preserving the similarity structure (the pairwise distance matrices for the real-valued GloVe and binarized vectors were well correlated: r = 0.68, $p < .0001$). On average 33.61% of units were active for any word's semantic representation (SD = 7.27%).

[1] Letter spacing was calculated as $\frac{N}{l+1} - exp(\frac{aN}{(l+1)-b})$ where N is the number of units over which the word is represented (30 for this simulation), l is the length of the word, and the parameters a and b were set to 1 and 9.1 so as to prevent the longest words from spilling beyond the available units and to maintain some letter overlap for short words.

3.1.2.4 *Training procedure*

The network was trained for 16,000 epochs using back-propagation with a learning rate of $5\text{x}10^{-8}$ (scaled by connection-specific factors adapted with the delta-bar-delta procedure; Jacobs, 1988) and momentum of 0.8. At the start of training, initial weight values prior to training were somewhat constrained: weights connecting the bias unit to the hidden layers were randomly initialized (mean weight value 0.25, range of 0.1), while weights to the output layer were all set to -1.0 to aid the networks suppression of output activation during initial training epochs. All other weight values were randomized with a mean of 0 and a range of 0.03. During training, the error and output unit error derivatives for each example were scaled by that words frequency in the TASA corpus. By the end of training, 99.99% of output activations were within 0.5 of their targets, and 84.93% were within 0.3 of their targets. Weight values were saved every 100 epochs up to 1000 epochs, and every 1000 epochs subsequently, to allow analysis during multiple phases of training.

3.1.3 Testing

3.1.3.1 *Prime-target testing pairs*

The stimuli used for initial network testing of semantic transparency priming effects were prime-target pairs used in the masked priming studies reported by Rastle et al. (2004) and Beyersmann et al. (2012a). The Rastle set consisted of 50 morphologically transparent pairs (e.g., TEACHER → TEACH), 50 morphologically opaque pairs (e.g., CORNER → CORN), and 47 form-related pairs, as well as a control pair with the same target but an unrelated prime for each experimental pair. The Beyersmann set consisted of 34 prime-target word pairs in each of the same three conditions (although they referred to these conditions as "true-suffixed", "pseudosuffixed" and "nonsuffixed"), as well as matched control pairs. Of these 498 original prime-target pairs, 9 pairs and their controls were dropped because they occurred in both studies, 101 pairs and their controls were excluded because

	Orthographic		Semantic	
Condition	Related	Control	Related	Control
Form	0.43	0.02	0.04	0.03
Opaque	0.45	0.03	0.04	0.01
Transp.	0.40	0.05	0.40	0.02

Table 3.1: Mean correlations of prime and target representations used for network testing. Calculated for both orthographic and semantic representations.

either a related prime or a target word was not present in the training vocabulary, and 14 unrelated pairs were altered to use an unrelated word present in the training vocabulary (selected randomly from the other primes used within that study). This left 278 prime-target pairs for testing procedures: 47 transparent pairs, 45 opaque pairs, 47 form pairs and 139 matched control pairs. The correlation between word representations for each testing pair was calculated to ensure that intended relations between primes and targets held true for our representations (see Table 3.1). A two-way ANOVA with an interaction term confirmed that orthographic correlations were significantly higher for related pairs than for controls ($t(2) = 33.04, p < 0.0001$) and this difference did not vary across conditions (ps > 0.25 for both interaction terms). A similar model with semantic correlations as the dependent variable and planned contrasts confirmed that semantic correlations were only higher for related relative to control pairs in the transparent condition ($t[2, 2] = 17.60, p < 0.0001$).

3.1.3.2 Testing procedure

In order to gather settling time data to compare with reaction times, during testing the network was run over multiple time steps, with all units in hidden and output layers integrating their inputs incrementally (with a time constant of 0.01). During each test trial, a prime-target word pair or a single word was presented to the network. In the case of prime-target word pairs, the orthographic representation of the prime was presented prior to the target for 100 time steps (or ticks), and activations initiated by this input propagated through the network. Then the orthographic representation of the prime was replaced with that of the target. The network continued to run until the average amount of change in output units activations between two ticks was less than 0.0001 (i.e., until the network settled to a semantic representation of the target word). The testing procedure for individual words (as opposed to prime-target pairs) was identical, except no priming input was presented before the target word.

The number of ticks needed for the output units to settle was recorded and used as the dependent variable in all analyses. This measure of processing time was used as a proxy for lexical decision reaction times measured in the modelled studies. On a single trial, the network could run for a maximum of 1500 ticks, but this maximum was never reached. Experimental pairs' settling times (e.g., TEACHER $\rightarrow$ TEACH) were subtracted from those of their corresponding control pairs (ROBBERY $\rightarrow$ TEACH) to calculate the magnitude of a priming effect for each item in each condition.

3.2 RESULTS

Simulated reaction times (number of ticks to settle to a semantic representation) for the testing stimuli taken from Rastle et al. (2004) and Beyersmann et al. (2012) are shown in the left panel of Figure 3.2. Priming magnitudes (prime-target pair reaction time subtracted from that of the matched control pair) are shown in the left panel of Figure 3.3.

A one-way ANOVA was run for testing performance after 1,000, 6,000, 11,000 and 16,000 epochs with priming magnitude as the dependent variable and condition as the independent variable. Despite showing a numerical trend consistent with predicted and experimental findings at every training phase, none of the tests showed a significant effect of condition (for 1,000 epochs: $F[2] = 2.03, p = 0.135$; for 6,000 epochs: $F[2] = 2.03, p = 0.135$; for 11,000 epochs: $F[2] = 1.45, p = 0.239$; for 16,000 epochs: $F[2] = 1.28, p = 0.281$).

Network reaction times increased over the course of training ($F[3] = 517.2, p < 0.0001$).

This of course does not reflect the developmental trajectory of lexical decision reaction times (children tend to respond more slowly than adults) but is rather an artifact of the testing procedure: as the network achieves better performance, the semantic representations the output units settle to are more differentiated (closer to 0 and 1 and farther from 0.5) and thus take longer to reach.

The most likely explanation for the lack of a significant effect of condition on priming magnitude is low power: there is only one network performance being simulated with these testing pairs, compared with dozens of participants in each experiment. Increasing the power by running multiple, slightly varied simulations would allow our results to be more comparable to these studies, but that would require introducing our own theory of what causes individual differences in word recognition and lexical decision reaction times. Instead, power was increased via post-hoc analyses in which prime-target pairs from 8 other studies were added to the testing stimuli.

3.2.1 Post-hoc analysis: adding more testing pairs for semantic transparency assessment

To increase the number of prime-target pairs being compared during testing, 8 peer-reviewed studies were identified that (1) included a morphologically transparent, morphologically opaque, and form/orthographic condition, and (2) listed all related pairs in the publication or in supplementary materials. No nonword stimuli were included. See Table 3.2 for a summary of the studies from which prime-target pairs were taken.

All prime-target pairs from these studies were considered for inclusion in post-hoc analyses. Pairs were not included if the related prime or the target word were not in the network's training vocabulary. If an unrelated prime wasn't present in the training vocabulary, it was replaced by a different randomly selected prime from the same study that was in the training vocabulary. For studies that did not list their paired control primes, they were generated by permuting the related primes and re-pairing them to the targets.

First Author	Year	Pairs Used	Unrelated Listed
Marslen-Wilson	2008	22	No
Jared	2017	114	No
Diependaele	2013	47	No
Feldman	2009	20	Yes
Beyersmann	2016	20	Yes
Li	2017	29	Yes
Rueckl	2008	13	No
Morris	2007	18	No

Table 3.2: Studies from which prime-target pairs were used for post-hoc analyses.

	Orthographic		Semantic	
Condition	Related	Control	Related	Control
Form	0.34	0.04	0.05	0.01
Opaque	0.38	0.05	0.04	-0.01
Transp.	0.37	0.05	0.37	0.01

Table 3.3: Mean correlations of prime and target representations in extended testing set.

Doing this increased the total number of related test pairs from 139 to 422 (170 transparent, 120 opaque, 132 form). All representation correlation effects described for the previous stimulus set also held true for the extended set.

The extended testing pairs were then analyzed using the same one-way ANOVA analyses, and a significant effect of condition was found at all four phases of training (all Fs > 7.5, all ps < 0.001). Planned contrasts showed a significant difference between the transparent and orthographic conditions at all phases of training (1,000 epochs: $t(419) = 4.66, p < 0.0001$; 6,000: $t(419) = 4.14, p < 0.0001$; 11,000: $t(419) = 4.02, p < 0.0001$; 16,000: $t(419) = 4.66, p < 0.0001$), as well as a marginal negative difference between the opaque and orthographic conditions at all phases of training (1,000 epochs: $t(419) = -1.89, p = 0.060$; 6,000: $t(419) = -1.82, p < 0.070$; 11,000: $t(419) = -1.79, p = 0.074$; 16,000: $t(419) = -1.73, p < 0.085$). In short, after increasing the number of testing pairs the network showed transparent morphological effects early and robustly, while opaque

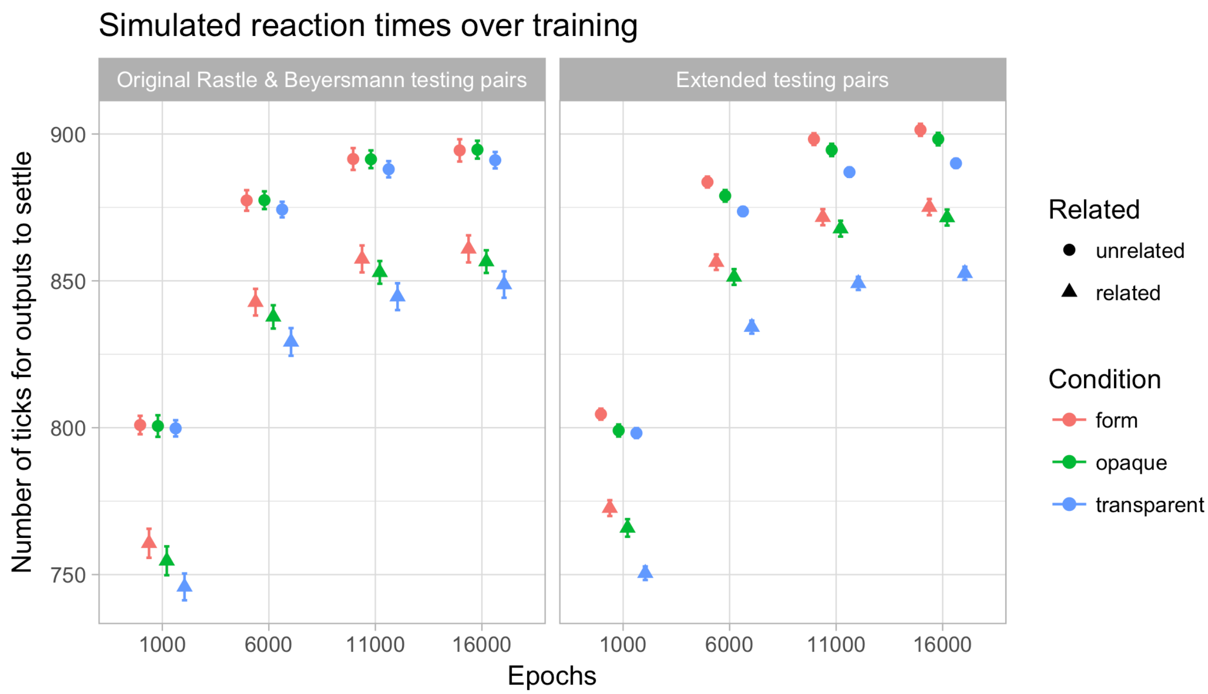

Figure 3.2: Mean number of ticks for experimentally related prime-target pairs and their matched controls. Error bars denote standard error.

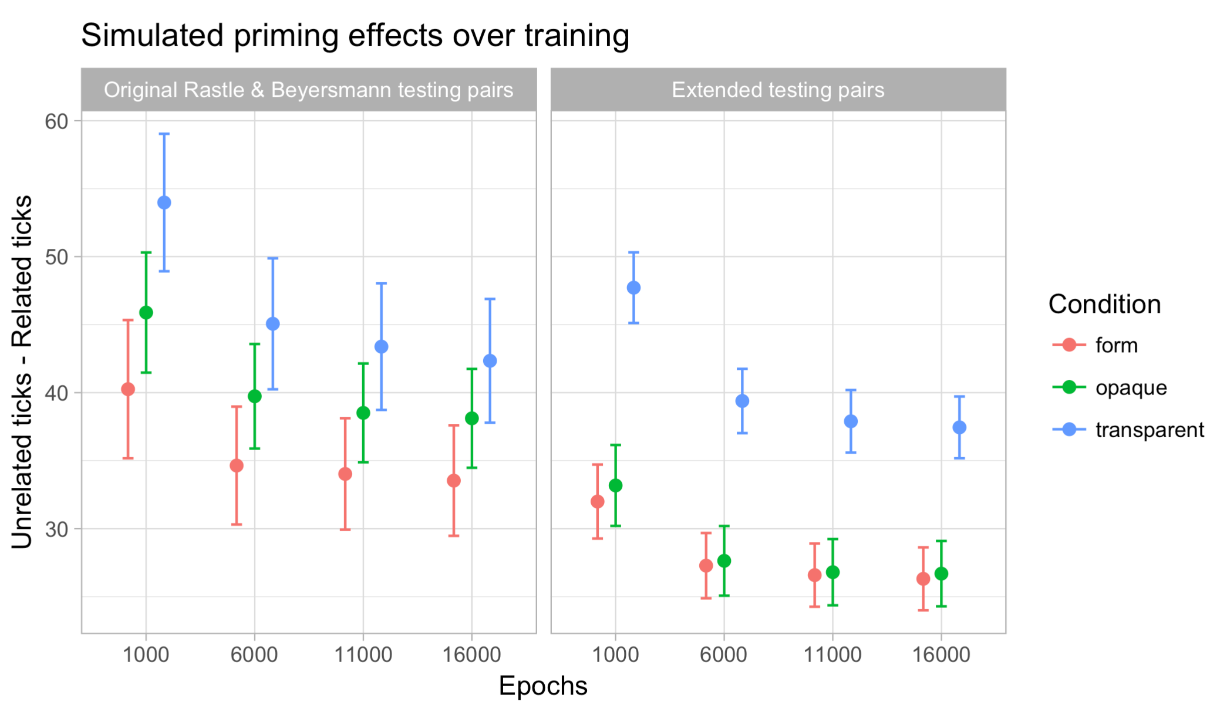

Figure 3.3: Simulated priming effects over training for stimuli from masked-priming lexical decision studies.

morphological effects do not appear.

3.3 DISCUSSION

In this simulation work, we explored whether a simple neural network trained on a developmental English vocabulary could capture the emergence of morphological priming effects. The trained network showed strong transparent morphological priming effects when tested on a large stimulus set. Beyond the NDL model described in the introduction (Baayen et al., 2011), this is the first computational demonstration to our knowledge that morphological sensitivity can emerge from authentic linguistic stimuli without any explicit knowledge of word structure being provided. All word representations were generated in an unbiased manner from letter strings and co-occurrence-based semantic vectors. However, the network did not capture the later, weaker increase in opaque morphological priming observed in experiments (e.g., Beyersmann et al., 2012a; Dawson et al., 2021). This suggests that while the regularities in form-to-meaning mappings can account for some aspects of morphological processing in English, a more complex model may be needed to capture others.

It is worth noting explicitly that certain aspects of the simulation depart notably from human performance: response latencies increase instead of decreasing as training progresses, and temporal dynamics such as the reversing impacts of orthographic and semantic priming with increased prime duration are not captured. The first aspect is a symptom of the criterion used to obtain simulated reaction times: output units reach activations farther from 0.5 before settling as the network becomes more trained, resulting in longer settling times and thus longer reaction times. The second aspect is unsurprising given the simple and entirely feed-forward training imposed on the network. However, it is important to consider how enabling dynamic processing of inputs over time during network training, as opposed to only during testing, would impact what connectivity structures the network learns. We can explore these questions via simulations with a

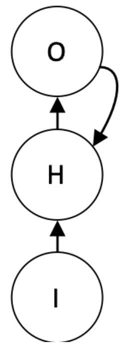

Figure 3.4: Example of a recurrent network.

recurrent neural network.

Including recurrence in a network structure entails allowing for connections from units to themselves or to previous layers, such that loops of activation propagation are possible. With this structure, activation must play out over multiple time points and find activation patterns to which the units "settle" for a given input state. For example, consider units I (input), H (hidden), and O (output) in Figure 3.4. They have feedforward connections, but unit O also feeds back to unit H. Assuming all units start at rest (0 activation), and then unit I turns on, the activation of unit H depends only on the activation of I at first, but after O has been activated by H, the activation of H in turn depends on O as well. Thus the activation of unit H must update multiple times in order to account for new activation of downstream units. Although this example is very simple, it demonstrates that including recurrence in a network structure requires changes in units' activation to play out over time, and the weights that are learned over training in this case are attuned to the dynamic nature of activation updates.

The dynamics of word processing are very relevant to detecting semantically opaque morphological priming in particular: as discussed in Chapter 2, this effect is only found in *masked* but not *overt* visual priming contexts in English. This suggests that morphological sensitivity to semantically opaque words like DEPARTMENT and CORNER only occurs early in word processing, and is later overwhelmed by activation of word-level and semantics-level information. Thus, training a recurrent network structure to map from orthography to semantics is the best next step for trying to simulate and understand the mechanisms of opaque priming. The effects of including recurrence, as well as other, more biologically plausible features of connectivity, in a neural network model of morphological processing are demonstrated and discussed in the next chapter.

CHAPTER 4

A Recurrent, Excitatory-Inhibitory Neural Network Model of Morphological Processing

Empirical evidence from a variety of sources suggest that the role of morphological information changes over the time-course of word processing (e.g, see Rastle et al., 2000; Longtin et al., 2003; Feldman et al., 2004; Quémart et al., 2011). In particular, orthographic characteristics appear to more strongly impact what information is represented early in processing, just after a word is seen, whereas semantic information becomes more dominant if the input is sufficiently strong or long-lasting. The modeling efforts with a feedforward neural network described in the previous chapter fell short of capturing how word recognition unfolds over time, as weights were learned in the context of generating a single, static pattern of activation in the semantic output layer. If, instead, a recurrent model is used, the inherently cyclical nature of activation propagation requires that learned weights accommodate multiple phases of information processing, rather than a single passage of information from input to output.

Prior efforts to simulate the dynamics of cognitive processing have found greater success when connectivity is constrained in a biologically plausible manner. Laszlo & Plaut (2012) successfully captured observed dynamics of the N400, an ERP component associated with semantic processing, with a recurrent neural network of written word comprehension by constraining connectivity among units to imitate the configuration of connectivity among actual neurons in the cortex. More specifically, weight values were constrained such that no unit can have both positive and negative outgoing connections (i.e., each unit is either excitatory or inhibitory) and negative weights can only occur within layers, not between layers (also see Cheyette & Plaut, 2017). Given prior neural-network simulations' success in capturing the dynamics of lexical processing, the separation of excitatory and inhibitory weights was also instituted in the current simulation work.

Recurrent networks take longer to learn than feedforward networks, and the added complexity of constraining weights to be either excitatory or inhibitory introduces further challenges to the training process. Thus, in the exploratory simulations presented in this chapter, neural networks were trained to map from orthography to semantics for words from small, artificial languages designed to vary in their morphological richness, comparable to those used in Plaut & Gonnerman (2000) (see Chapter 1 for a detailed summary of these simulations).

The current modeling work aimed to capture several general characteristics of complex word processing: an increasing importance of semantic characteristics and a diminishing role of orthographic characteristics for longer prime durations, as well as an overall advantage of morphologically related primes relative to other types of priming. Given that opaque priming is often cited as primary evidence of decomposition mechanism in morphological processing, the ability of recurrent, excitatory-inhibitory (E-I) networks to capture how opaque priming effects change with increasing prime duration, and for more or less morphologically rich languages, was also of particular interest. Finally, the current models' ability to account for developmental trends in morphological processing research, particularly how opaque priming emerges with reading experience, was assessed by testing the models at two earlier time points in training as well as after training was complete.

4.1 SIMULATION

As in Chapter 3, the LENS neural network simulator (Rohde, 2003) was used to build, train and test the network.

4.1.1 Network Architecture

The E-I recurrent network consisted of 65 input units, two 300-unit hidden layers, and a 200-unit output layer. Additionally, the hidden and output layers each had a corresponding inhibitory layer: 90 inhibitory units for each hidden layer, 50 inhibitory units for the

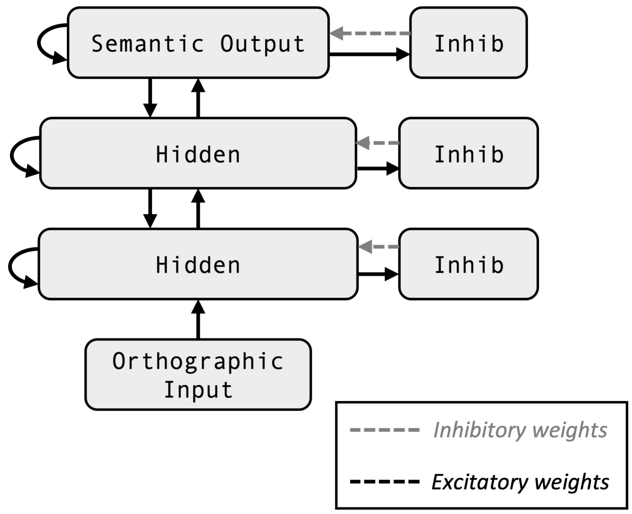

Figure 4.1: The network architecture trained to map from orthographic to semantic representations for morphologically impoverished and morphologically rich languages.

output layer (see Figure 4.1). All units except those in the orthographic layer also received input from a bias unit. All units used a rectified linear (ReLU) activation function, and the network was set to process 5 ticks (activation updates) per time interval.

Due to difficulties in training the network without it becoming stuck in a local minimum of error, only connections to and from the inhibitory layers were strictly enforced as positive or negative (positive if feeding into, negative if issuing out of, the inhibitory layer). All other weights were simply initialized on their respective side of 0, and over training the vast majority of weights remained on their initial side of 0.

Initial values for feedforward weights between non-inhibitory layers were positive, ranging from 0.04 to 0.35 in magnitude. Weights from a non-inhibitory layer to itself or to a previous layer were more weakly positive, ranging from 0.01 to 0.15. Weights from each hidden and output layer to their respective inhibitory layer ranged from 0.06 to 0.5 in magnitude, and were constrained to stay positive over training. Weights from inhibitory layers back to their primary layer had strong negative values (-0.4 to -0.5) and were constrained to stay negative. The hidden, output, and inhibitory layers also received strong inhibition from a bias unit (bias weights ranged from -1.6 to -2.4 for hidden and output layers, exactly -3.0 for inhibitory layers), and these connections were frozen such that they could not be updated during training.

The initial network structure and weight values were tailored to meet the following criteria: First, when the network received input, activation needed to reach the final output layer before being suppressed by the inhibitory units. Second, once hidden and output layers had turned off, inhibitory units needed to deactivate prior to the next example's input onset as well, so as to not suppress propagation of activation in a subsequent example.

4.1.2 Training

The network was trained to map from orthographic to semantic representations of words, as in Chapter 3. However, the words presented in this simulation were constructed specifically to contain varying degrees of systematicity in the mapping from orthography to semantics.

4.1.2.1 Artificial Languages

As in the Plaut & Gonnerman (2000) simulations, the network was trained on two artificial languages: one morphologically rich, one morphologically impoverished. Both languages included 15 possible letters, used to construct thirty 4-letter stems and twenty 2-letter second syllables. Eight of the second syllables were designated as affixes. Stems were designated as *transparent*, *quasi-transparent*, *opaque*, or experimental, and this designation informed how the meanings of words containing that stem were constructed (see the description of semantic representations below). In both languages, words constructed from the experimental stems consisted of two words ending in non-affix endings, and six words ending in affixes (two semantically transparent, two intermediate, and two semantically opaque). Words containing experimental stems were used exclusively for network testing, to ensure differences in effects were due to language differences and not word-specific or stem-specific differences.

In the morphologically impoverished language, 8 of the non-experimental stems were transparent, 6 were quasi-transparent, and 6 were opaque. In the morphologically rich language, 14 were transparent, 4 were quasi-transparent, and 2 were opaque. In each language, 240 words were constructed, eight words per stem. For transparent and quasi-transparent stems, 6 words were formed by combining the stem with an affix (all with meanings transparently or quasi-transparently related to stem meanings, respectively), and 2 were formed by combining the stem with a non-affix second syllable. For words formed from opaque stems, 2 were made by combining the stem with an affix (all with meanings unrelated to stem meanings), and 6 by combining the stem with a non-affix second syllable. Nonwords were random combinations of 6 letters of the 15 used in the language, removing any resulting strings that existed as words in the language or that were one letter off from matching a word in the language.

4.1.2.2 Representations

Orthographic representations used in the orthography-to-semantics training were learned from position-specific letter strings by a separate auto-encoder. Stems and affixes were constructed as random strings of length 4 (stem) or 2 (affix), selected from a set of 15 possible letters. The input (and output) of the auto-encoder consisted of 15 letter units at each of 8 positions. During training, a given 6-letter word was presented at each three positions: centered, or shifted one position to the left or right. For each of these presentations, the network learned a hidden representation (of 65 units) that could generate a corresponding position-specific representation at each of 3 possible output positions (indicated by the activation of one of three "position" units which projected only to the output side of the auto-encoder). For example, given the word GNPRNL starting from position 2 in the input layer (_GNPRNL_) it might be reproduced on a given trial starting from position 1 in the output layer (GNPRNL_ _). Given this training, the representations in the autoencoder's hidden layer learned to maintain information about letter identity and relative letter positions, while being consistent across 3 possible word positions. The 65-unit encoded representations for all three position presentations were used for training the E-I network, all mapping to the same semantic representation so that the network would be robust to variability in orthographic input. Nonword orthographic representations were generated by presenting nonword strings in a random starting position to the same autoencoder and saving the resulting hidden layer activation pattern.

Semantic representations were generated by randomly turning on 4% of 200 units for stem meanings, and 1% for affix meanings. The meanings of a given stem and affix were superimposed to make the meaning of the word they formed, and then transformed to a degree corresponding to the word's assigned transparency. For words formed from transparent stems, 6 words ended in affixes and no transformation of their meanings occurred after combining stem and affix meanings, while 2 words ending in non-affix word endings and had meanings completely unrelated to the stem (meanings were intentionally generating such that there was no overlap in units with a value of 1 between word meaning and stem meaning). The same was true for quasi-transparent stems, except the 6 words with affixes underwent partial transformation of the cumulative meaning (half of "on" units were swapped for other units). Opaque stems formed 6 words with non-affix endings and 2 with affix endings; all meanings were completely unrelated to the stem meaning. For "nonword" examples, all semantic targets were 0.

4.1.2.3 Procedure

Examples were presented to the network in a random order, and unit activations were not reset between examples, to imitate the demands of moving from recognizing one word to recognizing the next. Input units were soft-clamped (activations took on an intermediate value between activation at the previous tick and the unit's input), making input values change gradually over multiple *ticks*, or simulated time steps. At the beginning of each training example, no input was presented for a random duration between 2 and 15 ticks, making gaps between stimulus presentations variable. Orthographic inputs were turned on (gradually) for between 25 and 35 ticks, and targets were turned on (i.e., error was accumulated for output activations' distance from target values) for the final 20 ticks that target inputs were on. Thus, the network had to optimize performance for an unknown gap between words, and an unknown but occasionally brief amount of time between orthographic presentation and having to generate correct semantic activation.

Networks were trained using back-propagation, implemented with a bounded variation of momentum descent (`dougsMomentum` in Lens) with a learning rate of 0.05. A batch size of 2000 was used, in which approximately 1 in 5 examples was a nonword (drawn from 4,300 unique options) and the remaining examples were sampled with replacement from the 720 word inputs (240 words in 3 different positions). Weight values were saved every 100 epochs up to 1000 epochs, and every 1000 epochs subsequently, to allow analysis during multiple phases of training. Training ended when all output units across all examples were within 0.5 of their target value (during all ticks of the example for which target penalties were applied).

4.1.3 Testing

Networks were tested at two points during training, as well as at the end of training, via a procedure comparable to primed lexical decision. The actual process of deciding whether an input was a word or nonword was not directly simulated (unlike in, for example, Cheyette & Plaut, 2017). Instead, the decision was approximated as the moment when the overall activation for semantic units meant to be on for the target word passed a certain threshold activation.

4.1.3.1 Examples

Related prime-target pairs were selected using words built from the experimental stems, such that the two transparent words for each experimental stem were always used as targets, and paired with themselves (identity priming), each other (transparent morphological priming), the intermediate words for that stem (quasitransparent morphological priming), the opaque words for that stem (opaque morphological priming) and the unaffixed words for that stem (orthographic priming). Thus, the same 20 targets were used across all priming conditions. Unrelated primes were selected for each related prime-target pair to minimize orthographic and semantic relatedness.

4.1.3.2 Procedure

Prime-target pairs were presented to the network in a random order, without resetting units' output values between examples. The example began with 20 ticks of no input, followed by a prime whose duration varied from 1 to 10 ticks in length. The orthographic representation of the prime was presented immediately following that of the target (although, given the soft-clamping of input units, the transition from prime to target occurred smoothly over about 5 ticks) and was presented for 30 ticks.

The number of ticks it took for units in the output layer of the network with target values of 1 (corresponding to which units were activated in the semantic representation of the target word) to reach an average activation of 0.5 was used as a proxy for reaction time (see Figure 4.2).

4.2 RESULTS

4.2.1 Training

Training the network to meet the performance criterion required 32,900 and 17,200 epochs for the morphologically impoverished and rich languages, respectively. The greater time

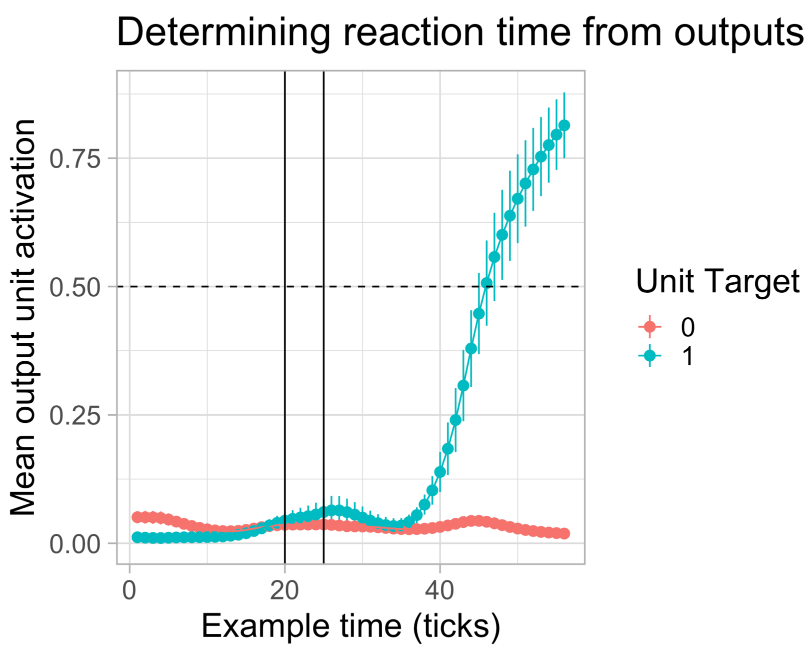

Figure 4.2: Mean activations of output layer for a single testing example with a prime duration of 5 ticks, separately for units whose target values are 1 or 0 for the primed word's semantic representation. Reaction times are estimated as the tick at which mean activation of output units with a target value of 1 surpass 0.5 (dotted line). The leftmost black line denotes onset of prime word input; the rightmost denotes onset of target word input.

needed to train the morphologically impoverished language is unsurprising given the network learned a less systematic mapping from orthography to semantics in this case. Two time points in training prior to reaching the criterion, one with overall network error closest to 20,000 and one with overall network error closest to 5,000, were selected for each language. Testing was conducted at these time points (15,000 and 23,000 epochs for the impoverished language; 5,000 and 10,000 epochs for the rich language) as well as the endpoint of training, to provide insight into patterns of effects during acquisition.

4.2.2 Testing

Examples for which the reaction time threshold was never reached, or for which the threshold was already surpassed prior to target presentation[1], were removed from analysis, resulting in the removal of 10,544, 6,417, and 4,030 out of 19,800 examples for the morphologically impoverished language at time points 1, 2 and 3, respectively. For the morphologically rich language, 1,913, 680, and 2,464 examples were removed at time points 1, 2, and 3, respectively. Simulated reaction times (number of ticks to cross the mean output activation threshold) for prime-target testing examples were fitted with linear mixed effect models, including random intercepts for target word. Changes in relative simulated priming magnitudes with increased prime duration were not monotonic; For instance, opaque priming appears to grow stronger and subsequently weaker relative to orthographic priming as prime duration increases, for the network fully trained on the morphologically rich language. To allow ease of model interpretation given this feature of model behavior, only the subset of data for which prime duration was 1, 4, 7, or 10 ticks was included, and prime duration was treated as categorical rather than a continuous.

Assessing model fit with an ANOVA revealed a significant four-way interaction of training language, priming condition, relatedness, and prime duration for all three points in acquisition ($F_{T1} = 1.74$, $p = 0.037$; $F_{T2} = 5.93$, $p < 0.0001$; $F_{T3} = 7.98$, $p < 0.0001$).

[1]The semantic output threshold was occasionally surpassed prior to presentation of the target if the prime was sufficiently semantically related to the target and presented for a sufficiently long duration.

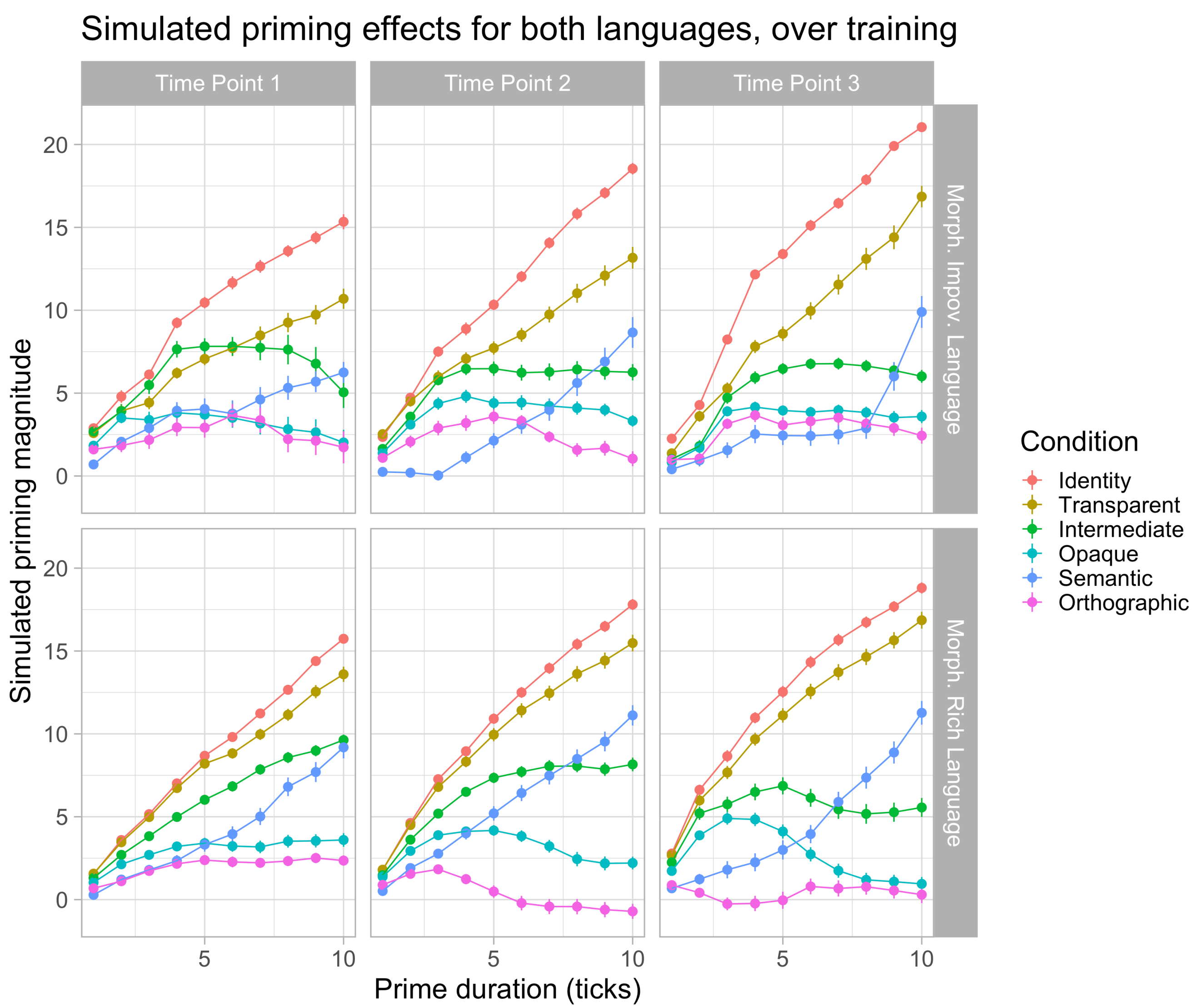

Figure 4.3: Mean simulated priming magnitudes from network testing with varying prime-target relations and prime durations, across languages and training time points. Error bars denote standard error.

Morph. impov.	Time 1			Time 2			Time 3		
prime dur. (ticks)	*4*	*7*	*10*	*4*	*7*	*10*	*4*	*7*	*10*
Transp, Id	-2.35	-3.57*	-4.06*	-2.00	-3.93*	-4.41*	-4.13*	-6.17*	-3.77*
Transp, Sem	1.38	1.39	0.56	3.18*	4.03*	1.90	3.27	5.84*	6.59*
Op, Orth	1.04	-0.24	0.81	1.25	1.29	2.41*	0.50	0.77	0.58
Morph. rich	Time 1			Time 2			Time 3		
prime dur. (ticks)	*4*	*7*	*10*	*4*	*7*	*10*	*4*	*7*	*10*
Transp, Id	-0.41	-1.13	-1.56*	-0.61	-1.33	-1.99*	-1.12	-1.69	-1.81*
Transp, Sem	2.66*	2.87*	2.62*	3.23*	3.99*	3.57*	5.73*	5.66*	3.59*
Op, Orth	0.72	0.82	1.02	2.48*	3.03*	2.24*	3.65*	0.36	-0.18

Table 4.1: Beta values for interaction of relatedness, prime duration, and condition in predicting model RTs for select pairs of conditions, across languages and training time points. Effects for each prime duration are relative to those for a prime duration of 1 tick. Asterisks denote effects with p-values below $\alpha = 0.00093$.

Smaller models were then fitted to investigate planned contrasts of priming conditions for each combination of language and training time point. The contrasts of transparent and identity, transparent and semantic, and opaque and orthographic priming conditions were each assessed in a three-way interaction with relatedness and prime duration[2]. The Bonferroni correction was applied to these analyses, to account for the assessment of 54 separate hypotheses ($\alpha = 0.00093$). Results from these analyses are shown in Table 4.1, but the most notable patterns of effects are discussed below. See Figure 4.3 for simulated priming magnitudes with varying priming condition, prime duration and time point in training, for both languages.

Several overall patterns of effects across both languages, as evident in Figure 4.3, align with prior empirical findings. Identity priming and semantically transparent morphological priming are consistently stronger than the other conditions (for the rich language, final time point: $\beta = 8.01$, $p < 0.0001$; for the impoverished language, final time point: $\beta = 6.75$, $p < 0.0001$). Facilitation by morphological primes with intermediate

[2]The effects of priming for the other three prime durations, 4 ticks, 7 ticks, and 10 ticks were assessed relative to the 1-tick prime duration. The 1-tick prime duration serves as an informative control, as the orthographic representation of the prime word is not presented for long enough in this case to yield sufficient differentiation between priming conditions.

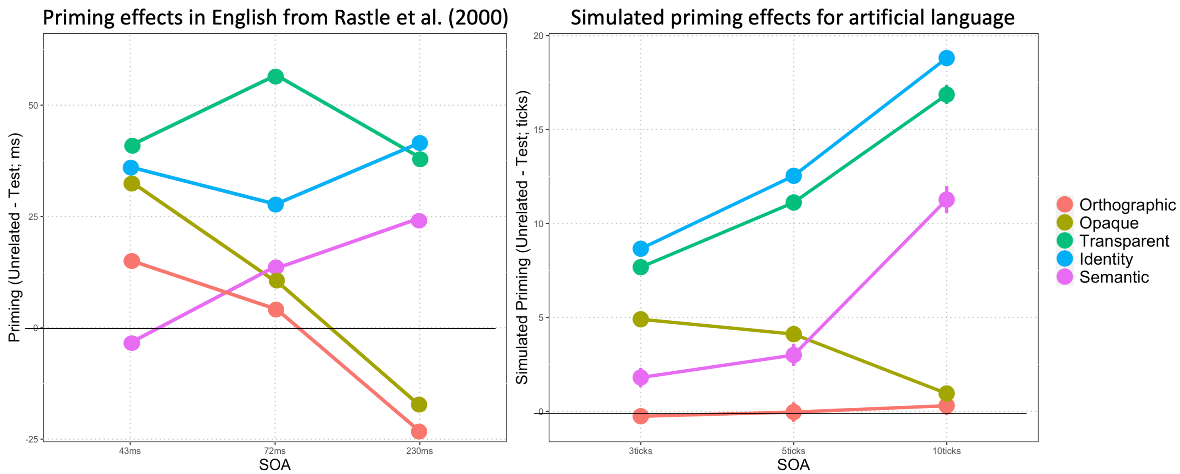

Figure 4.4: Direct comparison of empirical results from Rastle et al. (2000) and simulated results for the network fully trained on a morphologically rich language.

semantic transparency levels out to be more moderate with longer prime durations (see Gonnerman et al., 2007; Xu & Taft, 2015, for comparable empirical findings with quasi-transparent stems). Semantic priming grows stronger for longer prime durations (comparing 10 ticks with 4 ticks at final time point: $\beta = 9.46$, $p < 0.0001$ for rich language; $\beta = 6.63$, $p < 0.0001$ for impoverished language). More generally, and particularly by the end of training, semantically related primes (those in the identity, transparent, semantic, and intermediate conditions) show the greatest facilitation for speedy recognition of the target word when prime presentation duration is long, whereas opaque and orthographic priming facilitate more weakly if at all.

As evident in Figure 4.4, comparing simulated effects from the morphologically rich language with results from the time-course study reported by Rastle et al. (2000), the most noticeable difference between these two patterns of results is the lack of an inhibitory effect of orthgographic and opaque priming for long prime durations. This could be due to a lack of variety in artificial words' meanings, or a lack of sufficient distance within the semantic space made possible by a mere 200 output units; semantic processing cannot be brought too far off track in such a constrained context.

Some patterns of effects differ notably between the two languages. First, the difference between semantically transparent morphological priming and identity priming is less notable for the rich language than for the impoverished language. This trend is supported by the statistics shown in Table 4.1: in the rich language, transparent priming is only found to be significantly weaker than identity priming for the longest prime duration (relative to the shortest prime duration), while in this effect is larger and more often significant in the case of the impoverished language. Similarly, the difference between semantically transparent morphological priming and semantic priming is greater and more often significant (again, see Table 4.1). These two distinctions between the two languages can be summarized as an overall stronger effect of transparent priming in the morphologically rich language, such that it behaves less like semantic priming and more like identity priming.

The comparison of semantically opaque morphological priming and orthographic priming effects also reveals notable differences between languages. As understanding and contextualizing opaque priming across language, development, and task was a primary goal of this simulation work, these results are described in more detail. For the two fully-trained models, no advantage of opaque priming over orthographic priming is found in the case of the impoverished language, while a significant advantage is found for a 4-tick prime duration (relative to a 1-tick prime duration) in the case of the morphologically rich language. The network trained on the morphologically rich language demonstrates a pattern of priming effects comparable to that of a skilled reader of languages such as English (Rastle et al., 2000, 2004) or French (Longtin et al., 2003).

Turning to these effects' developmental emergence, opaque priming is not found to be significantly stronger than orthographic priming in either language at the earliest time point in training. Thus, the advantage of opaque priming over orthographic priming takes a good deal of training to emerge, in alignment with the findings of Beyersmann et al. (2012a) and Dawson et al. (2021). At the second time point, however, a significant advantage of opaque over orthographic priming is found for all three prime durations for the morphologically rich language, whereas such an advantage is only found for 10-ticks relative to 1-tick for the morphologically impoverished language. It is worth noting that, although opaque priming for the impoverished language did not reach significance with the Bonferroni-corrected α at 4-tick and 7-tick prime durations at time point 2, effects were greater in magnitude for these two prime durations than for any effects at time point 1 or time point 3; that is, the same pattern of opaque priming effects may be manifesting at time point 2 in both languages, to differing degrees. Thus, semantically opaque morphological priming in these networks appears to emerge first as a stronger and longer-lasting effect, before attenuating or being relegated to early processing with additional training. Although unexpected, these results do not necessarily conflict with the currently available results on morphological processing in developing readers (see chapter discussion).

One final noteworthy difference between the morphologically rich and impoverished languages is that, once fully trained, both opaque and orthographic priming appear to be more facilitatory at long prime durations for the impoverished language ($\beta = 2.71$,$p < 0.0001$) relative to the rich language ($\beta = 0.38$, $p = 0.15$). The network trained on the morphologically impoverished language, lacking aid from systematic morphological structure and subject to pressure to activate accurate semantics as quickly as possible, may learn to make use of orthographic similarity regardless of the prime's ending.

4.3 DISCUSSION

The neural network model presented in this chapter captures several general aspects of the time-course of morphological processing in skilled readers. The increasing role of semantics and diminishing aid of orthographic information for longer prime durations in this model echoes the results of morphological priming studies where prime duration has been varied (e.g., Rastle et al., 2000; Longtin et al., 2003; Feldman et al., 2004); to our knowledge, these aspects of complex word processing dynamics have not been simulated previously.

At the final training time point, the finding of significantly greater opaque priming relative to orthographic priming in the morphologically rich language but not in the morphologically impoverished language corresponds with our prediction that greater morphological richness should give rise to stronger opaque priming effects. However, it appears that the morphologically rich language in these simulations may be more likened to English or French (Longtin et al., 2003; Rastle et al., 2004) than to Hebrew, Arabic, or German (Frost et al., 2000; Boudelaa & Marslen-Wilson, 2001; Smolka et al., 2009), as opaque priming is found for the 4-tick prime duration but not for longer ones. These results motivate the examination of an even more morphologically rich language, to determine if overt opaque priming can be detected even in a fully-trained model in that case.

The current model better captures the time-course of complex word recognition by being trained in a dynamic and time-pressured context. The feedforward model (Chapter 3) was trained in a context where all orthographic input is immediately and fully processed. Thus, there was no phase of recognition where making use of information beyond than the full identity of the word was particularly advantageous. However, in the recurrent model presented in this chapter, the network is trained such that weights are pressured to take advantage of information at any level of complexity as it becomes available, such that words can be recognized under increasingly time pressured scenarios. Varying the amount of time before targets turned on among training examples allowed for slower processing techniques to yield some reduction of overall error during early phases of training, while motivating more efficient strategies, such as early sensitivity to orthographic structures that typically denote morphological complexity, to eventually play a more significant role. The developmental trajectory of these effects over training also provides insights regarding how such processes emerge.

Although the lack of opaque priming for both languages at training time point 1 is unsurprising, the stronger and long-lasting effects of opaque priming for both languages at time point 2 was not expected. It seems that at an intermediate phase of training, these networks become attuned to the orthographic structures typically associated with morphological relatedness, and these structures are made use of in a way that impacts how the prime word is represented even once it has been presented for a considerable duration (e.g., 10 ticks). However, processing complex words in this way will lead to new errors, due to the opaque and quasi-transparent words present in both languages. Thus, later in training the network attenuates the use of this type of sensitivity (in the case the impoverished language) or overrides its use with higher-level information as it becomes available (in the case of the rich language).

If this developmental phenomenon reflects the actual emergence of semantically opaque morphological priming, we would expect to see opaque priming emerging for both masked and overt priming contexts simultaneously, followed by a weakening of opaque priming in the overt context in languages in which skilled readers don't show opaque, overt priming. In English, masked priming by semantically opaque words does not typically emerge until developing readers have gained quite a bit of reading experience (Beyersmann et al., 2012a; Dawson et al., 2021); however, no developmental studies of morphological processing in English have looked at the emergence of sensitivity to semantically opaque words in an overt priming context.

To our knowledge, only one developmental study of morphological processing, conducted in French, has systematically manipulated prime duration: Quémart et al. (2011) investigated sensitivity to transparent, opaque, orthographic, and semantic priming across three prime durations (60 ms, 250 ms, and 800 ms) and two age groups (3rd to 7th graders, and adults). For the 60 ms prime, semantically transparent and opaque primes significantly and equivalently facilitated lexical decision for both the children and the adults, whereas semantic and orthographic priming were not significant for either group. However, for the 250 ms prime, transparent and opaque priming were significant and equivalent for children, whereas adults showed only significant transparent priming and marginal semantic priming. These results have not been replicated, in French or any other language, but if they are robust they may mirror the simulated pattern of opaque priming emergence found here. See Chapter 8 for further discussion and investigation of longer-lasting opaque priming in early readers.

Although the exploratory modeling efforts described in this chapter are a step forward in our understanding of morphological processing dynamics, a good deal remains to be done. First, these networks do not entirely abide by the constraints of a true excitatory-inhibitory network, as the constraint of only positive weights within layers was relaxed in order to facilitate training. It is unclear whether enforcing such constraints more strictly would notably impact the patterns of effects reported here. Second, all words in the artificial languages presented to the network were complex, and all prime-target pairs consisted of two complex words, whereas the majority of empirical results to which these simulated findings are compared were conducted with a complex word priming a simple target. Whether such a distinction impacts the simulated pattern of effects found here must be explored in future work. Additionally, the process of lexical decision was not suitably captured in the current model. A large number of examples had to be removed from testing due to the nature of the semantic activation threshold used to approximate lexical decision, leading to an imbalance in number of testing examples across the two languages. This issue could be addressed by implementing a leaky competing integrator model of decision making (see Cheyette & Plaut, 2017; Usher & McClelland, 2001). Finally, and most notably, the languages presented to these networks were exceedingly small (only 240 words, each presented at three different positions) and obviously do not capture the true complexity of real language or the contexts in which it is encountered. Expanding the current model and speeding up training procedures such that the model can be trained on authentic language input, as done with the feedforward network in Chapter 3, would greatly enhance our ability to interpret the implications of simulated results for real word processing. Presenting realistically ordered examples (e.g., THE, DOG, BARKED instead of DOG, TREE, COULD) or incorporating phonological information (see Plaut et al., 1996) would also be worthwhile steps towards a more ecologically plausible future model.

CHAPTER 5

A Variation on the Orthography-Semantics Consistency Metric

The impact of morphemes on word processing is due at least in part to the role they play in informing a word's meaning: *dict* is not only helpful for recognizing *dictate* because of its frequency across contexts, but also because of the information it provides about the meaning of *dictate*. Thus, it would presumably be informative to keep track of how useful a given morpheme is as a clue to the word's meaning. Put another way, how semantically similar are words that contain that morpheme? Modern computational linguistics methods for obtaining distributed semantic vector representations of words (e.g., Landauer & Dumais, 1997; Pennington et al., 2014) have made it more feasible to approximate a morpheme's semantic utility (see Amenta et al., 2020b). The recently developed *orthography-semantics consistency* metric takes advantage of this.

Introduced by Marelli et al. (2015) and discussed further by Marelli & Amenta (2018) and (Amenta et al., 2020a), orthography-semantics consistency (OSC) refers to the frequency-weighted semantic similarity between a standalone stem and the words containing it. Stems that are overall closer in meaning to words containing them, especially high-frequency words containing them, will have a higher consistency value. For example, words that contain the letter sequence *trust* tend to be fairly similar in meaning to *trust* (*distrust*, *trustworthy*, *trustee*), while words that contain *whisk* are not very similar to *whisk* (*whisking*, *whisker*, *whiskey*). Thus, *trust* has high OSC while *whisk* has low OSC.

As mentioned in Chapter 2, the original purpose of calculating this consistency metric was to quantify a potential confound present in priming studies of semantic transparency. Marelli et al. (2015) demonstrated that the targets used in several previous masked priming studies of semantic transparency differ significantly in OSC across conditions, and they argued that this difference could be driving commonly observed faster response times to transparent targets than to opaque targets when preceded by unrelated primes. In other words, lower-OSC words like WHISK, often used in opaque conditions following WHISKER, are responded to more slowly than higher-OSC transparent targets like CHEER (primed by CHEERFUL), regardless of what prime precedes them. Because the targets in transparent conditions are already being recognized quickly, the degree to which their recognition could be further facilitated is reduced, constraining transparent priming magnitudes to a level more comparable with those of opaque priming. Following this paper, subsequent work on OSC by these authors refined its calculation to better predict lexical decision latencies (Marelli & Amenta, 2018) and investigated its relation to masked priming other than as a potential confound (Amenta et al., 2020a). Very recently, Siegelman et al. (2022) also explored the operational definition of OSC and demonstrated that effects of OSC are more evident in tasks that place more emphasis on word semantics relative to word phonology (i.e., lexical decision but not reading words aloud).

The Marelli & Amenta (2018) metric of a morpheme's OSC is calculated by identifying all of the words containing the stem (regardless of whether they're considered morphologically related to the stem), and calculating the frequency-weighted average cosine similarity between the semantics of each such word and that of the standalone stem. More formally:

$$\mathrm{OSC}_s(t) = \frac{\sum_{x=1}^{k} \cos(\mathbf{r}_t, \mathbf{r}_x) * f_x}{\sum_{x=1}^{k} f_x} \tag{5.1}$$

In this equation, $\mathbf{r}_t$ is a semantic vector for the target stem, $\mathbf{r}_x$ is a semantic vector for each word x containing the target stem (i.e., each "orthographic neighbor"), and f_x is the frequency of each word containing the target stem. Semantic vectors are obtained using distributional semantics techniques, which will be described in more detail in subsequent sections.

Marelli & Amenta (2018) demonstrated that their OSC metric predicts lexical decision times from the English Lexicon Project (Balota et al., 2007), after controlling for word frequency, orthographic length, and morphological family size. OSC also predicts the magnitude of facilitation effected by morphological masked priming (Amenta et al., 2020a), suggesting that the degree to which morphologically related words boost each others' recognition is in part determined by the OSC of the shared morpheme.

However, as a measure of a morpheme's semantic consistency across the words in which it appears, the OSC metric proposed by Marelli & Amenta (2018) has two notable shortcomings. First, distributed semantics models typically only provide semantic vectors for letter strings that exist as whole words. Consequently, by calculating all semantic distances relative to the standalone stem's meaning, OSC cannot be calculated for morphemes that do not exist as independent words (i.e., bound stems like *dict* and *cept*, and affixes like *pre* and *ness*). Second, using the standalone stem as the semantic distance reference point gives it a stronger role in determining OSC, relying on the assumption that the standalone stem reserves a unique status in contributing to a morpheme's semantic significance among words containing the string. While this may be true, we do not know of any evidence to suggest that it is, and furthermore there are at least some cases where the meaning of the stem word is aberrant from meanings of the rest of the words containing it (e.g., *man* can mean "male human" or "work or operate", as in "man the pumps", but the majority of words containing the word *man* relate primarily to the first meaning). Using the stem as the reference point to which all longer words are compared in such cases would result in a skewed metric of how informative that letter string is for the set of words not including the stem.

The current work proposes and evaluates a variation on the OSC metric. In this variation, all words containing the letter string of interest contribute to the calculation proportional to their frequency. Adapting how OSC is quantified to allow its computation for non-word letter strings advances the primary goals of this dissertation by expanding the implications and utility of this metric. The ability to calculate OSC for any string of letters is key to determining whether morphological sensitivity is driven by the utility of certain letter strings for cuing word meanings, as is most explicitly predicted by the distributed account of morphological processing. A refined OSC metric also facilitates adopting a graded view of morphology in psycholinguistics (at least in languages with concatenative morphology), as letter strings can be localized along a spectrum of semantic utility instead of being categorized as “morpheme” or “non-morpheme”.

5.1 METHODS

The primary innovation for this variation of OSC is the reference point to which semantic distance was calculated. Instead of calculating semantic distances from all words containing the letter string of interest (the stem’s “orthographic neighbors”: WHISKY, WHISKER, etc.) to the target on its own (WHISK), a frequency-weighted centroid for the cluster of meanings of words containing the string (shown as an “X” in Figure 5.1) is used as the distance reference. Consistency is then calculated as the mean cosine similarity between semantic vectors of words containing the string and this centroid. In formal terms,

$$\mathrm{OSC}_c(t) = \mathbb{E}_{x=1}^{n}\big[\cos(\mathbf{r_x}, \mathbf{c})\big] \tag{5.2}$$

In this case, $\mathbf{r_x}$ is a semantic vector for each word containing the target stem, including the standalone stem itself, if it exists. The variable $\mathbf{c}$ is the frequency-weighted centroid:

$$\mathbf{c} = \frac{\sum_{j=1}^{n} \mathbf{r_j} * f_j}{\sum_{j=1}^{n} f_j} \tag{5.3}$$

This variation of OSC is a metric of the tightness of the cluster of word meanings, or the semantic similarity of all words containing the letter string of interest. The letter string itself does not need to exist as a standalone word in order for OSC to be calculated, and the set of words containing the letter string is assessed holistically, not relative to a single word.

Although the most noteworthy change in the new OSC metric is the reference point for semantic distances, different frequencies and semantic vectors were also used. Instead

Figure 5.1: Reference points for calculating semantic relatedness via the original OSC metric (red +) and the new proposed variation (red X). Semantic space reduced to two dimensions for visualization using multidimensional scaling. Point size corresponds to word frequency.

	OSC_c	OSC_s	OSC_o
OSC_c	1.00	0.76	0.62
OSC_s	-	1.00	0.55
OSC_o	-	-	1.00

Table 5.1: Correlation matrix for the three OSC metrics.

of constructing original semantic spaces, as done by Marelli & Amenta (2018), the GLoVe 300-dimensional semantic space was used (Pennington et al., 2014). Frequency values based on the 17.9 billion word COBUILD English corpus were obtained from the CELEX database (Baayen et al., 1995). Additionally, words that contained the letter string of interest were identified from the set of words present in both the CELEX database and the SUBTLEX-US database (Brysbaert & New, 2009). Because of these differences, the new cluster-centered metric (OSC_c) was compared to a metric generated by applying the stem-centered calculation to CELEX frequencies and GLoVe representations (OSC_s), as well as to the original OSC values (OSC_o) from Marelli & Amenta (2018). Correlations between the three OSC metrics are shown in Table 5.1.

Model	OSC Effect	Variance Explained	AIC	BIC
Baseline		0.41	-30370.91	-30335.46
OSC_c	-0.0029	0.41	-30369.74	-30327.21
OSC_s	-0.015***	0.42	-30412.08	-30369.55
OSC_o	-0.011***	0.42	-30403.65	-30361.12

Table 5.2: Comparing OSC metrics as predictors of English Lexicon Project reaction times. Effects with *** had p-values less than 0.0001. OSC_c was not a significant predictor.

5.2 RESULTS

Analyses were conducted in R (version 4.0.3; R Core Team, 2017), and linear mixed-effects models were fit using the *lme4* package (version 1.1-26; Bates et al., 2015). The *lmerTest* package was used to obtain p-values for mixed effects terms (version 3.1-3; Kuznetsova et al., 2017).

The three OSC metrics were first assessed in terms of how well they captured variation in lexical decision reaction times. These analyses followed the comparisons described by Marelli & Amenta (2018), but using words and reaction times from the English Lexicon Project (Balota et al., 2007) instead of the British Lexicon Project (Keuleers et al., 2012). Similar to Marelli & Amenta (2018), OSC values were assessed relative to a baseline model including frequency (from the SUBTLEX-US database; Brysbaert & New, 2009), morphological family size (from CELEX; Baayen et al., 1995), and orthographic length as predictors. Words that were not found to be embedded in any other words were excluded, as were words not present in the GloVe database, and those for which family size or SUBTLEX-US frequency could not be obtained. This left 8,857 items for analysis.

Results are summarized in Table 5.2. The model that best explained variance in lexical decision reaction times was the one including the new stem-centered OSC metric, followed by the original OSC from Marelli & Amenta (2018). Variance explained did not differ dramatically across the models. The cluster-centered variation increased AIC and BIC values when added to the baseline model (indicating a worse fit), and did not significantly predict variations in reaction times beyond baseline predictors. Interestingly, the cluster-centered and stem-centered metrics explained variance in the other model's residuals (for OSC$_c$: $t = 0.0043$,$p = 0.06$; for OSC$_t$: $t = -0.0087$, $p < 0.001$), suggesting they capture different aspects of variance in reaction times.

The next analyses concerned data from previous studies of masked morphological priming, and followed procedures reported by Amenta et al. (2020a). Mean reaction times for each prime-target pair were taken from Rastle et al. (2000), Rastle et al. (2004), Andrews & Lo (2013), and Jared et al. (2017). Prime-target pairs for which the target OSC could not be calculated or for which frequencies, orthographic neighborhoods or family sizes could not be obtained were excluded, leaving average priming magnitudes for 676 related prime-target pairs (535 unique pairs, 513 unique targets). These pairs were categorized as orthographic, opaque, quasitransparent or transparent, corresponding to how they were categorized in the study from which they were taken.

Four models were fit to the priming data. All used mean priming magnitude (reaction times for unrelated minus related pairs) as a dependent variable and included target length, log-transformed morphological family size, log-transformed prime and target frequency, orthographic neighborhood (OLD20, obtained from Balota et al., 2007) and transparency (cosine similarity of the prime and target GloVe representations) as covariates. Interaction terms for OSC and the two frequency covariates, and random effects for target and study, were also included.

The OSC$_o$ model showed a marginal main effect of OSC ($t = 21.19, p = 0.08$), and a significant positive interaction with prime frequency ($t = 19.56, p = 0.04$). The OSC$_s$ model also yielded a significant positive interaction of OSC and prime frequency ($t = 37.51, p = 0.005$), although no main effect of OSC was found ($t = 6.37, p > 0.5$). The OSC$_c$ model yielded a strong, positive main effect on priming magnitudes ($t = 72.60, p = 0.004$) and, though lacking a positive interaction with prime frequency ($t = 7.16, p > 0.5$), showed a marginal negative interaction with target frequency ($t = -27.44, p = 0.09$).

The OSC_c model had the lowest AIC and BIC values (6879.81 and 6938.52, respectively), very closely followed by OSC_s (6880.85 and 6939.56), OSC_o (6881.04 and 6939.75), and the baseline model (6907.40 and 6952.56). Similar patterns of effects were observed when prime-target pairs assigned to the orthographic condition were excluded.

5.3 DISCUSSION

The cluster-centered OSC metric is a worse predictor of unprimed lexical decision RTs, but is as good or perhaps slightly better at explaining variance in priming magnitudes, relative to the two stem-centered OSCs. That the cluster-centered OSC as a poorer predictor of unprimed lexical decision makes sense, given its reduced focus on the meaning of the standalone stem. When calculating the stem-centered metrics, the meanings of all words containing it are compared to the stem's standalone meaning, but the cluster-centered calculation does not give the standalone stem's meaning a privileged status. Thus, the stem-centered OSC may be a better metric for the *word*, whereas the cluster-centered OSC better captures the consistency of the *letter string* across all its appearances.

Aligning with this interpretation, the main effect of OSC on priming magnitudes was significantly positive for the cluster-centered metric, while being marginal for the original metric from Marelli & Amenta (2018) and nonexistent for the re-calculated stem-centered metric. Taking a distributed view of morphological processing, one might expect that morphological priming magnitudes would be stronger for prime-target pairs with a more semantically consistent shared morpheme. This prediction follows from the fact that more consistent morphemes would be more useful entities to leverage in the process of mapping from orthography to semantics. Thus, the observation of a stronger main effect for the cluster-centered metric on priming magnitudes suggests that it better captures the letter string's consistency across its appearances (the letter string -TRUST- across TRUSTING, UNTRUSTWORTHY, DISTRUST), rather than that of the *word* (TRUST).

The cluster-centered OSC model showed a marginal negative interaction of OSC with target frequency, while the stem-centered OSC models yielded a positive interaction with stem frequency. The cause for this difference in observed interactions can be explained at least partially by the word-metric versus morpheme-metric distinction just discussed. If the stem-centered metric is a better measure of the word's consistency than of the morpheme's consistency, then the effect of this metric on speed of target recognition would be stronger, and the additional effect of the metric on priming magnitude would only be detectable when the prime is also high-frequency. Conversely, the negative interaction of the cluster-centered OSC metric with target frequency would suggest that the facilitating effect of OSC is stronger when the target is low-frequency, because high-frequency targets can only be recognized slightly more quickly than they already are and thus can only be facilitated to a small degree by their relatedness to the prime. Put more simply, an interaction with prime characteristics makes more sense if the predictor best describes the target word, while an interaction with target characteristics is expected if the predictor best describes the relationship between prime and target.

5.4 CONCLUSION

The revised metric of orthography-semantics consistency proposed in this chapter appears to better capture the consistency of a particular string of letters across its contexts (TRUSTING, UNTRUSTWORTHY, etc.), without giving undue priority to the meaning of the standalone word (TRUST). Additionally, the new approach to calculation can be applied to any letter string, not being limited to letter strings that exist as independent words. These improvements in how OSC is operationalized will enhance our ability to investigate morphological processes during visual word recognition, and their acquisition, in the context of a graded spectrum of semantic utility from non-morpheme to morpheme. The experiments conducted in Chapters 6 and 8 use this variation on OSC to inform stimuli selection and conduct statistical analyses.

CHAPTER 6

Effects of Orthography-Semantics Consistency on Morphological Processing in Skilled Adult Readers

In skilled adult readers, morphological processing mechanisms have been shaped by pressures to recognize the written words they are exposed to quickly and accurately. This learning context means that the brain must begin extracting and making use of information as soon as possible, without waiting until the identity of the presented word is certain. As discussed in Chapter 2, the masked primed lexical decision paradigm provides insight into what factors most strongly influence word perception before it is recognized, by presenting the prime word very briefly (about 50 milliseconds). In a masked priming context, the factors that appear to impact morphological priming magnitudes have more to do with the word's appearance, and the general roles of the word and its substrings in the language more broadly, as this is the type of information most notably represented when the target appears. A primary example of this is semantically opaque morphological priming in English. Rastle et al. (2004) demonstrated that prime-target pairs such as CORNER-CORN and DEPARTMENT-DEPART show significant facilitation in a masked priming context relative to merely orthographically related pairs (CASHEW-CASH; also see Feldman & Soltano, 1999; Marslen-Wilson et al., 2008; Fiorentino & Fund-Reznicek, 2009). As lexical processing continues, factors more particular to the word's whole form, and semantic context, become more relevant. For instance, Rastle et al. (2000) demonstrated that when primes are presented long enough to be consciously processed (e.g., 250 milliseconds), opaque priming effects become negligible or slightly inhibitory, while semantic priming (SKIP-MISS) strengthens and semantically transparent priming remains robustly facilitatory.

Semantically opaque morphological priming, as discussed previously, has become a linchpin result in support of the decomposition perspective (Rastle & Davis, 2008). The pattern of effects described above is clearly compatible with the idea that at some phase of complex word processing, the word is decomposed into morphological constituents prior to recognition or semantic access. However, of the many studies conducted on the topic of opaque morphological priming, few go beyond the question of *whether* it occurs to explore *for which items* it occurs.

From the distributed perspective, opaque priming provides evidence that some morphological paradigms are so consistently significant, the brain starts reacting to their presence to get a head start on lexical retrieval even in cases where their presence is misleading (i.e., the presence of DEPART in DEPARTMENT). From this explanation naturally follows the prediction that opaque morphological priming should be stronger for more semantically consistent stems. Armed with our revised, cluster-centered metric of OSC from Chapter 5, we can begin to test this prediction.

To our knowledge, the only other study that has investigated the effect of OSC on morphological priming magnitudes is Amenta et al. (2020a). They found that OSC interacts with prime frequency to predict morphological priming, such that priming increases with prime frequency for high-OSC stems, while priming decreases with prime frequency for low-OSC stems. Their explanation of this effect is well phrased, and aligns with our interpretation (p. 1553):

> ...in highly consistent families, a frequent prime is a blessing, because form is a reliable cue to meaning and thus primes do yield exploitable information on the target; thus, the stronger the prime, the easier target processing. When form to meaning mapping is instead quite inconsistent in the family, a frequent prime hinders target identification, as form is not a reliable cue to meaning and thus primes are not likely to yield information on the target. (p. 1553)

However, three shortcomings in this study make further investigation necessary. First, analyses used the stem-centered calculation of OSC, and thus are influenced by this metric's bias in favor of the meaning of the stand-alone stem. As described in Chapter 5, analysing the same data used in Amenta et al. (2020a) with our revised OSC metric in yielded a notably different pattern of results, including a significant interaction with target frequency but not prime frequency. Second, the effects described do not differentiate between transparent and opaque priming, and in fact orthographic prime-target pairs were included in this analysis as well. Third, these analyses were conducted with data from prior morphological priming studies; thus, OSC was not maximally varied across stimuli to optimize investigation of its effect on priming, and it was confounded with transparency (OSC was higher among targets of the transparent condition than among targets of the opaque condition; see Marelli et al., 2015). The present study was designed to address these issues with the aim of illuminating the effect of OSC on both semantically opaque and semantically transparent morphological priming.

6.1 STUDY 1: THE EFFECT OF OSC ON MASKED MORPHOLOGICAL PRIMING

This first study looks at the effect of OSC on two types of morphological priming, semantically transparent (GOVERNMENT-GOVERN) and semantically opaque (DEPARTMENT-DEPART), when the prime is masked and briefly presented to prevent recognition. Of primary interest is whether OSC positively predicts opaque masked priming magnitudes. Such an effect would support the distributed perspective: Word recognition processes are most likely to be sensitive to the presence of stems that are generally "semantically useful", and this would be easiest to detect when the meaning of the prime does not also provide semantic information about the target. By including a transparent condition as well, however, we gain an opportunity to observe whether the effect of OSC on morphological priming differs for semantically opaque versus semantically transparent prime-target pairs. Masked priming taps into representations during lexical processing that are less semantically-informed, and thus may allow for the effects of OSC on even transparent priming magnitudes to be detected. Alternatively, the utility of the prime's full word form for recognizing the target may lead to neutral, or even inhibitory effects of stem OSC.

6.1.1 Methods

6.1.1.1 Participants

Participants consisted of 62 undergraduates at a private university in Western Pennsylvania, completing the experiment for credit in a psychology course. All were required to be native speakers of English with normal or corrected-to-normal vision.

6.1.1.2 Stimuli

Prime-target pairs were selected from a set of 278 semantically opaque prime-target pairs and 326 semantically transparent prime-target pairs used in prior studies of morphological priming (Rastle et al., 2000; Beyersmann et al., 2012a; Marslen-Wilson et al., 2008; Jared et al., 2017; Diependaele et al., 2013; Feldman et al., 2009; Beyersmann et al., 2016c; Li et al., 2017; Rueckl & Aicher, 2008; Morris et al., 2007, see description of post-hoc analysis in Chapter 3 for detail). Frequencies, stem family sizes and affix family sizes for all words were obtained via CELEX (Baayen et al., 1995) and orthographic neighborhoods (OLD20) and bigram frequencies were taken from the English Lexicon Project (Balota et al., 2007). Orthography-semantics consistency (OSC_c) values were of the cluster-centered variation described and calculated in Chapter 5.

A publicly available software package for Stochastic Optimization of Stimuli (SOS!, Armstrong et al., 2012) was used to select two 100-item sets: one opaque and one transparent. Items were selected such that OSC_c varied maximally within each set, while remaining minimally correlated with prime and target frequencies, lengths, orthographic neighborhoods and bigram frequencies, as well as stem and affix family sizes, and the number of words in which the stem was found embedded during calculation of OSC. (See Appendix A for the complete set of experimental stimuli used.) Correlations between OSC and these other item metrics were low and non-significant (all rs < 0.2 in magnitude, all ps > 0.05) with the exception of length and orthographic neighborhood in the opaque condition, which were both significantly correlated with OSC (for length: $r = 0.33$, $p < 0.0001$; for OLD20: $r = 0.26$, $p < 0.01$). Length and OLD20 were strongly correlated across the stimulus set ($r = 0.67$, $p < 0.0001$) so we attributed this to the same shortcoming of the opaque set and accounted for this in analysis.

Real word targets were paired with both related primes (pre-determined based on prime-target pairings from prior studies) and unrelated primes (related primes reassigned such that prime and target had distinct consonant onsets and less than half of the letters in the prime were in the target). Each participant saw a given target once, with relatedness of the preceding prime counterbalanced across participants, and performed lexical decision on a total of 400 word and nonword targets.

One hundred thirty-five prime-target pairs (neither prime nor target was complex) were added as filler, to achieve a proportion of 30% related primes for real word targets (following the procedure of Rastle et al., 2000). Targets in this set were between 3 and 5 letters long, while primes were between 5 and 10 letters long, such that relative lengths were comparable to those of the experimental pairs.

Nonwords were obtained for the lexical decision task via three different methods. One hundred were created by permuting the consonant onsets with respect to the remaining strings of the target words until no string combinations created a real word, a near misspelling of a real word (SEAP), a common proper noun (DAVE) or a slang word (RAD). These nonwords were paired with real-word primes that were matched on frequency and length with the experimental primes. One hundred fifty additional nonwords, constrained to be within 3 to 6 letters in length, were taken from the English Lexicon Project (Balota et al., 2007), and paired with real words between 5 and 10 letters in length. Finally, 50 nonwords were extracted as the "stems" of pseudo-complex (e.g., BANT from BANTER or CAPT from CAPTION), and paired with their source word as a prime. Overall, each participant responded to 335 words (200 experimental and 135 filler) and 300 nonwords.

6.1.1.3 Procedure

All data collection was done remotely, using jsPsych (De Leeuw, 2015) for experiment presentation, the jspsych-psychophysics plugin (Kuroki, 2021) for masked priming trials presentation [1], and the jspsych-virtual-chinrest plugin to account for participants' screen size in determining stimulus font sizes (Li et al., 2020). Although participants sat at variable distances from the screen on which stimuli were presented, the height of presented text was consistently 0.8 cm; assuming distances between participants' eyes and their screens fell between 30 and 60 cm, this would yield a visual angle of between 0.8 and 1.5 degrees. The study was unmoderated: participants followed the link to the study at their convenience, and read instructions and completed the study in one sitting without experimenter supervision.

Following consent procedures and screen size calibration, participants read instructions that advised them to respond with whether each string of letters was a real word or not, as quickly and accurately as possible. The F and J keys were used for responses, and which key corresponded to "yes" was randomized for each participant. Sixteen practice trials were administered, after which participants received a report of their accuracy and reaction time, given feedback on their statistics (e.g., "Looks like you understand the task! But try to respond more quickly"), and given the option to repeat instructions and practice or begin the experiment. The mean reaction time threshold for recommending faster responses was 720 ms, and participants were advised to review the instructions if their accuracy was below 85%.

The main experiment was broken up by two optional 30-second breaks, such that

[1] The jspsych-psychophysics plugin allows visual stimuli presentations to synchronize with the refresh rate of the display. This is essential for ensuring that masked primes are displayed for the same or as similar as possible amounts of time across the various computers participants are using when they follow the experiment link. See the links at https://osf.io/pj4sb/wiki/home/ under "Histograms related to Table 4" for duration distributions of a stimulus set to last 50 ms across various computers and web browsers, with versus without the jspsych-psychophysics plugin.

participants made continuous lexical decisions for about 10 minutes at a time. The whole experiment lasted 35 minutes on average.

Lexical decision trials began with a fixation cross in the center of the screen for 500 ms, followed by a mask of hash marks (########) for 500 ms. The uppercase prime word was then flashed for 50 ms, and immediately followed by the lowercase target word. The target word was shown for 500 ms or until a response was given, and the trial ended when a response was given, or 2000 ms after the target word first appeared if no response was provided by that time. Participants were informed of the time limit for responses during instructions. A blank screen was shown for one second between trials. This masked priming procedure closely follows the procedure described in Jared et al. (2017).

Following the masked priming task, participants were asked about their experience of the masked priming trials as a check of the effectiveness of this paradigm in a remote context. They were first asked if they noticed anything else beyond the cross, mask, and target being shown; then they were asked to describe what they noticed, how often they noticed it, and finally how often they were able to recognize the uppercase word. Fifty-five percent of participants reported not noticing anything else other than the fixation cross, hash mask, and target word, indicating they were not aware of the 50 ms prime. Of the 25 participants who did detect the flashed stimuli, most only detected that there were letters being shown but were unsure if they formed actual words, and only 1 reported being able to consistently identify the prime word.

6.1.2 Results

Of the sixty-two participants recruited, 7 were excluded from analyses due to a high false positive rate (6 greater than 25%) or slow response times (1 greater than 900 ms). Eight targets, mis-categorized as nonwords by more than half of participants, were removed (1 from transparent set, 7 from opaque; see Appendix A). Trials where no response was saved due to the response time limit being exceeded (49 trials) and trials with incorrect

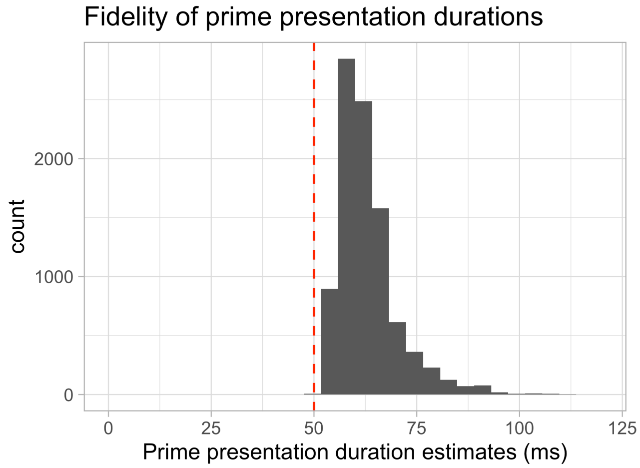

Figure 6.1: Histogram of prime presentation estimates for Study 1. The red dashed line denotes the intended prime duration (50 ms).

responses (1,050) were removed, as were reaction times less than 200 ms (20) and outliers (more than 3.5 standard deviations from by-participant mean; 61).

Estimates of prime presentation durations were generated using the by-trial `jsPsych` variable for amount of time elapsed since the participant began the study (see Figure 6.1). Looking at how well actual prime presentation durations aligned with designated duration (50 ms), three additional participants were excluded for inconsistent prime presentation durations (mean estimates greater than 75 ms or less than 50 ms). The participant who reported being able to consistently identify the prime was also removed, as described above. One hundred twelve trials were removed for having an estimated prime presentation of greater than 80 or less than 20 ms.[2] This left 8,581 responses to 99 transparent and 93 opaque targets, from 51 participants, for subsequent analyses.

The stimulus set was designed such that OSC was maximally varying, and minimally correlated with potentially confounding factors, *within* the two transparency conditions; however, the distribution of OSC values differed significantly between the transparent and opaque conditions. Accordingly, model fitting and statistical tests were conducted separately for these two sets of prime-target pairs. All models were fit using the `lme4`

[2]See Chapter 7 for a more detailed analysis of prime presentation fidelity.

(Bates et al., 2015) and `lmerTest` (Kuznetsova et al., 2017) packages in R (Version 4.0.3; R Core Team, 2020). Following the recommendation of Lo & Andrews (2015), reaction times were not transformed but instead were fitted with generalized linear mixed effects models, using an Inverse Gaussian distribution and an identity link function to capture the relationship between the predictors and the dependent variable.

6.1.2.1 OSC and semantically transparent morphological priming

The primary question of interest in this study is how the OSC of a morpheme impacts morphological priming magnitudes, and thus the interaction of OSC and priming magnitude is focal. However, as discussed in Chapter 5, target frequency appears to moderate the relationship between the newly calculated cluster-centered OSC and morphological priming magnitudes. Thus, the first model assessed for target responses in the semantically transparent condition included the three-way interaction of OSC, priming (related or unrelated) and target frequency. Morphological family size and target length were included as covariates (family size and target frequency were both taken from CELEX; Baayen et al., 1995), and the maximal random effects structure was used (Barr et al., 2013), including intercepts and slopes relative to the priming predictor for both participants and targets.[3] The three-way interaction was not found to be significant ($\beta = 20.12$, $p = 0.11$), so the model was re-fit with a two-way interaction of OSC and priming, including target frequency, morphological family size and length as covariates. The maximal random effect structure in this case only included random intercepts by participant. The simplified model showed a significant interaction of OSC and priming ($\beta = 34.05$, 0.017) as well as a significant main effect of priming ($\beta = 31.53$, $p < 0.0001$), but no main effect of OSC ($p > 0.5$). See Figure 6.2 for a visualization of this relationship.

[3]Most models did not converge using the maximal random effects structure: in these cases, random effects were simplified until the model did converge, following recommendations of Brauer & Curtin (2018).

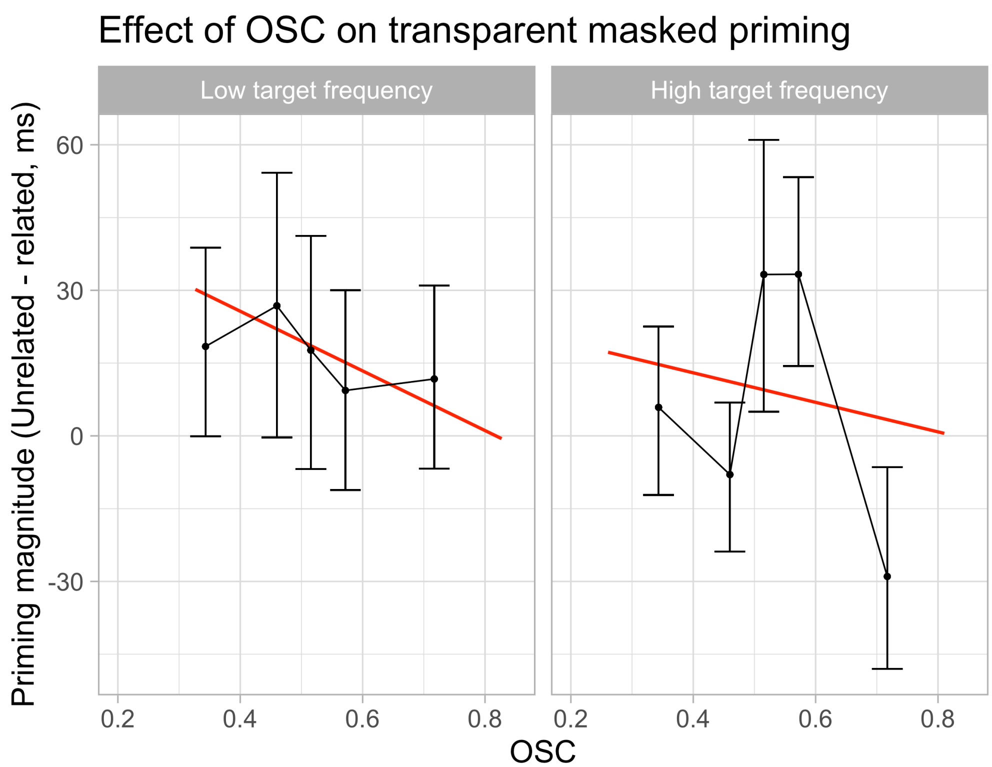

Figure 6.2: Relationship between cluster-centered OSC and transparent masked priming magnitudes in Study 1, median-split by target frequency. Each point represents mean priming magnitude by OSC quintile; error bars represent bootstrapped 95% confidence intervals. Red line shows line fit for by-item priming magnitudes. There was a significant effect of OSC on priming magnitude, and no significant interaction with target frequency.

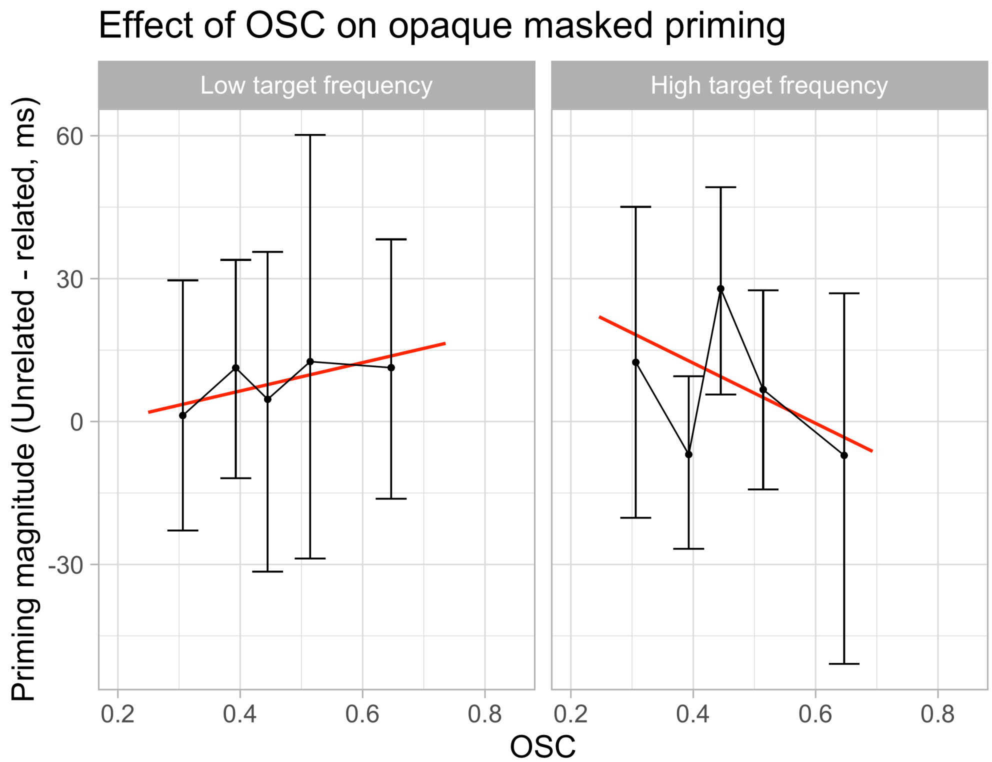

Figure 6.3: Relationship between OSC and opaque masked priming magnitudes, moderated by target frequency, in Study 1. The interaction with target frequency was significant in this case. See Figure 6.2 caption for details of figure generation.

6.1.2.2 *OSC and semantically opaque morphological priming*

Once again, the first model fit to responses for the semantically opaque prime-target pairs included a three-way interaction of OSC, priming, and target frequency, with family size and target length as covariates, and random slopes and intercepts for targets and participants. In this case, the three-way interaction was significant ($\beta = 44.47$, $p = 0.013$), as was the interaction of OSC and priming ($\beta = -86.56$,$p = 0.0028$), the main effect of priming ($\beta = 36.42$, $p = 0.015$), and the main effect of OSC ($\beta = 155.82$, $p < 0.0001$). This pattern of effects is depicted in Figure 6.3.

6.1.2.3 *Post-hoc comparisons for semantically opaque pairs*

To further investigate the nature of the OSC-priming interaction for lower-frequency targets (as shown on the left in Figure 6.3), models were fit to the lower-frequency and higher-frequency targets separately (median split). Target frequency was removed as a covariate, and the random effects maximal structure in both cases was random intercepts by participant. The interaction of OSC and priming magnitude was marginally significant for higher-frequency targets ($\beta = 38.56$, $p = 0.10$) and not significant for lower-frequency targets ($\beta = 21.44$, $p = 0.59$). It is worth noting that in these models, a positive interaction of OSC and priming signifies a reduction in priming magnitude with larger OSC values (as reaction times for related prime-target pairs are considered relative to unrelated pairs). Thus the direction of the marginally significant interaction for high-frequency targets aligns with the relationship depicted in Figure 6.3, while that of the non-significant interaction for low- frequency targets does not.

6.1.3 Discussion

The cluster-centered OSC metric is intended to quantify the degree to which a given letter string is semantically useful; in other words, how far on the spectrum from morpheme to non-morpheme does it fall? Given this view, we would expect morphological priming magnitudes to increase for prime-target pairs sharing higher OSC letter strings. This is not what we observe in the present study: OSC shows an overall negative effect on transparent priming magnitudes, and in the case of opaque priming, OSC interacts with target frequency to show neutral and negative effects for low- and high-frequency targets, respectively. How can we reconcile this pattern of effects with our conceptualization of morphological processing?

One perspective to consider is that this metric is orthographically-informed, and participants are all skilled adult readers. With experience, the process of recognizing a whole word form and retrieving its meaning becomes more and more efficient, and semantic and morphological information about the prime is accessed earlier. Particularly given that the prime presentations were overall slightly longer than intended (mostly falling between 55 and 70 ms), it is possible that the transparent prime is sufficiently processed for its semantic relation to the prime to have come into effect. If this is the case, a salient high-OSC substring within the prime may be a source of processing conflict, detracting from the prime's processing of whole-word characteristics that would most facilitate target recognition. A comparable interaction of stem- and word-level characteristics was demonstrated by Baayen et al. (2007) with prime and stem frequency: for high-frequency primes, stem frequency is inversely related to reaction times. The fact that there is a trend toward an interaction with target frequency ($\beta = 20.12$, $p = 0.11$) suggests the inhibitory effect of OSC is a bit weaker for higher-frequency primes. This makes sense as the amount of priming that can occur is constrained for high-frequency targets, as responses for these words are already quite fast.

Rapid access to semantic and morphological information about the prime would only clarify the pattern of results seen for transparent priming, not opaque priming. The negative effect of OSC on opaque priming magnitudes for high-frequency targets may have more to do with ceiling effects: given that high-frequency targets are already recognized quickly, ones that also have high OSC may be responded to even more quickly and thus be more difficult to prime. In the case of low-frequency opaque primes, a positive effect of OSC on priming magnitude could be canceling out or overcoming the increasing difficulty of priming higher-OSC targets, resulting in a neutral overall effect.

Despite these potential explanations, it is evident that the relationship of OSC and priming magnitudes is not a simple one, and the current pattern of results do not suggest that OSC generally boosts morphological priming. To tease apart what aspects of these results may be unique to the masked priming paradigm (and, correspondingly, to lower-level and more orthographically informed phases of word processing), we repeated this study with an overt, unmasked prime.

6.2 STUDY 2: THE EFFECT OF OSC ON OVERT MORPHOLOGICAL PRIMING

In this follow-up study, we look at how OSC interacts with both opaque and transparent morphological priming in the context of an *overt* prime, shown for 230 ms with no preceding mask.

6.2.1 Methods

6.2.1.1 Participants

Fewer participants were needed for this study, as patterns of overt priming are notably easier to detect than those of masked priming. Forty-three undergraduates were recruited from the same private university in Western Pennsylvania. As in Study 1, all participants were required to be native speakers of English with normal or corrected-to-normal vision, and they received course credit for participation.

6.2.1.2 Stimuli

Stimuli were identical to those used in Study 1.

6.2.1.3 Procedure

The procedure for this study was identical to that of Study 1, with the following exceptions:

- No hash mask was presented on any trial; instead, the prime word immediately followed the fixation cross.
- The prime was shown for 230 ms instead of 50 ms.
- Instructions emphasized that participants should categorize the second, lowercase string as a word or nonword, as the prime was now overtly visible.

- Participants were not asked about whether they noticed the prime word after finishing, as primes were presented for significantly longer with the intention of being perceived.

As in Study 1, the whole study took 35 minutes on average, with roughly 9 minute blocks of lexical decisions between breaks.

6.2.2 Results

Of the 42 participants recruited, 11 with false positive rates over 25% and 1 with mean reaction time over 900 ms were removed from further analysis. Twelve targets were removed for being responded to incorrectly by more than 50% of participants: 1 from the transparent condition and 11 from the opaque condition (see Appendix A). Trials where no responses was saved due to the response time limit being exceeded (70 trials) and trials with incorrect responses (600) were removed, as were reaction times less than 200 ms (1) and outliers (more than 3.5 sds from by-participant mean; 20).

Looking at prime presentation duration estimates, no additional participants needed to be excluded due to inconsistent prime presentation durations (all mean estimates between 200 and 300 ms). Sixty-one trials were removed for having an estimated prime presentation of greater than 300 ms or less than 200 ms. This left 4,889 responses to 99 transparent and 89 opaque targets, from 30 participants, for subsequent analysis.

Statistical models were fit following the same approach described for Study 1: separately for opaque and transparent pairs, using generalized linear mixed effects models with Inverse Gaussian distribution and identity link function.

6.2.2.1 OSC and semantically transparent morphological priming

The model of transparent morphological priming reaction times included a three-way interaction of OSC, priming, and target frequency, with morphological family size and target length as covariates, as in Study 1. The maximal random effects structure included

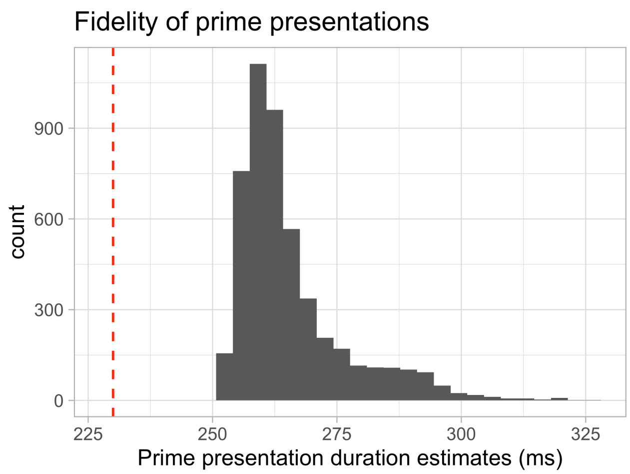

Figure 6.4: Histogram of prime presentation estimates for Study 2. The red dashed line denotes the intended prime duration (230 ms).

random intercepts by target and random intercepts and slopes relative to priming by participant.

The three-way interaction of OSC, priming and target frequency was significant ($\beta = -70.18$, $p < 0.0001$) as was the overall effect of OSC on priming magnitudes ($\beta = 264.50$, $p < 0.0001$) and the main effects of OSC ($\beta = -73.45$, $p = 0.022$) and priming ($\beta = -210.64$, $p < 0.0001$). See Figure 6.5 for a visualization of this pattern of effects.

6.2.2.2 *OSC and semantically opaque morphological priming*

Responses to opaque prime-target pairs were also fit with a model including the three-way interaction of OSC, priming, and target frequency, with family size and target length as covariates. The maximal random effects structure matched that for the transparent priming model, including random intercepts by target and random intercepts and slopes relative to priming by participant.

The three-way interaction was not significant ($p > 0.35$); a new model fit with target frequency as a covariate instead of a moderator revealed no interaction of OSC and priming, either ($p > 0.5$). However, a model fit to examine the main effect of priming, with OSC, target frequency, family size and target length as moderators, revealed a significant

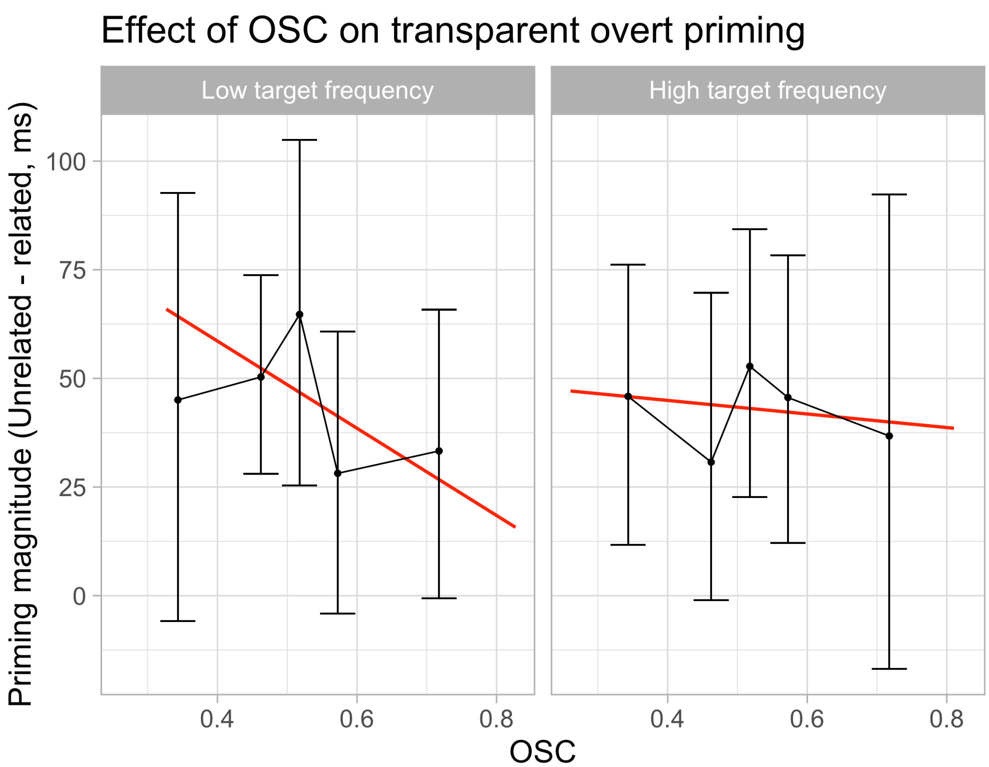

Figure 6.5: Relationship between OSC and transparent overt priming magnitudes, moderated by target frequency, in Study 2. The three-way interaction was significant. See Figure 6.2 caption for details of figure generation.

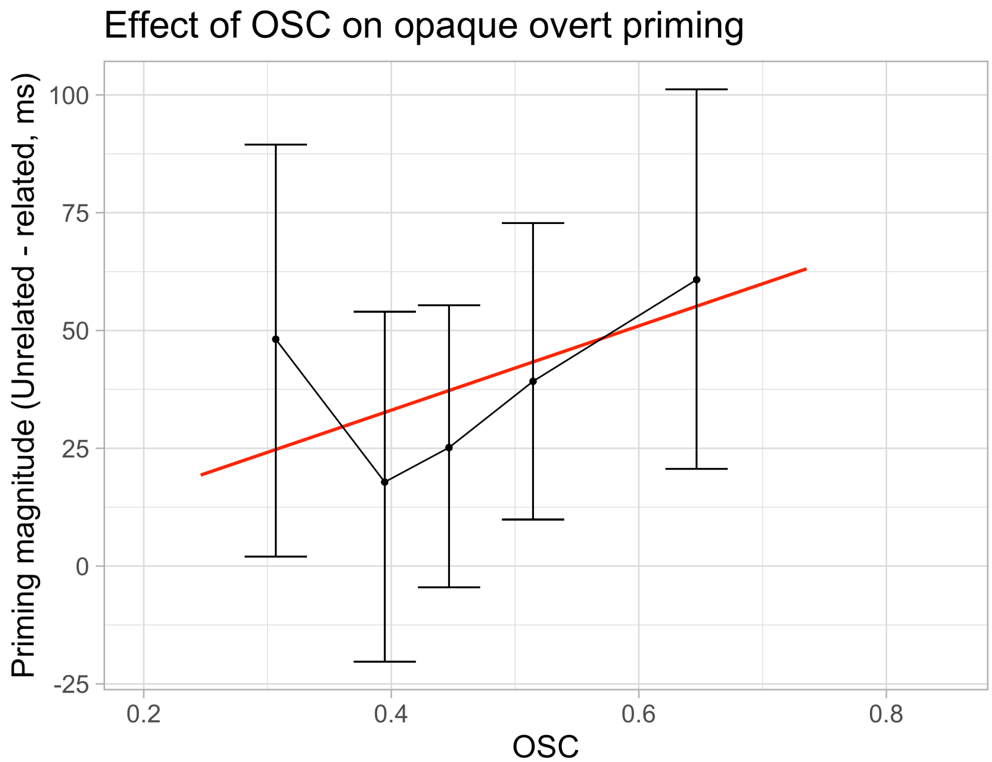

Figure 6.6: Relationship between OSC and opaque overt priming magnitudes in Study 2. The interaction of OSC and opaque priming was not significant, whether moderated by or unmoderated by target frequency. See Figure 6.2 caption for details of figure generation.

effect of opaque morphological priming ($\beta = -37.83$, $p < 0.0001$). See Figure 6.6 for a visualization of this pattern of effects.

6.2.3 Discussion

Once again, we see a negative relationship between transparent priming and OSC. It's possible that this effect is still due to an interference effect of stem OSC, detracting from beneficial whole-word processing, as discussed in relation to Study 1. However, the pattern in this case could also be explained entirely by target ceiling effects. Given that transparent primes are shown for longer in Study 2, priming may already be quite large regardless of OSC. Thus, for low-frequency targets, increasing OSC pushes target response times towards the response speed limit, due to this metric's significance for the target as opposed to the prime-target relationship, reducing room for priming to occur. High-frequency targets following overt transparent primes are already closer to that limit, and thus the negative effect of OSC on magnitudes is weaker for these targets, as there is less reduction in priming magnitude to be had.

Given the nature of the OSC metric, capturing a property of the stem across all its occurrences with no regard for morphological relatedness or semantic context, it seems unlikely that stem OSC should interfere with prime processing once the prime is presented for a full 260 milliseconds (given that prime presentations were slightly longer than intended). Additionally, this explanation is supported by the statistical model for the transparent overt data: target frequency had a strong negative main effect on reaction time (-35.79 ,$p < 0.0001$), whereas this effect was notably weaker for opaque overt priming data (For the three-way interaction model: 13.59, $p > 0.3$; for the main effect model: -14.59, $p = 0.04$).

OSC did not have any reliable effect on priming magnitudes in the case of opaque overt priming, regardless of target frequency. If our ceiling effects explanation for the transparent pattern of overt priming effects is correct, lack of an OSC effect here is probably due to the fact than an opaque prime is not as facilitatory as a target prime, and thus the additional constraint of target OSC on priming magnitudes is not detectable in this case. (This assumes a non-linear relationship between ceiling effects due to target OSC and those due to target frequency.) As noted above, OSC was not expected to facilitate priming in this case because it is a more orthographically-informed metric whose information would presumably be overrided by the prime's whole-word information after 230 ms (250 ms) of display time.

We did, however, observe a significant main effect of opaque priming, suggesting that sensitivity to opaque primes can in fact be detected in an overt priming context. This is surprising because opaque overt priming is typically found only in morphologically rich languages like Hebrew (Frost et al., 2000), Arabic (Boudelaa & Marslen-Wilson, 2001), and German (Smolka et al., 2015), and not in English (Rastle et al., 2000). Indeed, that expectation formed the premise of the modeling work presented in Chapter 4.

6.3 DISCUSSION

The two empirical studies reported in the current chapter examined the effects of OSC on morphological priming in adults, both under brief masked conditions (Study 1) and under overt, unmasked conditions (Study 2). Although results were somewhat difficult to interpret, this sequence of studies afforded several insights which will inform subsequent chapters of this thesis. First, patterns of results undeniably demonstrate sensitivity to the relatedness of the prime, even when it is masked, supporting this remote paradigm as a viable means of investigating masked priming effects. However, prime presentation times appear to be noisy, and participants are less motivated to stay on task (inferred from their slower reaction times and lower accuracy), so data cleaning and analyses must be conducted very carefully.

Second, the only OSC effects on priming magnitude that were observed were negative. We suspect that this finding has more to do with the effect of OSC on recognition of the target, and resulting ceiling effects, than with the effect of OSC on the prime-target relationship. The fact that the pattern of effects for transparent prime-target pairs is the same in both masked and overt priming contexts gives us reason to believe that the masked transparent effects can actually also be attributed to ceiling effects, instead of the stem versus whole-word interference explanation provided in the discussion for Study 1. Thus, facilitatory effects of OSC on morphological priming, if they exist, may be quite difficult to detect.

Taking an optimistic perspective, we note that the context in which we would most expect to see a facilitatory effect of OSC, opaque masked priming of low-frequency targets, did not show a negative effect. It is possible that in this case, a positive effect of OSC on morphological priming interacts with the ceiling effects described, resulting in no detectable effect. The issue of ceiling effects is certainly a challenge for interpreting the role of OSC; one possible solution is to investigate their effects in a population whose reaction times for lexical decision are slow. This, combined with the predicted role of OSC in acquiring morphological sensitivity, motivates the exploration of OSC effects on morphological priming in developing readers in Chapter 8.

Third, we found strong evidence of overt, opaque priming in Study 2. As mentioned, this is a surprising effect to encounter in a less morphologically rich language like English, and contradicts the results of Rastle et al. (2000). Given that there was no effect of OSC, the fact that stimuli were selected to maximize variance in OSC is unlikely to explain why opaque overt priming was found in this study and not in others. It's worth noting that only a few studies have demonstrated a lack of overt opaque priming effects in English. In addition to Rastle et al. (2000), Feldman et al. (2004) found an effect of semantic transparency on priming magnitudes and no effect of opaque priming for an overt visual prime (200 ms), and Feldman & Soltano (1999) found opaque and transparent priming for a 48 ms prime but only transparent priming for a 250 ms prime. Both of these studies used complex prime-target pairs with a shared stem or pseudo-stem (e.g., ACCORDANCE-ACCORDING versus ACCORDANCE-ACCORDION), as opposed to a complex prime paired with simple target (ACCORDANCE-ACCORD) as is done in the majority of morphological priming studies. Rueckl & Aicher (2008) also investigated semantic transparency effects, but in the context of long-term priming, which would be expected to differ from immediate priming. Besides these, most studies investigating semantic transparency effects on overt priming have been cross-modal, with the prime presented aurally (e.g., Marslen-Wilson et al., 1994; Gonnerman et al., 2007), and the patterns of effects may well differ when primes are visually or aurally presented. The main effect of overt priming in Study 2 suggests that it may be premature to conclude that there is no overt visual priming for opaque prime-target pairs. Given the central nature of opaque priming to the controversy between distributed and decomposition viewpoints, and consequently its central nature to the current thesis, we decided to investigate the robustness of the results reported by Rastle et al. (2000) with a close, remote replication study, reported in the next chapter.

CHAPTER 7

Clarifying the Time-Course of Morphological Processing in Skilled Adult Readers

Understanding how represented information changes over the time-course of visual word recognition provides a window into the mechanisms of this process. As discussed in Chapters 2 and 4, varying the duration for which a prime is presented prior to a target is one approach to identifying at what point in processing different prime-target relationships become more relevant, or in other words what aspect of the information carried by the prime is currently most dominant in how it is being represented.

The findings reported in Rastle et al. (2000) provided a beneficial overview of patterns of effects due to various word characteristics as a word is recognized, by assessing priming magnitudes for three different prime presentation durations: 43 ms, 72 ms, and 230 ms. They found strong and consistent priming by identity (STRONG-STRONG) and morphologically transparent (TEACHER-TEACH) pairs across all three SOAs, while orthographic priming (CASHEW-CASH) diminished and semantic priming (SKIP-MISS) strengthened with longer prime durations (see the left side of Figure 7.2 for a visualization of effects reported in this study). These patterns align with our intuition that earlier word processing is more sensitive to orthographic factors such as shared letters, while later word processing is more attuned to semantic characteristics. The pattern of effects for semantically opaque morphological prime-target pairs was particularly interesting, in that opaque priming was similar in magnitude to transparent priming for 43 ms SOA, but decreased to be comparable to orthographic priming at subsequent SOAs. Though opaque priming was not found to be significantly stronger than orthographic priming, it did diminish significantly with respect to transparent priming as SOA increased while the same interaction was not significant for orthographic with respect to transparent. This evidence for a short-lived sensitivity to semantically opaque morphological primes, more directly presented by Rastle et al. (2004), sparked new speculation regarding early stages of complex word processing.

The current chapter describes a close replication of Rastle et al. (2000), to assess the robustness of this pattern of results. Given the implications of these results for larger-scale interpretations of how word processing dynamics unfold, it is important to know whether they replicate, particularly given our reported findings of positive overt priming by semantically opaque morphological primes in Chapter 6. The most notable difference between this study and Rastle et al. (2000) is the fact that the current study was conducted remotely: thus, the current chapter also provides an opportunity to assess how similar results for remote masked primed lexical decision studies are to in-person instances with otherwise very similar stimuli and procedures.

7.1 METHODS

7.1.1 Participants

Participants were 96 undergraduate students at a private university in Western Pennsylvania, who received course credit in exchange for completing the study. All participants had normal or corrected-to-normal vision, and were native speakers of English. More participants were recruited than in the original study (24 more participants overall, 8 more per SOA), primarily due to the unmoderated remote nature of data collection as well as noisier stimulus presentation (demonstrated in Chapter 6).

7.1.2 Stimuli

Experimental prime-target pairs were taken directly from Rastle et al. (2000). This included 24 pairs in each of five conditions: 1) semantically transparent morphological, referred to as "morphologically, semantically, and orthographically related", or [+M+S+O], in Rastle et al. (2000); 2) semantically opaque morphological [+M-S+O]; 3) semantic [-M+S-O]; 4) orthographic [-M-S+O]; and 5) identical (e.g., cape-CAPE). During original stimulus selection by Rastle et al. (2000), morphological relationships between prime and target were established using the Oxford English Dictionary, and semantic relationships were confirmed using participant ratings and LSA ratings (Landauer & Dumais, 1997).

The exact unrelated primes, filler words, and nonwords used by Rastle et al. (2000) were not provided; as such, these items were generated following their described process. For each of 120 targets, an unrelated prime was selected from the SUBTLEX-US vocabulary (Brysbaert & New, 2009) that had no semantic or morphological relation to the target, matched the related prime in length, and did not share letters in the same position as the target. Sixty filler primes and targets with word lengths within the range of experimental primes and targets, frequencies were sampled from SUBTLEX-US, to reach a 30% ratio of related to unrelated primes among word targets.

One hundred eighty pairs with nonword targets were generated. Twenty four nonword targets were pseudo-stems of real words (e.g., BANT from BANTER or LEOP from LEOP), paired with their source word. One hundred fifty-six pronounceable nonwords were selected from the English Lexicon Project, length-matched to the real word targets (Balota et al., 2007); 24 were primed by themselves (slint-SLINT) and the rest were paired with an orthographically unrelated word, length-matched to real word primes (e.g., slither-BRAFT). This resulted in 360 trials for lexical decision: 180 nonwords and 180 words.

7.1.3 Procedure

Thirty-two participants were assigned to each of the three SOA conditions. In each SOA, 16 participants were presented with list 1 and 16 were presented with list 2.

As in the studies presented in Chapter 6, data collection was remote and unmoderated. Stimuli were presented using `jsPsych` (De Leeuw, 2015), with the `jspsych-psychophysics` plugin (Kuroki, 2021) for masked priming trials presentation and the `jspsych-virtual-chinrest` plugin to accommodate participants' varying screen sizes (Li et al., 2020).

Participants were instructed to decide for each target string whether it was a word

("Press the J key") or not ("Press the F key"). Participants assigned to the 230 ms SOA were told that they would briefly see a lowercase string before the uppercase string, and to only respond to the uppercase string. Participants assigned to the 43 ms and 72 ms SOAs were not informed of the presence of the lowercase prime. Sixteen practice trials were administered prior to beginning the main experiment. Participants were given feedback on the speed and accuracy of their practice performance, and given the option to repeat instructions and practice trials. The mean reaction time threshold for recommending faster responses was 720 ms, and participants were advised to review the instructions if their accuracy was below 85%.

Lexical decision trials began with a fixation cross in the center of the screen for 500 ms, followed by a mask of hash marks (########) for 500 ms. The prime word was then flashed for the specified time (set to 43, 72, or 230 ms) and immediately followed by the target word. The target word was shown until a response was given, or until the trial ended if no response was provided after 2 seconds of target display. Participants were informed of the time limit for responses during instructions. A blank screen was shown for one second between trials.

The study took 23 minutes to complete, on average. Lexical decisions were split into three 6-minute chunks with optional, 30 second breaks in between.

As in the masked priming study in Chapter 6, participants in the 43 and 72 ms conditions were asked about their experience of the flashed prime word. Sixty-three percent of participants in the 43 ms condition reported that they didn't notice anything else besides the fixation cross, hash mask, and target word. Of the 12 who did report noticing something, 9 noticed that the flashed stimuli was a character string of some sort, but all said they either never recognized the word or only recognized the word on a few trials. For the 72 ms condition, 41% of participants didn't notice the prime stimulus. Of the 19 who did, all noticed that the flashed stimulus was a word, 10 reported seeing it in all or most trials, and 4 reported being able to recognize the flashed word most of the time. As the 72 ms SOA is on the edge of prime durations we would expect to be fully masked, the participants who reported being able to recognize the word were not excluded from main analyses.

7.2 RESULTS

The analyses described below address both objectives outlined in the introduction to this chapter: to examine the robustness of the time-course effects reported by Rastle et al. (2000), and to assess the soundness of remote data collection for investigating morphological processing effects with the primed lexical decision paradigm.

7.2.1 Data cleaning

The approach to trimming data was slightly different from that of Rastle et al. (2000), as participant responses were overall slower and less accurate in the current study. This difference can likely be attributed to the remote context of data collection. One participant was removed for having a mean reaction time (RT) of over 1000 ms (this threshold was 800 ms in the original study), and 8 participants were removed for having a false positive rate of over 25% (original cut-off was 20%). This left 87 participants: 33, 26, and 28 for the 43 ms, 72 ms, and 230 ms SOAs, respectively. Reaction times were removed if the response was incorrect or not recorded due to exceeding the trial time limit (632 responses), or if they were greater than 1550 ms (20 responses, matching the removal of 0.25% of responses using a cut-off of 1400 ms in Rastle et al. (2000)). The recorded data was then examined to determine how faithfully prime presentation times on participants' computer screens adhered to their assigned duration (43, 72 or 230 ms).

7.2.2 Checking fidelity of prime presentation durations

The `JsPsych` package saves the average number of milliseconds spent on each frame (`avg_frame_time`) and the number milliseconds since the participant began the study

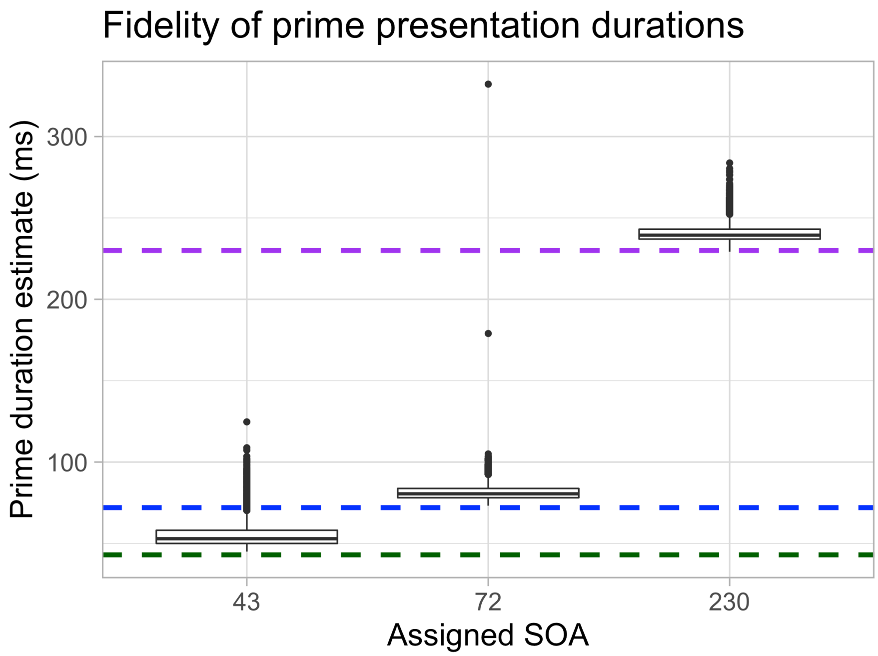

Figure 7.1: Confirmation that prime duration estimates mostly align with the desired value across SOA conditions. Trials with prime duration estimates more than 25 ms from the desired value (e.g., prime estimates greater than 68 ms in the 43 ms SOA) were removed from analysis.

(`time_elapsed`) for each trial (De Leeuw, 2015). The first variable was used to calculate the average duration of a single display frame for each computer used to complete the study. All participants' frame durations were at or below 16.67 ms except one individual in the 43 ms SOA (mean frame duration of 33.3 ms), who was removed from further analyses (leaving 86 participants total). Such a slow refresh rate would make it impossible for this individual's computer to reliably differentiate between 43 and 72 ms presentation durations.

Prime presentation duration was then estimated for each trial by using `time_elapsed` to infer trial duration, and then subtracting the amount of time needed to display pre-prime stimuli (2000 ms) and recorded reaction times. The resulting estimates of prime presentation times for each SOA are shown in Figure 7.1. Trials with prime duration estimates more than 25 ms from the assigned value (e.g., prime estimates greater than 68 ms for the 43 ms SOA) were removed from analysis, resulting in the removal of 11.5%, 4.5%, and 4.9% of trials from the 43 ms, 72 ms, and 230 ms SOAs, respectively.

Prime durations in the 43 ms SOA condition, though mostly concentrated around 50 ms, trended towards being shorter than those in the masked priming study with prime durations set to 50 ms, presented in Chapter 6 ($\beta = -3.95$, $p = 0.081$). For the 43 ms SOA, 2,696 prime estimates were less than 50 ms, relative to only 3 prime estimates less than 50 ms for the 50 ms study. Given that these values are estimates (relying on the assumption that all other stimuli — e.g., fixation cross, hash mask — are presented for exactly the intended duration), there may have been a higher proportion of trials for which primes were presented too briefly to be processed in the current study than in the prior study (this possibility is returned to in the discussion).

Applying the filters for speed, accuracy, and display precision described above resulted in 8,767 reaction times for use in subsequent analyses.

7.2.3 Main analyses

Following the analyses described in Rastle et al. (2000), the data was analyzed using a mixed-design ANOVA with four factors: SOA (3 levels), condition (5 levels), list (2 levels), and priming (2 levels). For the by-participants model, condition and priming were repeated factors; for the by-items model, SOA and priming were repeated factors.

The main effect of priming was significant for both the by-participants model ($F_p(1, 85) = 97.45$, $p < 0.0001$) and the by-items model ($F_i(1, 115) = 45.47$, $p < 0.0001$), indicating that response times for words presented after related primes differed significantly from those presented after unrelated primes. The interaction of condition and priming was also significant for both models ($F_p(4, 336) = 6.42$, $p < 0.0001$; $F_i(4, 111) = 3.21$, $p = 0.016$),[1] as was the three-way interaction of condition, priming, and SOA ($F_p(8, 328) = 2.90$,

[1]For all by-participants models other than those isolating main effects of priming, one participant was removed from analysis because they did not have trials for all combinations of condition and priming. Thus, results from these models based on 85 as opposed to 86 participants.

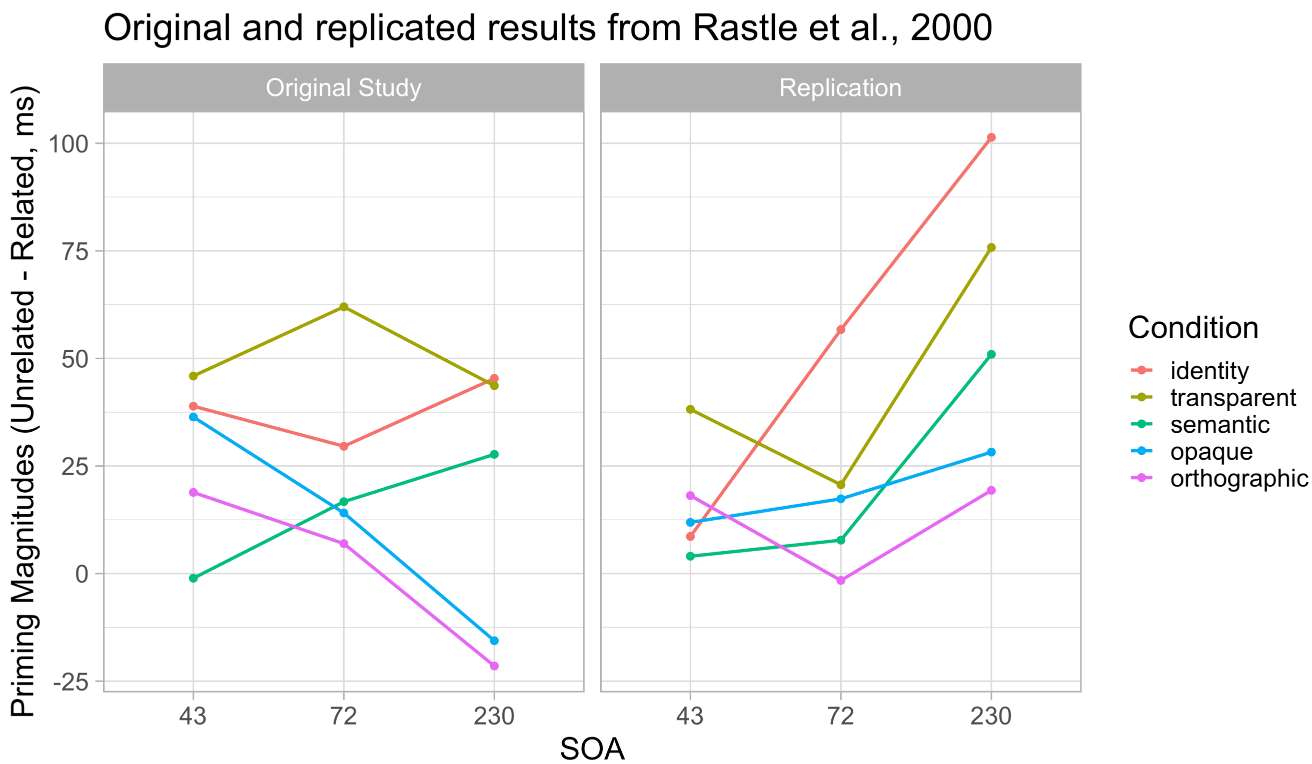

Figure 7.2: Comparison of priming magnitudes across condition and SOA from Rastle et al. (2000)] and the current, close replication study. The left axis represents amount of priming, calculated as mean reaction time for a target following a related prime, subtracted from mean reaction time for that target following an unrelated prime.

$p = 0.0039$; $F_i(8, 222) = 2.74$, $p = 0.0067$). There was no significant four-way interaction of condition, priming, SOA and list (both ps > 0.5). Thus, priming magnitudes varied significantly depending on the relationship between prime and target (transparent, orthographic, etc.) and this pattern of effects depended on the duration of prime presentation. See Figure 7.2 for patterns of effects found in the current study compared with those of Rastle et al. (2000).

A series of planned contrasts further illuminated the nature of the three-way interaction of priming, condition, and SOA:

- Looking only at the transparent and identity priming conditions, the interaction between priming and condition was not significant (both ps > 0.25), but there was a three-way interaction of priming, condition, and SOA ($F_p[2, 82] = 7.40$,$p = 0.0011$; $F_i[2, 88] = 5.16$, $p = 0.0076$), signifying that the difference between transparent and identity priming varied with SOA. This interaction differs from findings in Rastle et al. (2000), and likely captures the flip from stronger transparent priming to stronger identity priming from the 43 ms to the 72 ms prime duration. When data from participants who were shown primes for 43 ms were removed from this contrast, the three-way interaction was no longer present (both ps > 0.35) and the interaction of priming and condition became significant ($F_p(1, 53) = 9.87$,$p = 0.0028$; $F_i(1, 44) = 4.53$, $p = 0.039$), confirming that identity priming was consistently stronger than transparent priming for 72 ms and 230 ms SOAs.

- For transparent and semantic conditions, an interaction of condition and priming (significant by-participants: $F_p[1, 84] = 8.22$, $p = 0.0053$; marginal by-items: $F_i[1, 46] = 3.39$,$p = 0.072$) was found, indicating that transparent priming was overall stronger than semantic priming. The three-way interaction of priming, condition and SOA was not significant (both ps > 0.45). These results match findings from Rastle et al. (2000).

- Comparing transparent and orthographic conditions, there was a significant interaction of priming and condition ($F_p[1, 84] = 14.61$, $p = 0.00025$; $F_i[1, 45] = 5.93$, $p = 0.019$) such that transparent priming is overall stronger than orthographic priming. The three-way interaction of priming, condition, and SOA did not reach significance ($F_p[2, 82] = 2.02$, $p = 0.14$; $F_i[2, 90] = 1.93$, $p = 0.15$). Both results align with those of Rastle et al. (2000).

- Considering the transparent and opaque conditions, transparent priming was stronger than opaque priming overall (significant by-participants: $F_p[1, 84] = 6.43$, $p = 0.013$; marginal by-items: $F_i[1, 45] = 3.62$, $p = 0.064$), but this difference did not vary with SOA ($F_p[2, 82] = 1.11$, $p = 0.34$; $F_i[2, 90] = 1.82$, $p = 0.17$). Rastle et al. (2000)

similarly found stronger transparent priming than opaque priming overall, but also found opaque priming decreasing significantly relative to transparent priming with increasing SOA.

- For the opaque and orthographic conditions, a main effect of priming was found ($F_p[1,84] = 10.24$, $p = 0.0019$; $F_i[1,45] = 5.27$, $p = 0.026$) but neither the interaction of priming and condition nor that of priming, condition and SOA were significant (all ps > 0.25), mirroring results of Rastle et al. (2000). However, unlike in the original study, there was no change in priming magnitude with increased prime duration ($F_p(2,82) = 1.39$, $p = 0.26$; $F_i(2,90) = 0.88$, $p = 0.42$) for these conditions.

7.3 DISCUSSION

The current chapter reports an attempt to replicate Rastle et al.'s (2000) findings of the effects of SOA on morphological priming—and, in particular, opaque priming—but using remote testing. Overall, the primary differences between the results of Rastle et al. (2000) and those presented here are that, first, identity priming is oddly low for the 43 ms SOA and, second, opaque and orthographic priming do not decrease in magnitude with greater SOA. The identity priming difference may be due to the nature of remote masked priming presentation: overall, the effect of condition on priming magnitudes appears less dramatic than in the original results. With the masked prime set to 43 ms (instead of 50 ms as in Chapter 6), there may be a higher proportion of trials when the prime was shown too briefly to be processed, or no prime was actually shown, and our analysis procedures using the `time_elapsed` in `jsPsych` to estimate prime presentation times may not be precise enough to fully account for this. Alternatively, in the relaxed atmosphere of a remote study, participants may not be sufficiently focused to show consistent sensitivity to a 43 ms (as indicated by their slower responses and higher false positive rates relative to participants in Rastle et al., 2000). Either way, filtering for just the 43 ms trials in this study, the interaction of priming and condition showed only a trend toward significance ($F_p(4, 120) = 2.00$, $p = 0.099$; $F_i(4, 111) = 1.88$, $p = 0.12$). Given the larger sample size and more prime-target pairs per condition, this was not an issue for the masked priming study reported in Chapter 6, but it does suggest that the configuration of transparent and identity priming for 43 ms SOA should not be overinterpreted.

The lack of a decrease in opaque and orthographic priming is interesting, and was the primary purpose of the current replication study. These results suggest that, at least in the context of a complex word priming its simple stem (see Feldman & Soltano, 1999; Feldman et al., 2004, for alternatives), semantically opaque and orthographic effects in an overt visual priming context may be more facilitatory than the pattern found by (Rastle et al., 2000).

CHAPTER 8

The Role of Consistency in the Emergence of Morphological Processing in Developing Readers

The developmental literature on complex word reading primarily focuses on conscious reasoning about words' structure, referred to as *morphological awareness* (e.g., Nagy et al., 2006). This is a distinct area of research from morphological *processing*, which primarily focuses on processes occurring before the word has been recognized. Morphological processing research in developing readers is much more limited. In English, only two studies to our knowledge explore masked priming effects in young readers (Dawson et al., 2021; Beyersmann et al., 2012a), and these studies used the same set of stimuli. Both studies detected transparent morphological masked priming as early as 3rd grade (around 9 years old), whereas opaque morphological priming is not detected until late middle school or early high school (around 14 years old). Additionally, participants' word recognition skill better predicts when these effects emerge than does their age (Dawson et al., 2021).

As mentioned previously, several theories of morphological processing propose that morphological sensitivity emerges due to the detection of letter sequences that are associated with similar meanings across contexts, providing "islands of regularity" that can facilitate semantic access in some or most words in which they appear (e.g., Plaut & Gonnerman, 2000; Rastle & Davis, 2008). Thus, orthographic-semantic consistency (OSC) may be a useful metric to predict which morphemes impact visual processing earliest in developing readers.

In the current cross-sectional developmental study, readers from 4th to 9th grade (ages 9 to 16) completed a masked primed lexical decision task. Prime-target pairs were selected for maximally varying stem OSC, comparable to those used in Chapter 6, to allow examination of how the effect of OSC on priming changes with reading experience. Priming effects in less experienced readers are expected to exhibit greater sensitivity to the OSC of prime-target pairs' shared morpheme than those in more experienced readers, particularly in the case of opaque priming. In addition to providing insight regarding the emergence of morphological processing effects, the current study provides another opportunity to examine how (and whether) OSC interacts with morphological priming, in a sample that is potentially less susceptible to the ceiling effect issues encountered in Chapter 6.

The same stimuli were presented in an overt priming context to a separate cohort of 4th and 5th graders. This condition was included to more closely examine the time-course of morphological processing when sensitivity to morphology is less fully formed. French-reading children in 3rd grade have been found to show sensitivity to opaque primes (e.g., BAGUETTE, "breadstick", priming BAGUE, "ring") for both short (60 ms) and long (250 ms) prime duration (Quémart et al., 2011). However, 5th graders, and 7th graders, and adults show such priming only for the 60 ms prime duration. These findings contrast with results in English, where opaque masked priming is not detected until around 7th grade. The pattern of effects found in Quémart et al. (2011) raises the possibility that opaque priming does occur in English, but can only be detected for longer prime durations due to the slower processing of lexical information in less experienced readers. Additionally, the simulated developmental trajectory of opaque priming for the morphologically rich language in Chapter 4 supports the possibility that opaque priming emerges in overt priming context, or both masked and overt priming contexts, at first. Given that no other work on morphological processing acquisition has manipulated prime duration within the same study, we considered it worthwhile to include an overt condition for the youngest age group in the current design. An overt priming condition was not included for the older age groups (6th to 9th grade) due to greater difficulties recruiting enough participants in these age ranges.

8.1 METHODS

8.1.1 Participants

One hundred twenty-one participants between 4th and 9th grade were recruited for this study. This sample broke down as 59 4th and 5th graders (mean age: 10.2), 33 6th and 7th graders (mean age: 11.9), and 29 8th to 9th graders (mean age: 14.2). Participants had normal or corrected-to-normal vision, and all of them first learned to read in English. Four parents reported that their child had learned another language prior to English, but that their child had first learned to read in, and was currently more proficient at reading in, English. Given that recruitment was conducted via online advertisements as opposed to in a more uniform setting, such as a school, this sample was likely biased towards higher-performing readers for their age, who would be more willing to participate in a reading-related task in their free time.

Recruitment was conducted via Facebook (posting in groups related to parenthood, as well as sharing via personal pages) and the Children Helping Science website (https://childrenhelpingscience.com/). For 70% of participants, the parent met with the researcher via phone or video call to answer questions about their children's reading experience; the primary purpose of this call was to verify that they had a child who was eligible for the study.[1] Participants were sent a $10 gift card (via email to their parent) as a thank-you for their time and effort.

8.1.2 Stimuli

The process for selecting prime-target pairs was very similar to that described in Chapter 6: the Stochastic Optimization of Stimuli software (SOS!, Armstrong et al., 2012) was used to select items within each condition such that the OSC of the shared stem varied maximally, while remaining minimally correlated with prime and target frequencies, lengths,

[1] This protocol was put in place for all parents not personally known by the researcher to address issues with false and repeat participants, following difficulties in early data collection.

orthographic neighborhoods, and bigram frequencies, as well as stem and affix family sizes, and the number of words in which the stem was found embedded during calculation of OSC. The only characteristic significantly correlated with OSC in the final stimuli was target word length in the opaque set (0.38, $p = 0.015$). Correlations of orthographic target length and opaque prime length trended toward significance ($r = 0.21$, $p = 0.2$ and $r = 0.25$, $p = 0.12$, respectively), and no others were significant (all ps > 0.5).

The primary differences in the current stimuli selection procedure relative to that in Chapter 6 were: (1) the inclusion of an *orthographic* condition in addition to *transparent* and *opaque* morphological conditions, (2) the reduction of pairs per condition from 100 to 40 for a shorter experiment duration, and (3) the use of only higher-frequency targets (frequencies greater than 32 per million words according to the SUBTLEX-US frequency norms; Brysbaert & New, 2009), so that younger readers would be able to complete the task with sufficient accuracy. (See Appendix A for the experimental stimuli used.) Unrelated primes for each target were generated by shuffling related primes such that no unrelated prime had the same consonant onset as its target, or contained more than 40% of the letters present in the target. Each target was presented exactly once per participant, either with its related prime (e.g., BUZZARD-BUZZ) or with an unrelated prime (WALLOP-BUZZ); the pairing of each target with a related or or unrelated prime was counterbalanced across participants. Forty pairs of simple primes (5 to 10 letters long, to correspond with experimental primes) and simple targets (3 to 5 letters) were added as fillers, to reduce the ratio of real word targets presented with related primes to 37.5%. Filler primes and targets were identified as simple using the CELEX database (Baayen et al., 1995).

Sixty nonwords were created by swapping the onset consonant strings of experimental targets such that no nonword formed a word, including slang words (e.g., GOTH, RAD). These were primed by unused transparent, opaque, and orthographic primes from prior morphological priming studies, matched for length with experimental primes. Additionally, 20 nonwords were drawn from the English Lexicon Project (Balota et al., 2007) with lengths between 3 and 6 to be comparable with experimental targets, and paired with simple primes between 5 and 10 letters in length. Twenty pseudo-complex words and their nonword pseudo-stems (e.g., BANTER-BANT, KITCHEN-KITCH) were included as well.

Altogether, participants were presented with 260 trials: 160 words and 100 nonwords. This imbalance was tolerated for the sake of keeping overall experiment time as short as possible, while maximally varying OSC within each experimental condition. Trials were ordered such that no more than 5 trials went by without a nonword being presented.

8.1.3 Procedure

After returning the consent form and completing a 5-minute interview with the researcher, parents were provided with a participant ID and URL for their child, and instructed to ensure that the study be completed on computer or laptop, with at least 45 minutes available and minimal distractions. Parents were told that they may be in the same space when their child completed the task, but were asked not to "comment on how they are doing, give them answers, or watch their responses over their shoulder". As in Chapter 6, the entire experiment was presented using jsPsych (De Leeuw, 2015), with the jspsych-psychophysics plugin (Kuroki, 2021) for masked priming trials presentation and the jspsych-virtual-chinrest plugin to account for participants' screen size in determining stimulus font sizes (Li et al., 2020).

Study introduction and instructions contextualized the task as a means of defeating a super-villain intent on destroying the internet. Participants were told that, in order to discover a "master password," they would need to sort real words from fake words as quickly and accurately as possible.

A task designed to assess word recognition skill was administered first, followed by the main study, and finally the morphological awareness task. Participants were given four opportunities for optional, 30-second breaks: one after the word recognition measurement task, and one between four blocks of the primary experimental trials. During each break, their progress in cracking the code and defeating the evil villain was reported ("You've uncovered more of the password! So far you've uncovered: P L A _____"). The entire study took on average 33.5 minutes to complete.

8.1.3.1 Assessing word recognition ability

Prior to assessing developing readers' sensitivity to morphological priming, their word recognition skill was measured in order to provide a more accurate approximation of reading experience than age or grade level (see Dawson et al., 2021). One common approach to measuring word recognition ability is the Word Recognition in Isolation (WRI) task (Morris, 2013), in which the child reads aloud as many words as they can from a list of words spanning several reading levels within 90 seconds. Each word read and pronounced correctly adds a point towards their final word recognition score. To accommodate the remote and unmoderated data collection context, a variation on this task was devised that relies on lexical decision instead of reading aloud. Twenty-eight simple (i.e., non-complex) words spanning reading levels from grade 3 to grade 8 were selected from the WRI list, and 28 nonwords were selected from the English Lexicon Project (Balota et al., 2007) to match the words for length, orthographic neighborhood and bigram frequency. To make the overall experience of this set of lexical decision trials comparable to those of the morphological priming trials, a fixation cross (500 ms), a hash mask (800 ms), and a graphotactically illegal text string (XXOXOXX; 50 ms) were displayed before each target word appeared.[2] Targets were displayed for 5 seconds, and reaction times slower than 5 seconds were not recorded.

[2]Participants in the overt priming condition received an additional instructional interlude after this segment, informing them that they would now see two strings (one in uppercase before the one in lowercase to which they should respond) because the evil villain was trying to trick them by changing the task.

8.1.3.2 *Primary study: Masked priming*

Ten practice trials of the lexical decision task were administered prior to the main task. Each age group had its own speed threshold (1600 ms for 4th and 5th graders, 1000 ms for 6th and 7th graders, 750 ms for 8th and 9th graders), as more experienced readers would be expected to respond more quickly than less experienced readers. These thresholds were determined based on mean reaction times reported in Dawson et al. (2021). If responses were less than 80% accurate, or were on average slower than the threshold for that participant's age group, they were required to repeat the practice trials until they met criterion, or until they had repeated the practice three times[3].

Masked priming trials for the main experiment began with a fixation cross in the center of the screen for 500 ms, followed by a mask of hash marks (########) for 800 ms. The uppercase prime word was then flashed for 50 ms, and immediately followed by the lowercase target word. The target word was shown for 5 seconds or until a response was given, and reaction times slower than 5 seconds were not recorded. A blank screen was shown for one second between trials. This masked priming procedure closely follows the procedure used by Dawson et al. (2021), with the exception that primes were presented in uppercase rather than lowercase.

8.1.3.3 *Assessing morphological awareness*

At the end of the study, participants' morphological awareness was measured to ascertain whether children's ability to reason about words' morphological structure explains additional variance in morphological processing effects, after accounting for word recognition skill. Although it is possible that explicit knowledge of morphology may contribute to how children (and adults) process complex words, it was predicted that word recognition skill would be the stronger predictor of morphological priming magnitudes.

[3]Nine participants still had accuracy below 80% even after completing 3 iterations of practice, 7 of whom were in the youngest age group; this may indicate that the practice response speed threshold for the youngest age group was too fast.

Three subtests were selected from the University of Washington Language Battery, previously used with upper elementary and middle school students to investigate morphological awareness in this age range (Nagy et al., 2006). The first subtest was a sentence completion task with real words as response options; e.g., the options to complete the sentence "The doctor helps patients stay ___" were HEAL, HEALTH, HEALTHY, and HEALTHINESS (15 items). The second task was similar, but sentences were completed with pseudo-complex nonwords; e.g., the options to complete "The girl dances ___ " were SPRIDDERISH, SPRIDDERED, SPRIDDERLY, SPRIDDING (4 items). For the third task, participants were asked to decide for pairs of words whether one of the words comes from the other (e.g., REASONABLE – REASON or DOLLAR – DOLL; 50 items). No word pairs presented in the third task matched prime-target pairs from the main experiment. These subtests were selected because they were most focused on word-level and not sentence-level reasoning of those available in the Washington Language Battery, and because they were easily adapted for remote and unmoderated data collection.

8.2 RESULTS

A good deal of data needed to be removed before primary analyses were conducted (unsurprisingly, given the remote and unmoderated nature of this developmental study). Twenty-seven participants of the original sample were removed due to having a false positive response rate of over 30%. Four prime-target pairs were removed because the target was mislabeled as a nonword by more than half of participants (3 from the orthographic condition, 1 from the opaque condition). Forty-eight trials were removed due to a lack of response within the 5-second response interval, and 969 were removed due to incorrect response. Fifty responses were removed as by-participant outliers (greater than 3.5 sds from mean reaction time), and 11 were removed due to being faster than 200 ms.

Prime presentation durations were estimated following the same procedure described in Chapters 6 and 7. Three additional participants were removed: one due to slow aver-

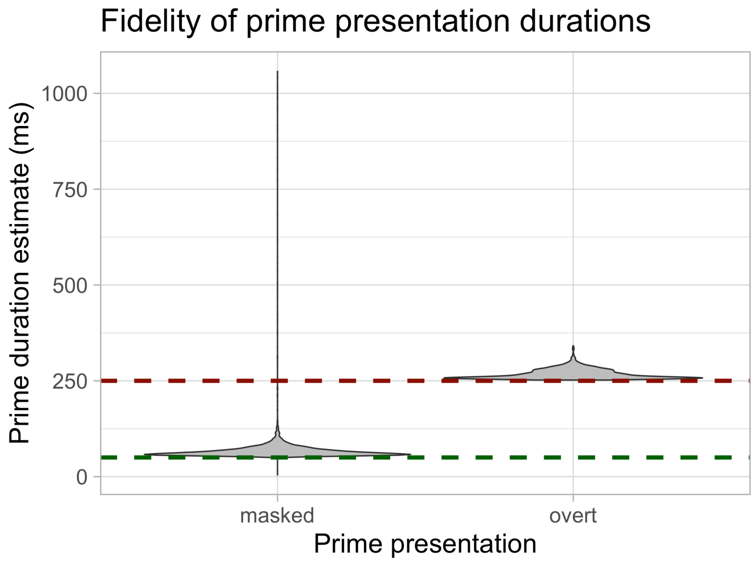

Figure 8.1: Estimates of prime duration for the two prime presentation conditions, prior to removal of excessively high or low estimates. Three times as many participants were assigned to masked priming than were assigned to overt priming.

age frame duration (over 30 ms) and two in the masked condition due to average prime durations greater than 100 ms. Trials with estimates greater than 90 ms or less than 10 ms were removed, for participants assigned to the masked prime presentation, and those with estimates greater than 300 ms or less than 200 ms were removed (see Figure 8.1). This resulted in a loss of an additional 3.2% of trials.

Following the data cleaning described above, 9,092 reaction times, in response to 116 prime-target pairs (37 orthographic, 39 opaque, and 40 transparent) from 90 participants remained for analysis. Twenty-two participants remained in the oldest group (8th and 9th grade), 28 in the middle group (6th and 7th grade), and 40 in the youngest group (4th and 5th grade), split between 21 viewing masked primes and 19 viewing overt primes.

8.2.1 Word recognition scores

Word recognition skill was estimated as accuracy of responses to real words. Accuracy was used instead of mean reaction time because it is a better estimate of participants'

Age Group	Prime Display	Lower WR	Higher WR
Younger	Masked	16	5
Younger	Overt	12	7
Middle	Masked	14	14
Older	Masked	4	18

Table 8.1: Assignment of participants to word recognition skill level from each age and prime display group, based on accuracy median-split.

word knowledge, particularly given the selection of vocabulary words spanning multiple grade levels, and because it is more distinct from the dependent variable (primed reaction times) that word recognition scores will be used to predict. Mean accuracy was 84.6%, and this metric was moderately correlated with participant age ($r = 0.40$, $p = 0.00010$).

Word recognition accuracy was used as the sole proxy variable for reading experience. This is due to the results of Dawson et al. (2021), in which their metric of word reading ability was found to significantly moderate the magnitude of opaque priming, while age did not. Although recruiting across a range of grade levels is an effective means of ensuring that word recognition skill varies sufficiently, a behavioral measure of word recognition serves as a more precise approximation of reading experience. Table 8.1 shows the distribution of participants in each age group that performed above or below the median word recognition score (79%). This table demonstrates that there were several high-performing readers in the youngest age group and similarly several low-performing readers in the oldest age group.

8.2.2 Changes in priming effects with reading acquisition

As in Chapter 6, all models reported in this and subsequent sections were fit to the transparent, opaque, and orthographic conditions separately because item characteristics were not matched across conditions during stimulus selection. Reaction times were fitted with generalized linear mixed effects models, with an Inverse Gaussian distribution and an identity link function. Due to concerns that certain effects were driven by outliers, model outliers (residuals more than 3.5 standard deviations from the mean) were removed and the model refit to the remaining data (Baayen & Milin, 2010). This data trimming never resulted in the removal of more than 0.7% of the data.

Morphological family size and target frequency were included as covariates in all models as they were found to be significant predictors of reaction times. Target length was also included in models of semantically opaque morphological priming which included OSC as a predictor, as target length and OSC were correlated for this subset of stimuli. All models included the maximal random effects structure (Barr et al., 2013).

For insight regarding how each type of priming changes with reading experience, the interaction of word recognition skill and priming (related versus unrelated) was assessed for each priming condition. For transparent morphological priming, there was no change in priming magnitude with word recognition skill ($\beta = 9.79$, $p > 0.6$). Removing the interaction term revealed main effects of both priming ($\beta = -23.55$, $p = 0.034$) and word recognition skill (-853.89, $p < 0.0001$). For opaque morphological priming, this interaction was significant ($\beta = -65.19$, $p = 0.0055$), suggesting stronger opaque priming in more experienced readers. For orthographic morphological priming, the interaction was significant in the opposite direction ($\beta = 117.69$, $p = 0.00029$) suggesting weaker, or more inhibitory, orthographic priming in more experienced readers.

8.2.3 Changing effects of OSC on priming with reading acquisition

For semantically transparent morphological prime-target pairs, the three-way interaction of word recognition accuracy, OSC, and priming was not significant ($\beta = -5.59$, $p > 0.8$). For semantically opaque morphological prime-target pairs, participants' word recognition accuracy did significantly moderate the effect of OSC on priming magnitude ($\beta = 726.95$, $p < 0.0001$). Orthographic priming also showed a change in the effect of OSC as word recognition accuracy increased, in the same direction as that for opaque priming

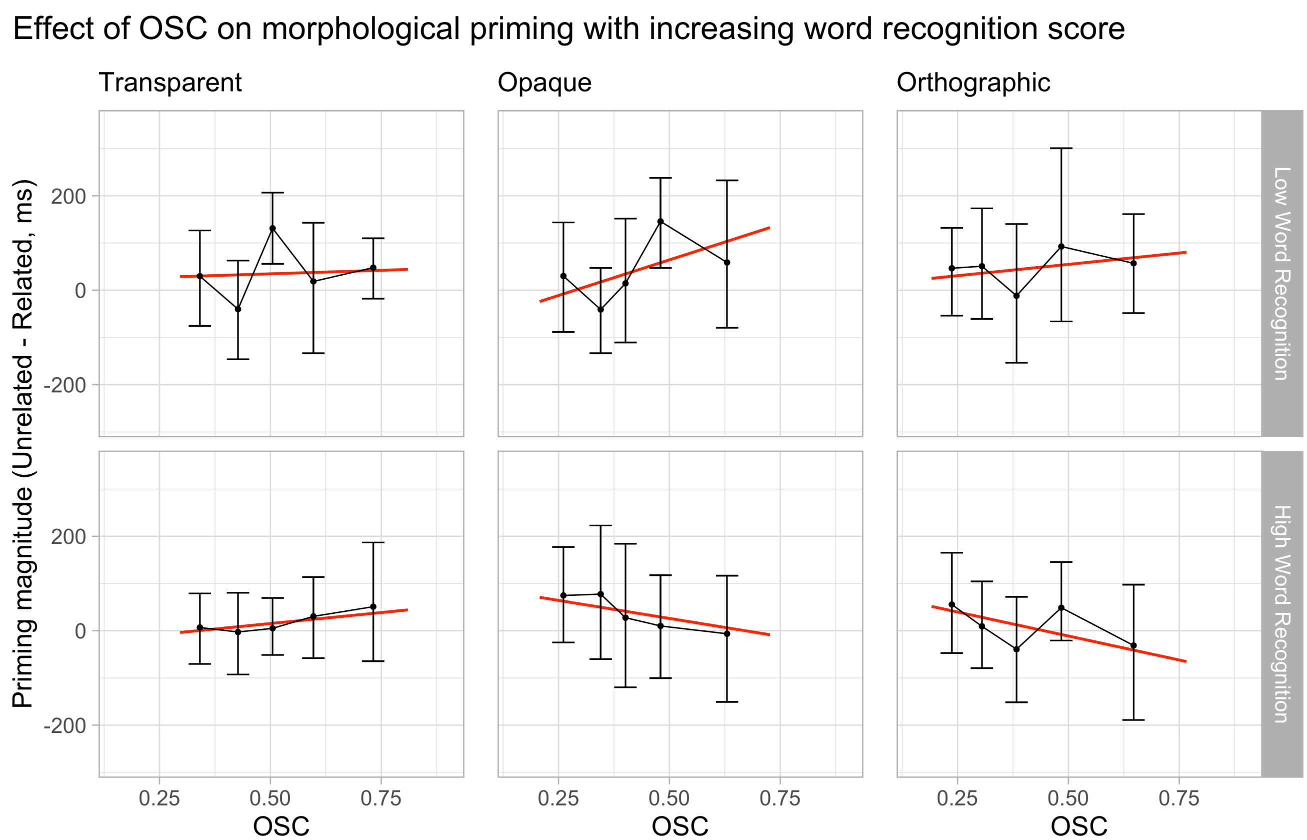

Figure 8.2: Priming magnitudes with increasing OSC, median-split by participant word recognition score. Each point represents mean priming magnitude by OSC quintile; error bars represent bootstrapped 95% confidence intervals. Red line shows line fit for by-item priming magnitudes.

($\beta = 577.46$, $p < 0.0001$).

To determine the direction of OSC effects on opaque priming for less and more experienced readers, the interaction of OSC and priming was examined separately for participants with word recognition accuracy at or below, versus above, the median (79%; see Figure 8.2). For orthographic priming, participants with lower word recognition accuracy showed a significant effect of OSC on priming such that priming magnitudes increased with greater OSC ($\beta = -119.77$, $p = 0.02$). A trend in the opposite direction was found for participants with higher word recognition accuracy ($\beta = 57.99$, $p = 0.09$). In the case of opaque priming, although neither interaction was significant, the strengthening effect of OSC on priming was greater in magnitude for participants with low word recognition ($\beta = -75.29$, $p = 0.28$) than those with high word recognition ($\beta = -27.87$, $p = 0.55$). This suggests that in both opaque and orthographic priming, the facilitation of stem OSC is stronger, or easier to detect, among less experienced readers.

8.2.4 Contrasting OSC effects on overt and masked priming in youngest age group

As only 4th and 5th graders were recruited for the overt condition of this study, word reading experience was not included as a moderator for models examining the implications of a longer prime duration. Instead, the three-way interaction of OSC, priming, and prime duration (masked or overt) was examined, and models were fit following the same procedure described above.

Prime duration significantly and positively moderated the effect of OSC on priming for all three conditions: transparent ($\beta = 177.469$, $p = 0.0012$), opaque ($\beta = 270.91$, 0.00185), and orthographic ($\beta = 510.35$, $p < 0.0001$). Looking at youngest-group participants receiving masked or overt priming separately, the effects of OSC on both transparent and opaque priming went from facilitatory for masked ($\beta = -158.11$, $p = 0.0018$ for transparent; $\beta = -108.32$, $p = 0.070$ for opaque) to neutral for overt ($\beta = 44.55$, $p > 0.4$ for transparent; $\beta = -20.39$, $p > 0.7$ for opaque). In the case of orthographic priming, the effect of OSC went from neutral when the prime was masked ($\beta = 12.25$, $p > 0.8$) to strongly inhibitory when the prime was overt ($\beta = 299.19$, $p < 0.0001$). Overall, in the case of transparent and opaque prime-target pairs, the effect of OSC was stronger and more facilitatory in masked relative to overt priming; for orthographic pairs, the effect of OSC was stronger and more inhibitory in overt relative to masked. (See Figure 8.3 for a visualization of these effects.)

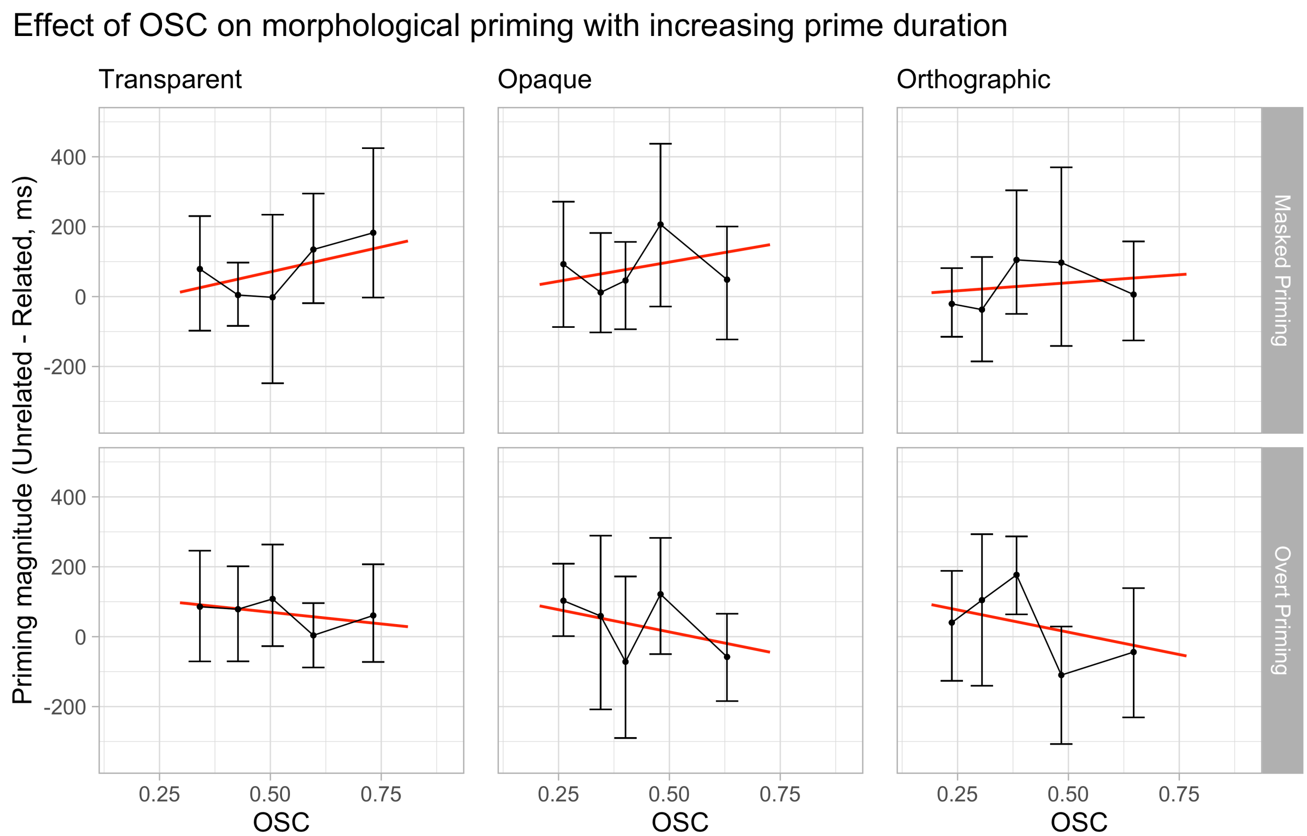

Figure 8.3: Priming magnitudes for masked versus overt priming with increasing OSC, for 4th and 5th grade participants. See Figure 8.2 for details of figure generation.

8.2.5 Investigating effects of morphological awareness

Performance on the three morphological awareness subtests administered at the end of the experiment were significantly correlated (all rs > 0.3, all ps < 0.001). Thus, accuracies on these subtests were combined for a cumulative metric of morphological awareness. Mean participant accuracy was 82.3%, and these scores were moderately correlated with participants' word recognition score ($r = 0.34$, $p = 0.0011$)

Analyses that included morphological awareness scores were conducted to answer two primary questions: First, does morphological awareness explain additional variance in morphological masked priming after word recognition has been accounted for? And second, does morphological awareness predict priming magnitudes differently in the context of overt priming relative to masked priming?

To address the first question, the residuals of the models fit to assess the effect of word recognition skill on overall priming in each condition (not including OSC as a predictor) were analyzed to examine whether the interaction of morphological awareness and priming explained additional variance. No significant effects of participants' morphological awareness performance on residualized priming magnitudes were found in any of the three priming conditions (all ps > 0.15).

To address the second question, the three-way interaction of morphological awareness, priming, and whether the prime was masked or overt was assessed for each priming condition. In all three conditions, the three-way interaction was significant such that better morphological awareness more strongly corresponded to greater priming in the overt priming context (for transparent priming: $\beta = -434.9$, $p < 0.0001$; for opaque priming: $\beta = -183.19$, $p = 0.00026$; for orthographic priming: $\beta = -258.22$, $p < 0.0001$). For contrast, with word recognition score in place of morphological awareness, there was no three-way interaction in the case of opaque priming ($p > 0.4$), while the interaction was negative for transparent priming ($\beta = -191.62$, $p < 0.0001$) but positive for orthographic priming ($\beta = 357.51$, $p < 0.0001$). Higher morphological awareness scores in young readers appear to more closely correspond to greater sensitivity to the presence of stems in overt primes than in masked primes, regardless of that prime's overall morphological structure.

8.3 DISCUSSION

As word characteristics (e.g., target and prime frequency) were not matched across conditions, the relative magnitudes of orthographic, opaque, and transparent priming cannot be compared with the current data. However, changes in magnitudes for each kind of priming with reading experience could be examined: as word recognition skill increased, transparent priming remained constant, opaque priming increased, and orthographic priming decreased (or rather became more inhibitory). These results align with those of Dawson et al. (2021), where opaque priming was found to significantly increase with word recognition experience relative to orthographic priming, but not relative to transparent priming. Their result could be attributed to an increase in opaque priming, a decrease in orthographic priming, or both, as found in the current study. This alignment of our results with the most robustly conducted prior study of morphological processing development in English gives us confidence in interpreting the rest of the findings presented.

For both semantically opaque morphological prime-target pairs and orthographic prime-target pairs, OSC was found to more strongly facilitate priming in less experienced readers than in more experienced readers. This result speaks to the primary objective of the current study, and aligns with our original predictions regarding the revised calculation of OSC: as a metric of how "meaningful" a particular letter string is across words, it should correspond to stronger and earlier-emerging sensitivity to the presence of morphological structure.

For participants with lower word recognition performance, the effect of OSC was detected only in cases where the word is not truly morphologically and semantically related.

It may be that the benefits of a high-OSC stem are overpowered when a more relevant, truly morphological relationship between prime and target is present. However, a significant positive effect of OSC on transparent priming *was* found in the analysis contrasting overt and masked primes, when only the youngest age group was included in analysis. A follow-up study with briefer prime presentation primes, or primes presented for the same duration but in an in-person, less variable context, might show OSC effects for transparent priming more consistently. The strong role of OSC in orthographic in addition to opaque priming also merits discussion; it appears that less experienced readers are sensitive to OSC in both cases, even though orthographic primes did not have an overall appearance of morphological structure. Given that OSC is calculated based on all words that contain the stem, regardless of their morphological structure, it is perhaps unsurprising that OSC might moderate orthographic priming as well as opaque. It is beyond the scope of the current study to determine whether OSC moderated opaque priming more strongly than orthographic priming (due to the lack of matched word characteristics across priming conditions), but this question may merit further investigation.

There are several potential explanations for the observed reduction of OSC's effect on opaque and orthographic priming in more experienced readers. Experienced readers process and recognize words faster: A 50-ms prime, which in young readers only allows access to more general characteristics of substrings independent of their current context, could allow for access to more context-sensitive information in older readers. This would explain both the decreasing role of orthographic-semantic consistency and the simultaneous strengthening of opaque priming and weakening of orthographic priming magnitudes found in this study as participants' word recognition skill increased. Alternatively, instead of sensitivity to OSC being simply easier to detect in less experienced readers, it could be that OSC is genuinely made use of more by these readers, whereas older readers have shifted away from OSC towards more informative or context-sensitive characteristics. Differentiating between the first explanation and this one would be difficult, but it could be potentially done with a more precisely implemented study, varying prime durations (e.g., 30, 50, and 70 ms) within multiple reading skill groups. Finally, it could simply be that, as readers recognize targets more quickly, there is less room for facilitation and thus effects are harder to detect. However, the overall magnitudes of facilitatory opaque priming did increase with word recognition skill, demonstrating that there was not simply a general ceiling effect with increasing skill.

The contrast of masked and overt priming for 4th and 5th grade participants demonstrated reduced facilitation of transparent and opaque morphological priming by OSC when the prime was overt. A weakened or nonexistent effect of OSC for longer prime durations is compatible with our previous characterization of the word processing time-course: Information associated with orthographic features and more general utility are dominant for briefly presented primes, whereas semantic and context-sensitive information can be activated by a longer (and thus stronger) input. For orthographically related prime-target pairs, however, the effect of OSC went from neutral for masked priming to inhibitory for overt priming. It could be that consciously processed orthographic primes are more "distracting" if the shared stem has higher OSC, whereas for masked orthographic primes the orthographically-informed metric is helpful or neutral. Alternatively, it could be that higher-OSC targets are easier to inhibit than lower-OSC targets, as they are responded to more quickly in general. The latter explanation may be more likely, particularly if we believe that the impact of OSC as a characteristic of the prime-target relationship is overridden in an overt priming context.

A puzzling discrepancy exists between results from analyses of the effect of OSC with varying word recognition skill, and those for varying prime duration. The effect of OSC on opaque masked priming is facilitatory for both the younger age group ($\beta = -108.32$, $p = 0.070$) and the lower word recognition group ($\beta = -75.29$, $p = 0.28$). However, the effects of OSC on orthographic and transparent priming vary between these two, overlapping groups. The effect of OSC on orthographic masked priming is neutral for participants from the younger age group ($\beta = 12.25$, $p > 0.8$), but it is strongly facilitatory for participants from the lower word recognition group ($\beta = -119.77$, $p = 0.02$). Additionally, opaque and orthographic priming behave more similarly with respect to OSC effects in the first analysis, and opaque and transparent priming behave more similarly in the second. We attribute these differences to the fact that, although the youngest age group and the lower word recognition group are overlapping, they are not the same (see Table 8.1). Switching from analyzing the lower word recognition group to the younger age group entails a loss of 18 lower-performing participants from the middle and older age groups, and the addition of 5 higher-performing participants from the lower age group. The change in OSC effects for orthographic priming could be due to a loss of power, and the change in transparent priming could be due to a slight overall increase in word recognition, although we do not have the data to support this speculation directly.

Finally, we had not expected to find that morphological awareness would explain additional variance in masked priming magnitudes after accounting for word recognition skill, and our findings confirmed this prediction. However, morphological awareness scores did better predict overt priming magnitudes relative to masked priming magnitudes, across all three priming conditions. This contrast aligns with the expectation that morphological awareness would play a greater role in how readers respond to primes that can be processed consciously, although the fact that the effect was found in all three priming conditions as opposed to just the transparent condition was unexpected. Apparently, participants with better morphological awareness were more generally sensitive to the presence of the stem in the overt prime, regardless of whether it was semantically transparent, semantically opaque, or followed by a non-affix ending. Perhaps this finding is a symptom of the young age of participants included in this contrast: a different pattern of effects may have been found if more experienced readers had also been presented with overt primes.

The primary goals of this study were (1) to further illuminate how morphological processing emerges with reading experience, (2) to investigate whether OSC predicts early emergence of morphological priming effects, and (3) to investigate how masked priming effects differ from overt priming effects in less experienced readers. Our results support those of Dawson et al. (2021) in showing a strengthening of opaque priming and a weakening of orthographic priming with reading experience. The finding that OSC more strongly facilitates opaque and orthographic priming for less experienced readers suggests that early sensitivity to morphological structure may be driven by higher-OSC stems. However, this was not unique to seemingly morphologically structured (i.e., opaque) primes, but occurred for orthographic primes as well. The emergence of sensitivity to opaque priming distinct from orthographic priming appears to coincide with a diminishing importance of OSC, indicating increased sensitivity to more wholistic morphological cues with reading experience. Regarding the differences between masked and overt effects in the youngest age group, the failure of OSC to predict transparent and opaque overt priming aligns with our conceptualization of OSC as characteristic with more bearing on early word processing. Although some of these effects are weak or variable, the overall pattern of results enriches our understanding of how morphological processing mechanisms are acquired by contextualizing the role of OSC within the acquisition of sensitivity to morphological structure.

The current study has limitations, and thus results must be interpreted with caution. Most notably, the unmoderated and remote nature of data collection resulted in the loss of a great deal of data, and stimulus presentation was less precise than would be expected in an in-person context. The unmoderated context also required the use of a less conventional approach to measuring word recognition skill (accuracy in a lexical decision task, as opposed to accuracy in reading words aloud as in Morris, 2013), although we feel that the use of words spanning a range of reading levels made our alternative a relatively sound approximation. Were developmental participants more easily accessible at the time the study was conducted, a robust assessment of word recognition skill would have been administered, and participants would have been assigned to conditions based on their reading skill as opposed to their age (this was only a concern for the assignment to the youngest group to overt or masked priming, as manipulations were otherwise within-participant). Similarly, the morphological awareness measure was also shortened and likely less precise than the full one conducted in Nagy et al. (2006). Finally, the study is cross-sectional, and thus we are not tracking the change in a certain set of individuals' word recognition processing with more experience, but rather comparing individuals with presumably more and less experience (as determined by their word recognition score). Given prior evidence for individual variability in morphological processing effects (e.g., Andrews & Lo, 2013) and the generally variable nature of these effects (as evidenced by the difference in OSC effects on transparent and orthographic priming for the youngest age group relative to the group with lower word recognition skill), a longitudinal variation of the current study may be worth conducting.

Individuals in the midst of early reading acquisition are forming morphological sensitivity to more general paradigms (e.g., the addition of -MENT to a verb) and particular morphological families (e.g., -TRUST-) at the same time. However, skilled readers can also acquire new sensitivities when they learn new words and integrate this knowledge into their already developed word processing mechanisms. For example, high-schoolers studying for the SAT may encounter several previously unknown words sharing the same Greek or Latin root (such as ANTHRO- or -MORPH-). The effect of OSC on experienced readers' acquisition of sensitivity to novel morphological families is discussed and explored in the next chapter.

CHAPTER 9

The Role of Consistency in Learning New Stems: A Training Study with Adult Readers

Although our findings in Chapter 8 demonstrate a stronger facilitatory role of stem OSC for less skilled word readers, this finding is simply correlational. Finding stronger facilitation of priming by OSC in less experienced word readers does not mean OSC is driving the emergence of morphological sensitivity; it could be, as noted in the prior chapter's discussion, that it is simply easier to detect such effects in young readers, or that these effects are stronger in young readers for other reasons. As such, pursuing the question of OSC's role in the emergence of morphological sensitivity may require studies better designed for investigation of causal relationships.

Artificial language learning studies with adults can provide important insights into how morphological sensitivity is acquired, while providing a much more convenient context for manipulation of word and morpheme characteristics. Tamminen et al. (2015) investigated affix acquisition by teaching participants artificial words containing novel affixes (e.g., SLEEPNULE, SAILAFE). The impact of changes in learning context and materials on generalization of those affixes to new words was measured during sentence reading, using both explicit (judgments of semantic congruency) and implicit (semantic congruency effects in speeded reading aloud) tasks. Using this paradigm, Tamminen et al. (2015) demonstrated that generalization of affix learning was stronger for affixes that appeared in a greater variety of trained words, and for affixes that altered stem meaning in a semantically consistent manner. Similarly, by training adults to recognize "words" constructed from morpheme-like chunks in an artificial script, Lelonkiewicz et al. (2020) demonstrated that sensitivity to letter-like chunks can be acquired even without mapping to meaning, supporting the role of mechanisms such as boundary detection and identifying frequent letter chunks (Rastle & Davis, 2008) in bolstering morphological sensitivity. Together, these findings emphasize that both the semantic utility and statistical regularities of an affix determine how well it is learned. Although initial acquisition of morphological sensitivity in developing readers may differ from how skilled readers learn a few new affixes, artificial language learning studies provide notable insights regarding favorable conditions for learning new morphological paradigms.

Tamminen et al. (2015) also investigated how the amount of time between learning and testing sessions impacted memory of, and sensitivity to, acquired word knowledge. They found that conducting the testing session a week later rather than immediately after training was necessary for semantic congruency effects to be detected during sentence reading, although performance with explicit judgments was unaffected (also see Merkx et al., 2011). The essential role of allowing time for *consolidation* of acquired knowledge has been found in several other artificial language learning studies as well (e.g., Merkx et al., 2011; Tamminen et al., 2012). These findings support a complementary learning systems framework of episodic and semantic memory (McClelland et al., 1995). The importance of consolidation is not sufficiently discussed in either decomposition-based or distributed accounts of morphological acquisition (but see Kumaran et al., 2016, for a discussion of a neural network implementation of complementary learning systems in a more general context).

As described above, Tamminen et al. (2015) did investigate the role of semantic consistency on morpheme learning in the context of an artificial language learning study, and thus we already have evidence that semantic consistency impacts how easily morphological knowledge is acquired with this paradigm. However, there are several aspects of their study that make additional investigation desirable. First, although the meaning transformations performed by each affix were chosen to mimic those of actual affixes, the actual word definitions generated were oddly specific. For example, adding -NULE implied a person related to the stem meaning, and BRICKNULE was defined as "The labourer who operates the oven which hardens clay to brick" (p. 9). FOXNULE was defined as "Foxnule is someone who looks after a fox harmed in a car accident". It may be that the less-plausible nature of these definitions impacted participants' learning experience, necessitating suspension of disbelief and allowing less linguistically-informed learning processes to take over.

Second, Tamminen et al. (2015) focused on the acquisition of new affixes; however, as skilled readers we are more likely to learn new stems. For instance, consider a high school student in the United States, studying vocabulary words to prepare for the SAT. Studying Latin and Greek roots is a common strategy for enhancing this preparation, because low-frequency words are often composed of familiar affixes appended to unfamiliar roots (e.g., DEFENESTRATION, INDUBITABLY). There are fewer affixes than there are stems among linguistically acknowledged morphemes (Baayen et al., 1995), and thus it stands to reason that learning new stems is a more common process for skilled readers than learning new affixes. Additionally, it is possible that the role of semantic consistency impacts stem learning differently than affix learning, meriting a separate investigation of this process.

Finally, semantic consistency in Tamminen et al. (2015) was varied by either allowing each affix to play the same semantic role across all four words in which it appeared, or by assigning it to a distinct role for two out of those four words. Understandably, this manipulation is more reminiscent of how semantic consistency can vary in affixes, such as the varying roles of -ER in GREATER and BIGGER relative to TEACHER and READER. Such behavior can be observed in stems (e.g., STICK in STICKER versus CHOPSTICKS), but consistency can also be varied by varying the number words centered around a particular semantic theme (e.g., TRUST in TRUSTING, DISTRUST, and TRUSTWORTHY relative to WHISK in WHISKING, WHISKER, WHISKEY). It is possible that establishing two conflicting meaning associations has a different impact on morpheme learning than does establishing a single meaning association with various degrees of consistency.

The current study investigated the role of semantic consistency on the acquisition of morphological knowledge and sensitivity by teaching skilled readers novel, pseudo-complex words, and subsequently assessing responses to these words and their morphological structure.

9.1 METHODS

The study was entirely remote and unmoderated. Participants completed two sessions: during the *learning* phase, they learned form-to-meaning correspondences for 36 novel words with a complex structure (a novel stem, such as CLERN, paired with a familiar affix, such as -MENT). The *testing* phase occurred about a week later, providing time for consolidation of novel word knowledge, and consisted of a sequence of tasks designed to assess participants' memory of the novel words they learned. Overall, learning and testing phase procedures closely followed those described in Tamminen et al. (2015), with adjustments to accommodate the remote testing context.

9.1.1 Participants

Fourteen residents of the United States were recruited for this study. All were native speakers of English, with normal or corrected-to-normal vision, between the ages of 20 and 40 years old. All were recruited via shared public posts on Facebook.[1] Study links and basic instructions were sent via email. All participants received a $10 gift card for each session they completed.

9.1.2 Materials

See Appendix B for the complete set of materials used in this study.

[1] Social media was used for recruitment instead of a more traditional participant database due to repeated issues with individuals participating multiple times with different emails, or investing minimal effort as evidenced by lack of improvement during learning phase. Although some participants were personal acquaintances of the researcher, none had any knowledge of the experiment's context or purpose.

9.1.2.1 *Learning phase*

Eighteen novel stems were obtained by selecting nonwords in the English Lexicon Project Balota et al. (2007) between 4 and 6 letters in length that did not contain any symbols or punctuation and that did not closely resemble any existing English words or common slang. Stems were divided into six groups of three, such that nonword length, bigram frequency, and orthographic neighborhood were reasonably balanced across groups. All of these metrics were taken from Balota et al. (2007).

Nine suffixes were selected from the CELEX database (Baayen et al., 1995), with the requirement that they have a family size greater than 100 and an OSC value (using the new variation described in Chapter 5) of more than 0.3. To make the creation of words and their corresponding reasonable definitions straightforward, only suffixes that turned verbs into nouns or adjectives (e.g., adding -IVE to the verb SUPPORT yields the adjective SUPPORTIVE) were used.

Four words were constructed for each stem by pairing it with four distinct suffixes; consequently 12 words were created for each stem group. Within each stem group, six of the nine selected suffixes were paired with a stem once and three were paired with a stem twice. This procedure resulted in 72 novel, complex words.

Thirty-six definitions were written to establish semantic consistency of words within each novel morphological family. For the three sets of four words containing stems assigned to be semantically consistent, definitions for each word within the family centered around a particular semantic theme, such as “sleep” or “unity”. For the sets of words containing quasi-consistent stems, three words’ definitions related to a common theme, while one was unrelated. For sets of words containing inconsistent stems, the four words’ definitions were all unrelated. In all cases, the word’s definition was compatible with the typical semantic role of the word’s suffix according to the Merriam-Webster dictionary. For example, as the suffix -ABLE typically means “capable of, fit for, or worthy of” (https://www.merriam-webster.com/dictionary/-able), pairing this suffix with a stem assigned to be related to "sleep" yielded the definition, "Sufficiently comfortable for sleeping." Each stem was combined with two suffixes that formed nouns and two that formed adjectives.

Each participant trained on only 36 of the 72 words (thus learning only 9 of the 18 novel stems). Assignment of a particular stem group to a consistency level was counterbalanced across participants such that, across those learning words with a particular stem (e.g., CLERN), a third of participants mapped those words to a semantically consistent set of definitions, a third mapped them to a quasi-consistent set, and a third mapped them to an inconsistent set. Thus, any orthographic idiosyncrasy of an individual nonword stem or word did not confound the consistency manipulation.

9.1.2.2 Testing phase

The testing phase consisted of three distinct tasks: a recognition memory task, a sentence congruency task, and a definition matching task. No additional materials needed to be generated for the definition matching task.

For the recognition task, in addition to the words on which the participant had previously been trained, three types of foils were included. The first type consisted of untrained combinations of stems and suffixes that the participant had learned: for example, if they had learned the words DUTTABLE, DUTTMENT, TUSHELANCE, and TUSHELISM previously, an untrained combination could be DUTTISM or TUSHELABLE. The second type of foil consisted of untrained stems (which for other participants were trained stems) combined with suffixes that were seen in the learning phase (e.g., PHREWNMENT, given PHREWN- was not seen as a stem in the learning phase). The third foil type combined trained stems with suffixes that were not seen in the learning phase (e.g., DUTTIFY). Suffixes used for this type of foil were selected to have family sizes greater than 95 and OSC values greater than 0.3.

For the sentence congruency task, a sentence was presented that ended in a trained word (first block of trials) or untrained word (second block) that contained a trained stem. Sentences were written such that the final word was semantically congruent or non-congruent with the rest of the sentence; the final word's suffix was always compatible with the rest of the sentence. For example, given that a participant had learned to associate the word CLERNMENT with the definition "A sip; the action of taking a sip", a compatible sentence could be "The drink was so bitter, I couldn't manage more than a single clernment." A non-congruent sentence could be "My words seemed harsher when I saw them written down in the clernment." Congruent and non-congruent sentences were matched for word length, and each stem appeared in the final word for an equal number of congruent and noncongruent sentences. For the block of trials for which sentences ended in untrained words, semantic congruency was determined as compatibility with the common meaning (for consistent stems) or the dominant meaning (for quasi-consistent stems) of the stem across the words in which it appeared. For inconsistent stems, congruency was determined as compatibility with one of the four meanings of words containing that stem.

9.1.3 Procedure

Experiment instructions and stimuli were presented using jsPsych (De Leeuw, 2015) and the jspsych-virtual-chinrest plugin was used for size calibration to account for participants' varying screen sizes (Li et al., 2020). Session links were sent to participants via e-mail or Facebook, with the second link being made available to the participant 6 days after they completed the first session. Although the majority of instructions were provided via self-paced reading after following the study link, participants were informed when they received each session link that they should (1) complete the task on a laptop or computer (as opposed to a tablet or smartphone), (2) complete the study in a minimally distracting environment, and (3) not listen to music. Following completion of each phase, participants were sent a $10 gift card, and they were also debriefed regarding the study purpose after the testing phase was completed.

9.1.3.1 Learning phase

Participants were informed that the first phase would last about an hour. They were asked not to take notes or create any other reminder of words or word meanings as they completed the learning phase. The learning phase consisted of three tasks, completed in the same order three times in a row. They were provided the opportunity for an optional, 30-second break after every three blocks. Throughout the learning phase, trials featuring words containing the same stem were never presented one after the other.

In the first task, participants saw a novel word and its definition, and they were prompted to type the word in a text box. If the word they typed was incorrect (even by a single letter, although extraneous spaces and variant capitalization were permitted), participants were alerted of the error: "You typed CLSRNMENT which does not match CLERNMENT. Be careful to type the word correctly!". They were then shown a real word (e.g., BIRD), and prompted "Is this related in meaning to the word you just typed?" During instructions prior to beginning the task, participants were informed that words they should consider to be related could be synonyms, antonyms, or otherwise related in meaning. The comparison word's relation to the novel word's definition was varied in this way to prevent participants from thinking about the word's meaning too simplistically or narrowly. The number of related and unrelated words presented for this purpose were balanced within each block of the task, and participants compared each word with a related word and an unrelated word at least once over the task's three iterations. Participants were given feedback on their response after the relatedness judgment ("Incorrect. SWALLOW is related to CLERNMENT which means 'A sip; the action of taking a sip"'). The typing task and the relatedness judgment were designed to encourage attention to details of the word's form and meaning, respectively.

In the second task, the word was shown and three definitions were provided below it.

The participant was tasked with selecting the definition they had learned corresponded to that word. The two foils were definitions for other words the participant was currently learning. To make the selection more challenging, foil definitions corresponded to words that were the same part of speech as the target word in all cases, and to a word with the same suffix as the target word as much as possible. Each definition was used as a foil response during the definition selection task exactly three times. Following their selection of a definition, participants were given feedback and shown the correct definition for the target word.

In the third task, the definition of a word was shown and the participants were asked to type the word to which it corresponded in a text box. Participants were reassured that this task was challenging and they should simply do their best to remember as many words as they can. Feedback was provided on their typed input: "Incorrect. You typed TUSHELFUL. 'A sip; the action of taking a sip' is the definition of CLERNMENT."

The three tasks described above were then repeated, in the given order, twice more. Participants took on average 58 minutes to complete the learning phase ($s.d.$ = 20 minutes).

9.1.3.2 *Testing phase*

Participants were e-mailed the link to the testing phase 6 days after completing the training phase, and asked to complete the session within 48 hours of receiving the link. The particular stimuli presented to them for each testing task were specific to the form-to-meaning correspondences they had trained on during the learning phase.

In the *recognition memory* task, participants were presented with one of the trained or untrained words, designed as described above, and asked whether it was one of the words they had seen during the learning phase. Words were presented in a randomly generated order. Participants pressed the J key for words they believed they had trained on, and F for words they did not. They were instructed to try to respond as quickly and accurately as possible, and informed that they only had 8 seconds to respond before they were moved on to the next trial. Prior to each word presented, a fixation cross was shown for 500 ms, and a blank screen was shown between trials for 1000 ms.

Two blocks of the *sentence congruency* task were administered next: first, a block of sentences ending in trained words, then a block of sentences ending in untrained words. Within each block, sentences were presented in a random order. In each trial, the sentence without the last word was displayed first (e.g., "The drink was so bitter, I couldn't manage more than a single..."). Participants were told to press the Enter key once they had read the first part of the sentence, and they were not able to move past the sentence until they had viewed it for at least 3 seconds. After they had pressed Enter, a blank screen was shown for 500 ms, and then the final word was presented (e.g., CLERNMENT). Participants pressed J if they thought the word made sense in the sentence context, and F if they thought it did not make sense. They had 8 seconds to respond, and after their response a blank screen was shown for one second before beginning the next trial. The same procedure was repeated in the next block, but with sentences designed for and ending in untrained words as opposed to trained words.

Finally, participants completed a *definition matching* task, for which a definition and a trained word were presented, and they were instructed to determine if the definition and the word matched. Participants pressed J if the definition and word matched, and F if it didn't. Non-matching definitions were randomly selected from the foil definitions corresponding to that word that were used during the third learning phase task. Again, participants had 8 seconds to respond, and a blank screen was shown for one second between trials.

The entire testing phase took 18 minutes to complete, on average ($s.d.$ = 4 minutes).

9.2 RESULTS

One participant was included only in analyses of the learning phase, due to accidentally repeating the learning phase instead of completing the testing phase. The other thirteen were included in all analyses reported below. The mean delay between the learning phase and the testing phase was 7.5 days ($s.d.$ = 1.5).

9.2.1 Learning phase

By the third iteration of learning phase tasks, mean accuracy on the definition selection task was 88.4%, while mean accuracy for typing the word when given the definition was 44.6%. See Figure 9.1 for a visualization of improvement on word knowledge tasks over the course of the learning phase.

9.2.2 Recognition memory

For analysis of this task, in which participants decided whether each string was one they had learned previously, there were two questions of interest. First, were participants faster or more accurate in accepting trained words with consistent stems than those with inconsistent stems? And second, were participants slower or less accurate in rejecting untrained words with consistent stems than those with inconsistent stems? Either or both of these effects would imply that stem consistency impacted how words and morphemes from the first phase of the study were learned and remembered.

For this analysis, and those of all subsequent testing phase tasks, trials for which no response was recorded (due to exceeding the trial duration limit of 8 seconds) were removed.

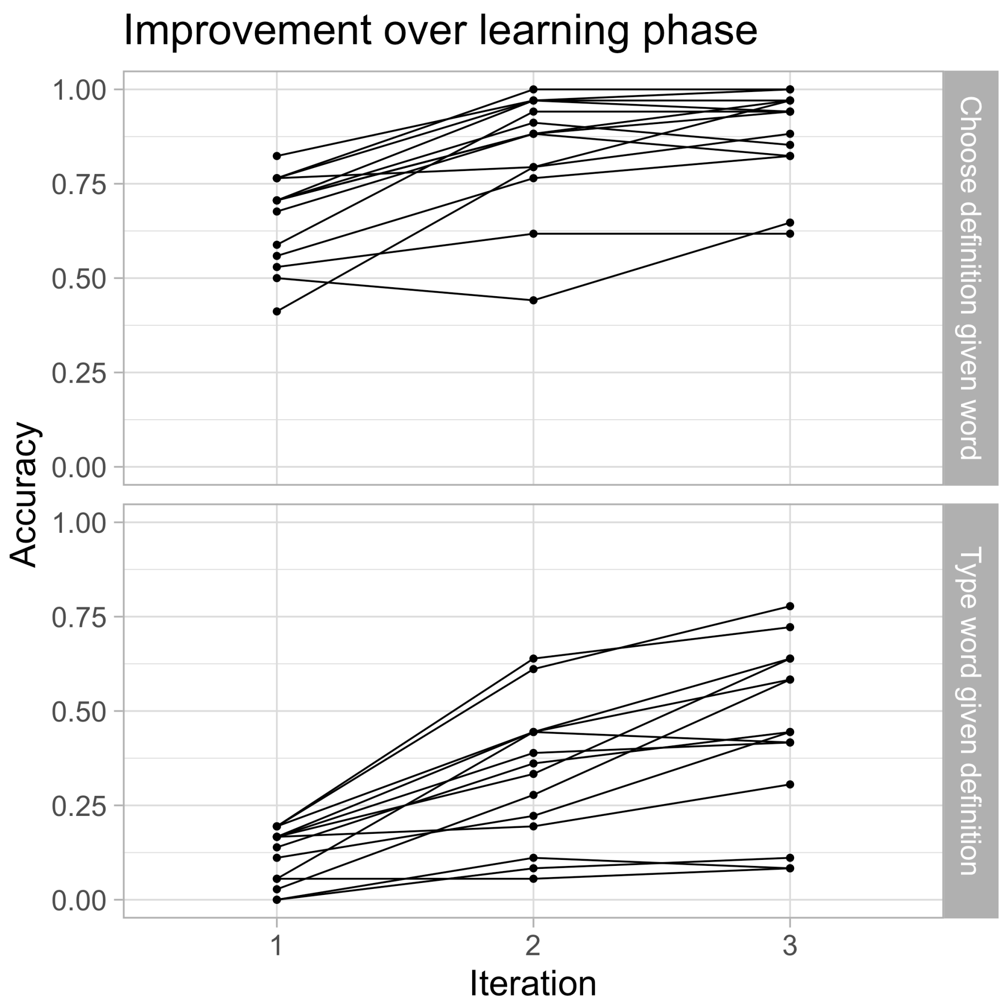

Figure 9.1: Accuracy for definition selection and word typing tasks, over the three iterations during learning phase.

9.2.2.1 Response accuracy

The effect of consistency on accuracy of responses was assessed using generalized linear mixed effects models with a binomial distribution. Models for these and all subsequent analyses included random intercepts by participant and stem, except in cases where such a model did not converge. In these cases, random effects structures were simplified until the model fit converged, following the recommendation of Brauer & Curtin (2018).

Models of accuracy were fit separately for trained words (466 responses), untrained words consisting of novel combinations of trained stems and trained suffixes (hitherto referred to as "untrained words with trained suffixes"; 351 responses), and untrained words consisting of trained stems with untrained suffixes ("untrained words with untrained suffixes"; 231 responses).[2] No main effects of consistency were found for any of the three models (all ps > 0.15).

As seen in Figure 9.1, there was a good deal of variability in how well participants learned the words over the course of the learning phase. To look more closely at whether performance during the learning phase impacted participants' sensitivity to stem consistency, participants' accuracy in the final iteration of the definition selection task was added to the model as a moderator of the effect of consistency.[3] The interaction was not significant for any of the three models (all ps > 0.2), except for the contrast of trained words with consistent stems and those with quasi-consistent stems ($\beta = 12.65$, $p = 0.0023$). Participants with higher performance in the final learning phase definition selection task were more accurate in accepting words with quasi-consistent stems than those with consistent stems. This result was unexpected, particularly considering that no such effect was found between trained words with consistent and inconsistent stems.

[2]As responses to words containing untrained stems could not be assessed for effects of consistency, these were not modeled.

[3]Accuracy on the definition selection task was used, rather than that the typing task or some combination of the two tasks, due to concerns that performance on the typing task might be partially determined by other characteristics of the participant, such as typing ability or testing anxiety.

Word type	Accuracy	Correct RT	Incorrect RT
Trained	87.4%	1064 (572)	1876 (1282)
Trained stem, trained suffix	42.2%	1699 (987)	1085 (641)
Trained stem, untrained suffix	69.7%	1379 (734)	1198 (838)
Untrained stem, trained suffix	92.9%	879 (402)	1419 (1839)

Table 9.1: Accuracy, mean reaction times for correct responses and mean reaction times for incorrect responses across the four types of words presented to participants during the recognition memory task. Numbers in parentheses denote standard deviations.

9.2.2.2 *Reaction times*

Trials with incorrect responses (333) and by-participant outliers (9 responses more than 3.5 standard deviations from the participant's mean reaction time) were removed prior to model fitting, leaving responses to 408 trained words, 148 untrained words with trained suffixes, and 162 untrained words with untrained suffixes for analysis. As in Chapters 6 and 8, reaction times were fitted with generalized linear mixed effects models, with an Inverse Gaussian distribution and an identity link function.

For trained words and untrained words with untrained suffixes, no main effects of consistency were found on correct response latencies (all ps > 0.7). However, correct rejections of untrained words with trained suffixes were slower for consistent stems relative to quasi-consistent stems ($\beta = -331.8$, $p = 0.016$) and relative to inconsistent stems ($\beta = -320.5$, $p = 0.0074$).

As was done for accuracy, the interaction of participants' learning accuracy with stem consistency was added to the statistical models of response latencies (see Figure 9.2). A significant interaction was found in all three models, and patterns of effects for trained words and untrained words with trained suffixes supported the possibility that participants with higher learning accuracy were more sensitive to stem consistency. For trained words, higher learning accuracy corresponded to slower acceptance of words with quasi-consistent stems ($\beta = 381.43$, $p < 0.0001$) and inconsistent stems ($\beta = 586.61$, $p < 0.0001$) relative to acceptance of those with consistent stems. For untrained words with trained

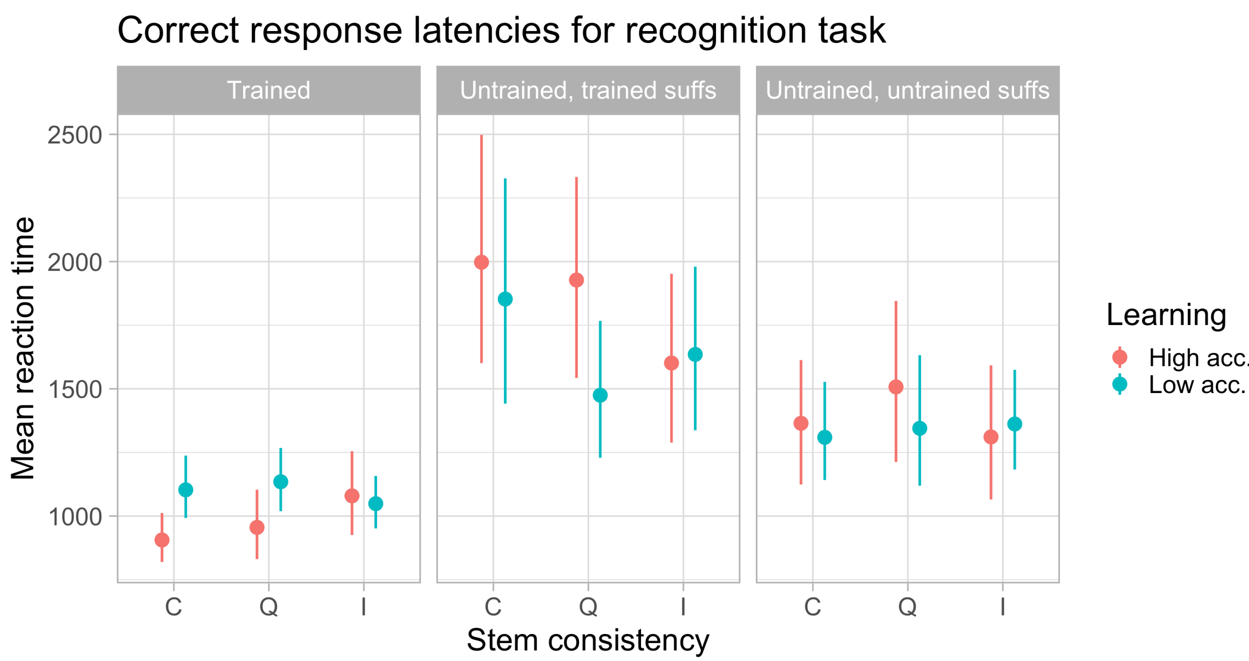

Figure 9.2: Interaction of participant learning accuracy (accuracy on final iteration of definition selection during learning phase) and stem consistency in predicting correct recognition response latencies. Note that the left facet shows latencies for correct acceptance of the word as one previously learned, while middle and right facets show latencies for correct rejection. Error bars denote bootstrapped 95% confidence intervals.

suffixes, higher learning accuracy corresponded to faster rejection of words with quasi-consistent ($\beta = -1124.15$, $p < 0.0001$) and inconsistent ($\beta = -252.89$, $p = 0.063$) stems relative to those with consistent stems.

For untrained words with untrained suffixes, participants with lower learning accuracy rejected words with inconsistent stems more quickly than consistent stems ($\beta = 281.13$, $p = 0.014$), though this was not the case for the contrast of quasi-consistent and consistent stems ($\beta = -6.02$, $p > 0.9$). It is possible the lack of an effect of consistency in the case of participants with higher learning accuracy is due to their greater sensitivity to which suffixes were trained (i.e., they may be rejecting words with untrained suffixes quickly enough that stem consistency effects cannot be detected).

9.2.3 Sentence congruency

Participants were presented with a sentence and asked to decide if the final word made sense in the context of the sentence. The task was conducted in two blocks of trials, first for sentences ending in trained words, then for sentences ending in untrained words with trained suffixes. On average, participant responses to sentences with trained final words were 76% accurate ($s.d.$ = 10%), and responses to those with untrained final words were 72% accurate ($s.d.$ = 12%). See Table 9.2 for a summary of accuracies and reaction times for this task.

Accuracy was determined by alignment of the sentence with the word's learned definition in the case of trained words; for untrained words, accuracy was determined by alignment of the sentence with the stem's common meaning (in the case consistent stems), the stem's dominant meaning (in the case of quasi-consistent stems), or the meaning of one of the words the stem occurred in (in the case of inconsistent). By design, inconsistent stems did not have an associated "meaning" across its occurrences; thus, the determination of congruency for sentences ending in untrained words with inconsistent stems was somewhat subjective. This should be kept in mind when interpreting the results reported

Congruent	Final word	Accuracy	Correct RT	Incorrect RT
Yes	Trained	85.9%	1815 (1337)	2091 (1493)
No	Trained	66.2%	2196 (1401)	2254 (1410)
Yes	Untrained	63.7%	1353 (1071)	2252 (1438)
No	Untrained	79.9%	2085 (1343)	1766 (1423)

Table 9.2: Accuracy, mean reaction times for correct responses and mean reaction times for incorrect responses for congruent and noncongruent trials of the sentence congruency task. Results for sentences ending in trained words (second block of testing phase) and untrained words (third block) are both shown.

below.

9.2.3.1 Response accuracy

As for the recognition task, response accuracy was modeled using generalized linear mixed effects models with a binomial distribution. Models were fit separately for sentences ending in trained words (465 responses) and those ending in untrained words (466 responses).

For trials ending in trained words, the overall effect of stem consistency on accuracy trended towards significance, with a marginal difference in accuracy was found for words with inconsistent stems relative to those with inconsistent stems ($\beta = -0.45$, $p = 0.095$), while no difference between consistent and quasi-consistent stems was detected ($\beta = 0.14$, $p = 0.62$). Similarly, for trials ending in untrained words, words with consistent stems were responded to more accurately than those with inconsistent stems ($\beta = -0.92$, $p = 0.00054$), while the difference between consistent and quasi-consistent stems was not significant ($\beta = -0.33$, $p > 0.24$).

Adding an interaction with participant learning accuracy to the model (as done for recognition task analyses), moderation of the consistency effect was found for both sentences ending in trained words and those ending in untrained words. In both cases, participants with higher learning accuracy showed greater differentiation due to stem consistency, such that responses were less accurate for words with inconsistent stems than for words with consistent stems (though this effect only trended towards significance in the trained words case: $\beta = -3.94$, $p = 0.064$ for trained final words; $\beta = -4.23$, $p = 0.044$ for untrained final words). Learning accuracy did not moderate the difference in accuracy between quasi-consistent and consistent stems in either case ($\beta = -3.94$, $p = 0.14$ for trained final words; $\beta = -2.32$, $p = 0.28$ for untrained final words).

9.2.3.2 *Reaction times*

Trials with incorrect responses (239) and by-participant outliers (9 responses more than 2.5 standard deviations from the participant's mean reaction time) were removed prior to model fitting, leaving 352 responses for sentences ending in trained words and 333 responses for sentences ending in untrained words for analysis.[4] Reaction times were again fitted with generalized linear mixed effects models, with an Inverse Gaussian distribution and an identity link function.

In these analyses, sentences ending in trained and untrained words had notably different patterns of effects (see Figure 9.3). For sentences ending in trained words, those ending in words with quasi-consistent stems were responded to marginally faster ($\beta = -167.21$, $p = 0.080$), while no effect was found between inconsistent and consistent stems ($p > 0.8$). Adding in participant learning accuracy as a moderator, the effect of faster responses for quasi-consistent relative to consistent stems was stronger for participants with higher learning accuracy ($\beta = -458.85$, $p < 0.0001$), while the opposite effect was significant for the contrast of inconsistent and consistent stems ($\beta = 1823.35$, $p < 0.0001$).

For sentences ending in untrained words, no main effect of consistency was found (both ps > 0.5). Adding participant learning accuracy as a moderator, we see a significant interaction of learning accuracy and consistency for both quasi-consistent relative to

[4]The criterion for outlier removal was 2.5 standard deviations, instead of 3.5 as used in analysis for prior studies and chapters, so that the statistical models would converge. With 3.5 as the cutoff, some models did not converge, even for the simplest possible random effects structure, while those that did converge showed the same pattern of results as with a 2.5 criterion.

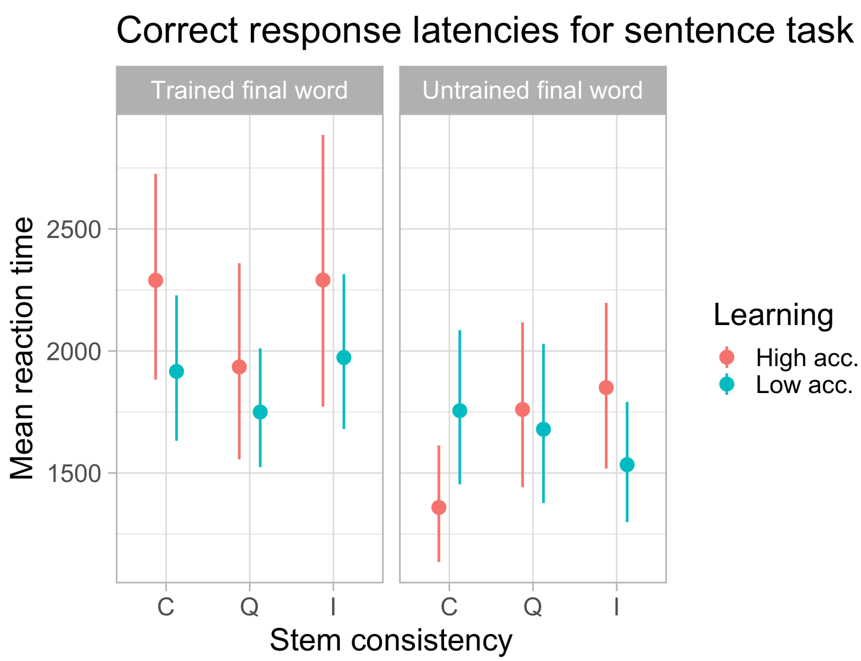

Figure 9.3: Interaction of participant learning accuracy and stem consistency in predicting response latencies for the sentence congruency task. Error bars denote bootstrapped 95% confidence intervals.

consistent ($\beta = 789.55$, $p < 0.0001$) and inconsistent relative to consistent ($\beta = 671.90$, $p < 0.0001$). This pattern of effects indicates that participants with higher performance at the end of the learning phase responded more quickly to sentences ending in untrained words with consistent stems, than to those ending in untrained words with quasi-consistent or inconsistent stems. Such a difference was either weaker or nonexistent in participants with lower learning accuracy.

9.2.4 Definition selection

In this final task of the testing phase, a word and a definition were presented simultaneously, and the participant determined whether the pairing was correct. Participant accuracy averaged 86% ($s.d.$ = 10%), with a minimum accuracy of 64% and a maximum accuracy of 97%. See Table 9.3 for a breakdown of accuracy and reaction times across matching and non-matching trials.

Matching	Accuracy	Correct RT	Incorrect RT
Yes	89.7%	2206 (1320)	3284 (2030)
No	82.9%	2212 (1231)	2412 (1435)

Table 9.3: Accuracy, mean reaction times for correct responses and mean reaction times for incorrect responses for the definition matching task, separated by trials for which definition and word are matching or non-matching.

9.2.4.1 *Response accuracy*

Response accuracy was again modeled with generalized linear mixed effects models with a binomial distribution, and all responses were fit with one model (468 responses).

The main effect of consistency was not significant (both ps > 0.3). Including an interaction of consistency and participant learning accuracy yielded no significant effects, although the contrast of consistent and inconsistent stems trended towards significant ($\beta = -3.89$, $p = 0.16$ for inconsistent stems, $\beta = -1.39$, $p = 0.61$ for quasi-consistent stems, relative to consistent stems). Thus, no significant effects of stem consistency on accuracy were found for the definition task.

9.2.4.2 *Reaction times*

Incorrect responses (64) were excluded, leaving 404 reaction times for analysis; no by-participant outliers were removed as all reaction times were within 3.5 standard deviations of participant means. Reaction times were again fitted with generalized linear mixed effects models, with an Inverse Gaussian distribution and an identity link function.

There was no main effect of consistency on reaction times (both ps > 0.8). However, adding participant learning accuracy as a moderator revealed a significant positive interaction for words containing inconsistent ($\beta = 194.67$, $p = 0.0083$) and quasi-consistent stems ($\beta = 278.65$, $p < 0.0001$), relative to words containing consistent stems. The direction of this effect suggests that participants with higher learning accuracy responded more slowly to words containing inconsistent and quasi-consistent stems relative to those with consistent stems, while such an effect was weaker or not present for participants with lower learning accuracy.

9.3 DISCUSSION

Results from this study suggest that overall, consistent stems were learned more easily, and had a stronger impact on recognition and processing of novel words. In the first task of the testing phase, in which participants were asked to determine for each trial whether they had learned the displayed word in their previous session, participants were faster to reject untrained words with trained suffixes compared with ones they hadn't learned if the stem was quasi-consistent or inconsistent. Additionally, participants with higher accuracy during the learning phase showed greater sensitivity to stem consistency, both in accepting trained words and in rejecting untrained words, relative to participants with lower learning accuracy. Similarly, when asked if a certain word and definition were correctly paired, participants with greater learning accuracy were slower to respond correctly to words with inconsistent or quasi-consistent stems than to words with consistent stems.

The interaction of learning accuracy and consistency effects in these two tasks suggests that participants who learned the entire set of words more effectively also became more sensitive to the presence of consistent stems relative to inconsistent stems. Whether this means that better learning leads to greater stem sensitivity or vice versa is unclear. However, given that participants' learning accuracy was assessed prior to consolidation, and implicit effects are found to arise only after consolidation (Tamminen et al., 2015), the former direction of causality is more likely. The detection of this pattern of effects in the first task implies that stem sensitivity impacts recognition of the word form itself, whereas the extension of such findings to the definition selection task implies this sensitivity also impacts the memory of form-to-meaning mappings.

Results from the sentence congruency task were less clear cut, which is unfortunate since this task comes closest to the authentic experience of reading. The effect of consistency on task accuracy, and the interaction of consistency and participant learning accuracy on task accuracy, did align with what we would expect: Participants were less accurate for responses to sentences ending in words with an inconsistent stem, than for words with a consistent stem. This held for sentences ending in both trained and untrained words, although for untrained words the concept of a "correct" meaning of a word containing a inconsistent stem is problematic (as noted previously). For latencies of responses to sentences ending in trained words, however, patterns of effects showed a unexpected advantage for words with quasi-consistent stems relative to consistent stems. The fact that a similar effect—or at least one in the same direction—was not found for inconsistent stems calls into question whether this effect is due to the consistency manipulation, or to some other characteristic of sentences or meanings used in the quasi-transparent condition. On the other hand, for latencies of responses to sentences ending in *untrained* words, facilitation for consistent stems relative to quasi-consistent and inconsistent stems was found. Again, this effect is more interpretable in the case of quasi-consistent stems than inconsistent stems, as determination of semantic congruency is more straightforward in this case.

Some shortcomings of the design of study materials may have contributed to the noisy and counterintuitive results found in the sentence congruency task. Due to the effort required to generate sentence reading materials, each word was paired with only one sentence, either semantically congruent or non-congruent with its learned definition, and thus words were only ever in the congruent or the non-congruent condition instead of this assignment being counterbalanced across participants. Additionally, the lack of counterbalancing semantic themes across the consistency conditions may have also impacted the pattern of results: Although the assignment of words to meanings was counterbalanced across participants, the assignment of meanings to consistency conditions was not. Thus, words with consistent stems always related to the concepts of sleeping, exiting, or unit-ing, while words with quasi-consistent stems primarily related to the concepts of drinking, being profitable, or feeling sick. This potential confound of semantic categories with semantic consistency may have played a greater role in the sentence congruency task than in the other two tasks due to the richer and more complex semantic processes involved in completing this task.

Even if we set aside the findings from the sentence congruency task entirely, the current study still strengthens the findings of Tamminen et al. (2015) by demonstrating that the forms and meanings of new words are more effectively remembered and internalized if the words contain semantically consistent stems. Future studies making use of the artificial language learning paradigm might investigate how this manipulation interacts with the pairing of more or less productive affixes, or affixes versus non-affix letter strings, while being sure to carefully control semantic categories and sentence materials across consistency conditions.

CHAPTER 10

General Discussion and Conclusions

10.1 SUMMARY

In this dissertation, we have made use of several approaches to investigate the mechanisms of morphological processing within a distributed framework. Following an introductory chapter, Chapter 2 provided a thorough review of the literature on known effects of morphological processing, discussing the implications of findings in each case for the contrast of distributed and decomposition based theories. Variations in morphological processing phenomena across languages and over the time course of reading recognition were emphasized. While there are some results that seem more directly indicative of a decomposition-based account, the broader range of findings better aligns with the distributed view, and the occurrence of decomposition-like behavior can be contextualized within a distributed framework. We argued that a better understanding of the nature of semantically opaque morphological priming (e.g., CORNER–CORN) within the distributed framework, and more emphasis on the graded nature of morphemes as semantic cues, are essential for a more complete picture of morphological processing mechanisms.

Chapter 3 presented simulation work with a feedforward network that was trained to map from orthographic to semantic representations for words in an authentic and developmentally-plausible English vocabulary. Orthographic representations were based only on letter configuration, and semantic representations were based on word co-occurrence statistics: No explicit morphological information was provided to the model. Testing the model with prime-target pairs from prior studies of semantic transparency and morphological priming revealed network sensitivity to semantically transparent morphological priming comparable to empirical findings, but no sensitivity to semantically opaque morphological primes.

In Chapter 4, a recurrent model with (mostly) separate excitation and inhibition was trained to simulate the dynamics of morphological processing. The network was trained on one of two artificial languages: one morphologically rich and one morphologically impoverished. Presenting the network with prime-target pairs from the trained language yielded response behaviors that reflected general patterns of processing dynamics seen in studies of the time-course of complex word processing, such as increasing semantic priming for longer prime durations. Patterns of opaque priming effects for the network fully trained on the morphologically rich language mirrored those found in languages such as English and French, and analyses regarding the emergence of these effects led to interesting predictions regarding patterns of opaque priming in developing readers.

To enable empirical investigation of the full spectrum of semantic utility of various letter strings in lexical processing, a novel approach to calculating orthography-semantics consistency was proposed in Chapter 5. Instead of using the stem word as a reference point, the frequency-weighted centroid of the cluster of words containing the stem was used to calculate semantic distances. This makes OSC a metric of the semantically similarity of all words containing the given stem or letter string, removing bias in favor of stem meaning and allowing generalization to letter strings that are not words in their own right. Using this new metric to predict reaction times from prior studies demonstrated that, although the new OSC metric is a somewhat worse predictor of single word judgments than the original, it is as good or better at predicting morphological priming magnitudes.

Chapter 6 investigated the effects of OSC on morphological priming; OSC was intentionally varied as much as possible within two sets of morphological prime-target pairs, one semantically transparent and the other semantically opaque. Results suggested that adults are sensitive to stem OSC, although interactions with target frequency and the possibility of ceiling effects made it difficult to tell whether the OSC of the shared morpheme impacted facilitation of target recognition by the prime. The overt priming variation of this study (Study 2), revealed an inverse effect of OSC on priming magnitudes, likely due to ceiling effects in responses, but also showed robust priming for not only transparent, but also opaque priming, in contradiction with prior findings in English.

Chapter 7 presented a close replication of the study of the time-course of complex word processing reported in Rastle et al. (2000). Patterns of results generally mirrored those found in the original study, validating the use of the masked priming paradigm in a remote context, with the most notable divergence being the stronger opaque priming for prime durations of 250 ms (as found in the previous chapter).

Chapter 8 investigated OSC effects in developing readers, spanning 4th to 9th grade (ages 9 to 16). Facilitation of morphological priming by the OSC of the shared stem was strongest in readers with lower word recognition performance, whereas the differentiation of opaque and orthographic priming was evident only for readers with better word recognition. This supports the hypothesis that sensitivity to morphological structure emerges earlier for letter strings with higher OSC (i.e., letter strings that are more semantically useful), and that this emergence can be decoupled from sensitivity to a word's more global morphological structure.

Finally, Chapter 9 presented a study in which adult readers were trained on a novel vocabulary with varying morphological structure. The semantic consistency of artificial stems was manipulated by pairing them with four different affixes and defining the resulting words in a manner that was semantically consistent, quasi-consistent, or inconsistent. Participants were trained to associate these novel words with their definitions, and a week later were tested on their memory of and sensitivity to word structures. Testing session results revealed greater memory of and sensitivity to more consistent stems in both the recognition and definition-matching tasks, although results for the sentence reading task were less straightforward.

Taken together, these results advance our knowledge of how morphological processing occurs during visual word recognition, and how such sensitivity is acquired, within a distributed view of cognition. Simulation results demonstrate that explicit morphological representations or input are not required for a system to successfully reproduce known findings, even opaque priming effects, particularly when recurrence and biologically plausible constraints on connectivity are added to the model. Empirical results demonstrate that although effects of orthography-semantics consistency on morphological priming magnitudes are difficult to detect in skilled adult readers, there is support for a role of OSC in the emergence of morphological sensitivity, both for early readers whose word recognition mechanisms are still maturing, and in more skilled readers learning new morphological families.

10.2 IMPLICATIONS

Complex words make up a large proportion of the words we encounter in day-to-day reading; additionally, morphological structure provides an essential avenue for rapid recognition and comprehension of words we have never seen before. Thus, the mechanisms of morphological processing are a key contributor to successful word recognition and, more generally, successful reading. The current work contributes indirectly to the field of literacy education, as more detailed characterization of how such cognitive processes unfold and how they are acquired with reading experience will enhance our ability to understand the ways in which they "go wrong" for struggling readers. Additionally, better understanding of how complex words are recognized underscores the importance of gaining extensive exposure to diverse written language, given the complexity of the processes demonstrated by skilled readers for efficient word recognition. If distinctive profiles of word processing behavior could be established for various levels of reading proficiency, these patterns of effects could even be used for as a formative assessment of word recognition ability, instead of only having research-specific implications.

10.3 LIMITATIONS

10.3.1 Empirical limitations

The primary limitation of the work presented in this dissertation is the fact that all empirical results were obtained via remote, unmoderated studies. Masked priming effects in particular are very sensitive to variation in task and environment, and implementing this paradigm in a context in which stimulus presentation was more noisy, and participants were more likely to be less attentive, presents issues that do not have clear solutions. As far as possible, concerns regarding stimulus presentation precision were mitigated in the data cleaning phase of these studies (Chapters 6–8) in which trials with prime duration estimates outside of the desired range, as well as participants whose displays appeared more volatile. Additionally, replication of results reported by Rastle et al. (2000), as well as obtaining generally reasonable findings, such as faster responses to words following related relative to unrelated primes, makes us confident that data collected in this manner, while challenging, can nonetheless inform our understanding of morphological processing mechanisms.

However, the reduced attentiveness of participants, demonstrated by lower accuracy and slower reaction times in the current set of studies relative to prior ones, may have impacted certain patterns of findings. Surprisingly strong facilitation by opaque primes in overt priming contexts, in particular, may be a symptom of the unique context of remote testing, as this effect is found across all studies presented here, and is not a pattern established in previous work. It is possible that the processing of overt primes is more impacted by lack of attentiveness than the processing of masked primes, as behaviors under the participant's conscious control are more relevant in such a context. The effect of attentiveness on how overtly presented words are processed could be an interesting avenue for future research, particularly given the educational implications of such a manipulation.

One prediction of interest for the developmental study (Chapter 8) was whether opaque overt priming could be found for less experienced readers (before weakening or disappearing in more experienced readers). This prediction stems from the pattern of priming effects found during the emergence of opaque priming in the E-I network trained on a morphologically rich language (Chapter 4), as well the French developmental results found by Quémart et al. (2011). However, due to difficulties recruiting older children and adolescents (to contrast with younger participants) in the overt priming condition, and due to the fact that opaque overt priming was found even in English-reading adults in the current studies, the evidence presented here is not sufficient to test that prediction. Future studies investigating morphological processing in developing readers could help resolve this matter by including multiple prime durations. A large number of participants will be needed for such a developmental study with so many conditions; even in Quémart et al. (2011), significant effects were found only when 3rd, 5th, and 7th graders were grouped together, and compared with an equally large sample of adults.

Another note regarding limitations of empirical results is that all morphologically related word pairs in the presented priming studies consisted of complex primes paired with simple targets (e.g., TRUTHFUL-TRUTH). This is limiting in that we are really interested in the role of morphologically informed representations and morphological sensitivity more generally, but most findings in the literature are narrowly focused on this one case. The current studies were conducted with complex-simple pairs both for the sake of ease of interpretation (most prior priming studies have been conducted with complex-simple pairs, making it easier to contextualize current results within the literature), and for convenience during stimulus generation (selecting prime target pairs that from those used in prior studies). The recurrent model (Chapter 4) and the training study (Chapter 9) did explore the processing of complex words containing bound stems to some degree. However, conducting more studies with complex-complex pairs (see, e.g., Feldman et al., 2004; Feldman & Soltano, 1999) will help determine to what degree the patterns of priming effects established for complex-simple pairs generalizes to other types of morphological relatedness. The new variation of OSC makes investigation of OSC effects for these types of prime-target pairs possible, and it may be that moving away from simple words as targets would make effects of metrics like OSC easier to detect (see Feldman et al., 2004).

10.3.2 Computational limitations

The simulations presented in this thesis are very limited in their capacity for generalization to real language. Expanding and improving the E-I network (Chapter 4) such that it can capture sensitivity to form-to-meaning mappings in authentic vocabularies, and reflect findings across multiple tasks and contexts, will greatly enhance the utility of these simulations to the subfield of morphological processing in visual word recognition. Several additional limitations of the current computational models, beyond their inherent simplicity, are discussed below.

First, the manner in which lexical decision is approximated in Chapters 3 and 4 merits scrutiny. For the feedforward network, reaction time is estimated as the amount of time needed for activation in the output layer to settle, such that change in activation from the previous tick to the current tick is negligible. In the recurrent network, model reaction time is the amount of time needed for output units to reach within a certain range of their target values. Both approaches yield shortcomings in model results: the first causes reaction times to increase with model training, instead of decreasing as do the reaction times of actual readers with experience. This simulated developmental trend occurs because the more trained model is settling to output values that are more accurate, and farther from 0.5, which take longer to reach. The second approach causes many prime-target pair examples, for which semantic activation never reaches criterion, to be excluded from analysis. The selective sampling thus introduced may impact the pattern of testing results. Additionally, using semantic settling or closeness to targets to obtain reaction times implies that semantic retrieval must be well underway in order for a decision to be made regarding whether the input is a word or not. This assumption is not necessarily accurate: sufficient information to make a lexical decision may be available well before semantic processing of the input is complete. To address the above concerns, alternatives to the methods used to obtain reaction times in Chapters 3 and 4 should be explored, to determine whether the current patterns of results persists under more rigorous and realistic approximations of lexical decisions. For example, Plaut & Booth (2000) made use of "semantic stress", a measure of the degree to which semantic representations are binary, as a basis for networks' lexical decision (also see Plaut, 1997), while Laszlo & Plaut (2012) implemented a trained response system with input from the primary orthography-to-semantics structure to determine lexicality.

Second, the model put forward in Chapter 4 did not adhere to true separation of excitatory and inhibitory connections. Although the inhibitory connections were constrained to stay negative throughout training, the excitatory connections were merely initialized as positive but not constrained to remain that way. This condition was loosened due to difficulties in training the network without becoming stuck in local minima of error. In this sense, the network's structure did not fully implement the neurally plausible constraints discussed in Cheyette & Plaut (2017) and Laszlo & Plaut (2012), making it less comparable to what we know of connectivity within and between cortical layers. Alternative approaches to training such as implementing layer normalization (Lei Ba et al., 2016) may allow for these simulations to be re-done with more strictly implemented constraints.

Another extension of the current simulation work that would inform our interpretation is to explore how variations in size and number of hidden layers impacts testing results. For the sake of providing support for the distributed view of morphological processing, one network structure that can yield the predicted pattern of results is less reassuring than a variety of network structures that yield those same results. Similarly, exploring the degree to which different characteristics of the recurrent model, such as constraints on positive or negative connections, determine the resulting pattern of effects would be informative (as done in Laszlo & Plaut, 2012).

Finally, as discussed in Chapter 4, although the current models focus entirely on successful mapping from orthography to semantics, word recognition within the context of authentic reading is a more complex undertaking. For instance, variations in where the reader's eye fixates, or in the predictability of the current word given preceding text, may impact how processing unfolds. Incorporating such complexities into future simulations will also bring us closer to a full and accurate understanding of morphological processing in the context of reading text.

10.4 CONCLUSIONS

The research reported in this dissertation combines neural network modeling, computational linguistics methods, and behavioral data from both developing readers and skilled adult readers to investigate the mechanisms of morphological processing. The findings contribute to a better understanding of the phenomena surrounding semantically opaque priming—particularly that its emergence can be accounted for within a distributed framework given a dynamic, biologically plausible, and time-pressured learning environment. We additionally provide a novel metric of orthography-semantics consistency that better assesses morphemes' semantic utility across words, and we establish that more semantically consistent morphological families impact recognition processes earlier in acquisition. Finally, we make specific predictions regarding the trajectory of opaque priming emergence with reading experience, although fully adequate testing of these predictions remains for future work.

As initially discussed at the end of Chapter 2, contextualizing complex word reading within a distributed view of processing makes our understanding of these effects easier to integrate with research in other domains of cognitive science, including those concerning higher-level reading operations (e.g., comprehension and inference) and those concerning more general visual processing (e.g., object recognition). Interpreting any given subfield within the context of a more general theory of cognition allows us to step away from one particular pattern of effects to see the bigger picture.

APPENDIX A

Priming Stimuli

A.1 STIMULI FOR ASSESSING EFFECT OF OSC ON MORPHOLOGICAL PRIMING IN SKILLED ADULT READERS

Semantically transparent morphological priming			
Target	Prime	Target	Prime
hang	hanger	strain	strainer
king	kingdom	digit	digital
bush	bushy	idol	idolize
gang	gangster	cheer	cheerful
hero	heroic	bulb	bulbous
green	greenery	bold	boldness
dead	deadly	govern	government
harp	harpist	broil	broiler
march	marcher	magnet	magnetic
vend	vendor	meek	meekly
herb	herbal	fool	foolish
gold	golden	mood	moody
hunt	hunter	resist	resistant
worth	worthless	cool	coolant
suit	suitable	depend	dependable
angel	angelic	vast	vastly
raid	raider	bulk	bulky
block	blockade	heal	healer
acre	acreage	view	viewer

quick	quickly	report	reporter
speak	speaker	critic	critical
lone	lonely	dirt	dirty
tour	tourism	protect	protection
clear	clearance	dark	darkness
quiet	quietly	final	finalize
mock	mockery	weak	weaken
lump	lumpy	paint	painter
soft	soften	slow	slowly
duck	duckling	farm	farmer
cloud	cloudless	fear	fearsome
eject	ejection	goal	goalie
acid	acidic	fizz	fizzle
erupt	eruption	differ	difference
flesh	fleshy	self	selfish
north	northern	brave	bravely
cream	creamy	mourn	mourner
find	finder	select	selective
stalk	stalker	calm	calmness
trash	trashy	predict	predictable
cheap	cheaply	zeal*	zealous
loaf	loafer	blend	blender
knock	knocker	loot	looter
guilt	guilty	peel	peeler
passion	passionate	fail	failure
neat	neatly	deep	deeply

sharp	sharpen	music	musical
react	reaction	renew	renewable
human	humanity	teach	teacher
tight	tighten	debt	debtor
chill	chilly	buy	buyer
Semantically opaque morphological priming			
all	allure	later	lateral
stock	stocky	brig*	brigade
carp	carpal	bash	bashful
pill	pillage	fruit	fruitless
dam	damage	organ	organize
raft	rafter	fair	fairy
fig	figment	broth	brother
pond	ponder	brand	brandy
temp	temper	wick	wicker
scar	scary	crypt	cryptic
ration	rational	trump	trumpet
gall*	gallant	custom	customer
iron	irony	earn	earnest
trait	traitor	slim	slimy
butt	buttery	hind	hinder
spin	spinster	whole	wholesome
barb	barber	helm	helmet
putt	putty	flour	flourish
tract	tractable	colon	colonist
earl	early	snip	sniper

sever	several	liquid	liquidate
bread	breadth	hung	hunger
touch	touchy	husk	husky
plum	plumage	author	authorize
poster	posterity	square	squarely
treat	treaty	limb	limber
sir	siren	skew	skewer
rend	render	audit	audition
glut*	gluten	chart	charter
posit*	positive	marsh	marshal
host	hostage	access	accessory
grate	grateful	respect	respective
quest	question	coast	coaster
rook	rookie	invent	inventory
show	shower	feud	feudal
feat	feature	thick	thicket
dole*	doleful	dorm	dormant
dice	dicey	puck	pucker
moth	mother	pluck	plucky
mass	massage	brisk	brisket
numb	number	glow	glower
welt	welter	wand	wander
shift	shiftless	wonder	wonderful
craft	crafty	brace	bracelet
crate	crater	blaze	blazer
bowl	bowler	batter	battery

poll	pollen	buzz	buzzard
troll	trolley	beak	beaker
seed	seedy	flick	flicker
draw	drawer	bloom	bloomer

Table A.1: Simuli used for adult priming study report in Chapter 6. Asterisks denote targets removed from analyses due to low accuracy across participants.

A.2 STIMULI FOR EXAMINING EMERGENCE OF OSC EFFECTS ON MORPHOLOGICAL PRIMING IN DEVELOPING READERS

Semantically transparent morphological priming			
bush	bushy	lime	limeade
gang	gangster	dead	deadly
gold	golden	near	nearly
quick	quickly	speak	speaker
worth	worthless	block	blockade
flesh	fleshy	mock	mockery
cheap	cheaply	herb	herbal
sly	slyly	cream	creamy
trick	trickery	sharp	sharpen
bulb	bulbous	cloud	cloudless
dirt	dirty	passion	passionate
strain	strainer	dense	densely
final	finalize	view	viewer
self	selfish	depend	dependable
rare	rarely	fear	fearsome

mix	mixer	heal	healer
select	selective	agree	agreement
report	reporter	deeply	deep
magic	magical	renew	renewable
buy	buyer	debt	debtor
Semantically opaque morphological priming			
lad	lady	count	counter
gall*	gallant	pig	pigment
board	boarder	raft	rafter
dam	damage	pill	pillage
clam	clamor	miss	mission
butt	buttery	home	homely
treat	treaty	touch	touchy
bread	breadth	moth	mother
barb	barber	ramp	rampant
quest	question	plum	plumage
poll	pollen	come	comely
poster	posterity	snip	sniper
brand	brandy	seed	seedy
fair	fairy	slim	slimy
bowl	bowler	organ	organic
hung	hunger	coast	coaster
respect	respective	limb	limber
wand	wander	buzz	buzzard
depart	department	beak	beaker
flick	flicker	pluck	plucky

Orthographic priming			
add	address	pal	palace
bun	bunch	too	tooth
rack	racket	bill	billow
wall	wallop	tin	tinsel
tick	tickle	fee	feeble
quart	quartz	nick	nickel
stand	standard	tend	tendon
chap	chaplain	harp	harpoon
stern	sternum	extra	extract
shall	shallow	enter	enterprise
rabbi*	rabbit	curt*	curtain
troll	trollop	surf	surface
plain	plaintiff	ether*	ethereal
cash	cashew	scrap	scrape
pack	packet	vow	vowel
text	textile	jerk	jerkin
stir	stirrup	smug	smuggle
want	wanton	won	wonder
proper	property	code	codeine
monk	monkey	exam	example

Table A.2: Simuli used for developmental priming study report in Chapter 8. Asterisks denote targets removed from analyses due to low accuracy across participants.

APPENDIX B

Training Study Stimuli

B.1 WORDS

Words learned in the training study described in Chapter 9. Participants either learned words from list one or list two, and assignment of words to definitions was counterbalanced across participants.

List 1: THRIPABLE, THRIPMENT, THRIPION, THRIPAL, HEPTMENT, HEPTION, HEPTLESS, HEPTAL, JEVERIVE, JEVERANCE, JEVERISM, JEVERFUL, GRIESTABLE, GRIESTMENT, GRIESTION, GRIESTAL, KUNKMENT, KUNKION, KUNKLESS, KUNKAL, POATIVE, POATANCE, POATISM, POATFUL, DUTTABLE, DUTTMENT, DUTTION, DUTTAL, RALONMENT, RALONION, RALONLESS, RALONAL, TUSHELIVE, TUSHELANCE, TUSHELISM, TUSHELFUL

List 2: PHREWNABLE, PHREWNMENT, PHREWNION, PHREWNAL, NESHMENT, NESHION, NESHLESS, NESHAL, FUMPIVE, FUMPANCE, FUMPISM, FUMPFUL, EGENABLE, EGENMENT, EGENION, EGENAL, SCOOKMENT, SCOOKION, SCOOKLESS, SCOOKAL, LAULIVE, LAULANCE, LAULISM, LAULFUL, CLERNABLE, CLERNMENT, CLERNION, CLERNAL, DRACKMENT, DRACKION, DRACKLESS, DRACKAL, SNAMIVE, SNAMANCE, SNAMISM, SNAMFUL

B.2 DEFINITIONS

affix	P.O.S.	definition	stem type
able	adj.	Sufficiently comfortable for sleeping	consistent
ment	noun	A brief interval of rest	consistent
ion	noun	Exhaustion, need of rest	consistent
al	adj.	Tired or sleepy	consistent

ment	noun	An exit or means of leaving	consistent
ion	noun	The act of departing	consistent
less	adj.	Difficult to get out of, or lacking exits	consistent
al	adj.	Likely to leave early, or unexpectedly	consistent
ive	adj.	Causing people to unite behind a common cause	consistent
ance	noun	An event that brings people closer together	consistent
ism	noun	The belief that all religions have the same goal	consistent
ful	adj.	Able to gather large groups of people	consistent
able	adj.	Suitable for drinking, not poisonous	quasi-consistent
ment	noun	A sip; the action of taking a sip	quasi-consistent
ion	noun	The written form of a statement or declaration	quasi-consistent
al	adj.	Appealing to drink, or thirst-quenching	quasi-consistent
ment	noun	The total income of a business	quasi-consistent
ion	noun	An item which can be sold for a good profit	quasi-consistent
less	adj.	Yielding no gain or benefit	quasi-consistent
al	adj.	Cold or freezing	quasi-consistent

ive	adj.	Nauseous, likely to vomit	quasi-consistent
ance	noun	The act of getting sick, regurgitation	quasi-consistent
ism	noun	A chaotic or confusing event	quasi-consistent
ful	adj.	Disgusting, or causing people to feel ill	quasi-consistent
able	adj.	Easily folded or collapsed to a smaller size	inconsistent
ment	noun	Honesty or directness	inconsistent
ion	noun	An understanding or idea	inconsistent
al	adj.	Having to do with forests or trees	inconsistent
ment	noun	The act of making a trade or transaction	inconsistent
ion	noun	Feeling of anger or rage	inconsistent
less	adj.	Unappealing, lacking in charm	inconsistent
al	adj.	Likely to take more than their fair share	inconsistent
ive	adj.	Disrespectful towards a superior	inconsistent
ance	noun	An event that was unexpected or surprising	inconsistent
ism	noun	The belief that plants can feel pain	inconsistent
ful	adj.	Tending to or likely to make noise	inconsistent

Table B.1: Suffixes, parts of speech, and definitions for novel words learned by participants in the training study described in chapter 9. Each chunk of four definitions corresponds to words with the same stem (assignment of stem to definition groups was counterbalanced across participants).

B.3 TESTING STIMULI

B.3.1 Suffixes for untrained words with trained stems and untrained suffixes

These suffixes were added to trained stems to form novel words with trained stems and untrained suffixes, for the recognition memory task presented in the testing phase: -LY, -URE, -IFY, -NESS, -IST, -EST

B.3.2 Sentences for trained words

sentence	congruent	stem type
I had to sleep on the living room sofa, but the extra pillows made it...	yes	consistent
The break room gives tired hospital employees a place for a...	yes	consistent
After hours of discussion and nothing decided, I was annoyed by the group's...	no	consistent
She couldn't reach the shelf where we keep spices because she's too...	no	consistent
They didn't know the commander's orders because they never received his...	no	consistent
The final school bell meant it was time for...	yes	consistent
When the disguised door was sealed shut, the escape room appeared to be...	yes	consistent
Even though I eat healthily, I need a lot of exercise to stay...	no	consistent
By angering everyone regardless of their political party, the president's actions turned out to be...	yes	consistent

Wanting to bond with her friends, Amelia organized a camping trip as a...	yes	consistent
Until I saw their intense work ethic firsthand, I didn't believe their reputation for...	no	consistent
The structure is well-built and safe; there's no need to be overly...	no	consistent
Before drinking the river water, we filtered it to make it...	yes	quasi-consistent
The drink was so bitter, I couldn't manage more than a single ...	yes	quasi-consistent
The wood floors were too slippery after they applied a layer of...	no	quasi-consistent
He's got a sweet tooth; he likes candy, pastries, and anything else that is...	no	quasi-consistent
My words seemed harsher when I saw them written down in the...	no	quasi-consistent
Needing extra cash, he was excited to learn a vintage record he had saved was now a....	yes	quasi-consistent
Although it was never published, the author's hours spent working on that novel were not...	yes	quasi-consistent
Please arrive early; with such a long meeting agenda we'll need to be...	no	quasi-consistent
The mountains looked small from a distance, but up close they were actually...	no	quasi-consistent
Side effects of this medication include nausea and possible...	yes	quasi-consistent

The celebrity's positive posts about a totalitarian government made people think she supported...	no	quasi-consistent
The rotting hamburger someone had left in the basement was...	yes	quasi-consistent
Since I'll have to fit this chair in a suitcase, it needs to be...	yes	inconsistent
Don't mislead or exaggerate when pitching your idea to him; he appreciates...	yes	inconsistent
The team hoped for good weather, but they ended up playing in the middle of a...	no	inconsistent
Her brother had us laughing all through dinner; we found him very...	no	inconsistent
I've been trying to get rid of a rash on my arm; the doctor recommended this...	no	inconsistent
She kept her voice calm, but her clenched fists betrayed her feelings of...	yes	inconsistent
We kept sanding the tabletop until it was perfectly...	no	inconsistent
We only have enough pizza for everyone to get two slices, so please don't be...	yes	inconsistent
When the boy mocked his teacher, she sent a note to his parents saying he'd been...	yes	inconsistent
The underdog team's narrow victory was certainly a...	yes	inconsistent
We had no infrastructure for removing waste, due to the city's...	no	inconsistent
People would be more willing to admit their mistakes to her if she weren't so...	no	inconsistent

Table B.2: Sentences used for the first block of sentence congruency trials, during the testing phase of the training study described in chapter 9. Each chunk of four sentences ends with trained words containing the same stem.

B.3.3 Sentences for untrained words

sentence	congruent	stem type
He was too worried about the presentation to sleep; for him, last night was...	yes	consistent
Since we discovered that taking naps at regular intervals boosts productivity, we've begun advocating...	yes	consistent
When I saw her shouting, I realized she needs another outlet for her...	no	consistent
After years of war, the country's new peace is the result of several hardworking diplomats'...	no	consistent
A magician must succcesfully escape several complex traps before becoming a certified...	yes	consistent
With untrained guards and flimsy doors, the prison was surprisingly...	yes	consistent
From her fidgeting fingers and tapping toes, I could tell the pills had made her...	no	consistent
We should leave decisions on policy and welfare to the...	no	consistent
The street performer's tricks and jokes built a large crowd quickly, demonstrating his...	yes	consistent
Since the holidays are best spent with other people, it is a good time to be...	yes	consistent

As the contortionist touched her nose to the back of her knee, I wondered how she became so...	no	consistent
The porridge was flavorless, so I asked the cook to add some...	no	consistent
They weren't very good hosts: other than cups of water, their party was...	yes	quasi-consistent
There are very limited options for what you can drink once you transition to...	yes	quasi-consistent
I thought she was just jealous of my singing talent, so I said I was tired of her...	no	quasi-consistent
The barber cuts my hair very precisely; I appreciate his careful approach to...	no	quasi-consistent
The entrepreneur drove many others out of business, marking himself a merciless...	yes	quasi-consistent
With the housing market so unpredictable right now, I'm not sure if your condo will be...	yes	quasi-consistent
Instead of just correcting the grammar mistakes in my essay, I wish his feedback had been more...	no	quasi-consistent
Let's continue talking about this tomorrow; I've enjoyed our...	no	quasi-consistent
Even before the storm started, the ship's tilting motion gave him...	yes	quasi-consistent
Thanks for taking care of me when I was sick last night; I don't usually get so...	yes	quasi-consistent
His elegant and concise writing style inspired them to make their own essays more...	no	quasi-consistent

When they emerged from the dark cave, they blinked painfully at the sun's sudden...	no	quasi-consistent
We were able to disassemble some furniture to make space, but this stone bench is...	yes	inconsistent
When she didn't react at all to eating an extremely hot pepper, they admired her...	no	inconsistent
The team hasn't had any new ideas since she left; they had been relying on her...	yes	inconsistent
Dozens of newly constructed houses were the results of the company's sudden...	no	inconsistent
Don't try to do tricks or jumps on a motorbike until you are an experienced...	no	inconsistent
It was easy to make him angry; he is known for being...	yes	inconsistent
I thought he wouldn't mind that I ate the carrot off the floor, but his expression was...	yes	inconsistent
Even years after the factory closed, many workers still hadn't found new...	no	inconsistent
When the boy didn't say goodbye before leaving, his father was shocked by his...	yes	inconsistent
The optimistic applicant left his interview feeling...	no	inconsistent
Since the new city law was passed, selling alcohol from food trucks is now...	no	inconsistent
The party guests screamed when the balloon popped; they weren't expecting a...	yes	inconsistent

Table B.3: Sentences used for the second block of sentence congruency trials, during the testing phase of the training study described in chapter 9. Each chunk of four sentences ends with untrained words (novel combinations of trained stems and trained suffixes) containing the same stem.

BIBLIOGRAPHY

Alegre, M., & Gordon, P. (1999). Frequency Effects and the Representational Status of Regular Inflections. *Journal of Memory and Language*, *40*(1), 41–61.

Alhama, R. G., Siegelman, N., Frost, R., & Armstrong, B. C. (2019). The role of information in visual word recognition: A perceptually-constrained connectionist account. In *Proceedings of the 41st Annual Conference of the Cognitive Science Society*.

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*(JUL), 1–12.

Amenta, S., Crepaldi, D., & Marelli, M. (2020a). Consistency measures individuate dissociating semantic modulations in priming paradigms: A new look on semantics in the processing of (complex) words. *Quarterly Journal of Experimental Psychology*, *73*(10), 1546–1563.

Amenta, S., Günther, F., & Marelli, M. (2020b). A (distributional) semantic perspective on the processing of morphologically complex words. *The Mental Lexicon*, *15*(1), 62–78.

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.

Anderson, S. R. (1992). *A-morphous morphology*. 62. Cambridge University 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.

Armstrong, B. C., & Plaut, D. C. (2016). Disparate semantic ambiguity effects from semantic processing dynamics rather than qualitative task differences. *Language, Cognition and Neuroscience*, *31*(7), 940–966.

Armstrong, B. C., Watson, C. E., & Plaut, D. C. (2012). Sos! an algorithm and software for the stochastic optimization of stimuli. *Behavior research methods*, *44*(3), 675–705.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of memory and language*, *62*(1), 67–82.

Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge, MA: MIT Press.

Baayen, R. H. (2010). The directed compound graph of english: An exploration of lexical connectivity and its processing consequences. *New impulses in word-formation*, *17*, 383–402.

Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., & Blevins, J. P. (2019). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de) composition but in linear discriminative learning. *Complexity*.

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language*, *37*(1), 94–117.

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., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28.

Baayen, R. H., Milin, P., Ređić, D. F., Hendrix, P., & Marelli, M. (2011). An Amorphous Model for Morphological Processing in Visual Comprehension Based on Naive Discriminative Learning. *Psychological Review*, *118*(3), 438–481.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The celex lexical database (release 2). *Distributed by the linguistic data consortium, University of Pennsylvania*.

Baayen, R. H., & Smolka, E. (2020). Modeling Morphological Priming in German With Naive Discriminative Learning. *Frontiers in Communication*, *5*(April).

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.

Badecker, W., & Allen, M. (2002). Morphological parsing and the perception of lexical identity: A masked priming study of stem homographs. *Journal of Memory and Language*, *47*(1), 125–144.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The english lexicon project. *Behavior research methods*, *39*(3), 445–459.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255–278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bentin, S., & Feldman, L. B. (1990). The contribution of morphological and semantic relatedness to repetition priming at short and long lags: Evidence from hebrew. *The quarterly journal of experimental psychology*, *42*(4), 693–711.

Berg, K., & Aronoff, M. (2017). Self-organization in the spelling of english suffixes: The emergence of culture out of anarchy. *Language*, *93*(1), 37–64.

Bergen, B. K. (2004). The psychological reality of phonaesthemes. *Language*, *80*(2), 290–311.

Bertram, R., Baayen, R. H., & Schreuder, R. (2000). Effects of Family Size for Complex Words. *Journal of Memory and Language*, *42*(3), 390–405.

Beyersmann, E., Castles, A., & Coltheart, M. (2012a). Morphological processing during visual word recognition in developing readers: Evidence from masked priming. *The Quarterly Journal of Experimental Psychology*, *65*(7), 1306–1326.

Beyersmann, E., Cavalli, E., Casalis, S., & Colé, P. (2016a). Embedded stem priming effects in prefixed and suffixed pseudowords. *Scientific Studies of Reading*, *20*(3), 220–230.

Beyersmann, E., Coltheart, M., & Castles, A. (2012b). Parallel processing of whole words and morphemes in visual word recognition. *Quarterly Journal of Experimental Psychology*, *65*(9), 1798–1819.

Beyersmann, E., & Grainger, J. (2018). Support from the morphological family when unembedding the stem. *Journal of experimental psychology. Learning, memory, and cognition*, *44*(1), 135–142.

Beyersmann, E., Mousikou, P., Schroeder, S., Javourey-Drevet, L., Ziegler, J. C., & Grainger, J. (2021). The dynamics of morphological processing in developing readers: A cross-linguistic masked priming study. *Journal of Experimental Child Psychology*, *208*.

Beyersmann, E., Ziegler, J. C., Castles, A., Coltheart, M., Kezilas, Y., & Grainger, J. (2016b). Morpho-orthographic segmentation without semantics. *Psychonomic bulletin & review*, *23*(2), 533–539.

Beyersmann, E., Ziegler, J. C., Castles, A., Coltheart, M., Kezilas, Y., & Grainger, J. (2016c). Morpho-orthographic segmentation without semantics. *Psychonomic Bulletin and Review*, *23*(2), 533–539.

Bhide, A., Schlaggar, B. L., & Barnes, K. A. (2014). Developmental differences in masked form priming are not driven by vocabulary growth. *Frontiers in psychology*, *5*, 667.

Blevins, J. P. (2016). *Word and paradigm morphology*. Oxford University Press.

Boudelaa, S., & Marslen-Wilson, W. D. (2001). Morphological units in the Arabic mental lexicon. *Cognition*, *81*(1), 65–92.

Bradley, D. (1979). Lexical representation of derivational relation. In M. Aronoff, & M. L. Kean (Eds.) *Juncture*. Cambridge, Mass: Academic Press.

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, *23*(3), 389.

Brontë, C. (2002). *Jane Eyre*. Mineola, NY: Dover Publications, Inc. (Original work published 1847).

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., & Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

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.

Burani, C., Dovetto, F. M., Spuntarelli, A., & Thornton, A. M. (1999). Morpholexical access and naming: The semantic interpretability of new root–suffix combinations. *Brain and Language*, *68*(1-2), 333–339.

Burani, C., Marcolini, S., & Stella, G. (2002). How early does morpholexical reading develop in readers of a shallow orthography? *Brain and language*, *81*(1-3), 568–586.

Burani, C., Salmaso, D., & Caramazza, A. (1984). Morphological structure and lexical access.

Burani, C., & Thornton, A. M. (2011). The interplay of root, suffix and whole-word frequency in processing derived words. In *Morphological structure in language processing*, (pp. 157–208). De Gruyter Mouton.

Cao, K., & Rei, M. (2016). A joint model for word embedding and word morphology. In *Proceedings of the 1st Workshop on Representation Learning for NLP*, (pp. 18–26).

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, *28*(3), 297–332.

Carden, J. R., Barreyro, J. P., Segui, J., & Jaichenco, V. (2019). The fundamental role of position in affix identity. *The Mental Lexicon*, *14*(3), 357–380.

Casalis, S., Dusautoir, M., Colé, P., & Ducrot, S. (2009). Morphological effects in children word reading: A priming study in fourth graders. *British Journal of Developmental Psychology*, *27*(3), 761–766.

Casalis, S., Quémart, P., & Duncan, L. G. (2015). How language affects children's use of derivational morphology in visual word and pseudoword processing: Evidence from a cross-language study. *Frontiers in Psychology*, *6*, 452.

Castles, A. (1999). Neighbourhood effects on masked form priming in developing readers. *Language and Cognitive processes*, *14*(2), 201–224.

Castles, A., Davis, C., Cavalot, P., & Forster, K. I. (2007). Tracking the acquisition of orthographic skills in developing readers: Masked priming effects. *Journal of Experimental Child Psychology*, *97*(3), 165–182.

Chee, Q. W., & Yap, M. (2022). Are there task-specific effects in morphological processing? examining semantic transparency effects in semantic categorization and lexical decision. *Quarterly Journal of Experimental Psychology*.

Cheyette, S. J., & Plaut, D. C. (2017). Modeling the n400 erp component as transient semantic over-activation within a neural network model of word comprehension. *Cognition*, *162*, 153–166.

Ciaccio, L. A., Kgolo, N., & Clahsen, H. (2020). Morphological decomposition in Bantu: a masked priming study on Setswana prefixation. *Language, Cognition and Neuroscience*, *0*(0), 1–15.

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of german inflection. *Behavioral and brain sciences*, *22*(6), 991–1013.

Clahsen, H., & Ikemoto, Y. (2012). The mental representation of derived words: An experimental study of–sa and–mi nominals in japanese. *The Mental Lexicon*, *7*(2), 147–182.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: a parallel distributed processing account of the stroop effect. *Psychological review*, *97*, 332–361.

Colé, P., Beauvillain, C., & Segui, J. (1989). On the representation and processing of prefixed and suffixed derived words: A differential frequency effect. *Journal of Memory and language*, *28*(1), 1–13.

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

Crepaldi, D., Rastle, K., & Davis, C. J. (2010b). Morphemes in their place: Evidence for position-specific identification of suffixes. *Memory & cognition*, *38*(3), 312–321.

Dasgupta, T., Sinha, M., & Basu, A. (2015). Computational Modeling of Morphological Effects in Bangla Visual Word Recognition. *Journal of Psycholinguistic Research*, *44*(5), 587–610.

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 and Review*, *17*(5), 749–755.

Dawson, N., Rastle, K., & Ricketts, J. (2018). Morphological effects in visual word recognition: Children, adolescents, and adults. *Journal of Experimental Psychology: Learning Memory and Cognition*, *44*(4), 645–654.

Dawson, N., Rastle, K., & Ricketts, J. (2019). Individual differences in morphological processing in developing and skilled readers. *Society for the Scientific Study of Reading*, *116*, 1–3.

Dawson, N., Rastle, K., & Ricketts, J. (2021). Finding the man amongst many: A developmental perspective on mechanisms of morphological decomposition. *Cognition*, *211*, 104605.

De Grauwe, S., Lemhöfer, K., & Schriefers, H. (2019). Processing derived verbs: the role of motor-relatedness and type of morphological priming. *Language, Cognition and Neuroscience*, *34*(8), 973–990.

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*(1-3), 555–567.

De Jong, N. H., Schreuder, R., & Baayen, H. R. (2000). The morphological family size effect and morphology. *Language and cognitive processes*, *15*(4-5), 329–365.

De Leeuw, J. R. (2015). jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior research methods*, *47*(1), 1–12.

De Rosa, M., & Crepaldi, D. (2021). Letter chunk frequency does not explain morphological masked priming. *Psychonomic bulletin & review*, (pp. 1–11).

Deacon, H., Tong, X., & Mimeau, C. (2019). Morphological and semantic processing in developmental dyslexia across languages: A theoretical and empirical review. In L. Verhoeven, C. A. Perfetti, & K. R. Pugh (Eds.) *Developmental Dyslexia Across Languages and Writing Systems: The Big Picture*, (pp. 327–349). Cambridge: Cambridge University Press.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321.

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. In *The Cambridge Handbook of Psycholinguistics*, 30, (pp. 311–350). Cambridge University Press.

Diependaele, K., Morris, J., Serota, R., Bertrand, D., & Grainger, J. (2013). Breaking boundaries: Letter transpositions and morphological processing. *Language and Cognitive Processes*, *28*(7), 988–1003.

Diependaele, K., Sandra, D., & Grainger, J. (2009). Semantic transparency and masked morphological priming: The case of prefixed words. *Memory and Cognition*, *37*(6), 895–908.

Drews, E., & Zwitserlood, P. (1995). Morphological and orthographic similarity in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(5), 1098.

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.

Duñabeitia, J. A., Laka, I., Perea, M., & Carreiras, M. (2009). Is milkman a superhero like batman? constituent morphological priming in compound words. *European Journal of Cognitive Psychology*, *21*(4), 615–640.

Duñabeitia, J. A., Perea, M., & Carreiras, M. (2007). Do transposed-letter similarity effects occur at a morpheme level? evidence for morpho-orthographic decomposition. *Cognition*, *105*(3), 691–703.

Duñabeitia, J. A., Perea, M., & Carreiras, M. (2008). Does darkness lead to happiness? Masked suffix priming effects. *Language and Cognitive Processes*, *23*(7-8), 1002–1020.

Ehri, L. C. (2005). Learning to Read Words: Theory, Findings, and Issues,. *Scientific Studies of Reading*, *9*(2), 167–188.

Elman, J. L., Bates, E. A., & Johnson, M. H. (1996). *Rethinking innateness: A connectionist perspective on development*, vol. 10. MIT press.

Feldman, L., O'Connor, P., & del Prado Martin, 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. (1994). Beyond orthography and phonology: Differences between inflections and derivations. *Journal of Memory and Language*, *33*, 442–470.

Feldman, L. B., Barac-Cikoja, D., & Kostić, A. (2002). Semantic aspects of morphological processing: Transparency effects in Serbian. *Memory and Cognition*, *30*(4), 629–636.

Feldman, L. B., Milin, P., Cho, K. W., Moscoso del Prado Martín, F., & O'Connor, P. A. (2015). Must analysis of meaning follow analysis of form? A time course analysis. *Frontiers in Human Neuroscience*, *9*(March), 1–19.

Feldman, L. B., & Soltano, E. G. (1999). Morphological priming: The role of prime duration, semantic transparency, and affix position. *Brain and Language*, *68*, 33–39.

Feldman, L. B., Soltano, E. G., Pastizzo, M. J., & Francis, S. E. (2004). What do graded effects of semantic transparency reveal about morphological processing? *Brain and Language*, *90*(1-3), 17–30.

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

Fiorentino, R., Politzer-Ahles, S., Pak, N. S., Martínez-García, M. T., & Coughlin, C. (2015). Dissociating morphological and form priming with novel complex word primes: Evidence from masked priming, overt priming, and event-related potentials. *The mental lexicon*, *10*(3), 413–434.

Fleischhauer, E., Bruns, G., & Grosche, M. (2021). Morphological decomposition supports word recognition in primary school children learning to read: Evidence from masked priming of german derived words. *Journal of Research in Reading*.

Floridi, L., & Chiriatti, M. (2020). Gpt-3: Its nature, scope, limits, and consequences. *Minds and Machines*, *30*(4), 681–694.

Ford, M. A., Davis, M. H., & Marslen-Wilson, W. D. (2010). Derivational morphology and base morpheme frequency. *Journal of Memory and Language*, *63*(1), 117–130.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of verbal learning and verbal behavior*, *12*(6), 627–635.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 680–698.

Frost, R., Deutsch, A., Gilboa, O., Tannenbaum, M., & Marslen-Wilson, W. D. (2000). Morphological priming: Dissociation of phonological, semantic, and morphological factors. *Memory and Cognition*, *28*(8), 1277–1288.

Frost, R., Forster, K. I., & Deutsch, A. (1997). What can we learn from the morphology of Hebrew? A masked-priming investigation of morphological representation. *Journal of Experimental Psychology: Learning Memory and Cognition*, *23*(4), 829–856.

Gilbert, S. J., & Shallice, T. (2002). Task switching: A pdp model. *Cognitive psychology*, *44*, 297–337.

Giraudo, H., Dal Maso, S., & Piccinin, S. (2016). The role of stem frequency in morphological processing. In *Mediterranean Morphology Meetings*, vol. 10, (pp. 64–72).

Giraudo, H., & Grainger, J. (2000). Effects of prime word frequency and cumulative root frequency in masked morphological priming. *Language and Cognitive Processes*, *15*(4-5), 421–444.

Gomez, P., Ratcliff, R., & Perea, M. (2008). The overlap model: A model of letter position coding. *Psychological review*, *115*(3), 577–600.

Gonnerman, L. M., Seidenberg, M. S., & Andersen, E. S. (2007). Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology. *Journal of Experimental Psychology: General*, *136*(2), 323–345.

Grainger, J., & Beyersmann, E. (2017). Edge-aligned embedded word activation initiates morpho-orthographic segmentation. In *Psychology of learning and motivation*, vol. 67, (pp. 285–317). Elsevier.

Grainger, J., & Beyersmann, E. (2020). Effects of lexicality and pseudo-morphological complexity on embedded word priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Grainger, J., Colé, P., & Segui, J. (1991). Masked morphological priming in visual word recognition. *Journal of Memory and Language*, *30*(1), 370–384.

Grainger, J., & Ziegler, J. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology*, *2*, 1–13.

Günther, F., & Marelli, M. (2018). Enter sandman: Compound processing and semantic transparency in a compositional perspective. *Journal of Experimental psychology. Learning, Memory, and Cognition*, *45*(10), 1872–1882.

Günther, F., Marelli, M., & Bölte, J. (2020). Semantic transparency effects in german compounds: A large dataset and multiple-task investigation. *Behavior Research Methods*, *52*(3), 1208–1224.

Günther, F., Smolka, E., & Marelli, M. (2019). 'Understanding' differs between English and German: Capturing systematic language differences of complex words. *Cortex*, *116*, 168–175.

Hannagan, T., Agrawal, A., Cohen, L., & Dehaene, S. (2021). Emergence of a compositional neural code for written words: Recycling of a convolutional neural network for reading. *Proceedings of the National Academy of Science USA*, *118*.

Hasenäcker, J., Solaja, O., & Crepaldi, D. (2021). Does morphological structure modulate access to embedded word meaning in child readers? *Memory & Cognition*, *49*(7), 1334–1347.

Hay, J. B., & Baayen, R. H. (2005). Shifting paradigms: Gradient structure in morphology. *Trends in cognitive sciences*, *9*(7), 342–348.

Hendrix, P., & Sun, C. C. (2021). A word or two about nonwords: Frequency, semantic neighborhood density, and orthography-to-semantics consistency effects for nonwords in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(1), 157.

Hockett, C. F., & Hockett, C. D. (1960). The origin of speech. *Scientific American*, *203*(3), 88–97.

Hu, M., & Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a foreign language*, *23*, 403–430.

Jacobs, R. A. (1988). Increased rates of convergence through learning rate adaptation. *Neural networks*, *1*(4), 295–307.

Jared, D., Jouravlev, O., & Joanisse, M. (2017). The effect of semantic transparency on the processing of morphologically derived words: Evidence from decision latencies and event-related potentials. *Journal of Experimental Psychology: Learning Memory and Cognition*, *43*(3), 422–450.

Kahraman, H., & Kırkıcı, B. (2021). Letter transpositions and morphemic boundaries in the second language processing of derived words: An exploratory study of individual differences. *Applied Psycholinguistics*, *42*(2), 417–446.

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.

Kgolo, N., & Eisenbeiss, S. (2015). The role of morphological structure in the processing of complex forms: evidence from Setswana deverbative nouns. *Language, Cognition and Neuroscience*, *30*(9), 1116–1133.

Kriegeskorte, N. (2015). Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annual review of vision science*, *1*, 417–446.

Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What learning systems do intelligent agents need? complementary learning systems theory updated. *Trends in cognitive sciences*, *20*(7), 512–534.

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.

Kuroki, D. (2021). A new jspsych plugin for psychophysics, providing accurate display duration and stimulus onset asynchrony. *Behavior Research Methods*, *53*(1), 301–310.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.

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.

Laszlo, S., & Plaut, D. C. (2012). A neurally plausible parallel distributed processing model of event-related potential word reading data. *Brain and Language*, *120*(3), 271–281.

Laudanna, A., Badecker, W., & Caramazza, A. (1989). Priming homographic stems. *Journal of Memory and Language*, *28*(5), 531–546.

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.

Law, J. M., & Ghesquière, P. (2021). Morphological processing in children with developmental dyslexia: A visual masked priming study. *Reading Research Quarterly*.

Lei Ba, J., Kiros, J. R., & Hinton, G. E. (2016). Layer normalization. *ArXiv e-prints*, (pp. arXiv–1607).

Leinonen, A., Grønholm-Nyman, P., Järvenpää, M., Søderholm, C., Lappia, O., Laine, M., & Krause, C. M. (2009). Neurocognitive processing of auditorily and visually presented inflected words and pseudowords: evidence from a morphologically rich language. *Brain Research*, *1275*, 54–66.

Lelonkiewicz, J. R., Ktori, M., & Crepaldi, D. (2020). Morphemes as letter chunks: Discovering affixes through visual regularities. *Journal of Memory and language*, *115*, 104152.

Li, J., Taft, M., & Xu, J. (2017). The processing of english derived words by chinese-english bilinguals. *Language Learning*, *67*(4), 858–884.

Li, Q., Joo, S. J., Yeatman, J. D., & Reinecke, K. (2020). Controlling for participants viewing distance in large-scale, psychophysical online experiments using a virtual chinrest. *Scientific reports*, *10*(1), 1–11.

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

Liu, F., Lu, H., Lo, C., & Neubig, G. (2017). Learning character-level compositionality with visual features. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, (pp. 2059–2068).

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in psychology*, *6*, 1171.

Longtin, C. M., & Meunier, F. (2005). Morphological decomposition in early visual word processing. *Journal of Memory and Language*, *53*(1), 26–41.

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

Lüdeling, A., & De Jong, N. (2002). German particle verbs and word-formation. *Verb-particle explorations*, (pp. 315–333).

Marelli, M., & Amenta, S. (2018). A database of orthography-semantics consistency (osc) estimates for 15,017 english words. *Behavior research methods*, *50*(4), 1482–1495.

Marelli, M., Amenta, S., & Crepaldi, D. (2015). Semantic transparency in free stems: The effect of Orthography-Semantics Consistency on word recognition. *Quarterly Journal of Experimental Psychology*, *68*(8), 1571–1583.

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., & Baroni, M. (2015). Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review*, *122*(3), 485–515.

Marelli, M., Gagné, C. L., & Spalding, T. L. (2017). Compounding as abstract operation in semantic space: Investigating relational effects through a large-scale, data-driven computational model. *Cognition*, *166*, 207–224.

Marelli, M., Traficante, D., & Burani, C. (2020). Reading morphologically complex words: Experimental evidence and learning models. In V. Pirrelli, I. Plag, & W. U. Dressler (Eds.) *Word knowledge and word usage: A cross-disciplinary guide to the mental lexicon*, (pp. 553–592). de Gruyter Mouton.

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.

Marslen-Wilson, W. D., 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., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*, 419–457.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*(5), 375–407.

McCormick, S. F., Brysbaert, M., & Rastle, K. (2009). Short article: Is morphological decomposition limited to low-frequency words? *Quarterly Journal of Experimental Psychology*, *62*(9), 1706–1715.

McCormick, S. F., Rastle, K., & Davis, M. H. (2008). Is there a 'fete' in 'fetish'? Effects of orthographic opacity on morpho-orthographic segmentation in visual word recognition. *Journal of Memory and Language*, *58*(2), 307–326.

McCutchen, D., Logan, B., & Biangardi-Orpe, U. (2009). Making Meaning: Children's Sensitivity to Morphological Information During Word Reading. *Reading Research Quarterly*, *44*(4), 360–376.

Merkx, M., Rastle, K., & Davis, M. H. (2011). The acquisition of morphological knowledge investigated through artificial language learning. *Quarterly Journal of Experimental Psychology*, *64*(6), 1200–1220.

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

Milin, P., Feldman, L. B., Ramscar, M., Hendrix, P., & Baayen, R. H. (2017). Discrimination in lexical decision. *PloS one*, *12*(2), e0171935.

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

Milin, P., Smolka, E., & Feldman, L. B. (2018). Models of lexical access and morphological processing. In E. M. Fernández, & H. S. Cairns (Eds.) *The Handbook of Psycholinguistics*, (pp. 240–268). John Wiley & Sons, Inc.

Morris, D. (2013). *Diagnosis and correction of reading problems*. Guilford Publications.

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

Morris, J., Grainger, J., & Holcomb, P. (2008). An electrophysiological investigation of early effects of masked morphological priming. *Language and Cognitive Processes*, *23*(7-8), 1021–1056.

Morris, J., Porter, J. H., Grainger, J., & Holcomb, P. J. (2011). Effects of lexical status and morphological complexity in masked priming: An erp study. *Language and cognitive processes*, *26*(4-6), 558–599.

Morton, J. (1979). Facilitation in word recognition: Experiments causing change in the logogen model. In *Processing of visible language*, (pp. 259–268). Springer, Boston, MA.

Moscoso del Prado Martın, F. (2003). Paradigmatic effects in morphological processing: Computational and cross-linguistic experimental studies. *MPI Series in Psycholinguistics. Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*.

Moscoso Del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R., & Harald Baayen, R. (2004). Morphological family size in a morphologically rich language: The case of finnish compared with Dutch and Hebrew. *Journal of Experimental Psychology: Learning Memory and Cognition*, *30*(6), 1271–1278.

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., Ernestus, M., & Baayen, R. H. (2004). Do type and token effects reflect different mechanisms? connectionist modeling of dutch past-tense formation and final devoicing. *Brain and Language*, *90*(1-3), 287–298.

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.

Mulder, K., Dijkstra, T., Schreuder, R., & Baayen, H. R. (2014). Effects of primary and secondary morphological family size in monolingual and bilingual word processing. *Journal of Memory and Language*, *72*, 59–84.

Murrell, G. A., & Morton, J. (1974). Word recognition and morphemic structure. *Journal of Experimental Psychology*, *102*(6), 963.

Myers, J., Huang, Y. C., & Wang, W. (2006). Frequency effects in the processing of chinese inflection. *Journal of Memory and Language*, *54*(3), 300–323.

Nagy, W., Berninger, V. W., & Abbott, R. D. (2006). Contributions of morphology beyond phonology to literacy outcomes of upper elementary and middle-school students. *Journal of educational psychology*, *98*(1), 134.

Nation, K., & Cocksey, J. (2009). Beginning readers activate semantics from sub-word orthography. *Cognition*, *110*(2), 273–278.

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing*, (pp. 1532–1543).

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

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and cognitive processes*, *12*(5-6), 765–806.

Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: empirical and computational support for a single-mechanism account of lexical processing. *Psychological review*, *107*(4), 786.

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.

Plaut, D. C., & McClelland, J. L. (1993). Generalization with componential attractors: Word and nonword reading in an attractor network. In *Proceedings of the 15th conference of the Cognitive Science Society*, (pp. 824–829).

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological review*, *103*(1), 56–115.

Quémart, P., & Casalis, S. (2015). Visual processing of derivational morphology in children with developmental dyslexia: Insights from masked priming. *Applied Psycholinguistics*, *36*(2), 345–376.

Quémart, P., Casalis, S., & Colé, P. (2011). The role of form and meaning in the processing of written morphology: A priming study in French developing readers. *Journal of Experimental Child Psychology*, *109*(4), 478–496.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rabovsky, M., & McClelland, J. L. (2020). Quasi-compositional mapping from form to meaning: a neural network-based approach to capturing neural responses during human language comprehension. *Philosophical Transactions of the Royal Society B*, *375*(1791).

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, *1*(8), 9.

Rapp, B. C. (1992). The nature of sublexical orthographic organization: The bigram trough hypothesis examined. *Journal of Memory and Language*, *31*, 33–53.

Rastle, K. (2019a). Eps mid-career prize lecture 2017: Writing systems, reading, and language. *Quarterly Journal of Experimental Psychology*, *72*(4), 677–692.

Rastle, K. (2019b). The place of morphology in learning to read in English. *Cortex*, *116*, 45–54.

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

Rastle, K., Davis, M. H., Marslen-Wilson, W. D., & Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes*, *15*(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 and Review*, *11*(6), 1090–1098.

Rohde, D. L. (2002). Methods for binary multidimensional scaling. *Neural Computation*, *14*(5), 1195–1232.

Rohde, D. L. (2003). *LENS: The light, efficient network simulator*. Version 2.63.

Rueckl, J., & Aicher, K. (2008). Are corner and brother morphologically complex? not in the long term. *Language and Cognitive Processes*, *23*(7–8), 972–1001.

Rueckl, J. G. (2010). Connectionism and the role of morphology in visual word recognition. *The mental lexicon*, *5*(3), 371–400.

Rueckl, J. G., & Raveh, M. (1999). The influence of morphological regularities on the dynamics of a connectionist network. *Brain and Language*, *68*, 110–117.

Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning internal representations by error propagation. In *Parallel distributed processing: explorations in the microstructure of cognition, vol. 1: foundation*, (pp. 318–362). Cambridge, MA: MIT Press.

Rumelhart, D., & McClelland, J. (1986). On learning the past tenses of english verbs. In *Parallel distributed processing: explorations in the microstructure of cognition, vol. 2: psychological and biological models*, (pp. 216–271). Cambridge, MA: MIT Press.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human perception and performance*, *3*(1), 1.

Schiff, R., Raveh, M., & Fighel, A. (2012). The Development of the Hebrew Mental Lexicon: When Morphological Representations Become Devoid of Their Meaning. *Scientific Studies of Reading*, *16*(5), 383–403.

Schiff, R., Raveh, M., & Kahta, S. (2008). The developing mental lexicon: Evidence from morphological priming of irregular Hebrew forms. *Reading and Writing*, *21*(7), 719–743.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, *95*(1), 26–43.

Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. *Morphological aspects of language processing*, *2*, 257–294.

Schreuder, R., & Baayen, R. H. (1997). How complex simplex words can be. *Journal of memory and language*, *37*(1), 118–139.

Seidenberg, M. S., & Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences*, *4*(9), 353–361.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523–568.

Seidenberg, M. S., & Plaut, D. C. (2014). Quasiregularity and its discontents: The legacy of the past tense debate. *Cognitive Science*, *38*(6), 1190–1228.

Shapiro, B. J. (1969). The subjective estimation of relative word frequency. *Journal of verbal learning and verbal behavior*, *8*(2), 248–251.

Siegelman, N., Rueckl, J. G., Lo, J. C. M., Kearns, D. M., Morris, R. D., & Compton, D. L. (2022). Quantifying the regularities between orthography and semantics and their impact on group-and individual-level behavior. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Silva, R., & Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism*, *11*(2), 245–260.

Smolka, E., Gondan, M., & Rösler, F. (2015). Take a stand on understanding: Electrophysiological evidence for stem access in German complex verbs. *Frontiers in Human Neuroscience*, *9*(FEB), 1–21.

Smolka, E., Komlósi, S., & Rösler, F. (2009). When semantics means less than morphology: The processing of German prefixed verbs. *Language and Cognitive Processes*, *24*(3), 337–375.

Stanners, R. F., Neiser, J. J., Hernon, W. P., & Hall, R. (1979). Memory representation for morphologically related words. *Journal of Verbal Learning and Verbal Behavior*, *18*(4), 399–412.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition*, *7*(4), 263–272.

Taft, M. (1985). The decoding of words in lexical access: A review of the morphographic approach. In D. Besner, T. Waller, & G. MacKinnon (Eds.) *Reading research: Advances in theory and practice*, (pp. 83–126). New York: Academic Press.

Taft, M. (1994). Interactive-activation as a Framework for Understanding Morphological Processing. *Language and Cognitive Processes*, *9*(3), 271–294.

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

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words.

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.

Tamminen, J., Davis, M. H., Merkx, M., & Rastle, K. (2012). The role of memory consolidation in generalisation of new linguistic information. *Cognition*, *125*(1), 107–112.

Tamminen, J., Davis, M. H., & Rastle, K. (2015). From specific examples to general knowledge in language learning. *Cognitive Psychology*, *79*, 1–39.

Tsang, Y. K., & Chen, H. C. (2013). Early morphological processing is sensitive to morphemic meanings: Evidence from processing ambiguous morphemes. *Journal of Memory and Language*, *68*(3), 223–239.

Tzur, B., & Frost, R. (2007). Soa does not reveal the absolute time course of cognitive processing in fast priming experiments. *Journal of memory and language*, *56*(3), 321–335.

Ulicheva, A., Harvey, H., Aronoff, M., & Rastle, K. (2020). Skilled readers' sensitivity to meaningful regularities in English writing. *Cognition*, *195*(August), 1–21.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological review*, *108*(3), 550.

Vannest, J., Bertram, R., Järvikivi, J., & Niemi, J. (2002). Counterintuitive cross-linguistic differences: More morphological computation in English than in Finnish. *Journal of Psycholinguistic Research*, *31*(2), 83–106.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems*, (pp. 5998–6008).

Velan, H., Frost, R., Deutsch, A., & Plaut, D. C. (2005). The processing of root morphemes in Hebrew: Contrasting localist and distributed accounts. *Language and Cognitive Processes*, *20*(1-2), 169–206.

Vidal, Y., Viviani, E., Zoccolan, D., & Crepaldi, D. (2021). A general-purpose mechanism of visual feature association in visual word identification and beyond. *Current Biology*, *31*(6), 1261–1267.

Viviani, E., & Crepaldi, D. (2019). Masked morphological priming tracks the development of a fully mature lexical system in L2. *Journal of Memory and Language*.

Watkins, C. (Ed.) (2000). *The American heritage dictionary of Indo-European roots*. Houghton Mifflin Harcourt.

Xu, J., & Taft, M. (2015). The effects of semantic transparency and base frequency on the recognition of english complex words. *Journal of Experimental Psychology: Learning Memory and Cognition*, *41*(3), 904–910.

Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, *19*(3), 356–365.

Zeno, S. M., Ivens, S. H., Millart, R. T., & Duwuri, R. (Eds.) (1995). *The educators word frequency guide*. Brewster, NY: Touchstone Applied Science Associate, Inc.

Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-level convolutional networks for text classification. *Advances in neural information processing systems*, *28*, 649–657.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological bulletin*, *131*(1), 3.

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? a connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(4), 1131–1161.

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